Mathar / Pfeifer

Stochastik für Informatiker

Stochastik für Informatiker

Von Prof. Dr. rer. nat. Rudolf Mathar, RWTH Aachen
und Prof. Dr. rer. nat. Dietmar Pfeifer, Universität Oldenburg

Prof. Dr. rer. nat. Rudolf Mathar

Geboren 1952 in Kalterherberg. Studium der Mathematik an der RWTH Aachen (1972 bis 1978), Promotion 1981, Habilitation 1985 an der RWTH Aachen. 1987 Lehrauftrag an der European Business School (Oestrich), 1988/1989 Umhabilitation und Privatdozent an der Universität Augsburg. Seit 1990 Professor für Stochastik, insbesondere Anwendungen in der Informatik, an der RWTH Aachen.

Prof. Dr. rer. nat. Dietmar Pfeifer

Geboren 1953 in Wuppertal-Elberfeld. Studium der Mathematik an der RWTH Aachen (1971 bis 1977), Promotion 1980, Habilitation 1984 an der RWTH Aachen. 1985 Gastprofessor an der University of North Carolina, Chapel Hill, USA. 1986/87 Heisenberg-Stipendiat der DFG. 1987 Gastaufenthalt an der University of California, Santa Barbara, USA. Seit 1988 Professor für Mathematik an der Universität Oldenburg.

Einen Prozeß, mit dem man überhaupt nie fertig werden könnte, wie das Zusammenzählen einer unendlichen Reihe, ermöglicht die Mathematik unter günstigen Umständen in wenigen Augenblicken zu vollziehen. Bis zu komplizierten Logarithmenrechnungen, ja selbst Integrationen macht sie es überhaupt schon mit der Maschine; die Arbeit des Heutigen beschränkt sich auf das Einstellen der Ziffern seiner Frage und auf das Drehen an einer Kurbel oder ähnliches. Der Amtsdiener einer Lehrkanzel kann damit Probleme aus der Welt schaffen, zu deren Auflösung sein Professor noch vor zweihundert Jahren zu den Herren Newton in London oder Leibniz in Hannover hätte reisen müssen.

Aus: Robert Musil, Der Mathematische Mensch (1913)

Vorwort

Das vorliegende Buch entstand aus einer Reihe von Vorlesungen, die wir an der Rheinisch–Westfälischen Technischen Hochschule Aachen, der European Business School, der Universität Oldenburg und der Universität Augsburg seit 1984 gehalten haben. Diese Vorlesungen wandten sich vor allem an Informatikstudenten und Mathematikstudenten mit Nebenfach Informatik mit dem Ziel, stochastische Grundbegriffe unter besonderer Berücksichtigung Informatik–spezifischer Aspekte zu vermitteln.
Unter den zahlreichen Einsatzfeldern stochastischer Methoden in der Informatik seien hier beispielhaft genannt:
Die Average–Case–Analyse von Algorithmen, die stochastische Automatentheorie, Anwendungen im Bereich des CAD (Bézier–Kurven und –Flächen), stochastische Informationstheorie und Codierungstheorie, Rechnernetze und Leistungsbewertung von Rechnersystemen (Warteschlangenprobleme), Bildverarbeitung (Computertomographie), automatische Spracherkennung (Hidden–Markov–Modelle), Expertensysteme (effiziente Berechnung von bedingten Wahrscheinlichkeiten), künstliche Intelligenz (Neuronale Netze), stochastische Optimierungs– und Suchverfahren (Simulated Annealing), stochastische Simulation, probabilistische Algorithmen u.v.a..
Die zum Verständnis benötigten theoretischen Grundlagen, die erfahrungsgemäß häufig weit über den in einführenden Veranstaltungen angebotenen Stoff hinausgehen, sind dementsprechend vielfältig und reichen von einfachen kombinatorischen Überlegungen bei einigen Problemen der Average–Case–Analyse von Algorithmen bis hin zu tiefliegenden Sätzen der axiomatischen Wahrscheinlichkeitstheorie, etwa bei den Markoff–Ketten und –Prozessen oder der Theorie der Punktprozesse im Bereich der Bildverarbeitung.
Ziel des Buches ist es daher, eine einheitliche und möglichst geschlossene Übersicht über die zum Verständnis benötigten Grundlagen zu geben. Der weitaus größte Teil des Textes kann dabei mit Kenntnissen der Mathematik etwa im Rahmen des Werkes von Kiyek & Schwarz (1989), das in derselben Reihe wie das vorliegende Buch erschienen ist, verstanden werden; bei tieferliegenden Problemen der Analysis wird gelegentlich auf die beiden Bände von Heuser (1989) verwiesen. Der Text wird darüberhinaus durch Übungsaufgaben ergänzt.
Das Buch ist so konzipiert, daß es sowohl im Rahmen einer einführenden Veranstaltung in die Stochastik als auch für weiterführende Vorlesungen eingesetzt werden kann. Trotz des überwiegenden Lehrbuchcharakters dieses Textes haben wir uns allerdings auch bemüht, neuere Entwicklungen, die z.T. bisher nur in Originalarbeiten vorliegen, mit einzubeziehen, um dort, wo es im Rahmen unseres Zugangs möglich ist, Anschluß an Fragestellungen der aktuellen Forschung zu erlangen. Dies betrifft insbesondere die Abschnitte 3.3 (Simulated Annealing) und 6.1 (Erzeugung von Zufallszahlen) sowie Kapitel 4 (Probabilistische Analyse von Algorithmen). Dementsprechend haben wir häufig auch Querverweise auf andere Literatur mit in den Text aufgenommen.

Bei der didaktischen Konzeption des Buches haben wir uns von dem Grundsatz leiten lassen, zur Einführung neuer Begriffe wenn möglich immer zuerst den einfachsten Zugang zu wählen und diesen dann schrittweise zu verallgemeinern, also z.B. vom Laplace'schen Wahrscheinlichkeitsbegriff der endlich-diskreten Gleichverteilung und seiner kombinatorischen Interpretation auszugehen, um dann über beliebige diskrete Verteilungen und ihre Eigenschaften den allgemeinen, an der Maßtheorie orientierten Wahrscheinlichkeitsbegriff einzuführen. Dem — geringfügigen — Nachteil einer Ausweitung des Textes steht demgegenüber der Vorteil, große Teile späterer Kapitel auch dann noch lesen und verstehen zu können, wenn man sich in den beiden einführenden Kapiteln lediglich mit den Grundlagen der diskreten Wahrscheinlichkeitsrechnung vertraut machen möchte. Allerdings sind zum Verständnis komplexerer Zusammenhänge etwa im Bereich der stochastischen Prozesse, deren Anwendung in den verschiedensten Bereichen der modernen Informatik zunehmend zu beobachten ist, auch gewisse maßtheoretische Denkweisen unabdingbar.
Die Motivation der behandelten Theorie haben wir als wichtigen Aspekt aus zwei Perspektiven betont. Einen Teil hiervon bilden die zahlreichen angewandten Problemstellungen, die mit den bereitgestellten Hilfsmitteln elegant angegriffen werden können. Zum anderen stößt man häufig relativ schnell bei gründlichem Durchdenken einiger einfacher Axiome auf tiefliegende Probleme, deren Lösung für eine befriedigende Darstellung der Theorie unvermeidbar ist. Wir haben dies an der Frage der Existenz einer Gleichverteilung auf dem Einheitsintervall genauer dargestellt. Von den zahlreichen Beispiele aus der Informatik, die wir zur Veranschaulichung der theoretischen Grundlagen in den Text aufgenommen haben, werden einige in verschiedenen Kapiteln wiederholt aufgegriffen. Hierdurch läßt sich der Einsatz verschiedener Methoden der Theorie an den gleichen Problemstellungen verdeutlichen.
Unseren herzlichen Dank möchten wir allen Freunden, Kollegen, Mitarbeitern und Studierenden aussprechen, die durch ihre Kritik und Unterstützung am Entstehen des Buches direkt oder indirekt mitgewirkt haben. Insbesondere danken wir den Herren Prof. Dr. O. Emrich, Dr. H.J. Witte, Dipl. Math. M. Scheiper, und Cand. Math. B. Roos für die gründliche Durchsicht von Teilen des Manuskripts, verbunden mit zahlreichen wertvollen Hinweisen, ebenso Herrn Prof. Dr. K. Floret für interessante Anregungen bei einigen maßtheoretischen Grundlagenfragen. Herr Dipl. Math. G. Brücks und Herr Dipl. Math. J. Meier haben uns tatkräftig mit Ihren Soft- und Hardware-Kenntnissen bei der Anfertigung einiger schwieriger Zeichnungen unterstützt. Frau Doetsch hat in Augsburg Teile des Manuskripts in TEX geschrieben. Die kompakte Darstellung einer Charakterisierung von Zuständen bei Markoff-Ketten geht auf Ideen von Herrn Prof. Dr. N. Gaffke zurück. Allen genannten möchten wir an dieser Stelle besonders herzlich danken.

Aachen und Oldenburg, im August 1990

Rudolf Mathar
Dietmar Pfeifer

Inhaltsverzeichnis

1. Grundbegriffe der Wahrscheinlichkeitstheorie

In beinahe allen Bereichen des täglichen Lebens treten Situationen auf, bei denen "Zufall" eine Rolle spielt. Eine Bewertung des "Zufalls" (genauer: "zufälliger" Ereignisse) geschieht dann üblicherweise durch die Angabe einer "Wahrscheinlichkeit", mit der das betreffende Ereignis eintritt. Dabei ist zwischen verschiedenen Qualitäten des Begriffs "Wahrscheinlichkeit" zu unterscheiden, wie man an den beiden Feststellungen

"Wahrscheinlich gibt es heute abend ein Gewitter."

oder

"Bei binärer Suche findet man ein Schlüsselelement in einem geordneten Feld der Länge $2^n - 1$ mit Wahrscheinlichkeit $2^{k-1}/2^n$ in genau k Schritten $(1 \leq k \leq n)$."

erkennen kann. Die erste Aussage benutzt nämlich den Begriff "Wahrscheinlichkeit" im Sinn eines subjektiven Gefühls für die Möglichkeit des Eintretens eines bestimmten Ereignisses, während die zweite Aussage eine Quantifizierung der "Wahrscheinlichkeit" in einem mathematischen Modell etwa aufgrund kombinatorischer Überlegungen vornimmt. Aussagen der ersten Art entziehen sich in der Regel einer rigorosen mathematischen Behandlung. Wir werden uns im folgenden mit solchen Wahrscheinlichkeitsaussagen beschäftigen, die im Rahmen eines geeigneten mathematischen Modells formulierbar sind.

Eine erste Schwierigkeit besteht darin, zu präzisieren, was man überhaupt unter einem "Ereignis", für dessen Eintreten eine Wahrscheinlichkeit angegeben werden soll, zu verstehen hat. Für die mathematische Modellbildung hat es sich als sinnvoll erwiesen, Ereignisse durch Teilmengen einer bestimmten Grundmenge zu beschreiben. Logische Verknüpfungen von Ereignissen, etwa mittels *"und"* (gleichzeitiges Eintreten), *"oder"* (wahlweises Eintreten) oder *"nicht"* (Nicht–Eintreten), lassen sich dann vermöge der Durchschnitts–, Vereinigungs– oder Komplementbildung beschreiben. "Ereignisse" werden damit aufgefaßt als Teilmengen $A, B, \ldots$ einer nicht–leeren Grundmenge Ω, denen mittels einer geeigneten Abbildung P, die in einigen Fällen auf der ganzen Potenzmenge $\mathfrak{P}(\Omega)$ definiert werden kann, Wahrscheinlichkeiten $P(A), P(B), \ldots$ zugeordnet werden; P heißt deshalb auch *Wahrscheinlichkeitsverteilung* über Ω. Das Symbol P für Wahrscheinlichkeit ist dabei historisch auf das entsprechende Wort "probabilité" (englisch: "probability") zurückzuführen.

Natürlich wird man verlangen, daß die Abbildung P gewisse Eigenschaften besitzt, die sich soweit als möglich mit den intuitiven Vorstellungen von "Wahrscheinlichkeit" decken. Neben der Normierungsbedingung

$$0 \leq P(A) \leq 1 \quad \text{für alle } A \in \mathfrak{P}(\Omega) \tag{1.0.1}$$

wird man beispielsweise für das "unmögliche" Ereignis $\emptyset$ bzw. das "sichere" Ereignis Ω fordern

$$P(\emptyset) = 0, \quad P(\Omega) = 1. \tag{1.0.2}$$

Weiter erscheint es sinnvoll, das Nicht–Eintreten eines Ereignisses A, also das Eintreten des komplementären Ereignisses A^c, mit der Wahrscheinlichkeit

$$P(A^c) = 1 - P(A) \quad \text{für alle } A \in \mathfrak{P}(\Omega) \tag{1.0.3}$$

zu bewerten. Für sich gegenseitig ausschließende Ereignisse A und B, d.h. mit der Eigenschaft $A \cap B = \emptyset$, wird allgemeiner die Additivität von P gefordert, nämlich

$$P(A \cup B) = P(A) + P(B), \quad A, B \in \mathfrak{P}(\Omega). \tag{1.0.4}$$

Will man auch die Verknüpfung abzählbar–unendlich vieler Ereignisse miterfassen, was z.B. bei bestimmten Wartezeitproblemen nötig ist, so kann man die letzte Eigenschaft auch auf den Fall abzählbar vieler, paarweise disjunkter Ereignisse $A_n \in \mathfrak{P}(\Omega), n \in \mathbf{N}$, $A_i \cap A_j = \emptyset, i \neq j$, ausdehnen ($\sigma$–Additivität von P):

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n), \quad A_n \in \mathfrak{P}(\Omega). \tag{1.0.5}$$

Allerdings bringt dies gewisse Probleme mit sich, wenn P auf $\mathfrak{P}(\Omega)$ definiert werden soll und die Grundmenge Ω überabzählbar ist, da dann möglicherweise keine Wahrscheinlichkeiten mehr mit eventuell geforderten zusätzlichen Eigenschaften existieren. Beispielsweise gibt es keine "*Gleichverteilung*" über der Potenzmenge des Intervalls $\Omega = (0, 1]$, d.h. eine Wahrscheinlichkeitsverteilung P auf $\mathfrak{P}(\Omega)$ mit der charakterisierenden Eigenschaft

$$P\big((a + h, b + h]\big) = P\big((a, b]\big) \quad \text{für alle } 0 \leq a < b \leq 1,\ -a \leq h \leq 1 - b \tag{1.0.6}$$

(jedes linkshalboffene Intervall besitzt — unabhängig von seiner Lage in $(0, 1]$ — dieselbe Wahrscheinlichkeit) bzw. allgemeiner

$$P(A + h) = P(A) \quad \text{für alle } A \in \mathfrak{P}(\Omega),\ A + h \in \mathfrak{P}(\Omega) \tag{1.0.7}$$

mit $A + h = \{a + h \mid a \in A\}$, $h \in \mathbf{R}$ (jede Teilmenge A von $(0, 1]$ besitzt dieselbe Wahrscheinlichkeit wie die um den Betrag h verschobene Menge $A + h$).

Beziehung (1.0.6) ist dabei wegen der σ–Additivität von P gleichbedeutend mit der Eigenschaft

$$P\big((a, b]\big) = b - a \quad \text{für alle } 0 \leq a < b \leq 1 \tag{1.0.8}$$

(vgl. Aufgabe 1.1).

Mit Methoden, die schon 1905 von Vitali entwickelt wurden, läßt sich nachweisen, daß eine solche allgemeine Gleichverteilung (d.h. eine Verteilung mit der Eigenschaft (1.0.7)) auf der Potenzmenge $\mathfrak{P}(\Omega)$ nicht existieren kann.

Wenn man auf die σ–Additivität von Wahrscheinlichkeitsverteilungen nicht verzichten will, muß man also notgedrungen den Definitionsbereich $\mathfrak{P}(\Omega)$ von P einschränken. Dies sollte so geschehen, daß einerseits noch genügend viele "interessante" Ereignisse übrig bleiben, andererseits hinreichend viele P mit bestimmten gewünschten Eigenschaften existieren. Man gelangt damit zu σ–Algebren von Ereignissen, welche im nachfolgenden Abschnitt besprochen werden.

1.1. σ–Algebren und Wahrscheinlichkeitsmaße

Lehrbücher über Wahrscheinlichkeitsrechnung und damit verbundene kombinatorische Probleme gibt es bereits seit dem 16. Jahrhundert, z.B. "Liber de ludo aleae" von Cardano (1501–1576), "De ratiociniis in ludo aleae" von Huygens (1629–1695), "Ars conjectandi" von J. Bernoulli (1654–1705), "The doctrine of chances" von de Moivre (1667–1754), "Théorie analytique des probabilités" von Laplace (1749–1827) oder "Recherches sur la probabilité des jugements en matière criminelle et en matière civile" von Poisson (1781–1840). Obgleich sich diese frühen Lehrbücher nicht auf eine axiomatische Theorie der Wahrscheinlichkeitsrechnung im heutigen Sinn gründen, gibt es doch Ansätze zur Entwicklung eines Wahrscheinlichkeitsbegriffs (insbesondere durch Laplace), der im wesentlichen den oben gestellten Forderungen wenigstens für endliche Grundmengen Ω genügt. Wegen seiner grundsätzlichen Bedeutung im Zusammenhang mit gewissen kombinatorischen Problemen soll dieser Wahrscheinlichkeitsbegriff hier kurz vorgestellt werden.

Definition 1.1.1. *(Laplace'scher Wahrscheinlichkeitsbegriff)*
Es sei Ω eine nicht–leere endliche Menge, $\#(A)$ bezeichne die Anzahl der Elemente einer Teilmenge $A \in \mathfrak{P}(\Omega)$. Dann wird vermöge

$$P(A) = \frac{\#(A)}{\#(\Omega)}, \quad A \in \mathfrak{P}(\Omega), \tag{1.1.1}$$

eine Wahrscheinlichkeitsverteilung P auf $\mathfrak{P}(\Omega)$ definiert, welche normiert und (σ–) additiv ist. P heißt auch diskrete Gleichverteilung (oder Laplace–Verteilung) über Ω, in Zeichen: $P = \mathfrak{L}(\Omega)$.

In praktischen Beispielen beschreibt Ω die Menge aller "möglichen" und A die Menge aller "günstigen" Ergebnisse.

Beispiel 1.1.1. (binäre Suche — `binary search`)
In dem eingangs erwähnten Beispiel der binären Suche könnte man z.B. als Grundmenge die Menge $\Omega = \{0, 1, 2, \ldots, 2^n - 1\}$ wählen, wobei $\omega \in \Omega$, $\omega \geq 1$, die mögliche Platznummer des Schlüsselelements bezeichnet. $\omega = 0$ bedeutet, daß das gesuchte Element nicht im Feld vorhanden ist.
Das Ereignis A_1, daß das Schlüsselelement in genau einem Schritt gefunden wird, läßt sich darstellen als $A_1 = \{2^{n-1}\}$, da das Schlüsselelement bei `binary search` in einem geordneten Feld genau dann im ersten Schritt gefunden wird, wenn es sich in der Mitte des Feldes, also auf dem Platz Nummer 2^{n-1} befindet. Entsprechend läßt sich das Ereignis A_2, daß das Schlüsselelement in genau zwei Schritten gefunden wird, darstellen als $A_2 = \{2^{n-2}, 3 \cdot 2^{n-2}\}$, da das Schlüsselelement genau dann in zwei Schritten gefunden wird, wenn es entweder in der Mitte des rechten oder in der Mitte des linken Restfeldes liegt. Allgemein läßt sich das Ereignis A_k, daß zum Auffinden des Schlüsselelementes k Schritte nötig sind, darstellen als

$$A_k = \{(2j-1) \cdot 2^{n-k} \mid 1 \leq j \leq 2^{k-1}\} \quad \text{für } 1 \leq k \leq n\,. \tag{1.1.2}$$

Es ist also $\#(A_k) = 2^{k-1}$ und damit

$$P(A_k) = \frac{2^{k-1}}{2^n} \quad \text{für } 1 \leq k \leq n \tag{1.1.3}$$

im Laplaceschen Sinn. ∎

Man beachte jedoch, daß hierbei stillschweigend eine Gleichverteilungsannahme getroffen wurde insofern, als alle möglichen Platznummern $\omega \in \Omega$, $\omega \geq 1$, und das Ereignis $A_0 = \{0\}$, daß das Schlüsselelement nicht in dem Feld vorhanden ist, mit derselben Wahrscheinlichkeit $1/\#(\Omega) = 1/2^n$ bewertet wurden.

Man kann sich natürlich auch für die Wahrscheinlichkeit zusammengesetzter Ereignisse interessieren, etwa für das Ereignis B_k, daß das Schlüsselelement in höchstens $k \leq n$ Schritten gefunden wird, d.h.

$$B_k = \bigcup_{i=1}^{k} A_i, \text{ mit } P(B_k) = \sum_{i=1}^{k} P(A_i) = \frac{2^k - 1}{2^n}, \quad 1 \leq k \leq n\,. \tag{1.1.4}$$

Man erkennt hieran, daß die Wahrscheinlichkeit dafür, das Schlüsselelement in weniger als n Schritten (dem ungünstigsten Fall) zu finden, $P(B_{n-1}) = (2^{n-1} - 1)/2^n$, kleiner als 1/2 ist! Dies erklärt, warum das durchschnittliche Verhalten des Algorithmus praktisch kaum vom schlechtesten Fall abweicht; für eine ausführlichere Diskussion hierzu sei auf Knuth (1973), Bd. 3, S. 410ff. verwiesen. Natürlich ist `binary search` schon im schlechtesten Fall sehr effizient.

Wir wollen an dieser Stelle gleich noch einen weiteren Wahrscheinlichkeitsbegriff erwähnen, der ebenfalls im Laplace'schen Sinn definiert werden kann, den Begriff der bedingten Wahrscheinlichkeit. Weiß man z.B., daß das Schlüsselelement in dem Feld vorhanden ist, benötigt man zur Analyse des Verfahrens lediglich die kleinere Teilmenge $B = \{1, 2, \ldots, 2^n - 1\}$. Die bedingte Wahrscheinlichkeit dafür, daß das Schlüsselelement dann in $k \leq n$ Schritten gefunden wird, muß sich demnach als Verhältnis der "günstigen" Fälle in Bezug auf die nunmehr möglichen Fälle darstellen. Dies gibt Anlaß zu folgender Definition.

Definition 1.1.2. *(Laplace'scher bedingter Wahrscheinlichkeitsbegriff)*
Es sei Ω eine nicht–leere endliche Menge und B eine nicht–leere Teilmenge von Ω. Dann wird auf $\mathfrak{P}(\Omega)$ durch

$$P(A \mid B) = \frac{\#(A \cap B)}{\#(B)}, \quad A \in \mathfrak{P}(\Omega), \tag{1.1.5}$$

eine (weitere) Wahrscheinlichkeitsverteilung $P(\cdot \mid B)$ definiert, welche als die elementare bedingte Wahrscheinlichkeitsverteilung unter (der Hypothese) B bezeichnet wird.

Wie man durch den Vergleich der Beziehungen (1.1.1) und (1.1.5) sofort feststellt, läßt sich die gerade definierte bedingte Wahrscheinlichkeit auch ausdrücken als

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathfrak{P}(\Omega), \tag{1.1.6}$$

was für spätere Zwecke die nützlichere Schreibweise ist.

Für das obige Suchbeispiel erhält man somit als bedingte Wahrscheinlichkeit dafür, daß das Schlüsselelement im Falle des Vorhandenseins in k Schritten gefunden wird, den Ausdruck

$$P(A_k \mid B) = \frac{2^{k-1}}{2^n - 1}, \quad 1 \leq k \leq n, \tag{1.1.7}$$

wie sich unmittelbar aus (1.1.3) und (1.1.5) ergibt.

Zur Berechnung Laplace'scher Wahrscheinlichkeiten ist es offensichtlich von Vorteil, systematische Methoden zur Bestimmung der Anzahl von Elementen bestimmter Mengen zur Verfügung zu haben. Solche Methoden stellt die *Kombinatorik* bereit. Wir werden im folgenden einige für die Stochastik wichtige Begriffe aus diesem Bereich behandeln.

Definition 1.1.3. *(Permutationen und Kombinationen)*
Es sei $\Omega = \{\omega_1, \ldots, \omega_n\}$ eine nicht–leere Menge, also $\#(\Omega) = n$, $n \in \mathbb{N}$. Für $1 \leq k \leq n$ definieren wir:

a) *Jedes Element $(\eta_1, \eta_2, \ldots, \eta_k) \in \Omega^k$ (k–faches kartesisches Produkt) heißt (k,n)–Permutation aus Ω (mit Wiederholung). Die Menge aller solcher Permutationen wird mit $\mathrm{Perm}_k^n(\Omega; m.W.)$ bezeichnet.*

b) *Jede (k,n)–Permutation $(\eta_1, \eta_2, \ldots, \eta_k)$ mit paarweise verschiedenen Komponenten heißt (k,n)–Permutation aus Ω ohne Wiederholung. Die Menge dieser Permutationen wird mit $\mathrm{Perm}_k^n(\Omega; o.W.)$ bezeichnet.*

c) *Jede (k,n)–Permutation $(\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k})$ mit $1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq n$ heißt (k,n)–Kombination aus Ω (mit Wiederholung). Die Menge dieser Kombinationen wird mit $\mathrm{Komb}_k^n(\Omega; m.W.)$ bezeichnet.*

d) *Jede (k,n)–Kombination mit paarweise verschiedenen Komponenten heißt (k,n)–Kombination aus Ω ohne Wiederholung. Die Menge dieser Kombinationen wird mit $\mathrm{Komb}_k^n(\Omega; o.W.)$ bezeichnet.*

Über die Mächtigkeit von Permutations- und Kombinationsmengen gibt das folgende Resultat Auskunft.

Lemma 1.1.1. *Mit den Bezeichnungen von Definition 1.1.3 gilt:*

$$\#\big(\mathrm{Perm}_k^n(\Omega; m.W.)\big) = n^k \tag{1.1.8}$$

$$\#\big(\mathrm{Perm}_k^n(\Omega; o.W.)\big) = (n)_k = \binom{n}{k} k! = n(n-1)\cdots(n-k+1) \tag{1.1.9}$$

$$\#\big(\mathrm{Komb}_k^n(\Omega; m.W.)\big) = \binom{n+k-1}{k} \tag{1.1.10}$$

$$\#\big(\mathrm{Komb}_k^n(\Omega; o.W.)\big) = \binom{n}{k}. \tag{1.1.11}$$

Beweis. Für jede Komponente η_i einer (k,n)–Permutation mit Wiederholung $(\eta_1, \ldots, \eta_k)$ gibt es jeweils $n = \#(\Omega)$ Wahlmöglichkeiten, woraus sich sofort (1.1.8) ergibt. Für eine entsprechende Permutation ohne Wiederholung gibt es n Wahlmöglichkeiten für die erste Komponente η_1, für die zweite Komponente η_2 gibt es

$(n-1)$ Wahlmöglichkeiten, usw. bis schließlich $(n-k+1)$ Wahlmöglichkeiten für die letzte Komponente η_k. Dies ergibt (1.1.9). Zum Beweis von (1.1.11) beachte man, daß jede (k,n)-Kombination $(\omega_{i_1},\ldots,\omega_{i_k})$ ohne Wiederholung genau $k!$ verschiedene Permutationen ohne Wiederholung mit denselben Komponenten erzeugt. Die angegebene Formel ergibt sich aus (1.1.9). Beziehung (1.1.10) folgt aus (1.1.11), wenn man die Menge Ω um weitere $(k-1)$ Elemente $\omega_{n+1},\ldots,\omega_{n+k-1}$ zu einer Menge Ω^* ergänzt und die Bijektion $(\omega_{i_1},\omega_{i_2},\ldots,\omega_{i_k}) \mapsto (\omega_{i_1},\omega_{i_2+1},\ldots,\omega_{i_k+k-1})$ von $Komb_k^n(\Omega;m.W.)$ auf $Komb_k^{n-k+1}(\Omega^*;o.W.)$ betrachtet. ∎

Beispiel 1.1.2. (Hashing)
Kombinatorische Fragestellungen der obigen Art treten in der Informatik z.B. im Zusammenhang mit dem sogenannten Hashing auf; hierbei handelt es sich um das Problem, Teilmengen eines Universums mit Hilfe geeigneter Tafeln (sog. Hashtafeln) abzuspeichern oder auch auf derartig abgespeicherte Mengen zuzugreifen (etwa um Einfügungen oder Streichungen durchzuführen). Bei "rein zufälligem" Ablegen von k Daten in einem Feld der Länge n $(k \le n)$ können dabei Mehrfachbelegungen desselben Speicherplatzes vorkommen (Kollision). Bezeichnen wir dieses Ereignis mit $A_{k,n}$, so läßt sich das komplementäre Ereignis $A_{k,n}^c$ offensichtlich mit der Klasse $Perm_k^n(\Omega;o.W.)$ identifizieren, wobei Ω die Menge der zur Verfügung stehenden Speicherplätze repräsentiert und die Komponente ω_i einer entsprechenden (k,n)-Permutation mit Wiederholung anzeigt, auf welchem Speicherplatz das Datenelement i abgelegt wird. Gemäß Definition 1.1.1 gilt demnach

$$\begin{aligned} P(A_{k,n}^c) &= \frac{\#\big(Perm_k^n(\Omega;o.W.)\big)}{\#\big(Perm_k^n(\Omega;m.W.)\big)} = \frac{(n)_k}{n^k} = \prod_{i=0}^{k-1}\Big(1-\frac{i}{n}\Big) \\ &= \exp\Big(\sum_{i=0}^{k-1}\ln(1-\frac{i}{n})\Big) \le \exp\Big(-\sum_{i=0}^{k-1}\frac{i}{n}\Big) \\ &= \exp\Big(-\frac{(k-1)k}{2n}\Big). \end{aligned} \tag{1.1.12}$$

Hierbei wurde die Ungleichung $\ln(1-x) \le -x$, $x<1$, benutzt. Wählt man für vorgegebenes $p \in (0,1)$ k in Abhängigkeit von p und n, so erhält man z.B.

$$P(A_{k,n}^c) \sim \exp\big(\frac{-k^2}{2n}\big) \sim p, \quad \text{falls } k \sim \sqrt{2\cdot n\cdot|\ln p|} \quad \text{mit } n\to\infty \tag{1.1.13}$$

aufgrund der Taylor–Entwicklung der ln–Funktion. ∎

Eine weitergehende Diskussion zu diesem Problemkreis findet man etwa in Mehlhorn (1988), S. 138.

Für $n = 365$ ist das obige Problem auch als "Geburtstags–Paradoxon" bekannt, da nach (1.1.12) die Wahrscheinlichkeit dafür, daß unter bereits nur 23 zufällig ausgesuchten Personen wenigstens zwei am selben Tag Geburtstag haben, schon größer als 1/2 ist. (Zum Vergleich: der asymptotische Ausdruck auf der rechten Seite von (1.1.13) ergibt $k \approx 22.49$, also eine bereits recht gute Näherung.)

Beispiel 1.1.3. (Suchbäume)
Kombinationen treten häufig auch im Zusammenhang mit der Auswahl von Teilmengen der Grundmenge Ω auf, z.B. bei digitalen Suchbäumen. Dabei werden die Elemente einer Datei durch Folgen von Ziffern dargestellt vermöge eines Baumes, dessen Knoten die Präfixe der betreffenden Elemente repräsentieren. Das Verhalten der Zugriffszeit auf solche Dateien hängt dabei wesentlich von der Größe der abgespeicherten Menge ab. Benutzt man etwa die Tatsache, daß jede k-elementige Teilmenge von Ω mit einer (k,n)-Kombination ohne Wiederholung identifiziert werden kann, so gibt es $\binom{n}{k}$ verschiedene k-elementige Teilmengen von Ω, $1 \leq k \leq n$. Greift man nun aus einer n-elementigen Grundmenge Ω "zufällig" eine k-elementige Teilmenge heraus ($1 \leq k \leq n$), so läßt sich die Wahrscheinlichkeit für das Ereignis $A_{1,n}$, daß die ausgesuchte Teilmenge ein bestimmtes Schlüsselelement ω_1 enthält, berechnen zu

$$P(A_{1,n}) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}, \tag{1.1.14}$$

wobei formal

$$A_{1,n} = \{(\omega_1, \eta) \mid \eta \in Komb_{k-1}^{n-1}(\Omega \setminus \{\omega_1\}; o.W.)\} \tag{1.1.15}$$

gewählt werden kann und die Bezugsmenge zur Berechnung von $P(A_1)$ nach Definition 1.1.1 die ganze Klasse $Komb_k^n(\Omega; o.W.)$ ist. (Für $k = 1$ ist (1.1.15) zu lesen als $A_{1,n} = \{(\omega_1)\}$.)
Entsprechend erhält man als Wahrscheinlichkeit für das Ereignis $A_{r,n}$, daß in der ausgewählten Teilmenge die Schlüsselelemente $\omega_1, \ldots, \omega_r$ enthalten sind,

$$P(A_{r,n}) = \frac{\binom{n-r}{k-r}}{\binom{n}{k}} = \frac{(k)_r}{(n)_r}, \quad 1 \leq r \leq k \leq n,$$

mit

$$A_{r,n} = \{(\omega_1, \ldots, \omega_r, \eta) \mid \eta \in Komb_{k-r}^{n-r}(\Omega \setminus \{\omega_1, \ldots, \omega_r\}; o.W.)\}.$$

Läßt man r fest aber k (in Abhängigkeit von n) und n so gegen ∞ streben, daß das Verhältnis k/n gegen eine Zahl $p \in (0,1)$ strebt, so gilt offensichtlich

$$P(A_{r,n}) \sim p^r \quad (n \to \infty).$$

Will man also mit mehr als 50%iger Wahrscheinlichkeit alle r Schlüsselelemente in der ausgewählten Teilmenge vorfinden, so muß man den Auswahlumfang asymptotisch mindestens so groß wie $k \sim n \cdot 2^{-1/r}$ wählen. ∎

Für eine weiterführende Diskussion im Zusammenhang mit digitalen Suchbäumen sei auf Mehlhorn (1988), S. 99ff. verwiesen.

Die in den obigen Beispielen auftretenden Grenzwahrscheinlichkeiten lassen sich offensichtlich nicht unmittelbar aus einfachen Laplace-Experimenten mit endlichen Grundmengen gewinnen. Dies ist evident z.B. für den Fall, daß p eine irrationale Zahl ist. Andererseits vereinfachen sie in sinnvoller Weise die komplizierten kombinatorischen Ausdrücke für die Laplace-Wahrscheinlichkeiten aus Definition 1.1.1, so daß eine modellmäßige Erfassung auch solcher Grenzsituationen wünschenswert ist.

In der folgenden allgemeineren Vorgehensweise wird daher der Laplace'sche Wahrscheinlichkeitsbegriff zunächst auf abzählbare Grundmengen Ω erweitert.

Definition 1.1.4. *(diskrete Wahrscheinlichkeitsverteilung)*
Es sei Ω eine nicht–leere, abzählbare Menge. Eine Abbildung $P\colon \mathfrak{P}(\Omega) \to [0,1]$ mit den Eigenschaften

$$P(\Omega) = 1 \tag{1.1.16}$$

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \tag{1.1.17}$$

für jede Familie $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathfrak{P}(\Omega)$ paarweise disjunkter (p.d.) Mengen heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitsmaß über Ω.

Die Eigenschaft (1.1.17) ist dabei gerade die σ–Additivität der Wahrscheinlichkeitsverteilung P.

Offensichtlich genügt der Laplace'sche Wahrscheinlichkeitsbegriff den Bedingungen (1.1.16) und (1.1.17), so daß Definition 1.1.4 in der Tat eine Erweiterung des Laplace'schen Begriffs vornimmt.

Eine Wahrscheinlichkeitsverteilung P nach Definition 1.1.4 erzeugt für Ereignisse $B \in \mathfrak{P}(\Omega)$ mit $P(B) > 0$ in der folgenden Weise bedingte Verteilungen über Ω.

Definition 1.1.5. *(elementare bedingte Wahrscheinlichkeit)*
Unter den Voraussetzungen von Definition 1.1.4 wird für jedes Ereignis $B \in \mathfrak{P}(\Omega)$ mit $P(B) > 0$ durch

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathfrak{P}(\Omega), \tag{1.1.18}$$

eine Wahrscheinlichkeitsverteilung $P(\cdot \mid B)$ über Ω definiert, die bedingte Verteilung unter (der Hypothese) B.
$P(A \mid B)$ heißt elementare bedingte Wahrscheinlichkeit von A unter (der Hypothese) B.

Die beiden letzteren Definitionen werfen die Frage auf, wie Wahrscheinlichkeitsverteilungen auf abzählbaren Grundmengen beschrieben werden können, ohne die Wahrscheinlichkeit für jedes einzelne Ereignis $A \in \mathfrak{P}(\Omega)$ angeben zu müssen. Hierauf gibt das folgende Resultat Antwort.

Lemma 1.1.2. *(Darstellung diskreter Verteilungen)*
Es sei Ω eine nicht–leere abzählbare Menge und P eine Wahrscheinlichkeitsverteilung über Ω. Dann ist P bereits vollständig durch die Elementarwahrscheinlichkeiten $P(\{\omega\})$ festgelegt vermöge

$$P(A) = \sum_{\omega\in A} P(\{\omega\}), \quad A \in \mathfrak{P}(\Omega). \tag{1.1.19}$$

Beweis. Es ist $A = \bigcup_{\omega\in A}\{\omega\}$, so daß sich die Aussage aus der σ–Additivität (1.1.17) von P ergibt. ∎

Für bedingte Wahrscheinlichkeiten gilt entsprechend

$$P(A \mid B) = \frac{\sum_{\omega \in A \cap B} P(\{\omega\})}{\sum_{\omega \in B} P(\{\omega\})} = \sum_{\omega \in A} P(\{\omega\} \mid B) \tag{1.1.20}$$

für $A, B \in \mathfrak{P}(\Omega)$ mit $P(B) > 0$.

Umgekehrt läßt sich durch Festlegung von Elementarwahrscheinlichkeiten p_ω für $\omega \in \Omega$ mit $p_\omega \geq 0$, $\sum_{\omega \in \Omega} p_\omega = 1$ durch $P(A) = \sum_{\omega \in A} p_\omega$ jede beliebige Verteilung über Ω gewinnen.

Leider reichen zur Beschreibung vieler stochastischer Vorgänge selbst abzählbar–unendliche Grundmengen Ω nicht aus, wie wir im folgenden einsehen werden. Wir benötigen hierzu den zentralen Begriff der stochastischen Unabhängigkeit von Ereignissen.

Definition 1.1.6. *(stochastische Unabhängigkeit von Ereignissen)*
Es seien $A_1, \ldots, A_n \in \mathfrak{P}(\Omega)$, $n \in \mathbb{N}$, Ereignisse und P eine Wahrscheinlichkeitsverteilung. $A_1, \ldots, A_n$ heißen stochastisch unabhängig, wenn für alle Mengen $B_i \in \{A_i, A_i^c\}$, $1 \leq i \leq n$, gilt

$$P(\bigcap_{i=1}^{n} B_i) = \prod_{i=1}^{n} P(B_i). \tag{1.1.21}$$

Eine Familie von Ereignissen $\{A_n\}_{n \in \mathbb{N}}$ heißt stochastisch unabhängig, wenn die Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig sind für alle $n \in \mathbb{N}$.

Wie einfache Rechnungen zeigen, ist (1.1.21) für $n = 2$ äquivalent zu

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \tag{1.1.22}$$

und falls $P(A_2) > 0$ bzw. $P(A_1) > 0$ ist, auch äquivalent zu

$$P(A_1 \mid A_2) = P(A_1) \text{ bzw. } P(A_2 \mid A_1) = P(A_2). \tag{1.1.23}$$

Die letztere Beziehung besagt, daß die Wahrscheinlichkeit für das Eintreten des Ereignisses A_2 nicht von dem Eintreten des Ereignisses A_1 beeinflußt wird und umgekehrt, daß die Ereignisse in diesem Sinn also unabhängig voneinander eintreten.

Allgemeiner folgt bei beliebigem $n \in \mathbb{N}$ aus (1.1.21), daß für alle Auswahlen $1 \leq i_1 < \ldots < i_k \leq n$, $k \leq n$, mit $P(B_{i_1} \cap \ldots \cap B_{i_k}) > 0$ und $1 \leq l \leq n$, $l \notin \{i_1, \ldots, i_k\}$

$$P(A_l \mid B_{i_1} \cap \ldots \cap B_{i_k}) = P(A_l) \tag{1.1.24}$$

gilt. Dies kann entsprechend interpretiert werden: die Wahrscheinlichkeit für das Eintreten eines Ereignisses wird weder von dem Eintreten noch dem Nichteintreten der anderen Ereignisse beeinflußt.

Insbesondere Folgen von stochastisch unabhängigen Ereignissen spielen in der Stochastik bei der Modellbildung eine zentrale Rolle; sie werden z.B. in der Simulation im Zusammenhang mit sogenannten Verwerfungsverfahren benötigt, die in Kapitel 6 behandelt werden.

Das folgende Resultat zeigt, daß solche Strukturen allerdings in wichtigen Fällen nicht mehr mit Hilfe abzählbarer Grundmengen Ω beschrieben werden können.

Satz 1.1.1. *(Folgen unabhängiger Ereignisse)*
Es sei Ω eine nicht–leere, abzählbare Menge, P eine Wahrscheinlichkeitsverteilung über Ω und $p \in (0,1)$ beliebig. Dann existiert keine Folge stochastisch unabhängiger Ereignisse $\{A_n\}_{n\in\mathbf{N}} \subseteq \mathfrak{P}(\Omega)$ mit $P(A_n) = p$ für alle $n \in \mathbf{N}$.

Einen Beweis dieses Satzes werden wir später führen, wenn Rechenregeln für Wahrscheinlichkeiten von Grenzwerten von Ereignisfolgen bereitgestellt sind.

In Abschnitt 1.4 wird gezeigt, daß auf gewissen überabzählbaren Grundmengen Ω mit geeigneten σ–Algebren Folgen stochastisch unabhängiger Ereignisse (z.B. mit gleicher Wahrscheinlichkeit p, $0 < p < 1$) existieren.

Zu überabzählbaren Grundmengen Ω wird man außerdem in natürlicher Weise geführt, wenn man Zufallsexperimente betrachtet, deren Ausgang reelle Zahlen sind. Beispiele hierfür sind zufällige Wartezeiten, etwa eines Programms in der Warteschlange, oder zufällig schwankende Schichtdicken bei der Chipherstellung. Jede reelle Zahl innerhalb eines bestimmten Intervalls kann hierbei als möglicher Wert auftreten, und das sind bekanntlich überabzählbar viele.

Um eine genügend reichhaltige Theorie entwickeln zu können, muß man also auch überabzählbare Grundmengen Ω in die Betrachtungen mit einbeziehen. Will man den bisher eingeschlagenen Weg naiv fortsetzen und Wahrscheinlichkeitsmaße auf der ganzen Potenzmenge von Ω definieren, stößt man bald auf Schwierigkeiten. Erinnert sei an die Nichtexistenz einer Gleichverteilung, charakterisiert durch Bedingung (1.0.6) bzw. (1.0.7), auf der gesamten Potenzmenge $\mathfrak{P}(\Omega)$ mit $\Omega = (0,1]$.

Ein Ausweg aus dieser Situation ist, den Definitionsbereich für die Wahrscheinlichkeitsmaße P einzuschränken, und zwar so, daß genügend viele Wahrscheinlichkeitsverteilungen mit zusätzlichen, charakteristischen Eigenschaften existieren, andererseits aber nicht zu viele interessante Ereignisse verlorengehen. Eine Möglichkeit zur Bewältigung dieser Problematik bietet der maßtheoretische Zugang, wie er im wesentlichen von Kolmogoroff seit 1933 entwickelt wurde. Dieser basiert auf dem Begriff von σ–Algebren von Ereignissen, den wir nun ausführlicher behandeln wollen.

Definition 1.1.7. *(σ–Algebra von Ereignissen)*
Es sei Ω eine nicht–leere Menge und $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathcal{A}$ heißt σ–Algebra (von Ereignissen) über Ω, wenn gilt:

$$\Omega \in \mathcal{A} \tag{1.1.25}$$

$$A \in \mathcal{A} \implies A^c \in \mathcal{A} \tag{1.1.26}$$

$$\{A_n\}_{n\in\mathbf{N}} \subseteq \mathcal{A} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}. \tag{1.1.27}$$

Eine σ–Algebra von Ereignissen ist also abgeschlossen gegenüber der Bildung von Komplementen und abzählbaren Vereinigungsbildungen. Aufgrund der De–Morgan–Regeln und (1.1.26) erhält man sofort auch die Abgeschlossenheit gegenüber abzählbaren Durchschnittsbildungen:

$$\{A_n\}_{n\in\mathbf{N}} \subseteq \mathcal{A} \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}. \tag{1.1.28}$$

Damit erlaubt eine σ–Algebra jedenfalls die gewünschten logischen Verknüpfungen von Ereignissen mittels "*und*", "*oder*" und "*nicht*". Insbesondere ist $\mathfrak{P}(\Omega)$ stets eine σ–Algebra über Ω, und zwar die feinste, d.h. die größte in Bezug auf Kardinalität. $\mathcal{A} = \{\emptyset, \Omega\}$ ist offensichtlich die gröbste σ–Algebra, d.h. die kleinste bezüglich Kardinalität.

Wir erweitern nun noch einmal den Begriff einer (bedingten) Wahrscheinlichkeitsverteilung, welcher auch die Behandlung überabzählbarer Grundmengen Ω erlaubt.

Definition 1.1.8. *(allgemeine Wahrscheinlichkeitsverteilung)*
Es sei $\mathcal{A}$ eine σ–Algebra von Ereignissen über einer nicht–leeren Grundmenge Ω. Eine Abbildung $P\colon \mathcal{A} \to [0,1]$ mit den Eigenschaften

$$P(\Omega) = 1 \tag{1.1.29}$$

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \tag{1.1.30}$$

für jede Familie $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ paarweise disjunkter (p.d.) Mengen heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitsmaß auf $\mathcal{A}$. Das Tripel $(\Omega, \mathcal{A}, P)$ heißt Wahrscheinlichkeitsraum.

Man beachte, daß die ein Wahrscheinlichkeitsmaß definierenden Bedingungen (1.1.29) und (1.1.30) mit (1.1.16) und (1.1.17) übereinstimmen. Definition 1.1.8 ist eine Erweiterung von Definition 1.1.4 in dem Sinn, daß als Definitionsbereich für P eine beliebige σ–Algebra — nicht nur die Potenzmenge $\mathfrak{P}(\Omega)$ — zugelassen ist. Die Abzählbarkeit der Grundmenge Ω spielt jetzt keine Rolle mehr. Alle Aussagen, die wir im folgenden für Wahrscheinlichkeitsräume $(\Omega, \mathcal{A}, P)$ erhalten, gelten natürlich auch für die speziellere Situation $\mathcal{A} = \mathfrak{P}(\Omega)$ aus Definition 1.1.4.

Aufgrund der σ–Additivität (1.1.30) erhält man

$$P(\emptyset) = 0,$$

indem man $A_1 = \Omega$ und $A_n = \emptyset$ für alle $n \in \mathbb{N}$, $n \geq 2$, wählt. Diese Eigenschaft gilt mit obiger Bemerkung auch für den diskreten Wahrscheinlichkeitsbegriff aus Definition 1.1.4.

In Analogie zu Definition 1.1.5 erhält man entsprechend:

Definition 1.1.9. *(elementare bedingte Wahrscheinlichkeitsverteilung)*
Unter den Voraussetzungen von Definition 1.1.8 wird für jedes $B \in \mathcal{A}$ mit $P(B) > 0$ durch

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{A}, \tag{1.1.31}$$

eine Wahrscheinlichkeitsverteilung $P(\cdot \mid B)$ auf $\mathcal{A}$ definiert, die bedingte Verteilung unter (der Hypothese) B.
$P(A \mid B)$ heißt elementare bedingte Wahrscheinlichkeit von A unter (der Hypothese) B.

Definition 1.1.6 für stochastische Unabhängigkeit von Ereignissen mit den nachfolgenden Aussagen (1.1.22) bis (1.1.24) überträgt sich unmittelbar auf allgemeine Wahrscheinlichkeitsverteilungen. Es ist lediglich zu beachten, daß in (1.1.21)

die Ereignisse $A_1, \ldots, A_n$ jetzt Elemente einer σ–Algebra $\mathcal{A}$ sind, ein eventuell kleineres Mengensystem als $\mathfrak{P}(\Omega)$.

Wir werden nun zeigen, daß die Normierungsbedingung (1.1.29) zusammen mit der σ–Additivität (1.1.30) plausible Rechenregeln für Wahrscheinlichkeiten wie z.B. (1.0.3) oder (1.0.4) implizieren. Darüberhinaus erhält man jedoch auch Eigenschaften von Wahrscheinlichkeitsverteilungen im Zusammenhang mit unendlichen Verknüpfungen von Ereignissen. Solche Grenzwerte von Ereignisfolgen werden in der folgenden Definition betrachtet.

Definition 1.1.10. *(Limites von Ereignisfolgen)*
Es sei $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Ereignissen aus einer σ–Algebra $\mathcal{A}$ über einer nichtleeren Grundmenge Ω. Gilt für alle $n \in \mathbb{N}$

$$A_n \subseteq A_{n+1},$$

so heißt die Folge monoton wachsend. Gilt

$$A_{n+1} \subseteq A_n,$$

so heißt die Folge monoton fallend. Für monoton wachsende bzw. monoton fallende Ereignisfolgen heißt jeweils

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \quad \textit{bzw.}$$

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

der Limes der Ereignisfolgen $\{A_n\}$. Für beliebige (nicht notwendig monotone) Ereignisfolgen $\{A_n\}$ heißen

$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty} \Big(\bigcup_{k=n}^{\infty} A_k \Big) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \textit{bzw.} \tag{1.1.32}$$

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty} \Big(\bigcap_{k=n}^{\infty} A_k \Big) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \tag{1.1.33}$$

Limes superior bzw. Limes inferior der Ereignisfolge $\{A_n\}$.

Der Limes superior einer Ereignisfolge $\{A_n\}$ ist gerade die Menge der $\omega \in \Omega$, die in unendlich vielen der A_n liegen; er beschreibt, daß unendlich viele der Ereignisse A_n eintreten. Der Limes inferior enthält alle $\omega \in \Omega$, die in fast allen der A_n (d.h. allen bis auf endlich viele Ausnahmen) liegen; anschaulich ist das das Ereignis, daß fast alle der Ereignisse A_n eintreten. Wegen (1.1.27) und (1.1.28) gehören lim sup und lim inf von Ereignisfolgen $\{A_n\}$ mit $A_n \in \mathcal{A}$ ebenfalls zur entsprechenden σ–Algebra $\mathcal{A}$.

Lemma 1.1.3. *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $A, B, \{A_n\}_{n\in\mathbb{N}}$ seien Ereignisse in $\mathcal{A}$. Dann gilt:*

$$P(A^c) = 1 - P(A) \tag{1.1.34}$$

$$P(A \cup B) = P(A) + P(B), \quad \text{falls } A \cap B = \emptyset \tag{1.1.35}$$

$$P(A) \leq P(B), \quad \text{falls } A \subseteq B \quad (\textit{Monotonie von } P) \tag{1.1.36}$$

$$P(B \setminus A) = P(B) - P(A), \quad \text{falls } A \subseteq B \quad (\textit{Subtraktivität von } P) \tag{1.1.37}$$

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n), \tag{1.1.38}$$

falls $\{A_n\}$ monoton wachsend (Stetigkeit von P von unten)

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n), \tag{1.1.39}$$

falls $\{A_n\}$ monoton fallend (Stetigkeit von P von oben)

$$P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty} A_k\Big) \tag{1.1.40}$$

$$P(\liminf_{n\to\infty} A_n) = \lim_{n\to\infty} P\Big(\bigcap_{k=n}^{\infty} A_k\Big) \tag{1.1.41}$$

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} P(A_n) \quad (\textit{Subadditivität von } P) \tag{1.1.42}$$

$$\begin{aligned} P\Big(\bigcup_{k=1}^{n} A_k\Big) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{(i_1,\ldots,i_k)\in \mathrm{Komb}_k^n(\{1,\ldots,n\};o.W.)} P\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) \\ &= \sum_{k=1}^{n} P(A_k) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n} A_k\Big) \end{aligned} \tag{1.1.43}$$

(Siebformel von Poincaré – Sylvester)

$$\sum_{k=1}^{n} P(A_k) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) \leq P\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \sum_{k=1}^{n} P(A_k). \tag{1.1.44}$$

(Bonferroni – Ungleichung)

Beweis. Beziehung (1.1.34) ergibt sich aus (1.1.35) mit $B = A^c$; letztere folgt aus der σ-Additivität (1.1.30) durch Wahl von $A_1 = A$, $A_2 = B$, $A_n = \emptyset$ für $n \geq 3$. Wegen der Nichtnegativität von P folgt die Monotonieeigenschaft (1.1.36) aus (1.1.37), welche sich wiederum aus (1.1.35) ergibt, wenn man dort B durch die Differenzmenge $B \setminus A$ ersetzt.

Zum Beweis von (1.1.38) bemerken wir, daß für die durch $B_1 = A_1$, $B_{n+1} = A_{n+1} \setminus A_n$, $n \in \mathbb{N}$, definierte Ereignisfolge gilt: $\{B_n\}$ ist paarweise disjunkt mit $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Wegen der σ–Additivität von P ist somit

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(B_n) = P(A_1) + \lim_{m\to\infty} \sum_{k=1}^{m} P(A_{k+1} \setminus A_k)$$

$$= P(A_1) + \lim_{m\to\infty} \sum_{k=1}^{m} \big(P(A_{k+1}) - P(A_k)\big) = \lim_{m\to\infty} P(A_m),$$

wobei im vorletzten Schritt die Subtraktivität (1.1.37) ausgenutzt wurde.

(1.1.39) folgt aus (1.1.38) durch Anwendung der De–Morgan–Regeln auf $\bigcap_{n=1}^{\infty} A_n$, d.h. $P\big(\bigcap_{n=1}^{\infty} A_n\big) = 1 - P\big(\bigcup_{n=1}^{\infty} A_n^c\big)$, wobei nun $\{A_n^c\}$ monoton wachsend ist.

Die Beziehungen (1.1.40) bzw. (1.1.41) ergeben sich unmittelbar aus (1.1.39) bzw. (1.1.38) nach Definition 1.1.10. Beziehung (1.1.42) kann aus der rechten Ungleichung in (1.1.44) gefolgert werden, indem man die rechte Seite durch $\sum_{k=1}^{\infty} P(A_k)$ abschätzt und beachtet, daß die Ereignisfolge $\big\{\bigcup_{k=1}^{n} A_k\big\}_{n\in\mathbb{N}}$ monoton wachsend ist, so daß wir mit der Stetigkeitseigenschaft (1.1.38) erhalten: $P\big(\bigcup_{n=1}^{\infty} A_n\big) = P\big(\lim_{n\to\infty} \bigcup_{k=1}^{n} A_k\big) = \lim_{n\to\infty} P\big(\bigcup_{k=1}^{n} A_k\big)$.

Die Bonferroni–Ungleichung (1.1.44) folgt ihrerseits aus der Siebformel (1.1.43), welche man z.B. mit vollständiger Induktion folgendermaßen beweisen kann:

Für $n = 1$ ist die Formel trivial gültig. Für $n = 2$ erhält man unter Ausnutzung der Subtraktivität (1.1.37):

$$\begin{aligned} P(A_1 \cup A_2) &= P\Big(\big((A_1 \cup A_2) \setminus (A_1 \cap A_2)\big) \cup (A_1 \cap A_2)\Big) \\ &= P\big((A_1 \cup A_2) \setminus (A_1 \cap A_2)\big) + P(A_1 \cap A_2) \\ &= P\big(A_1 \setminus (A_1 \cap A_2)\big) + P\big(A_2 \setminus (A_1 \cap A_2)\big) + P(A_1 \cap A_2) \\ &= P(A_1) + P(A_2) - 2 \cdot P(A_1 \cap A_2) + P(A_1 \cap A_2) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2). \end{aligned} \tag{1.1.45}$$

Dies ist aber gerade die Siebformel für $n = 2$.

Unter der Induktionsvoraussetzung, daß (1.1.43) für ein $n \in \mathbb{N}$ gilt, erhält man unter Verwendung von (1.1.45):

$$P\Big(\bigcup_{k=1}^{n+1} A_k\Big) = P\Big(\bigcup_{k=1}^{n} A_k\Big) + P(A_{n+1}) - P\Big(\bigcup_{k=1}^{n} (A_k \cap A_{n+1})\Big)$$

$$= \sum_{k=1}^{n+1} P(A_k) - \sum_{1\le i_1<i_2\le n} P(A_{i_1} \cap A_{i_2}) + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n} A_k\Big)$$

$$-\Bigg(\sum_{k=1}^{n} P(A_k \cap A_{n+1}) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}\cap A_{n+1} + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n+1} A_k\Big)\Bigg)$$

$$= \sum_{k=1}^{n+1} P(A_k) - \sum_{1\le i_1<i_2\le n+1} P(A_{i_1}\cap A_{i_2}) + - \cdots + (-1)^{n+2} P\Big(\bigcap_{k=1}^{n+1} A_k\Big),$$

d.h. die Siebformel für $n+1$. Hierbei wurde die Siebformel für n zusätzlich auf $P\Big(\bigcup_{k=1}^{n}(A_k \cap A_{n+1})\Big)$ angewandt. ∎

Weitere Bonferroni–Ungleichungen ergeben sich durch Abbruch der Siebformel nach Termen gerader bzw. ungerader Ordnung. Wegen des Alternierens der Reihe erhält man für $P\big(\bigcup_{k=1}^{n} A_k\big)$ eine obere Schranke bei Abbruch nach einem Summanden mit positivem Vorzeichen bzw. eine untere Schranke bei Abbruch nach einem Summanden mit negativem Vorzeichen.

Wir kommen nun zu dem angekündigten Beweis des Satzes 1.1.1. Man beachte, daß $\mathfrak{P}(\Omega)$ stets eine σ-Algebra ist. Folglich ist die in Satz 1.1.1 verwendete diskrete Wahrscheinlichkeitsverteilung eine Wahrscheinlichkeitsverteilung im Sinn von Definition 1.1.8 und die Aussagen von Lemma 1.1.3 können angewendet werden.

Sei etwa $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathfrak{P}(\Omega)$ eine Folge von stochastisch unabhängigen Ereignissen der abzählbaren Menge Ω mit $P(A_n) = p$, $0 < p < 1$. Für jedes $\omega \in \Omega$ definieren wir

$$B_n(\omega) = \begin{cases} A_n & \text{für } \omega \in A_n \\ A_n^c & \text{für } \omega \in A_n^c. \end{cases}$$

Es gilt $\omega \in B_n(\omega)$ für alle $n \in \mathbb{N}$, also auch $\omega \in \bigcap_{n=1}^{\infty} B_n(\omega)$. Wegen der Stetigkeit von P von oben (1.1.39), der Monotonie und der Unabhängigkeit folgt für alle $\omega \in \Omega$

$$P(\{\omega\}) \le P\Big(\bigcap_{k=1}^{\infty} B_k(\omega)\Big) = P\Big(\lim_{n\to\infty} \bigcap_{k=1}^{n} B_k(\omega)\Big) = \lim_{n\to\infty} P\Big(\bigcap_{k=1}^{n} B_k(\omega)\Big)$$

$$= \lim_{n\to\infty} \prod_{k=1}^{n} P(B_k(\omega)) \le \lim_{n\to\infty} q^n = 0,$$

wobei $q = \max\{p, 1-p\} < 1$ ist.
Wegen der Abzählbarkeit von Ω gelangt man damit zu folgendem Widerspruch

$$1 = P(\Omega) = P\Big(\bigcup_{\omega\in\Omega} \{\omega\}\Big) = \sum_{\omega\in\Omega} P(\{\omega\}) = 0,$$

woraus die Behauptung folgt. ∎

Der obige Beweis zeigt, daß Satz 1.1.1 auch dann noch gültig bleibt, wenn für die Ereignisse $\{A_n\}$ nur gefordert wird, daß mit $p_n = P(A_n)$, $n \in \mathbb{N}$, gilt: $\lim_{n\to\infty} \prod_{k=1}^{n} q_k = 0$ bzw. äquivalent dazu $\sum_{k=1}^{\infty}(1 - q_k) = \infty$, wobei wieder $q_k = \max\{p_k, 1 - p_k\}$ zu setzen ist.

Mit Hilfe der Eigenschaften (1.1.34) und (1.1.35) läßt sich die folgende zu (1.1.21) äquivalente Beziehung für stochastische Unabhängigkeit von Ereignisen $A_1, \ldots, A_n$ zeigen. Sie ist in vielen Fällen leichter zu überprüfen als die Definition, da die Wahrscheinlichkeiten der Komplemente nicht benötigt werden.

Satz 1.1.2. *(Stochastische Unabhängigkeit von Ereignissen)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. $A_1, \ldots, A_n \in \mathcal{A}$ seien Ereignisse, $n \in \mathbb{N}$. $A_1, \ldots, A_n$ sind genau dann stochastisch unabhängig, wenn für alle Auswahlen $1 \le i_1 < \ldots < i_k \le n$, $1 \le k \le n$, gilt

$$P\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) = \prod_{j=1}^{k} P(A_{i_j}). \tag{1.1.46}$$

Beweis. Wenn (1.1.21) für je n Mengen $B_1, \ldots, B_n$ gilt, so auch für je k Mengen $B_{i_1}, \ldots, B_{i_k}$, $1 \le i_1 < \ldots < i_k \le n$. Dies sieht man wie folgt. Für jedes $l \in \{1, \ldots n\}$ gilt

$$\bigcap_{\substack{i=1\\ i\neq l}}^{n} B_i = (B_l \cup B_l^c) \cap \Big(\bigcap_{\substack{i=1\\ i\neq l}}^{n} B_i\Big) = \Big(B_l \cap \bigcap_{\substack{i=1\\ i\neq l}}^{n} B_i\Big) \cup \Big(B_l^c \cap \bigcap_{\substack{i=1\\ i\neq l}}^{n} B_i\Big),$$

also nach Voraussetzung und (1.1.35)

$$\begin{aligned} P\Big(\bigcap_{\substack{i=1\\ i\neq l}}^{n} B_i\Big) &= P(B_l) \cdot \prod_{\substack{i=1\\ i\neq l}}^{n} P(B_i) + P(B_l^c) \cdot \prod_{\substack{i=1\\ i\neq l}}^{n} P(B_i) \\ &= \big(P(B_l) + P(B_l^c)\big) \prod_{\substack{i=1\\ i\neq l}}^{n} P(B_i) = \prod_{\substack{i=1\\ i\neq l}}^{n} P(B_i). \end{aligned}$$

(1.1.46) folgt hieraus mit Induktion, wobei $B_i = A_i$, $1 \le i \le n$, gesetzt wird. Umgekehrt können in (1.1.46) die Mengen A_{i_j} beliebig durch ihre Komplemente ersetzt werden. Ist $i_1 < \ldots < i_k$ eine Auswahl von Indizes und $l \in \{1, \ldots, k\}$, so gilt nämlich

$$P\Big(\bigcap_{\substack{j=1\\ j\neq l}}^{k} A_{i_j}\Big) = P\Big(A_{i_l} \cap \bigcap_{\substack{j=1\\ j\neq l}}^{k} A_{i_j}\Big) + P\Big(A_{i_l}^c \cap \bigcap_{\substack{j=1\\ j\neq l}}^{k} A_{i_j}\Big),$$

also wegen (1.1.46)

$$\begin{aligned} P\Big(A_{i_l}^c \cap \bigcap_{\substack{j=1\\ j\neq l}}^{k} A_{i_j}\Big) &= \prod_{\substack{j=1\\ j\neq l}}^{k} P(A_{i_j}) - \prod_{j=1}^{k} P(A_{i_j}) \\ &= \big(1 - P(A_{i_l})\big) \cdot \prod_{\substack{j=1\\ j\neq l}}^{k} P(A_{i_j}) = P(A_{i_l}^c) \cdot \prod_{\substack{j=1\\ j\neq l}}^{k} P(A_{i_j}). \end{aligned} \tag{1.1.47}$$

Durch iterierte Anwendung (Induktion) von (1.1.47) folgt jetzt (1.1.21), wobei $k = n$ und $i_j = j$, $1 \le j \le n$, gesetzt wird. ∎

Während mit Definition 1.1.8 zunächst nur Wahrscheinlichkeiten für Vereinigungen disjunkter Ereignisse berechnet werden können, erlaubt die Siebformel (1.1.43) auch die Berechnung von Wahrscheinlichkeiten beliebiger, endlicher Vereinigungen von Ereignissen, sofern die Wahrscheinlichkeiten aller Durchschnitte bekannt sind. Beispielsweise läßt sich das sogenannte Rencontre–Problem mit Hilfe der Siebformel bzw. der Bonferroni–Ungleichung elegant lösen. Eine Formulierung dieses Problems im Bereich der Informatik lautet etwa folgendermaßen:

Beispiel 1.1.4. (Sortierprobleme)
Gegeben sei ein Feld der Länge n. Wie groß ist die Wahrscheinlichkeit dafür, daß mindestens $k \le n$ Elemente des Feldes schon an der richtigen Stelle stehen, wenn die Elemente bezüglich eines ordinalen Merkmals sortiert werden sollen. Für alle $n!$ möglichen Anordnungen von Elementen wird hierbei dieselbe Wahrscheinlichkeit $1/n!$, also eine Gleichverteilung, vorausgesetzt. Solche Fragen spielen bei der Untersuchung der Effizienz von Sortierverfahren eine Rolle. (`qicksort` verhält sich z.B. bei bereits teilsortierten Eingabefolgen extrem schlecht (vgl. etwa Mehlhorn (1988), S. 51).)
Wir betrachten zunächst den Fall $k = 1$, d.h. die Wahrscheinlichkeit dafür, daß wenigstens ein Element bereits an der richtigen Stelle im Feld steht. Nehmen wir als Grundmenge $Perm_n^n(\{1,\dots,n\}; o.W.)$ und bezeichnen wir mit $A_\ell = \{\eta \in \Omega \mid \eta_\ell = \ell\}$, $1 \le \ell \le n$, die Menge aller Eingabefolgen η, deren ℓ–tes Element bereits richtig sortiert ist, so muß $P\big(\bigcup_{\ell=1}^n A_\ell\big)$ berechnet werden. Für beliebige Auswahlen $1 \le i_1 < i_2 < \dots < i_j \le n$, $1 \le j \le n$, ist $\bigcap_{l=1}^j A_{i_l} = \{\eta \in \Omega \mid \eta_{i_1} = i_1, \dots, \eta_{i_j} = i_j\}$, folglich gilt $\#\big(\bigcap_{l=1}^j A_{i_l}\big) = (n-j)!$ für jede solche Auswahl und somit nach (1.1.1)

$$P\Big(\bigcap_{l=1}^{j} A_{i_l}\Big) = \frac{(n-j)!}{n!} = \left(j! \cdot \binom{n}{j}\right)^{-1}, \quad 1 \le j \le n. \tag{1.1.48}$$

Die Summanden im l–ten Term der Siebformel (1.1.43) sind hiermit konstant; ihre Anzahl berechnet sich mit (1.1.11) zu $\binom{n}{l}$, so daß

$$\begin{aligned} P\Big(\bigcup_{\ell=1}^{n} A_\ell\Big) &= \sum_{\ell=1}^{n} P(A_\ell) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) + - \dots \\ &\qquad + (-1)^{n+1} P\Big(\bigcap_{\ell=1}^{n} A_\ell\Big) \\ &= \frac{n}{n} - \frac{\binom{n}{2}}{2!\binom{n}{2}} + \frac{\binom{n}{3}}{3!\binom{n}{3}} - + \dots + (-1)^{n+1} \frac{\binom{n}{n}}{n!\binom{n}{n}} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - + \dots + \frac{(-1)^{n+1}}{n!} \\ &= 1 - \Big(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + - \dots \frac{(-1)^n}{n!}\Big). \end{aligned} \tag{1.1.49}$$

Der Ausdruck in der Klammer ist die n–te Partialsumme der Taylorentwicklung von $e^{-1} = 1/e$, so daß folgt:

$$\lim_{n\to\infty} P\Big(\bigcup_{\ell=1}^{n} A_\ell\Big) = 1 - \frac{1}{e} = 0.6321\ldots\,.$$

Man erhält also die erstaunliche Aussage, daß die Wahrscheinlichkeit für richtige Vorsortierung wenigstens eines Feldelementes für große n praktisch unabhängig von n etwa 0.63 beträgt. ∎

Der Abbruch der Reihe $1 - \frac{1}{2!} + \frac{1}{3!} - + \ldots + \frac{(-1)^{n+1}}{n!}$ in (1.1.49) nach Gliedern mit negativem bzw. positivem Vorzeichen liefert Bonferroni – Ungleichungen, etwa zweiter und dritter Ordnung

$$\frac{1}{2} = 1 - \frac{1}{2} \le P(\bigcup_{\ell=1}^{n} A_\ell) \le 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}.$$

Wir betrachten nun allgemein den Fall, daß mindestens k, $2 \le k \le n$, Elemente des Feldes schon richtig sortiert sind. Dieses Ereignis wird beschrieben durch die Menge $\bigcup_{1\le i_1<i_2<\ldots<i_k\le n} \bigcap_{l=1}^{k} A_{i_l}$. Die Anwendung der oberen Bonferroni–Schranke (1.1.44) auf die Wahrscheinlichkeit dieses Ereignisses liefert mit (1.1.48) und (1.1.11) die Abschätzung

$$P\Big(\bigcup_{1\le i_1<i_2<\ldots<i_k\le n} \bigcap_{l=1}^{k} A_{i_l}\Big) \le \sum_{1\le i_1<i_2<\ldots<i_k\le n} P\Big(\bigcap_{l=1}^{k} A_{i_l}\Big) = \frac{\binom{n}{k}}{k!\binom{n}{k}} = \frac{1}{k!},$$

welche für den Fall $k = n$ sogar mit Gleichheit gilt. Für $k \ge 2$ ist die gesuchte Wahrscheinlichkeit damit kleiner oder gleich $1/2$, für $k \ge 3$ sogar kleiner oder gleich $1/6$. Sie fällt allgemein bei wachsendem k mindestens so schnell wie $1/k!$. Dies zeigt, daß bei Gleichverteilung aller Eingabefolgen die Wahrscheinlichkeit für bereits mindestens k vorsortierte Elemente rasch sehr klein wird.

Die Wahrscheinlichkeit, daß in einem Feld der Länge n bereits genau k, $k \le n$, Elemente richtig sortiert sind, ist durch den Ausdruck

$$p_{nk} = \frac{1}{k!}\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + - \ldots \frac{(-1)^{n-k}}{(n-k)!}\Big) \tag{1.1.50}$$

gegeben.

Dies sieht man wie folgt. Es gibt $\binom{n}{k}$ Möglichkeiten, das Gesamtfeld in eine Gruppe von k und die $(n-k)$ verbleibenden Elemente aufzuteilen. Durch Multiplikation der Komplementärwahrscheinlichkeit von (1.1.49) mit $(n-k)!$ erhält man die Anzahl der Anordnungen im Restfeld, bei denen kein Element an der richtigen Stelle steht, zu $(n-k)!\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + - \ldots \frac{(-1)^{n-k}}{(n-k)!}\Big)$.

Insgesamt gibt es also $\binom{n}{k}(n-k)!\Big(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+-\ldots\frac{(-1)^{n-k}}{(n-k)!}\Big)$ Möglichkeiten, genau k Elemente richtig anzuordnen und die $(n-k)$ restlichen alle falsch. Die

Wahrscheinlichkeit (1.1.50) ergibt sich durch Division dieses Ausdrucks durch $n!$, der Anzahl aller Permutationen des Feldes.

Asymptotisch folgt damit für die Wahrscheinlichkeit q_k, daß in einem (großen) Feld der Länge n bereits mindestens k Elemente richtig sortiert sind,

$$q_k \approx \frac{1}{e}\sum_{j=k}^{\infty}\frac{1}{j!} = 1 - \frac{1}{e}\sum_{j=0}^{k-1}\frac{1}{j!} = q_k^*.$$

Die folgende Tabelle enthält einige Werte von q_k^*, die man mit den oben angegebenen Bonferroni–Schranken vergleichen möge.

k	1	2	3	4	5
q_k^*	0.6321	0.2642	0.0803	0.0190	0.0037

Eine ausführliche Behandlung dieses Sortierproblems findet man auch bei Knuth (1973), Bd. 1, S. 178.

Für unabhängige Ereignisse $A_1, \ldots, A_n$ läßt sich die Wahrscheinlichkeit für die Vereinigung mit Hilfe der De–Morgan–Regeln darstellen als

$$P\Big(\bigcup_{k=1}^{n} A_k\Big) = 1 - P\Big(\bigcap_{k=1}^{n} A_k^c\Big) = 1 - \prod_{k=1}^{n}\big(1 - P(A_k)\big).$$

Die in dem vorigen Beispiel auftretenden Ereignisse A_i sind jedoch nicht stochastisch unabhängig. Dies sieht man daran, daß $P(A_1 \cap A_2) = \big(2!\binom{n}{2}\big)^{-1} = \frac{1}{n(n-1)}$ für $n \geq 2$ verschieden ist von $P(A_1)\cdot P(A_2) = \frac{1}{n\cdot n}$. Allerdings sind bei festem k und großem n die Ereignisse $A_1, \ldots, A_k$ nach Satz 1.1.2 "fast" unabhängig, da für alle Auswahlen $1 \leq i_1 < \ldots < i_l \leq k$

$$P\Big(\bigcap_{j=1}^{l} A_{i_j}\Big) = \frac{(n-l)!}{n!} \approx \frac{1}{n^l} = \prod_{j=1}^{l} P(A_{i_j})$$

gilt. Würde man statt der Grundmenge $Perm_n^n(\{1,\ldots,n\}; o.W.)$ die Grundmenge $Perm_n^n(\{1,\ldots,n\}; m.W.)$ verwenden, so wären die Ereignisse $A_1, \ldots, A_n$ tatsächlich unabhängig mit $P(A_i) = \frac{1}{n}$ für $1 \leq i \leq n$.

Die gerade durchgeführten Rechnungen lassen vermuten, daß bei stochastischer Unabhängigkeit Wahrscheinlichkeiten für Limes superior und Limes inferior von Ereignisfolgen mit relativ einfachen Mitteln berechnet werden können. Ein wichtiges Hilfsmittel stellt hierbei das folgende Ergebnis dar.

Satz 1.1.3. *(Borel–Cantelli–Lemma)*
Es sei $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Ereignissen in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann gilt:

a) $$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Longrightarrow \quad P(\limsup_{n\to\infty} A_n) = 0.$$

Ist $\{A_n\}$ zusätzlich stochastisch unabhängig, so gilt

$$b) \qquad \sum_{n=1}^{\infty} P(A_n) = \infty \quad \Longrightarrow \quad P(\limsup_{n\to\infty} A_n) = 1.$$

Beweis. a) Nach (1.1.42) folgt für alle $n \in \mathbb{N}$

$$P\Big(\bigcup_{k=n}^{\infty} A_k\Big) \leq \sum_{k=n}^{\infty} P(A_k) \to 0 \quad (n \to \infty),$$

also

$$P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty} A_k\Big) = 0.$$

b) Es ist

$$P(\limsup_{n\to\infty} A_n) = 1 - P(\liminf_{n\to\infty} A_n^c) = 1 - \lim_{n\to\infty} P\Big(\bigcap_{k=n}^{\infty} A_k^c\Big)$$

$$= 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} \big(1 - P(A_k)\big).$$

Aus $\limsup_{n\to\infty} P(A_n) = 1$ folgt trivialerweise $\prod_{k=n}^{\infty} \big(1 - P(A_k)\big) = 0$ für alle $n \in \mathbb{N}$, also die Behauptung. Ist andererseits $\limsup_{n\to\infty} P(A_n) = L < 1$, so gilt $\prod_{k=n}^{\infty} \big(1 - P(A_k)\big) = \exp\big\{\sum_{k=n}^{\infty} \ln\big(1 - P(A_k)\big)\big\} \leq \exp\big\{-\sum_{k=n}^{\infty} P(A_k)\big\} = 0$ für alle genügend großen $n \in \mathbb{N}$, also ebenfalls die behauptete Gleichheit. ∎

Unter Verwendung der De-Morgan-Regeln kann man Satz 1.1.3 sofort um folgende Identitäten ergänzen:

$$c) \qquad \sum_{n=1}^{\infty} \big(1 - P(A_n)\big) < \infty \quad \Longrightarrow \quad P(\liminf_{n\to\infty} A_n) = 1.$$

Ist $\{A_n\}$ zusätzlich stochastisch unabhängig, so gilt

$$d) \qquad \sum_{n=1}^{\infty} \big(1 - P(A_n)\big) = \infty \quad \Longrightarrow \quad P(\liminf_{n\to\infty} A_n) = 0.$$

Wir wollen uns nun näher mit der Frage beschäftigen, welche σ-Algebren geeignet sind, um etwa das Existenzproblem einer Gleichverteilung im Sinn von (1.0.6) und (1.0.7) zu lösen. Offensichtlich ist dabei mindestens zu fordern, daß eine solche σ-Algebra alle Teilintervalle $(a, b]$, $0 \leq a < b \leq 1$ von $(0, 1]$ enthält. Man gelangt damit in natürlicher Weise zum Begriff des Erzeugers einer σ-Algebra.

Lemma 1.1.4. *(Erzeuger einer σ–Algebra)*
Es sei Ω eine nicht–leere Menge und $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ ein System von Teilmengen von Ω. Dann ist

$$\sigma(\mathcal{E}) = \bigcap \{\mathcal{A} \mid \mathcal{A} \text{ ist } \sigma\text{–Algebra über } \Omega,\ \mathcal{E} \subseteq \mathcal{A}\}$$

eine σ–Algebra über Ω, die $\mathcal{E}$ enthält, und zwar die kleinste (bzgl. Mengeninklusion) mit dieser Eigenschaft. $\mathcal{E}$ heißt Erzeuger von $\sigma(\mathcal{E})$.

Beweis. Man verifiziert leicht anhand von (1.1.25) bis (1.1.27), daß $\sigma(\mathcal{E})$ eine σ–Algebra über Ω ist, denn die entsprechenden Eigenschaften übertragen sich auf den Durchschnitt von σ–Algebren. Der Schnitt ist außerdem nicht–leer, da stets $\mathcal{A} = \mathfrak{P}(\Omega)$ eine σ-Algebra ist, die $\mathcal{E}$ enthält. Ist ferner $\mathcal{B}$ eine σ–Algebra über Ω, die $\mathcal{E}$ enthält, so ist diese für die Durchschnittsbildung zulässig, und somit gilt $\sigma(\mathcal{E}) \subseteq \mathcal{B}$. Damit ist $\sigma(\mathcal{E})$ in der Tat minimal im angegebenen Sinn. ∎

Aufgrund von Lemma 1.1.4 ist also die Existenz einer kleinsten, das Mengensystem $\mathcal{E}$ enthaltenden σ–Algebra sichergestellt. Die Borel'sche σ–Algebra wird auf diesem Weg wie folgt durch ein Erzeugendensystem definiert.

Definition 1.1.11. *(Borel'sche σ–Algebra)*
Es sei $\Omega = \mathbf{R}$ und $\mathcal{E} = \{(a, b] \mid a, b \in \mathbf{R},\ a < b\}$. Dann heißt die durch $\mathcal{E}$ erzeugte σ–Algebra $\sigma(\mathcal{E})$ die Borel–σ–Algebra über **R**, *in Zeichen: $\mathcal{B}^1 = \sigma(\mathcal{E})$.*

Wir werden im folgenden auch Borel'sche σ–Algebren über echten Teilmengen von **R** benötigen; sie lassen sich mit Hilfe von $\mathcal{B}^1$ folgendermaßen konstruieren.

Lemma 1.1.5. *(Spur–σ–Algebra)*
Es seien $\Omega_2 \subseteq \Omega_1$ nicht–leere Mengen und $\mathcal{A}$ eine σ–Algebra über Ω_1. Dann wird durch

$$\Omega_2 \cap \mathcal{A} = \{\Omega_2 \cap A \mid A \in \mathcal{A}\} \tag{1.1.51}$$

eine σ–Algebra über Ω_2 definiert. Diese heißt Spur–σ–Algebra von Ω_2 in $\mathcal{A}$. Ist ferner $\Omega_2 \in \mathcal{A}$, so gilt

$$\Omega_2 \cap \mathcal{A} = \{A \in \mathcal{A} \mid A \subseteq \Omega_2\}. \tag{1.1.52}$$

Beweis. Wir zeigen zunächst, daß $\Omega_2 \cap \mathcal{A}$ eine σ–Algebra über Ω_2 ist. Offenbar ist $\Omega_2 = \Omega_2 \cap \Omega_1$ mit $\Omega_1 \in \mathcal{A}$, also $\Omega_2 \in \Omega_2 \cap \mathcal{A}$. Dies ist (1.1.25).
Zu einer Menge $B \in \Omega_2 \cap \mathcal{A}$ existiert per Definition eine Menge $A \in \mathcal{A}$ mit $B = \Omega_2 \cap A$. Damit ist aber $\Omega_2 \setminus B = \Omega_2 \setminus A = \Omega_2 \cap (\Omega_1 \setminus A)$ mit $\Omega_1 \setminus A \in \mathcal{A}$, also (1.1.26) erfüllt.
Zu einer Folge $\{B_n\} \subseteq \Omega_2 \cap \mathcal{A}$ existiert analog eine Folge $\{A_n\} \subseteq \mathcal{A}$ mit $B_n = \Omega_2 \cap A_n$, $n \in \mathbf{N}$, so daß $\bigcup_{n=1}^{\infty} B_n = \Omega_2 \cap \bigcup_{n=1}^{\infty} A_n$ mit $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, also auch (1.1.27) erfüllt ist. Damit ist $\Omega_2 \cap \mathcal{A}$ eine σ–Algebra über Ω_2.
Zum Beweis von (1.1.52) bemerken wir, daß $\Omega_2 \cap A \in \mathcal{A}$ gilt, wenn Ω_2 und $A \in \mathcal{A}$ ist. Damit gilt jedenfalls $\Omega_2 \cap \mathcal{A} \subseteq \{A \in \mathcal{A} \mid A \subseteq \Omega_2\}$. Ist umgekehrt $A \in \mathcal{A}$ mit $A \subseteq \Omega_2$, so ist $A = \Omega_2 \cap A$. Also gilt $\Omega_2 \cap \mathcal{A} \supseteq \{A \in \mathcal{A} \mid A \subseteq \Omega_2\}$ und somit Gleichheit. ∎

Damit bilden die Borel'schen Teilmengen des Intervalls $(0,1]$, welche für das Gleichverteilungsproblem (1.0.6) und (1.0.7) wichtig sind, in der Tat eine σ–Algebra über $(0,1]$, nämlich gerade die entsprechende Spur–σ–Algebra.

Die in Lemma 1.1.4 betrachteten Erzeuger von σ–Algebren sind keineswegs eindeutig bestimmt. Beispielsweise sind im Fall der Borel'schen σ–Algebra $\mathcal{B}^1$ die folgenden Mengensysteme — neben anderen — Erzeuger von $\mathcal{B}^1$:

$$\mathcal{E}_1 = \{(a,b] \mid a,b \in \mathbf{R},\ a < b\} \tag{1.1.53}$$
$$\mathcal{E}_2 = \{[a,b) \mid a,b \in \mathbf{R},\ a < b\} \tag{1.1.54}$$
$$\mathcal{E}_3 = \{(a,b) \mid a,b \in \mathbf{R},\ a < b\} \tag{1.1.55}$$
$$\mathcal{E}_4 = \{[a,b] \mid a,b \in \mathbf{R},\ a < b\} \tag{1.1.56}$$
$$\mathcal{E}_5 = \{(-\infty,b] \mid b \in \mathbf{R}\} \tag{1.1.57}$$
$$\mathcal{E}_6 = \{(-\infty,b) \mid b \in \mathbf{R}\} \tag{1.1.58}$$
$$\mathcal{E}_7 = \{G \subseteq \mathbf{R} \mid G \text{ offen}\} \tag{1.1.59}$$
$$\mathcal{E}_8 = \{F \subseteq \mathbf{R} \mid F \text{ abgeschlossen}\} \tag{1.1.60}$$
$$\mathcal{E}_9 = \{K \subseteq \mathbf{R} \mid K \text{ kompakt}\}. \tag{1.1.61}$$

Dabei können in den Beziehungen (1.1.53) bis (1.1.58) sowie in Definition 1.1.11 die Ausdrücke "$a,b \in \mathbf{R}$" bzw. "$b \in \mathbf{R}$" noch durch die Ausdrücke "$a,b \in \mathbf{Q}$" bzw. "$b \in \mathbf{Q}$" (oder statt $\mathbf{Q}$ eine andere, in $\mathbf{R}$ dichte Teilmenge) ersetzt werden.

Zum Beweis dieser Beziehungen reicht es offensichtlich, nachzuweisen, daß die Elemente der verschiedenen Erzeuger aus geeigneten Intervallen bzw. Borel'schen Mengen durch abzählbare Mengenoperationen der Art (1.1.26) bis (1.1.28) konstruiert werden können und umgekehrt. Wir wollen dies in einigen Fällen exemplarisch ausführen.

Für die Elemente des Erzeugers $\mathcal{E}_3$ gilt etwa

$$(a,b) = \bigcup_{n=1}^{\infty} (a, b-(b-a)/2n],$$

für die Elemente des Erzeugers $\mathcal{E}_5$

$$(-\infty,b] = \bigcup_{n=1}^{\infty} (-n,b].$$

Für die Elemente des Erzeugers $\mathcal{E}_7$ benutzt man die Tatsache, daß jede in $\mathbf{R}$ offene Menge G als abzählbare disjunkte Vereinigung offener Intervalle darstellbar ist, für die Elemente des Erzeugers $\mathcal{E}_8$ gilt, daß sie gerade die Komplemente der offenen Mengen bilden.

Umgekehrt lassen sich Intervalle durch andere Erzeugermengen z.B. folgendermaßen darstellen: im Fall des Erzeugers $\mathcal{E}_3$ etwa durch

$$(a,b] = \bigcap_{n=1}^{\infty} (a, b+1/n),$$

im Fall des Erzeugers $\mathcal{E}_5$ durch

$$(a,b] = (-\infty,b] \setminus (-\infty,a].$$

Der Erzeuger $\mathcal{E}_7$ enthält insbesondere die offenen Intervalle, aus denen man die Intervalle aus $\mathcal{E}_1$ auf die gerade gezeigte Weise erhält. Analog argumentiert man mit dem Erzeuger $\mathcal{E}_8$, der gerade die Komplemente der offenen Mengen, also der Mengen von $\mathcal{E}_7$ enthält.

Sind ferner $a < b \in \mathbf{R}$, so gibt es monotone Folgen $\{a_n\}$ und $\{b_n\}$ rationaler Zahlen mit $a_n \uparrow a$, $b_n \downarrow b$, so daß $[a,b] = \bigcap_{n=1}^{\infty}[a_n, b_n]$ ist. Also wird die Borel'sche σ–Algebra auch durch Intervalle aus $\mathcal{E}_4$ mit rationalen Endpunkten erzeugt. Man sagt, $\mathcal{B}^1$ sei abzählbar erzeugt. Entsprechend argumentiert man für die übrigen Erzeuger.

Insbesondere gilt, daß die einelementigen Mengen $\{x\}$ mit $x \in \mathbf{R}$ Borel'sch sind, da $\{x\} = \bigcap_{n=1}^{\infty}(x - 1/n, x]$ gilt. Damit sind auch alle abzählbaren Teilmengen von $\mathbf{R}$ Borel'sche Mengen.

Allerdings ist das Mengensystem $\mathcal{E}_0 = \{\{x\} \mid x \in \mathbf{R}\}$ kein Erzeuger von $\mathcal{B}^1$, sondern von der kleineren σ–Algebra $\mathcal{A} = \{B \in \mathcal{B}^1 \mid B \text{ oder } B^c \text{ ist abzählbar}\}$.

Es soll noch bemerkt werden, daß man in natürlicher Weise auch Erzeuger von Spur–σ–Algebren $\Omega_2 \cap \mathcal{A}$ (vgl. Lemma 1.1.5) erhält, wenn $\mathcal{E}$ ein Erzeuger von $\mathcal{A}$ ist, etwa durch $\mathcal{E}^* = \Omega_2 \cap \mathcal{E} = \{\Omega_2 \cap E \mid E \in \mathcal{E}\}$. In diesem Sinn ist beispielsweise $\mathcal{E}^* = \{[a,b] \mid 0 \le a < b \le 1\}$ ein Erzeuger von $[0,1] \cap \mathcal{B}^1$.

Wir kommen jetzt auf das Problem zurück, wie man Wahrscheinlichkeitsverteilungen auf σ–Algebren erhalten kann, wenn man lediglich Wahrscheinlichkeiten für Erzeugermengen kennt, wie sie z.B. im Fall der Gleichverteilung auf $\Omega = (0,1]$ durch Bedingung (1.0.6) festgelegt werden. Wir wollen dabei das allgemeine Vorgehen am Beispiel der Gleichverteilung lediglich skizzieren, da eine ausführliche Darstellung wegen des maßtheoretischen Hintergrunds über den Rahmen dieses Buches hinausgeht. Der interessierte Leser sei auf Bauer (1978), §5, oder Billingsley (1986), Section 3, verwiesen.

Bedingung (1.0.6) legt die zu bestimmende Gleichverteilung auf $(0,1] \cap \mathcal{B}^1$ zunächst nur auf dem Erzeugersystem $\Omega \cap \mathcal{E}_1$, $\Omega = (0,1]$, $\mathcal{E}_1$ aus (1.1.53), fest. Das Mengensystem $\mathcal{R} = \left\{\bigcup_{i=1}^{n} E_i \mid E_i \in \Omega \cap \mathcal{E}_1,\ n \in \mathbf{N}\right\}$ ist dann ein sogenannter *Ring*, d.h. es besitzt die Eigenschaften

$$\begin{aligned} \emptyset &\in \mathcal{R}, \\ A, B \in \mathcal{R} &\Longrightarrow A \setminus B \in \mathcal{R}, \\ A, B \in \mathcal{R} &\Longrightarrow A \cup B \in \mathcal{R}. \end{aligned}$$

Insbesondere besitzt jede Menge $R \in \mathcal{R}$ eine Darstellung

$$R = \bigcup_{i=1}^{n} D_i \tag{1.1.62}$$

mit paarweise disjunkten Mengen $D_i \in \Omega \cap \mathcal{E}_1$. Für eine solche Darstellung von $R \in \mathcal{R}$ liefert dann

$$P(R) = \sum_{i=1}^{n} P(D_i) \tag{1.1.63}$$

eine Fortsetzung von P auf $\mathcal{R}$, die unabhängig von der speziellen Darstellung von R mit zerlegenden Mengen D_i und sogar σ–additiv auf $\mathcal{R}$ ist, d.h. für beliebige paarweise disjunkte Mengenfolgen $\{A_n\} \subseteq \mathcal{R}$ mit $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$ gilt (1.0.5). Definiert

man nun das sogenannte *äußere Maß* P^* auf $\mathfrak{P}(\Omega)$ durch

$$P^*(A) = \inf\left\{ \sum_{n=1}^{\infty} P(A_n) \mid \{A_n\} \subseteq \mathcal{R},\ A \subseteq \bigcup_{n=1}^{\infty} A_n \right\},\ A \in \mathfrak{P}(\Omega), \qquad (1.1.64)$$

so läßt sich zeigen, daß P^* auf $\mathcal{R}$ mit P übereinstimmt und daß das Mengensystem

$$\mathcal{A}^* = \left\{ A \in \mathfrak{P}(\Omega) \mid P^*(B) = P^*(B \cap A) + P^*(B \cap A^c) \text{ für alle } B \in \mathfrak{P}(\Omega) \right\} \quad (1.1.65)$$

eine σ–Algebra bildet, welche $\Omega \cap \mathcal{B}^1$ umfaßt und auf der das äußere Maß P^* ebenfalls σ–additiv ist.

Durch Einschränkung von P^* auf $\Omega \cap \mathcal{B}^1$ erhält man dann die gewünschte Gleichverteilung P auf $\Omega \cap \mathcal{B}^1$, die darüberhinaus eindeutig bestimmt ist und für Borel'sche Teilmengen der Bedingung (1.0.7) genügt.

Diese Methode zur Konstruktion einer Fortsetzung funktioniert auch ganz allgemein. Hat man eine Wahrscheinlichkeitsverteilung P auf einem Ring $\mathcal{R}$, so läßt sich diese über die Schritte (1.1.64) und (1.1.65) zu einer Wahrscheinlichkeitsverteilung auf der von $\mathcal{R}$ erzeugten σ–Algebra $\mathcal{A} = \sigma(\mathcal{R})$ fortsetzen.

Die σ–Algebren $\mathcal{A}^*$ und $\Omega \cap \mathcal{B}^1$ unterscheiden sich dabei nur unwesentlich; genauer gilt

$$\mathcal{A}^* = \left\{ A \cup N \mid A \in \Omega \cap \mathcal{B}^1,\ N \in \mathcal{N} \right\},$$

wobei $\mathcal{N} = \left\{ N \in \mathfrak{P}(\Omega) \mid \text{ es existiert } B \in \Omega \cap \mathcal{B}^1 \text{ mit } N \subseteq B \text{ und } P(B) = 0 \right\}$ das System der sogenannten *P–Nullmengen* ist. $\mathcal{A}^*$ entsteht also aus $\Omega \cap \mathcal{B}^1$ durch Hinzunahme aller Teilmengen N von Mengen $B \in \Omega \cap \mathcal{B}^1$ mit $P(B) = 0$, für die dann natürlich $P(N) = 0$ gelten muß.

Die Frage nach der Mächtigkeit solcher σ–Algebren läßt sich erstaunlicherweise nicht im Rahmen der klassischen axiomatischen Mengenlehre (etwa nach Zermelo–Fraenkel) beantworten. Will man z.B. nachweisen, daß im Fall der Gleichverteilung P über $\Omega = (0,1]$ $\mathcal{A}^* \neq \mathfrak{P}(\Omega)$ gilt, so benötigt man das

Auswahlaxiom: *Sind die Indexmenge I und Mengen M_i nicht–leer für alle $i \in I$, so ist das kartesische Produkt dieser Mengen $\bigtimes_{i \in I} M_i$ nicht–leer; es existiert also eine Auswahlfunktion*

$$f : I \longrightarrow \bigcup_{i \in I} M_i$$

mit $f(i) \in M_i$ für alle $i \in I$.

Anschaulich besagt das Auswahlaxiom, daß man aus beliebig vielen nicht–leeren Mengen M_i "gleichzeitig" ein Element auswählen kann; es ist in der Tat unabhängig von den übrigen Axiomen der klassischen Mengenlehre.

Mit Hilfe des Auswahlaxioms wurde eine Menge M, die nicht zu $\Omega \cap \mathcal{B}^1$ (und damit auch nicht zu $\mathcal{A}^*$) gehört, schon 1905 von Vitali auf die folgende Weise konstruiert (vgl. auch Bauer (1978), Satz 8.4, Floret (1981), S. 117 oder Benedetto (1976), S.49):

Auf $\mathbf{R}$ wird eine Äquivalenzrelation $\sim$ definiert durch "$x \sim y$ genau dann, wenn $x - y \in \mathbf{Q}$, $x, y \in \mathbf{R}$".

$\mathbf{R}$ zerfällt damit in die zugehörigen Äquivalenzklassen $K(x)$, $x \in \mathbf{R}$, d.h. $K(x) = \{y \in \mathbf{R} \mid x - y \in \mathbf{Q}\}$. Insbesondere ist mit x auch die reelle Zahl $y(x) = \begin{cases} x - \lfloor x \rfloor & \text{für } x \notin \mathbf{Z} \\ 1 & \text{sonst} \end{cases}$ Element von $K(x)$, wobei $\lfloor x \rfloor$ die größte ganze Zahl kleiner oder gleich x bezeichne. Wegen $y(x) \in (0,1]$ gibt es also nach dem Auswahlaxiom eine Repräsentantenmenge $M \subset (0,1]$, welche mit jeder Äquivalenzklasse $K(x)$ genau ein Element gemeinsam hat. Dann gilt: $M \notin \Omega \cap \mathcal{B}^1$.

Verzichtet man allerdings auf die Gültigkeit des Auswahlaxioms, so kann man mit Methoden, die auf Cohen und Soloway 1964 zurückgehen, ein widerspruchsfreies Modell angeben, in dem alle Teilmengen von Ω zu $\mathcal{A}^*$ gehören, also $\mathcal{A}^* = \mathfrak{P}(\Omega)$ gilt; vgl. etwa Benedetto (1976), S. 50 oder Floret (1981), S. 118.

Da im allgemeinen die σ–Algebra $\mathcal{A}^*$ wegen der zusätzlichen Nullmengen von P abhängt, diese jedoch für die Modellierung stochastischer Vorgänge praktisch keine Rolle spielen, werden wir hier nur die Borel'sche σ–Algebra $\mathcal{B}^1$ bzw. entsprechende Spur–σ–Algebren betrachten.

Bei der Frage, ob und wie weit z.B. die stetige Gleichverteilung über die σ–Algebra $(0,1] \cap \mathcal{B}^1$ hinaus fortgesetzt werden kann, kommt man sehr schnell zu ähnlichen Grundlagenproblemen wie schon bei der Beurteilung der Mächtigkeit der σ–Algebra $\mathcal{A}^*$. Bei Annahme des Auswahlaxioms läßt sich das das Vitali–Beispiel heranziehen, um die Unmöglichkeit der translationsinvarianten Fortsetzung (d.h. unter Beibehaltung der Eigenschaft (1.0.7)) auf die ganze Potenzmenge von $(0,1]$ zu zeigen; verzichtet man auf das Auswahlaxiom, läßt sich die Fortsetzbarkeit der Gleichverteilung ohne weitere Forderungen im Rahmen der klassischen Mengenlehre nicht entscheiden.

Allerdings existiert nicht einmal unter der schwächeren Forderung (1.0.6) eine Fortsetzung der Gleichverteilung auf ganz $\mathfrak{P}(\Omega)$, wenn man z.B. die Gültigkeit der sogenannten Kontinuumshypothese (die ebenfalls unabhängig von den klassischen Axiomen der Mengenlehre und auch unabhängig vom Auswahlaxiom ist) postuliert, wie man aus Arbeiten von Banach, Kuratowski und Ulam aus den Jahren 1929 und 1930 schließen kann. Sie lautet:

Kontinuumshypothese: *Jede überabzählbare Teilmenge A von $\mathbf{R}$ besitzt dieselbe Mächtigkeit wie $\mathbf{R}$, d.h.*

$$A \subseteq \mathbf{R}, \quad \operatorname{card}(A) > \aleph_0 \quad \Longrightarrow \quad \operatorname{card}(A) = \operatorname{card}(\mathbf{R}).$$

Das letzte Resultat läßt sich sogar noch dahingehend verallgemeinern, daß man — unter Verwendung der Kontinuumshypothese — zeigen kann, daß *überhaupt keine* nicht–diskreten Verteilungen P auf ganz $\mathfrak{P}(\mathbf{R})$ existieren. Allerdings läßt sich z.B. die Gleichverteilung P noch über die mit Nullmengen vervollständigte σ–Algebra $\mathcal{A}^*$ hinaus auf größere σ–Algebren $\mathcal{A} \supset \mathcal{A}^*$, $\mathcal{A} \subset \mathfrak{P}(\mathbf{R})$ fortsetzen, wobei sogar die Bedingung (1.0.7) eingehalten wird. Der interessierte Leser sei hier auf die Ausführungen und Literaturangaben in Benedetto (1976), S. 40 verwiesen.

Auch eine Beschränkung auf *additive* Wahrscheinlichkeitsverteilungen, d.h. solche, bei denen Beziehung (1.1.17) lediglich für *endliche* (statt abzählbar unendliche) Vereinigungen gefordert wird, bringt hier keine Vorteile: Banach hat nämlich 1923 gezeigt, daß selbst unter der restiktiveren Forderung (1.0.7) zwar Lösungen des Problems existieren, diese aber nicht eindeutig bestimmt und daher für praktische Zwecke nicht geeignet sind. Ein analoges Problem für Dimensionen ≥ 3

ist sogar wieder unlösbar (vgl. Bauer (1978), S. 50, Floret (1981), S. 119ff. oder Benedetto (1976), S. 50f.).

Die vorangehenden Überlegungen rechtfertigen also im Nachhinein noch einmal den hier beschrittenen Weg der eventuellen Einschränkung der Potenzmenge $\mathfrak{P}(\Omega)$ auf kleinere σ–Algebren als Definitionsbereich für Wahrscheinlichkeitsverteilungen.

Für das Fortsetzungs– bzw. Existenzproblem von Wahrscheinlichkeitverteilungen auf σ–Algebren ist es nach obigem also wesentlich, daß diese auf geeigneten Erzeugern sinnvoll erklärt sind (etwa wie in (1.0.6)), während die Fortsetzung dieser Verteilungen über Ringe zu äußeren Maßen dann wie in (1.1.62) bis (1.1.65) erfolgt. Der Erzeuger $\mathcal{E}_1$ der Borel'schen σ–Algebra oder entsprechender Spur–σ–Algebren spielt dabei eine besondere Rolle. Er führt zum Begriff der Verteilungsfunktion, die ein analytisches Hilfsmittel zur Darstellung von Wahrscheinlichkeitsmaßen auf $\mathcal{B}^1$ darstellt.

1.2. Verteilungsfunktionen und Dichten

In diesem Abschnitt wollen wir uns mit der Frage beschäftigen, wie man auf möglichst einfache Weise Wahrscheinlichkeitsverteilungen P auf $\mathcal{B}^1$ bzw. geeigneten Spur–σ–Algebren vollständig beschreiben kann. Im Fall diskreter Verteilungen genügt es offensichtlich gemäß Lemma 1.1.2, die Elementarwahrscheinlichkeiten $P(\{\omega\})$ zu spezifizieren. Es gibt dann eine kleinste (abzählbare) Menge $T \subset \mathbf{R}$ mit $P(T) = 1$, d.h. T enthält genau diejenigen Elementarereignisse ω mit $P(\{\omega\}) > 0$. T heißt auch *Träger* der Verteilung P. Im Fall beliebiger Verteilungen ist diese Vorgehensweise i.a. nicht sinnvoll, da $P(\{\omega\}) = 0$ gelten kann für alle $\omega \in \Omega$, wie das Beispiel der Gleichverteilung zeigt: $\{\omega\} \subset (0,1]$ läßt sich nämlich darstellen als $\{\omega\} = \bigcap_{n=1}^{\infty}(\omega - \frac{\omega}{n}, \omega]$, so daß mit der Stetigkeit von oben von P nach (1.1.40) folgt: $P(\{\omega\}) = \lim_{n\to\infty} P((\omega - \frac{\omega}{n}, \omega]) = \lim_{n\to\infty} \frac{\omega}{n} = 0$. Die Verteilung P kann also nicht über die Spezifikation von Elementarwahrscheinlichkeiten beschrieben werden. Andererseits zeigt die Vorgehensweise am Ende des vorigen Abschnitts, daß es ausreicht, eine Verteilung P etwa auf dem Erzeuger $\mathcal{E}_1$, also den Intervallen $(a,b] \subset \mathbf{R}$, $a < b$ in geeigneter Weise festzulegen. Wegen $(a,b] = (-\infty, b] \setminus (-\infty, a]$ und der Subtraktivität von P genügt es demnach sogar, P lediglich auf dem Erzeuger $\mathcal{E}_5$ sinnvoll zu definieren. Man gelangt damit zu der folgenden Begriffsbildung.

Definition 1.2.1. *(Verteilungsfunktion)*
Es sei P ein Wahrscheinlichkeitsmaß auf der Borel'schen σ–Algebra $\mathcal{B}^1$. Die durch

$$F_P(x) = P((-\infty, x]),\ x \in \mathbf{R}, \tag{1.2.1}$$

definierte Abbildung F_P heißt die zu P gehörige Verteilungsfunktion.

Aufgrund der in Lemma 1.1.3 angegebenen Eigenschaften von Wahrscheinlichkeitsmaßen erhält man sofort entsprechende Eigenschaften für Verteilungsfunktionen, von denen sich einige als fundamental für die Beschreibung von Wahrscheinlichkeitsverteilungen herausstellen werden.

Lemma 1.2.1. *Es sei $F = F_P$ die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung über $\mathcal{B}^1$. Dann gilt:*

$$F(x) \le F(y) \text{ für alle } x, y \in \mathbf{R},\ x \le y \text{ (Monotonie von } F) \tag{1.2.2}$$

$$P((a,b]) = F(b) - F(a) \text{ für alle } a,b \in \mathbf{R},\ a \le b \tag{1.2.3}$$

$$\lim_{x \downarrow y} F(x) = F(y),\ y \in \mathbf{R} \text{ (*rechtsseitige Stetigkeit von* } F) \tag{1.2.4}$$

$$\lim_{x \uparrow y} F(x) = F(y) \iff P(\{y\}) = 0,\ y \in \mathbf{R} \tag{1.2.5}$$

$$P(\{y\}) = F(y) - \lim_{x \uparrow y} F(x),\ y \in \mathbf{R} \tag{1.2.6}$$

$$\lim_{x \to \infty} F(x) = 1,\ \lim_{x \to -\infty} F(x) = 0. \tag{1.2.7}$$

Beweis. Die Beziehungen (1.2.2) und (1.2.3) ergeben sich unmittelbar aus den Beziehungen (1.1.36) und (1.1.37), wenn man dort $A = (-\infty, x]$ und $B = (-\infty, y]$ bzw. $A = (-\infty, a]$ und $B = (-\infty, b]$ wählt. (1.2.4) folgt entsprechend aus (1.1.39), indem man für beliebige Folgen $\{x_n\}$ mit $x_n \downarrow y$ $A_n = (-\infty, x_n]$ wählt. Die Beziehung (1.2.5) ist eine Konsequenz aus Beziehung (1.2.6), welche sich aus (1.1.38) ergibt, wenn man wieder für Folgen $\{x_n\}$ mit $x_n \uparrow y$ (d.h. insbesondere $x_n < y$) $A_n = (-\infty, x_n]$ wählt und beachtet, daß $\bigcup_{n=1}^{\infty} A_n = (-\infty, y)$ und $F(y) - P(\{y\}) = P((-\infty, y))$ gilt. Die letzte Beziehung (1.2.7) erhält man, wenn jeweils Mengenfolgen $A_n = (-\infty, x_n]$ mit $x_n \uparrow \infty$ bzw. $A_n = (-\infty, x_n]$ mit $x_n \downarrow -\infty$ gewählt werden, da dann im ersten Fall $\lim_{n\to\infty} A_n = \mathbf{R}$, im zweiten Fall $\lim_{n\to\infty} A_n = \emptyset$ gilt. ∎

Zur eindeutigen Beschreibung von Verteilungen P durch Verteilungsfunktionen F_P werden allerdings nicht alle diese Eigenschaften benötigt. Vielmehr gilt:

Satz 1.2.1. *(Fortsetzungssatz)*
Es sei $F : \mathbf{R} \longrightarrow [0,1]$ eine Abbildung mit den Eigenschaften (1.2.2), (1.2.4) und (1.2.7). Dann existiert eine Wahrscheinlichkeitsverteilung P auf $\mathcal{B}^1$ derart, daß F genau die zu P gehörige Verteilungsfunktion darstellt, d.h. es gilt $F = F_P$.

Beweis. Zunächst läßt sich die gewünschte Verteilung P auf dem Erzeuger $\mathcal{E}_1$ konstruieren vermöge

$$P((a,b]) = F(b) - F(a),\ a,b \in \mathbf{R},\ a < b.$$

Ist nun $\mathcal{R}$ wieder der von $\mathcal{E}_1$ erzeugte Ring, d.h. ist $\mathcal{R} = \left\{ \bigcup_{i=1}^{n} E_i \mid E_i \in \mathcal{E}_1,\ n \in \mathbf{N} \right\}$, so läßt sich in einem weiteren Schritt P auf $\mathcal{R}$ fortsetzen, indem man für jede Menge $R \in \mathcal{R}$ mit der — stets möglichen — Darstellung $R = \bigcup_{i=1}^{n} D_i$ mit paarweise disjunkten Mengen $D_i \in \mathcal{E}_1$ analog (1.1.63) setzt

$$P(R) = \sum_{i=1}^{n} P(D_i). \tag{1.2.8}$$

Mit Hilfe der rechtsseitigen Stetigkeit von F läßt sich dann weiter zeigen, daß das so erhaltene P sogar σ-additiv auf $\mathcal{R}$ ist, also wegen der Normierungsbedingungen (1.2.7) damit eine Wahrscheinlichkeitsverteilung auf $\mathcal{R}$ darstellt. Über die Schritte (1.1.64) und (1.1.65) erhält man schließlich das zu P gehörige äußere Maß P^*, dessen Einschränkung auf $\mathcal{B}^1$ dann die gewünschte Wahrscheinlichkeitsverteilung liefert. ∎

Satz 1.2.1 enthält noch keine Aussage zur *eindeutigen* Konstruktion von Wahrscheinlichkeitsverteilungen aus Abbildungen F mit den angegebenen Eigenschaften (1.2.2), (1.2.4) und (1.2.7). Diese ergibt sich allerdings aus der sogenannten *Durchschnittsstabilität* des Erzeugers $\mathcal{E}_1$, d.h. der Eigenschaft

$$E, F \in \mathcal{E}_1 \quad \Longrightarrow \quad E \cap F \in \mathcal{E}_1. \tag{1.2.9}$$

Für einen formalen Beweis der Eindeutigkeit der im Beweis zu Satz 1.2.1 angegebenen Konstruktion wird allerdings noch die Begriffsbildung eines Dynkin–Systems benötigt, insbesondere deshalb, weil für σ–Algebren die Abgeschlossenheit bezüglich beliebiger Vereinigungsbildungen — siehe (1.1.27) — gefordert wird, die σ–Additivität einer Wahrscheinlichkeitsverteilung sich aber auf *disjunkte* Vereinigungen — siehe (1.0.6) — bezieht.

Definition 1.2.2. *(Dynkin–System)*
Es sei Ω eine nicht–leere Menge und $\mathcal{D} \subseteq \mathfrak{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathcal{D}$ heißt Dynkin–System über Ω, wenn gilt:

$$\Omega \in \mathcal{D} \tag{1.2.10}$$

$$D \in \mathcal{D} \quad \Longrightarrow \quad D^c \in \mathcal{D} \tag{1.2.11}$$

$$D_n \in \mathcal{D},\ n \in \mathbb{N},\ \textit{paarweise disjunkt} \quad \Longrightarrow \quad \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}. \tag{1.2.12}$$

Die Eigenschaften eines Dynkin–Systems sind also mehr auf die Eigenschaften einer Wahrscheinlichkeitsverteilung abgestimmt als diejenigen einer σ–Algebra. Es bleibt daher zu klären, in welcher Beziehung diese beiden Mengensysteme zueinander stehen. Offensichtlich ist jede σ–Algebra auch ein Dynkin–System. Die Umkehrung ist i.a. nicht richtig; es gilt aber:

Lemma 1.2.2. *Jedes durchschnittsstabile Dynkinsystem ist eine σ–Algebra.*

Beweis. Es sei $\mathcal{D}$ ein durchschnittsstabiles Dynkin–System über Ω. Dann ist $\mathcal{D}$ vereinigungsstabil, d.h. mit $D, E \in \mathcal{D}$ ist auch

$$D \cup E = \big(D \cap E^c\big) \cup \big(D \cap E\big) \cup \big(E \cap D^c\big) \in \mathcal{D}$$

(man beachte, daß die drei vereinigten Mengen paarweise disjunkt sind), sowie differenzstabil, d.h. mit $D, E \in \mathcal{D}$ ist auch

$$E \setminus D = E \cap D^c \in \mathcal{D}.$$

Es bleibt also lediglich die Eigenschaft (1.1.27) einer σ–Algebra nachzuweisen. Sei dazu $\{D_n\} \subseteq \mathcal{D}$ eine beliebige (nicht notwendig paarweise disjunkte) Mengenfolge. Für $n \in \mathbb{N}$ setzen wir

$$F_0 = \emptyset,\ F_n = \bigcup_{k=1}^{n} D_k.$$

Es ist dann

$$\bigcup_{n=1}^{\infty} D_n = \bigcup_{n=0}^{\infty} (F_{n+1} \setminus F_n)$$

eine Vereinigung paarweise disjunkter Mengen $F_{n+1} \setminus F_n$ aus $\mathcal{D}$, also $\bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$, was zu zeigen war. ∎

Analog zu der Begriffsbildung einer von einem Mengensystem $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ erzeugten σ–Algebra $\sigma(\mathcal{E})$ kann man auch von einem von $\mathcal{E}$ erzeugten Dynkin–System sprechen, d.h. dem kleinsten Dynkin–System $\mathcal{D}$ über Ω, welches $\mathcal{E}$ enthält; i.Z.: $\mathcal{D} = \delta(\mathcal{E})$. Bezüglich der eindeutigen Charakterisierung von Wahrscheinlichkeitsverteilungen durch Verteilungsfunktionen ist vor allem das folgende Resultat von Bedeutung.

Satz 1.2.2. *Es sei $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ ein durchschnittsstabiles Mengensystem. Dann fällt das von $\mathcal{E}$ erzeugte Dynkin–System $\delta(\mathcal{E})$ mit der von $\mathcal{E}$ erzeugten σ–Algebra $\sigma(\mathcal{E})$ zusammen, d.h. es gilt $\delta(\mathcal{E}) = \sigma(\mathcal{E})$.*

Beweis. Es reicht, nachzuweisen, daß $\delta(\mathcal{E})$ eine σ–Algebra über Ω ist, da $\sigma(\mathcal{E})$ jedenfalls ein Dynkin–System über Ω ist. Gemäß Lemma 1.2.2 ist dafür lediglich zu zeigen, daß $\delta(\mathcal{E})$ durchschnittsstabil ist. Sei nun $E \in \delta(\mathcal{E})$. Das System $\mathcal{D}_E = \{F \in \delta(\mathcal{E}) | E \cap F \in \delta(\mathcal{E})\}$ ist ebenfalls ein Dynkin–System über Ω, wie man durch einfaches Nachprüfen der Beziehungen (1.2.10) bis (1.2.12) leicht feststellt. Nach Voraussetzung umfaßt aber $\mathcal{D}_E$ jedenfalls $\mathcal{E}$, so daß $\delta(\mathcal{E})$ als kleinstes $\mathcal{E}$ umfassendes Dynkin–System in $\mathcal{D}_E$ enthalten ist; d.h. aber wegen $\mathcal{D}_E \subseteq \delta(\mathcal{E})$: $\mathcal{D}_E = \delta(\mathcal{E})$. Es ist also $E \cap F \in \delta(\mathcal{E})$ für alle $F \in \delta(\mathcal{E})$. Da $E \in \delta(\mathcal{E})$ aber beliebig war, ist somit $\delta(\mathcal{E})$ als durchschnittsstabil nachgewiesen, was zu zeigen war. ∎

Nunmehr können wir auch die noch offene Eindeutigkeitsaussage in Satz 1.2.1 beweisen.

Satz 1.2.3. *(Eindeutigkeitssatz)*
Es sei $F : \mathbf{R} \longrightarrow [0,1]$ eine Abbildung mit den Eigenschaften (1.2.2), (1.2.4) und (1.2.7). Die nach Satz 1.2.1 existierende Wahrscheinlichkeitsverteilung P auf $\mathcal{B}^1$, für die $F_P = F$ gilt, ist dann eindeutig bestimmt.

Beweis. Seien P und Q zwei Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^1$ mit derselben Verteilungsfunktion F. Das System $\mathcal{D} = \{D \in \mathcal{B}^1 \mid P(D) = Q(D)\}$ ist ein Dynkin–System über $\Omega = \mathbf{R}$, denn es gilt:

$$P(\Omega) = Q(\Omega) = 1 \tag{1.2.13}$$

$$P(D^c) = 1 - P(D) = 1 - Q(D) = Q(D^c) \tag{1.2.14}$$

für alle $D \in \mathcal{D}$ sowie

$$P\Big(\bigcup_{n=1}^{\infty} D_n\Big) = \sum_{n=1}^{\infty} P(D_n) = \sum_{n=1}^{\infty} Q(D_n) = Q\Big(\bigcup_{n=1}^{\infty} D_n\Big) \tag{1.2.15}$$

für alle paarweise disjunkten Mengenfolgen $\{D_n\} \subseteq \mathcal{D}$ aufgrund der σ–Additivität von P bzw. Q.
Da aber $\mathcal{B}^1$ von dem durchschnittsstabilen Mengensystem $\mathcal{E}_1 \cup \{\emptyset\}$ (oder auch $\mathcal{E}_5$) erzeugt wird, und wegen

$$P((a,b]) = F(b) - F(a) = Q((a,b]),\ a,b \in \mathbf{R},\ a < b,$$

$\mathcal{E}_1$ in $\mathcal{D}$ enthalten ist, folgt nach Satz 1.2.2: $\mathcal{B}^1 = \sigma(\mathcal{E}_1) = \delta(\mathcal{E}_1) \subseteq \delta(\mathcal{D}) = \mathcal{D} \subseteq \mathcal{B}^1$, also $\mathcal{D} = \mathcal{B}^1$, was aber gerade bedeutet, daß $P(D) = Q(D)$ gilt für alle Borel'schen Mengen $D \in \mathcal{B}^1$. Die Wahrscheinlichkeitsverteilung P ist somit eindeutig bestimmt. ∎

Entsprechende Aussagen lassen sich für den Fall formulieren, daß die Wahrscheinlichkeitsverteilung P auf einer Spur-σ-Algebra $\Omega \cap \mathcal{B}^1$ definiert ist mit $\Omega \subset \mathbf{R}$. Falls Ω dabei selbst eine Borel'sche Menge ist (was in praktischen Fällen stets der Fall ist, wie die obigen Ausführungen zur Konstruktion nicht-Borel'scher Mengen zeigen), kann man die Wahrscheinlichkeitsverteilung P auch gleich auf die gesamte Borel'sche σ-Algebra $\mathcal{B}^1$ zu einer Wahrscheinlichkeitsverteilung P' fortsetzen vermöge

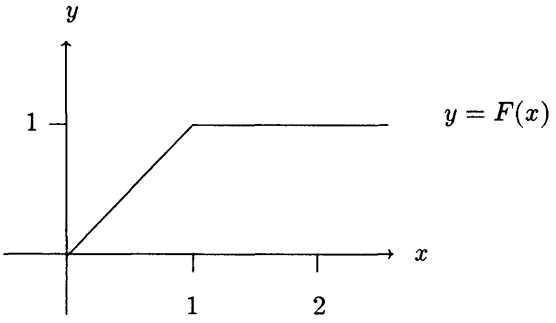
$$P'(B) := P(\Omega \cap B) \quad \text{für alle } B \in \mathcal{B}^1. \tag{1.2.16}$$

Dies bedeutet, daß den Mengen B außerhalb von Ω, also mit der Eigenschaft $B \cap \Omega = \emptyset$, die Wahrscheinlichkeit Null zugewiesen wird. In diesem Sinne kann also z.B. die eingangs betrachtete Gleichverteilung über $(0,1]$ auch als Wahrscheinlichkeitsverteilung auf ganz $\mathcal{B}^1$ angesehen werden. Zur Vereinfachung der Schreibweise werden wir deshalb im folgenden stets davon ausgehen, daß alle betrachteten Wahrscheinlichkeitsverteilungen P auf ganz $\mathcal{B}^1$ definiert sind.

Die Verteilungsfunktion der Gleichverteilung über $(0,1]$ ist in diesem Sinne gegeben durch

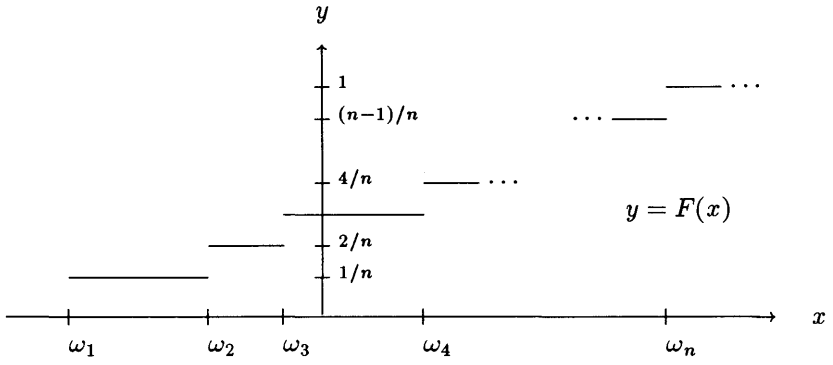
$$F(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ x & \text{für } 0 < x \leq 1 \\ 1 & \text{für } x > 1 \end{cases} \tag{1.2.17}$$

und damit eine auf ganz $\mathbf{R}$ stetige Funktion; man spricht deshalb auch von der *kontinuierlichen* oder *stetigen* Gleichverteilung über $(0,1]$.



Bezeichnet dagegen P die *diskrete* Gleichverteilung $\mathfrak{L}(\Omega)$ auf einer n-elementigen Grundmenge $\Omega = \{\omega_1, \ldots, \omega_n\} \subset \mathbf{R}$ mit $\omega_1 < \omega_2 < \ldots < \omega_n$, so ist die zugehörige Verteilungsfunktion F gegeben durch

$$F(x) = \begin{cases} 0 & \text{für } x < \omega_1 \\ \frac{k}{n} & \text{für } \omega_k \leq x < \omega_{k+1},\ 1 \leq k < n \\ 1 & \text{für } x \geq \omega_n. \end{cases}$$



Die sich hier ergebende Treppenform der Verteilungsfunktion ist typisch für diskrete Verteilungen mit *isolierten* Trägerpunkten; die Höhe der Sprünge an den Stellen ω_k entspricht wegen (1.2.6) dabei gerade den Elementarwahrscheinlichkeiten $P(\{\omega_k\})$. Im allgemeinen braucht jedoch eine diskrete Verteilungsfunktion keine Treppengestalt zu besitzen. Beispielsweise wird durch

$$F(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ \displaystyle\sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{n2^n} & \text{für } 0 < x \leq 1 \\ 1 & \text{für } x > 1 \end{cases} \tag{1.2.18}$$

eine (schwach) monoton wachsende, rechtsseitig stetige Funktion mit der Eigenschaft (1.2.7) definiert, d.h. F ist die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung P über $\Omega = (0,1]$; dabei gilt:

$$P(\{\omega\}) = \frac{1}{q} \ln\left(\frac{2^q}{2^q - 1}\right)$$

für alle rationalen Zahlen $\omega \in \Omega$ in teilerfremder Darstellung $\omega = \frac{p}{q}$ mit $p, q \in \mathbf{N}$ (vgl. Aufgabe 1.9). Die Verteilung P besitzt also den Träger $T = \mathbf{Q} \cap (0,1]$, wodurch die zugehörige Verteilungsfunktion F unstetig in allen rationalen, aber — wegen (1.2.5) — stetig in allen irrationalen Zahlen des Intervalls $(0,1]$ ist!

In praktischen Anwendungen spielen bei den kontinuierlichen Wahrscheinlichkeitsverteilungen vor allem solche eine Rolle, deren Verteilungsfunktion gewisse Differenzierbarkeitseigenschaften besitzen. Dies führt zu der im folgenden behandelten Begriffsbildung einer Verteilungsdichte.

Definition 1.2.3. *(Verteilungsdichte)*
Es sei $f : \mathbf{R} \longrightarrow \mathbf{R}^+$ eine über $\mathbf{R}$ uneigentlich Riemann-integrierbare Funktion mit

$$\int_{-\infty}^{\infty} f(x)\, dx = 1. \tag{1.2.19}$$

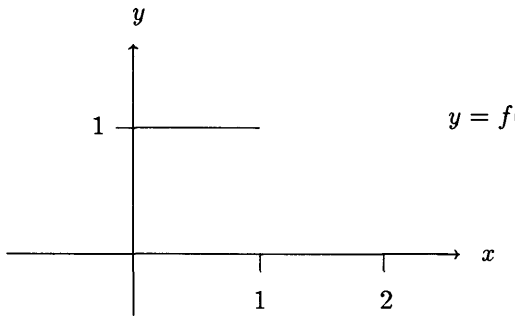
Dann wird durch

$$F(x) := \int_{-\infty}^{x} f(u)\, du, \quad x \in \mathbf{R}, \tag{1.2.20}$$

die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung P auf $\mathcal{B}^1$ definiert. Die Funktion f heißt (Verteilungs-)Dichte der Wahrscheinlichkeitsverteilung P.

Nach dem Hauptsatz der Differential- und Integralrechnung läßt sich also eine Dichte f in ihren Stetigkeitspunkten x darstellen vermöge $f(x) = F'(x)$. Im Fall der kontinuierlichen Gleichverteilung läßt sich auf diese Weise eine Dichte f bestimmen zu

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \text{ oder } x > 1 \\ 1 & \text{für } 0 \leq x \leq 1. \end{cases} \tag{1.2.21}$$



Offensichtlich ist eine Verteilungsdichte nicht eindeutig bestimmt; insbesondere gibt es gewisse Wahlfreiheiten in den Unstetigkeitspunkten x von f, d.h. den Punkten x, in denen F nicht differenzierbar ist. Beispielsweise könnte man in (1.2.21) der Dichte f an den Stellen 0 bzw. 1 auch den Funktionswert 0 zuweisen, ohne daß die Dichte–Eigenschaft verloren geht. Man kann daher die Gleichverteilung über dem Intervall $[0,1]$ ebenso gut mit der Gleichverteilung über $(0,1)$, $[0,1)$ oder $(0,1]$ identifizieren.

Ist allgemeiner $A \in \mathcal{B}^1$ eine Menge mit $c = \int_A dx > 0$, so ist die (stetige) *Gleichverteilung über A* gegeben durch die Dichte

$$f(x) = \begin{cases} 0 & \text{für } x \notin A \\ \frac{1}{c} & \text{für } x \in A, \end{cases} \tag{1.2.22}$$

in Zeichen: $\mathcal{R}(A)$ ("Rechteckverteilung" über A). Ist beispielsweise $A = [a,b]$ mit $a, b \in \mathbf{R}, a < b$ ein Intervall, so ist $c = b - a$; die zugehörige Dichte f hat dann die Form eines Rechtecks, woraus sich die Namensgebung ableitet.

Bei diskreten Verteilungen P spricht man anstatt von Elementarwahrscheinlichkeiten $P(\{\omega\})$, $\omega \in \Omega$ auch von *Zähldichten* $f(x) = P(\{x\})$, $x \in \mathbf{R}$, wegen der zu (1.2.20) formalen Analogie

$$F(x) = \sum_{u \leq x} f(u), \quad x \in \mathbf{R}, \tag{1.2.23}$$

für die zugehörige Verteilungsfunktion F. Man beachte dabei, daß die Summe in (1.2.23) wohldefiniert ist, da höchstens abzählbar viele Summanden von Null verschieden sind.

Außer den bisher behandelten diskreten und kontinuierlichen Wahrscheinlichkeitsverteilungen existieren noch sogenannte *singuläre* Wahrscheinlichkeitsverteilungen, das sind im wesentlichen solche, deren zugehörige Verteilungsfunktion zwar stetig ist, die aber keine Integraldarstellung (1.2.20) besitzen. Beispielsweise läßt sich für $p_0, p_1 \in (0,1)$ mit $p_0 + p_1 = 1$ und $p_0, p_1 \neq \frac{1}{2}$ eine Verteilungsfunktion F mit $F(0) = 0$, $F(1) = 1$ angeben mit der Eigenschaft

$$F\Big(\frac{k+1}{2^n}\Big) - F\Big(\frac{k}{2^n}\Big) = \prod_{i=1}^{n} p_{j_i} \tag{1.2.24}$$

für alle $n \in \mathbb{N}$, $0 \leq k < 2^n$, wobei $(j_1\ j_2\ \ldots\ j_n) \in \{0,1\}^n$ die Binärdarstellung von k bedeute, also

$$k = \sum_{i=1}^{n} j_i 2^{n-i}$$

gilt. Nach (1.2.24) ist F streng monoton auf $[0,1]$ und darüberhinaus stetig; dennoch gilt $F'(x) = 0$ für *jeden* Punkt $x \in \mathbb{R}$, in dem F differenzierbar ist, weshalb (1.2.20) hier nicht gelten kann. Da solche Verteilungen praktisch nicht von Interesse sind, werden wir sie im folgenden nicht weiter behandeln. Der interessierte Leser sei aber auf Billingsley (1985), Example 31.1 verwiesen.

Zur Unterscheidung singulärer Wahrscheinlichkeitsverteilungen P mit stetiger Verteilungsfunktion F von kontinuierlichen Wahrscheinlichkeitsverteilungen P mit der Eigenschaft (1.2.20) nennt man letztere auch *absolut–stetige* Verteilungen. Wir werden hier allerdings die Begriffe stetige und absolut–stetige Verteilung synonym verwenden.

Für absolut–stetige Verteilungen P mit Verteilungsdichte f wollen wir im folgenden auch die Schreibweise

$$\int_B f(u)\, du := P(B),\ B \in \mathcal{B}^1 \tag{1.2.25}$$

verwenden, was für Intervalle B mit der üblichen Integral–Schreibweise zusammenfällt. Tatsächlich wird durch die Beziehung (1.2.25) bzw. der dahinterstehenden Maßfortsetzung im Sinne von (1.1.64) und (1.1.65) aber eine Erweiterung des Riemann'schen Integrationsbegriffs vorgenommen, indem nunmehr Integrationen über beliebige Borel'sche Mengen möglich sind. Hierauf werden wir später im Zusammenhang mit dem sogenannten Lebesgue–Integral noch einmal zurückkommen.

1.3. Zufallsvariablen und ihre Verteilung

Bisher haben wir uns auf die analytische Beschreibung von Wahrscheinlichkeitsverteilungen auf der Borel'schen σ–Algebra $\mathcal{B}^1$ beschränkt, ohne jedoch näher darauf einzugehen, wie man im allgemeinen geeignete stochastische Modelle zur Beschreibung konkreter Situationen, in denen Zufall eine Rolle spielt, erhalten kann. Eine solche allgemeine Vorgehensweise muß insbesondere dem Umstand Rechnung tragen, daß häufig die zu beschreibenden zufälligen Ergebnisse in Form reeller Zahlen vorliegen, wie z.B. das Problem der Bestimmung der Verteilung der (zufälligen) Schrittzahl des binären Suchens — siehe Beispiel 1.1.1 — zeigt. In der Regel will man dabei zugleich auch arithmetische Verknüpfungen solcher Ergebnisse mitberücksichtigen, etwa die durchschnittliche Schrittzahl des Suchverfahrens bei n–maliger Ausführung, um hieraus Güteeigenschaften des Algorithmus abzuleiten. Ein adäquates Hilfsmittel zur Modellierung solcher Situationen stellen die sogenannten Zufallsvariablen dar, die im folgenden ausführlicher behandelt werden.

Definition 1.3.1. *(Zufallsvariable)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbf{R}$ eine Abbildung mit der Eigenschaft

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A} \text{ für alle } B \in \mathcal{B}^1. \tag{1.3.1}$$

Dann heißt X Zufallsvariable auf Ω.

Die durch (1.3.1) beschriebene Eigenschaft der Abbildung X nennt man auch *Meßbarkeit* von X. Diese Eigenschaft wollen wir im folgenden durch die Schreibweise

$$X : (\Omega, \mathcal{A}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$$

zum Ausdruck bringen. Man spricht dabei von den Paaren $(\Omega, \mathcal{A})$ und $(\mathbf{R}, \mathcal{B}^1)$, also Grundmengen mit darüber definierten σ–Algebren auch von *Meßräumen*.

Will man die Meßbarkeit von Zufallsvariablen nachweisen, braucht man allerdings Beziehung (1.3.1) nicht für alle Mengen $B \in \mathcal{B}^1$ zu verifizieren; es reicht vielmehr aus, Mengen B eines (beliebigen) Erzeugers $\mathcal{E}$ von $\mathcal{B}^1$, z.B. $\mathcal{E} = \mathcal{E}_5$, auszuwählen. Das System $\mathcal{Q}$ aller Mengen $B \in \mathfrak{P}(\mathbf{R})$ mit $X^{-1}(B) \in \mathcal{A}$ ist nämlich eine σ–Algebra über $\mathbf{R}$, so daß $\mathcal{B}^1 \subseteq \mathcal{Q}$ gilt, wenn $\mathcal{E} \subseteq \mathcal{Q}$ gilt, was aber mit Beziehung (1.3.1) zusammenfällt, wenn dort $\mathcal{B}^1$ durch $\mathcal{E}$ ersetzt wird.

Durch eine Zufallsvariable X erhält man in kanonischer Weise eine Wahrscheinlichkeitsverteilung Q auf $\mathcal{B}^1$ vermöge

Definition 1.3.2. *(Verteilung einer Zufallsvariablen)*
Es sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann wird durch

$$Q(B) := P(\{X \in B\}) = P(X^{-1}(B)), \quad B \in \mathcal{B}^1, \tag{1.3.2}$$

eine Wahrscheinlichkeitsverteilung Q auf $\mathcal{B}^1$ definiert. Diese heißt die Verteilung von X, in Zeichen: $Q = P^X$.

In der Tat wird durch (1.3.1) eine Wahrscheinlichkeitsverteilung auf $\mathcal{B}^1$ definiert, da für paarweise disjunkte Mengenfolgen $\{B_n\} \subseteq \mathcal{B}^1$ aufgrund der Urbildeigenschaften stets gilt

$$\{X \in \bigcup_{n=1}^{\infty} B_n\} = X^{-1}\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \bigcup_{n=1}^{\infty} X^{-1}(B_n) = \bigcup_{n=1}^{\infty} \{X \in B_n\} \tag{1.3.3}$$

— mit den paarweise disjunkten Mengen $\{X \in B_n\}$) — und damit Q tatsächlich σ–additiv ist. $Q = P^X$ heißt auch *Bildmaß* von P unter X.

Die Zufallsvariable X repräsentiert durch ihre Werte $X(\omega)$, $\omega \in \Omega$, die möglichen, durch Zufallseinfluß auftretenden Ergebnisse in dem zu beschreibenden Modell. Die Struktur des zugrundeliegenden Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, P)$ oder die explizite Definition der Abbildung X spielt dabei in der Regel keine Rolle; vielmehr interessiert die über dem Wertebereich von X gegebene Verteilung $Q = P^X$. Gibt man sich eine solche Verteilung Q auf $\mathcal{B}^1$ vor, so kann man stets durch die Wahl

$$\Omega = \mathbf{R},\ \mathcal{A} = \mathcal{B}^1,\ X(\omega) = \omega \text{ für alle } \omega \in \Omega \tag{1.3.4}$$

eine Darstellung der Verteilung Q als Bildmaß P^X einer geeigneten Zufallsvariablen X erreichen. Wir werden deshalb im folgenden grundsätzlich davon ausgehen, daß Zufallsvariablen immer auf geeigneten Grundmengen definiert sind, ohne jeweils diese Mengen oder die zugehörigen Abbildungsvorschriften genauer zu spezifizieren.

Gelegentlich ist es jedoch möglich, in einfacher Weise explizit Zufallsvariablen auf der Grundmenge anzugeben, etwa im Beispiel 1.1.1 des binären Suchens.

Beispiel 1.3.1. (`binary search`)
Es bezeichne X die Zufallsvariable, welche die benötigte Anzahl der Schritte bis zum Abbruch des Verfahrens angibt. X läßt sich mit den Bezeichnungen aus Beziehung (1.1.2) ausdrücken durch

$$X(\omega) = \begin{cases} k & \text{für } \omega \in A_k,\ 1 \le k < n \\ n & \text{für } \omega \in A_0 \cup A_n, \end{cases} \tag{1.3.5}$$

wobei wieder $A_0 = \{0\}$ zu setzen ist. Die Abbildung X ordnet also wie gewünscht allen Platznummern ω, die in $k \in \{1, \ldots, n-1\}$ Schritten erreichbar sind, den Wert k zu; der Wert n wird angenommen für den Fall, daß die maximale Schrittzahl benötigt wird (ungünstigster Fall) bzw. das Schlüsselelement nicht in dem Feld vorhanden ist. Da X nur endlich viele Werte $\{1, \ldots, n\}$ annimmt, ist die Verteilung P^X diskret und daher eindeutig bestimmt durch die Angabe

$$P(X = k) = \begin{cases} P(A_k) & = 2^{k-1-n} & \text{für } 1 \le k < n \\ P(A_0 \cup A_n) & = P(A_0) + P(A_n) = \frac{1}{2^n} + \frac{1}{2} & \text{für } k = n \end{cases} \tag{1.3.6}$$

der zugehörigen Elementarwahrscheinlichkeiten. ∎

Hierbei haben wir die kürzere und prägnantere Schreibweise $P(X = k)$ für das umständlichere $P(\{X \in \{k\}\})$ gewählt. Entsprechend kann das Ereignis B_k aus (1.1.4), daß das Schlüsselelement in höchstens $k < n$ Schritten gefunden wird, beschrieben werden durch $\{X \le k\}$ bzw.

$$P(X \le k) = \sum_{i=1}^{k} P(X = i) = \sum_{i=1}^{k} \frac{2^{i-1}}{2^n} = \frac{2^k - 1}{2^n}.$$

Die Wahrscheinlichkeit dafür, daß das Schlüsselelement in genau n Schritten *gefunden* wird, läßt sich dagegen *nicht* über X berechnen, da das Ereignis $\{X = n\} =$

$A_n \cup A_0$ sowohl das Auffinden des Schlüsselelements im n–ten Schritt als auch das Nichtvorhandensein des Elements im Feld beschreibt. Will man das Ereignis des Auffindens des Schlüsselelements in genau n Schritten durch eine Zufallsvariable beschreiben, so kann man etwa

$$Y(\omega) = \begin{cases} X(\omega) & \text{für } \omega \in \bigcup_{i=1}^{n-1} A_i \\ 0 & \text{für } \omega \in A_0 \\ n & \text{für } \omega \in A_n \end{cases} \tag{1.3.7}$$

wählen; es ist dann

$$P(Y = n) = P(A_n) = \frac{1}{2}. \tag{1.3.8}$$

Das Ereignis $\{Y = 0\}$ beschreibt den Fall, daß das Schlüsselelement nicht in dem Feld vorhanden ist.

Entsprechend lassen sich über Zufallsvariable auch bedingte Wahrscheinlichkeiten berechnen. Ist man z.B. an der Wahrscheinlichkeit dafür interessiert, daß das Schlüsselelement im Falle des Vorhandenseins (d.h. $Y \geq 1$) in $k \leq n$ Schritten gefunden wird (vgl. (1.1.7)), so ergibt sich

$$\begin{aligned} P(Y = k \mid 1 \leq Y \leq n) &= \frac{P(Y = k)}{P(1 \leq Y \leq n)} = \frac{P(Y = k)}{1 - P(Y = 0)} \\ &= \frac{2^{k-1-n}}{1 - 2^{-n}} = \frac{2^{k-1}}{2^n - 1}. \end{aligned} \tag{1.3.9}$$

Im allgemeinen ist bei Modellierungen stochastischer Vorgänge die gleichzeitige Betrachtung verschiedener Zufallsvariablen auf demselben Wahrscheinlichkeitsraum nötig, wie das vorangehende Beispiel zeigt. Diese Zufallsvariablen sind dabei in der Regel nicht unabhängig voneinander — siehe die Definition von Y in (1.3.7). Man ist deshalb auch an der *gemeinsamen* Verteilung solcher Zufallsvariablen interessiert, in obigem Beispiel also an den Wahrscheinlichkeiten $P(X = k, Y = j) := P(\{X = k\} \cap \{Y = j\})$ für $1 \leq k \leq n, 0 \leq j \leq n$. Aus den Beziehungen (1.3.5), (1.3.6) und (1.3.7) erhält man hier etwa

$$P(X = k, Y = j) = \begin{cases} 0 & \text{für } 1 \leq k, j < n, k \neq j \\ P(X = k) = 2^{k-1-n} & \text{für } 1 \leq k = j < n \\ P(Y = 0) = 2^{-n} & \text{für } k = n, j = 0 \\ P(Y = n) = \frac{1}{2} & \text{für } k = j = n. \end{cases} \tag{1.3.10}$$

Insbesondere zeigt sich hier, daß die Ereignisse $\{X = k\}$ und $\{Y = j\}$ für $1 \leq k \leq n, 0 \leq j \leq n$, nicht stochastisch unabhängig voneinander (im Sinne von Definition 1.1.6) sind. Beispielsweise gilt

$$P(Y = 0 \mid X = n) = \frac{P(X = n, Y = 0)}{P(X = n)} = \frac{2^{-n}}{2^{-n} + 2^{-1}} > \frac{1}{2^n} = P(Y = 0) \tag{1.3.11}$$

für $n > 1$, wobei $P(Y = 0 \mid X = n)$ die Wahrscheinlichkeit dafür angibt, daß das Schlüsselelement nicht in dem Feld vorhanden ist, wenn bereits die maximale Anzahl n von Suchschritten ausgeführt wurde. Die Ungleichung in (1.3.11) spiegelt

dabei die intuitiv einleuchtende Tatsache wider, daß die (bedingte) Wahrscheinlichkeit für das Nichtvorhandensein des Schlüsselelements mit der Anzahl der bereits ausgeführten Schritte wächst; genauer gilt hier

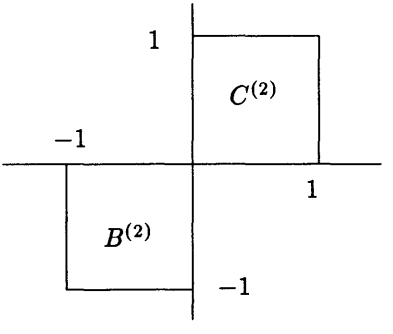
$$P(Y=0 \mid X \geq k) = \frac{P(Y=0)}{1-P(X<k)} = \frac{2^{-n}}{1-\sum_{i=1}^{k-1} 2^{i-1-n}}$$
$$= \frac{1}{2^n - 2^{k-1}+1} > \frac{1}{2^n - 2^{k-2}+1} = P(Y=0 \mid X \geq k-1)$$

für $1 < k \leq n$. Umgekehrt ist natürlich $P(X=n \mid Y=0) = 1$, da zur Feststellung des Nichtvorhandenseins des Schlüsselelements in dem Feld in jedem Fall n Suchschritte ausgeführt werden müssen.

Tatsächlich wird durch (1.3.10) eine diskrete Wahrscheinlichkeitsverteilung Q über der neuen Grundmenge $\Omega^{(2)} = \{1,\ldots,n\} \times \{0,\ldots,n\} \subset \mathbf{R}^2$ definiert vermöge $Q(\{(k,j)\}) = P(X=k, Y=j)$, $(k,j) \in \Omega^{(2)}$. Entsprechend gelangt man zu Verteilungen über Grundmengen $\Omega^{(m)} \subseteq \mathbf{R}^m$, wenn m Zufallsvariablen auf der ursprünglichen Grundmenge Ω betrachtet werden. Offensichtlich macht dies keine Probleme, wenn die verwendeten Zufallsvariablen diskrete Verteilungen besitzen; die daraus abgeleiteten Verteilungen über $\Omega^{(m)}$ sind dann nämlich ebenfalls diskret und können wie in (1.3.10) durch Angabe entsprechender Elementarwahrscheinlichkeiten eindeutig beschrieben werden. Probleme treten allerdings dann auf, wenn eine oder mehrere der betrachteten Zufallsvariablen stetige Verteilungen besitzen, da man dann über der Menge $\mathbf{R}^m$ eine geeignete σ–Algebra finden muß, um eine "gemeinsame" Verteilung der Zufallsvariablen beschreiben zu können. Es ist naheliegend, für eine solche σ–Algebra zu verlangen, daß sie zumindest kartesische Produkte (eindimensionaler) Borel'scher Mengen, also Mengen der Art

$$B^{(m)} = \bigtimes_{i=1}^{m} B_i \quad \text{mit} \quad B_1,\ldots,B_m \in \mathcal{B}^1$$

enthält. Man macht sich sofort klar, daß das System aller so gebildeten Mengen *keine* σ–Algebra über $\mathbf{R}^m$ bildet, da es z.B. nicht vereinigungsstabil ist. Dies zeigt etwa das Beispiel $m=2$, $B^{(2)} = [-1,0] \times [-1,0]$, $C^{(2)} = [0,1] \times [0,1]$; offenbar ist $B^{(2)} \cup C^{(2)}$ *kein* kartesisches Produkt zweier Teilmengen von $\mathbf{R}$, also insbesondere auch nicht zweier Borel'scher Teilmengen von $\mathbf{R}$.



In dem nachfolgenden Abschnitt werden wir uns deshalb mit geeigneten (sog. Produkt–)σ–Algebren über $\mathbf{R}^m$, $m \in \mathbf{N}$, sowie der gemeinsamen Verteilung endlich wie auch unendlich vieler Zufallsvariablen beschäftigen.

1.4. Produkträume und Zufallsvektoren

In diesem Abschnitt wollen wir uns zunächst mit dem Problem beschäftigen, wie man über dem m–dimensionalen Raum $\mathbf{R}^m$, $m \in \mathbf{N}$, bzw. allgemeiner über einem kartesischen Produkt $\Omega^{(m)} = \bigtimes_{i=1}^{m} \Omega_i$ von Grundmengen Ω_i mit darüber definierten σ–Algebren $\mathcal{A}_i$ in konsistenter Weise umfassendere σ–Algebren definieren kann. Solche σ–Algebren sollten sinnvollerweise das System aller Produktmengen $\bigtimes_{i=1}^{m} A_i$ mit Mengen $A_i \in \mathcal{A}_i$, $1 \leq i \leq m$ – welches i.a. selbst keine σ–Algebra ist, wie die Bemerkungen am Ende des vorigen Abschnitts zeigen — enthalten. Wählt man die kleinste σ–Algebra, die dieses System umfaßt, gelangt man zu der folgenden Begriffsbildung.

Definition 1.4.1. *(Produkt–σ–Algebra)*
Es seien $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq m$, $m \in \mathbf{N}$, Meßräume und

$$\mathcal{E}^{(m)} = \left\{ \bigtimes_{i=1}^{m} A_i \mid A_i \in \mathcal{A}_i,\ 1 \leq i \leq m \right\} \tag{1.4.1}$$

das System aller kartesischen Produkte aus den Mengen der gegebenen σ–Algebren. Dann heißt die durch

$$\bigotimes_{i=1}^{m} \mathcal{A}_i = \sigma(\mathcal{E}^{(m)}) \tag{1.4.2}$$

definierte σ–Algebra über dem kartesischen Produkt $\Omega^{(m)} = \bigtimes_{i=1}^{m} \Omega_i$ der Grundmengen $\Omega_1, \ldots, \Omega_m$ die Produkt–σ–Algebra der σ–Algebren $\mathcal{A}_1, \ldots, \mathcal{A}_m$.

Das Paar $(\Omega^{(m)}, \bigotimes_{i=1}^{m} \mathcal{A}_i)$ heißt auch das Produkt der Meßräume $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq m$.

Die Bildung von Produkt–σ–Algebren ist dabei *assoziativ*, d.h. es gilt

$$\bigotimes_{i=1}^{m} \mathcal{A}_i = \bigotimes_{i=1}^{r} \mathcal{A}_i \otimes \bigotimes_{i=r+1}^{m} \mathcal{A}_i \tag{1.4.3}$$

für alle $1 \leq r < m$.

Dies ist eine Konsequenz aus dem folgenden allgemeinen

Lemma 1.4.1. *(Erzeugungslemma für Produkt–σ–Algebren)*
In der Situation von Definition 1.4.1 seien $\mathcal{E}_i$, $1 \leq i \leq m$, Erzeuger der σ–Algebren $\mathcal{A}_i$, $1 \leq i \leq m$, derart, daß jeder Erzeuger $\mathcal{E}_i$ eine monoton wachsende Mengenfolge $\{E_{ik}\}_{k \in \mathbf{N}}$ enthalte mit

$$\bigcup_{k=1}^{\infty} E_{ik} = \lim_{k \to \infty} E_{ik} = \Omega_i, \quad 1 \leq i \leq m.$$

Dann ist auch

$$\mathcal{F}^{(m)} = \Big\{ \bigtimes_{i=1}^{m} E_i \mid E_i \in \mathcal{E}_i,\ 1 \le i \le m \Big\} \tag{1.4.4}$$

ein Erzeuger der Produkt-σ-Algebra $\bigotimes_{i=1}^{m} \mathcal{A}_i$.

Beweis. Es ist nur zu zeigen, daß die Produkt-σ-Algebra $\bigotimes_{i=1}^{m} \mathcal{A}_i$ in der σ-Algebra $\sigma(\mathcal{F}^{(m)})$ enthalten ist. Dazu reicht der Nachweis, daß der Erzeuger $\mathcal{E}^{(m)}$ der Definition 1.4.1 in der letzteren σ-Algebra enthalten ist. Die Mengensysteme

$$\mathcal{A}_i^{(m)} = \Big\{ \bigtimes_{j=1}^{i-1} \Omega_j \times A_i \times \bigtimes_{j=i+1}^{m} \Omega_j \mid A_i \in \mathcal{A}_i \Big\}, \quad 1 \le i \le m,$$

(für $i = 1$ bzw. $i = m$ entfällt dabei das linke bzw. rechte (leere) kartesische Produkt) bilden offensichtlich σ-Algebren über $\Omega^{(m)}$, welche von den Mengensystemen

$$\mathcal{F}_i^{(m)} = \Big\{ \bigtimes_{j=1}^{i-1} \Omega_j \times E_i \times \bigtimes_{j=i+1}^{m} \Omega_j \mid E_i \in \mathcal{E}_i \Big\}, \quad 1 \le i \le m,$$

erzeugt werden. Nach Voraussetzung ist aber für jedes $E_i \in \mathcal{E}_i$

$$\bigtimes_{j=1}^{i-1} \Omega_j \times E_i \times \bigtimes_{j=i+1}^{m} \Omega_j = \bigcup_{k=1}^{\infty} \left[\bigtimes_{j=1}^{i-1} E_{jk} \times E_i \times \bigtimes_{j=i+1}^{m} E_{jk} \right],$$

so daß $\mathcal{F}_i^{(m)} \subseteq \sigma(\mathcal{F}^{(m)})$ und daher auch $\mathcal{A}_i^{(m)} \subseteq \sigma(\mathcal{F}^{(m)})$ für alle $1 \le i \le m$ folgt. Aufgrund von

$$\bigtimes_{i=1}^{m} A_i = \bigcap_{i=1}^{m} \left[\bigtimes_{j=1}^{i-1} \Omega_j \times A_i \times \bigtimes_{j=i+1}^{m} \Omega_j \right]$$

für alle $A_i \in \mathcal{A}_i$, $1 \le i \le m$, ist also auch $\mathcal{E}^{(m)} \subseteq \sigma(\mathcal{F}^{(m)})$ und somit $\bigotimes_{i=1}^{m} \mathcal{A}_i \subseteq \sigma(\mathcal{F}^{(m)})$, was zu zeigen war. ∎

Auf die einschränkenden Voraussetzungen an die Erzeuger in Lemma 1.4.1 kann nicht verzichtet werden, wie das Beispiel $\mathcal{A}_1 = \{\emptyset, \mathbf{R}\}$, $\mathcal{E}_1 = \{\emptyset\}$, $\mathcal{A}_2 = \mathcal{E}_2 = \mathcal{B}^1$ zeigt; $\mathcal{E}^{(2)}$ erzeugt nämlich lediglich die kleinere σ-Algebra $\{\emptyset, \mathbf{R}^2\}$, nicht aber die Produkt-σ-Algebra $\mathcal{A}_1 \otimes \mathcal{B}^1$, die z.B. alle Mengen der Form $\mathbf{R} \times B$ mit $B \in \mathcal{B}^1$ enthält.

Für den Fall $\Omega_i = \mathbf{R}$, $\mathcal{A}_i = \mathcal{B}^1$, $1 \le i \le m$, heißt die resultierende Produkt-σ-Algebra $\mathcal{B}^m = \bigotimes_{i=1}^{m} \mathcal{A}_i$ wieder die (m-dimensionale) Borel'sche σ-Algebra (über $\mathbf{R}^m$). Die durch (1.4.3) gegebene Assoziativitätseigenschaft drückt sich damit einfacher aus als

$$\mathcal{B}^r \otimes \mathcal{B}^s = \mathcal{B}^{r+s} \quad \text{für alle } r, s \in \mathbf{N}. \tag{1.4.5}$$

Die Struktur höherdimensionaler Borel'scher σ–Algebren ist naturgemäß vielfältiger als diejenige von $\mathcal{B}^1$; so enthält etwa $\mathcal{B}^2$ auch alle offenen und abgeschlossenen Kreise $K^o(x,y;r)$ bzw. $K^a(x,y;r)$ mit Mittelpunkt $(x,y) \in \mathbf{R}^2$ und Radius $r > 0$. Die erste Aussage ergibt sich z.B. aus der abzählbaren Darstellung

$$K^o(x,y;r) = \bigcup_{(q_1,q_2)\in(\mathbf{Q}\times\mathbf{Q})\cap K^o(x,y;r)} (q_1-\delta, q_1+\delta)\times(q_2-\delta,q_2+\delta) \tag{1.4.6}$$

mit $\delta = \delta(q_1,q_2,x,y,r) = \frac{1}{\sqrt{2}}\left(r - \sqrt{(q_1-x)^2+(q_2-y)^2}\right)$, wobei $(q_1-\delta,q_1+\delta)\times(q_2-\delta,q_2+\delta) \in \mathcal{B}^2$ ist. Abgeschlossene Kreise erhält man dann z.B. aus offenen vermöge

$$K^a(x,y;r) = \bigcap_{n=1}^{\infty} K^o(x,y;r+\frac{1}{n}).$$

Analog zu (1.1.53) bis (1.1.61) bilden die folgenden Mengensysteme Erzeuger von $\mathcal{B}^m$, $m \in \mathbf{N}$:

$$\mathcal{E}_1^{(m)} = \{\bigtimes_{i=1}^{m}(a_i,b_i] \mid a_i,b_i \in \mathbf{R},\ a_i < b_i,\ 1 \le i \le m\} \tag{1.4.7}$$

$$\mathcal{E}_2^{(m)} = \{\bigtimes_{i=1}^{m}[a_i,b_i) \mid a_i,b_i \in \mathbf{R},\ a_i < b_i,\ 1 \le i \le m\} \tag{1.4.8}$$

$$\mathcal{E}_3^{(m)} = \{\bigtimes_{i=1}^{m}(a_i,b_i) \mid a_i,b_i \in \mathbf{R},\ a_i < b_i,\ 1 \le i \le m\} \tag{1.4.9}$$

$$\mathcal{E}_4^{(m)} = \{\bigtimes_{i=1}^{m}[a_i,b_i] \mid a_i,b_i \in \mathbf{R},\ a_i < b_i,\ 1 \le i \le m\} \tag{1.4.10}$$

$$\mathcal{E}_5^{(m)} = \{\bigtimes_{i=1}^{m}(-\infty,b_i] \mid b_i \in \mathbf{R},\ 1 \le i \le m\} \tag{1.4.11}$$

$$\mathcal{E}_6^{(m)} = \{\bigtimes_{i=1}^{m}(-\infty,b_i) \mid b_i \in \mathbf{R},\ 1 \le i \le m\} \tag{1.4.12}$$

$$\mathcal{E}_7^{(m)} = \{G \subseteq \mathbf{R}^m \mid G \text{ offen}\} \tag{1.4.13}$$

$$\mathcal{E}_8^{(m)} = \{F \subseteq \mathbf{R}^m \mid F \text{ abgeschlossen}\} \tag{1.4.14}$$

$$\mathcal{E}_9^{(m)} = \{K \subseteq \mathbf{R}^m \mid K \text{ kompakt}\}. \tag{1.4.15}$$

Die Erzeugereigenschaft von $\mathcal{E}_i^{(m)}$, $1 \le i \le 6$, ergibt sich dabei unmittelbar aus Lemma 1.4.1 sowie den Beziehungen (1.1.53) bis (1.1.58). Ähnlich wie in (1.4.6) läßt sich ferner zeigen, daß jede offene Teilmenge des $\mathbf{R}^m$ als abzählbare Vereinigung von Mengen aus $\mathcal{E}_3^{(m)}$ darstellbar, also insbesondere Borel'sch ist und daher $\mathcal{B}^m = \sigma(\mathcal{E}_3^{(m)}) \subseteq \sigma(\mathcal{E}_7^{(m)}) \subseteq \mathcal{B}^m$, also die Erzeugereigenschaft von $\mathcal{E}_7^{(m)}$ folgt. Wegen $\mathcal{E}_8^{(m)} = \{G^c \mid G \in \mathcal{E}_7^{(m)}\}$ ist auch $\mathcal{E}_8^{(m)}$ ein Erzeuger von $\mathcal{B}^m$. Entsprechendes gilt

für $\mathcal{E}_9^{(m)}$, da im $\mathbf{R}^m$ die kompakten gerade die abgeschlossenen und beschränkten Mengen sind.

Ähnlich wie in Abschnitt 1.2 lassen sich Wahrscheinlichkeitsverteilungen über $\mathcal{B}^m$ wieder durch ihre Verteilungsfunktionen charakterisieren, die allgemeiner wie folgt definiert sind.

Definition 1.4.2. *(m–dimensionale Verteilungsfunktion)*
Es sei P ein Wahrscheinlichkeitsmaß auf der Borel'schen σ–Algebra $\mathcal{B}^m$, $m \in \mathbf{N}$. Die durch

$$F_P(x_1, \ldots, x_m) = P\Big(\bigtimes_{i=1}^{m} (-\infty, x_i] \Big), \quad (x_1, \ldots, x_m) \in \mathbf{R}^m, \tag{1.4.16}$$

definierte Funktion heißt die zu P gehörige (m–dimensionale) Verteilungsfunktion.

Analog zu Lemma 1.2.1 lassen sich auch für m–dimensionale Verteilungsfunktionen charakteristische Eigenschaften aus den allgemeinen Rechenregeln für Wahrscheinlichkeiten ableiten. Die übliche Monotonie–Eigenschaft eindimensionaler Verteilungsfunktionen läßt sich dabei allerdings nicht auf einfache Weise auf den mehrdimensionalen Fall (etwa durch komponentenweise Ordnung) übertragen; vielmehr wird sich der folgende allgemeine Monotonie–Begriff als wesentlich herausstellen.

Definition 1.4.3. *(Δ–Monotonie)*
Eine Funktion $g : \mathbf{R}^m \longrightarrow \mathbf{R}$, $m \in \mathbf{N}$ heißt Δ–monoton, wenn gilt

$$\Delta g_{\boldsymbol{x}}^{\boldsymbol{y}} := \sum_{(\epsilon_1, \ldots, \epsilon_m) \in \{0,1\}^m} (-1)^{\sum_{i=1}^m \epsilon_i} g(\epsilon_1 x_1 + (1-\epsilon_1) y_1, \ldots, \epsilon_m x_m + (1-\epsilon_m) y_m) \geq 0 \tag{1.4.17}$$

für alle $\boldsymbol{x} = (x_1, \ldots, x_m)$, $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbf{R}^m$ mit $x_i \leq y_i$, $1 \leq i \leq m$.

Für $m = 1$ fällt also die Δ–Monotonie mit der üblichen (schwachen) Monotonie zusammen.

Lemma 1.4.2. *(Eigenschaften der Verteilungsfunktion)*
Es sei $F = F_P$ die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung über $\mathcal{B}^m$, $m \in \mathbf{N}$. Dann gilt:

$$\Delta F_{\boldsymbol{x}}^{\boldsymbol{y}} \geq 0 \text{ für alle } \boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^m,\ x_i \leq y_i,\ 1 \leq i \leq m \quad (\Delta - \textit{Monotonie von } F) \tag{1.4.18}$$

$$P\Big(\bigtimes_{i=1}^{m} (a_i, b_i] \Big) = \Delta_{(a_1, \ldots, a_m)}^{(b_1, \ldots, b_m)} F,\ a_i, b_i \in \mathbf{R},\ a_i < b_i, 1 \leq i \leq m \tag{1.4.19}$$

$$\lim_{x_1 \downarrow y_1, \ldots, x_m \downarrow y_m} F(x_1, \ldots, x_m) = F(y_1, \ldots, y_m),\ (y_1, \ldots, y_m) \in \mathbf{R}^m \quad (\textit{rechtsseitige Stetigkeit von } F) \tag{1.4.20}$$

$$\lim_{x_1 \uparrow y_1, \ldots, x_m \uparrow y_m} F(x_1, \ldots, x_m) = F(y_1, \ldots, y_m)$$
$$\iff P(\{\boldsymbol{y}\}) = 0, \ \boldsymbol{y} \in \mathbf{R}^m \tag{1.4.21}$$

$$P(\{\boldsymbol{y}\}) = F(\boldsymbol{y}) - \lim_{x_1 \uparrow y_1, \ldots, x_m \uparrow y_m} F(x_1, \ldots, x_m), \ \boldsymbol{y} \in \mathbf{R}^m \tag{1.4.22}$$

$$\lim_{x_1 \uparrow \infty, \ldots, x_m \uparrow \infty} F(x_1, \ldots, x_m) = 1,$$
$$\lim_{x_i \downarrow -\infty} F(x_1, \ldots, x_m) = 0, \quad 1 \leq i \leq m. \tag{1.4.23}$$

Der Beweis dieses Lemmas ist völlig analog zum Beweis von Lemma 1.2.1, wobei hier Beziehung (1.4.19), die man leicht mittels vollständiger Induktion herleiten kann, eine wesentliche Rolle spielt.

Der in (1.4.19) gegebene enge Zusammenhang zwischen Wahrscheinlichkeiten und Δ– Monotonie der Verteilungsfunktion erklärt auch, warum die komponentenweise Monotonie der Verteilungsfunktion (welche trivialerweise aus der Δ–Monotonie folgt) i.a. nicht zur Charakterisierung mehrdimensionaler Verteilungsfunktionen ausreicht. Analog zu Satz 1.2.1 gilt hier:

Satz 1.4.1. *(Fortsetzungssatz für m–dimensionale Verteilungsfunktionen)*
Es sei $F : \mathbf{R}^m \longrightarrow [0,1]$, $m \in \mathbf{N}$ eine Abbildung mit den Eigenschaften (1.4.18), (1.4.20) und (1.4.23). Dann existiert eine Wahrscheinlichkeitsverteilung P auf $\mathcal{B}^m$ derart, daß F genau die zu P gehörige Verteilungsfunktion darstellt, d.h. es gilt $F = F_P$. Insbesondere gilt Beziehung (1.4.19).

Beweis. Vermöge der Δ–Monotonie von F läßt sich die gewünschte Verteilung P zunächst durch die Gleichung in (1.4.19) auf den Erzeuger $\mathcal{E}_1^{(m)}$ fortsetzen. Bezeichnet analog zur eindimensionalen Situation $\mathcal{R}^{(m)}$ den von $\mathcal{E}_1^{(m)}$ erzeugten Ring, so ist wieder $\mathcal{R}^{(m)} = \{ \bigcup_{i=1}^n E_i \mid E_i \in \mathcal{E}_1^{(m)}, \ n \in \mathbf{N} \}$ mit der Möglichkeit, jedes Ringelement $R \in \mathcal{R}^{(m)}$ als disjunkte Vereinigung von Mengen aus $\mathcal{E}_1^{(m)}$ darzustellen, etwa $R = \bigcup_{i=1}^n D_i$ mit $D_1, \ldots, D_n \in \mathcal{R}^{(m)}$. Durch

$$P(R) = \sum_{i=1}^n P(D_i)$$

ist dann analog zu (1.2.8) eine konsistente Fortsetzung von P auf $\mathcal{R}^{(m)}$ gegeben. Die rechtsseitige Stetigkeit von F garantiert wieder die σ –Additivität von P auf $\mathcal{R}^{(m)}$, so daß man über die Schritte (1.1.64) und (1.1.65) das zu P gehörige äußere Maß und somit durch Einschränkung auf $\mathcal{B}^m$ die gewünschte Wahrscheinlichkeitsverteilung auf $\mathcal{B}^m$ erhält. ∎

Auch im mehrdimensionalen Fall ist die Eindeutigkeit der Fortsetzung von Verteilungsfunktionen zu Wahrscheinlichkeitsverteilungen auf ganz $\mathcal{B}^m$ gewährleistet:

Satz 1.4.2. *(Eindeutigkeitssatz für m-dimensionale Verteilungen)*
Es sei $F : \mathbf{R}^m \longrightarrow [0,1]$, $m \in \mathbf{N}$ *eine Abbildung mit den Eigenschaften (1.4.18), (1.4.20) und (1.4.23). Die nach Satz 1.4.1 existierende Wahrscheinlichkeitsverteilung* P *auf* $\mathcal{B}^m$, *für die* $F_P = F$ *gilt, ist dann eindeutig bestimmt.*

Beweis. Für zwei Wahrscheinlichkeitsverteilungen P und Q auf $\mathcal{B}^m$ mit derselben Verteilungsfunktion F ist wieder analog (1.2.13) bis (1.2.15) das System $\mathcal{D} = \{D \in \mathcal{B}^1 \mid P(D) = Q(D)\}$ ein Dynkin–System. Das Mengensystem $\mathcal{E}_1^{(m)} \cup \{\emptyset\}$ bzw. $\mathcal{E}_5^{(m)}$ ist auch hier durchschnittsstabil, und wegen

$$P\Big(\bigtimes_{i=1}^{m}(a_i,b_i]\Big) = \Delta_{(a_1,\ldots,a_m)}^{(b_1,\ldots,b_m)}F = Q\Big(\bigtimes_{i=1}^{m}(a_i,b_i]\Big)$$

für alle $(a_1,\ldots,a_m)$, $(b_1,\ldots,b_m) \in \mathbf{R}^m$ mit $a_i \le b_i$, $1 \le i \le m$, ist $\mathcal{E}_1^{(m)}$ in $\mathcal{D}$ enthalten. Mit Satz 1.2.2 ergibt sich dann wieder die Beziehung $\mathcal{B}^m = \sigma(\mathcal{E}_1^{(m)}) = \delta(\mathcal{E}_1^{(m)})$, d.h. es gilt $P(D) = Q(D)$ für alle $D \in \mathcal{B}^m$, was zu zeigen war. ∎

Analog zur eindimensionalen Situation läßt sich auch der Begriff einer (absolut–)stetigen Verteilung auf $\mathcal{B}^m$ erklären, und zwar durch geeignete mehrfache Integration.

Definition 1.4.4. *(m-dimensionale Verteilungsdichte)*
Es sei $f : \mathbf{R}^m \longrightarrow \mathbf{R}^+$, $m \in \mathbf{N}$ *eine über* $\mathbf{R}^m$ *uneigentlich Riemann–integrierbare Funktion mit*

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_m)\,dx_1\ldots dx_m = 1. \tag{1.4.24}$$

Dann wird durch

$$F(x_1,\ldots,x_m) := \int_{-\infty}^{x_m}\cdots\int_{-\infty}^{x_1} f(u_1,\ldots,u_m)\,du_1\ldots du_m \tag{1.4.25}$$

für $(x_1,\ldots,x_m) \in \mathbf{R}^m$ *die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung* P *auf* $\mathcal{B}^m$ *definiert. Die Funktion* F *heißt (m–dimensionale) (Verteilungs–)Dichte der Wahrscheinlichkeitsverteilung* P; P *heißt (m–dimensionale, absolut–)stetige Verteilung.*

Daß durch (1.4.25) tatsächlich eine Verteilungsfunktion definiert wird, läßt sich leicht nachweisen; insbesondere ist F Δ–monoton wegen

$$\Delta_{\boldsymbol{x}}^{\boldsymbol{y}}F = \int_{x_m}^{y_m}\cdots\int_{x_1}^{y_1} f(u_1,\ldots,u_m)\,du_1\ldots du_m \ge 0$$

für $\boldsymbol{x} = (x_1,\ldots,x_m)$, $\boldsymbol{y} = (y_1,\ldots,y_m) \in \mathbf{R}^m$. Nach dem Hauptsatz der Differential– und Integralrechnung für Funktionen mehrerer Variabler erhält man die Dichte f in ihren Stetigkeitspunkten $\boldsymbol{x} = (x_1,\ldots,x_m)$ wieder durch partielles Differenzieren der Verteilungsfunktion zurück vermöge

$$f(x_1,\ldots,x_m) = \frac{\partial^m}{\partial x_1\ldots\partial x_m}F(x_1,\ldots,x_m). \tag{1.4.26}$$

In Analogie zu (1.2.22) ist beispielsweise die stetige Gleichverteilung $\mathcal{R}(A)$ für $A \in \mathcal{B}^m$, $m \in \mathbf{N}$, mit $c = \int \ldots \int_A dx_1 \ldots dx_m > 0$ (m-dimensionales Volumen der Menge A[1]) wieder gegeben durch die Dichte

$$f(x_1, \ldots, x_m) = \begin{cases} 0 & \text{für } \boldsymbol{x} \notin A \\ \frac{1}{c} & \text{für } \boldsymbol{x} \in A. \end{cases} \tag{1.4.27}$$

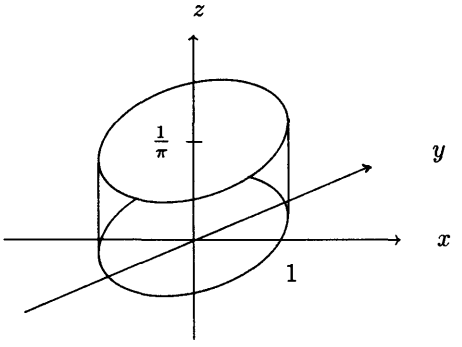
Bei höherdimensionalen stetigen Verteilungen ist die zugehörige Verteilungsfunktion i.a. nicht in einfacher, geschlossener Form darstellbar, weshalb solche Verteilungen in der Regel durch Angabe einer Dichte beschrieben werden. So ist etwa nach (1.4.27) durch

$$f(x,y) = \begin{cases} \frac{1}{\pi} & \text{für } x^2 + y^2 \leq 1 \\ 0 & \text{sonst} \end{cases} \tag{1.4.28}$$

eine (zweidimensionale) Dichte der stetigen Gleichverteilung $\mathcal{R}(K)$ über dem Einheitskreis $K = K^a(0,0;1)$ gegeben, da

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = \iint_K f(x,y)\, dx\, dy = \frac{1}{\pi} \iint_K dx\, dy = 1$$

gilt (zur Beachtung: $\iint_K dx\, dy$ ist gerade die Fläche des Einheitskreises, also π). Die zugehörige Verteilungsfunktion $F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\, du\, dv$, $x, y \in \mathbf{R}$, ist dagegen nur durch mehrfache Fallunterscheidung und mittels inverser trigonometrischer Funktionen darstellbar.



Das letzte Beispiel zeigt, daß man ähnlich wie im eindimensionalen Fall Verteilungen P auf Borel'schen Teilmengen $\Omega^{(m)} \subset \mathbf{R}^m$, $m \in \mathbf{N}$, wieder als Verteilungen auf ganz $\mathcal{B}^m$ auffassen kann vermöge der Fortsetzung

$$P'(B) = P(\Omega^{(m)} \cap B) \text{ für alle } B \in \mathcal{B}^m \tag{1.4.29}$$

[1] Das betrachtete Mehrfachintegral existiert im Riemann'schen Sinn zunächst nur für beschränkte, sog. *Jordan-Mengen* (vgl. auch (1.2.22)); sie sind dadurch charakterisiert, daß ihr Rand ∂A, d.h. die Differenzmenge aus der abgeschlossenen Hülle $\overline{A}$ und der Menge A° der inneren Punkte, eine Jordan-Nullmenge ist, d.h. zu jedem $\varepsilon > 0$ gibt es *endlich* viele abgeschlossene Intervalle des $\mathbf{R}^m$, die ∂A überdecken und ein Gesamtvolumen von höchstens ε besitzten. Jordan-Mengen sind stets auch Borel-Mengen, nicht aber umgekehrt; z.B. sind Intervalle, Kreise u.ä. Jordan-Mengen. Für Details sei etwa auf Heuser (1988), Abschnitt 201 verwiesen.

(vgl. Beziehung (1.2.16)).

Eine wichtige Klasse von Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^m$, $m \in \mathbf{N}$, ist die Klasse der sogenannten *Produktverteilungen*, welche in engem Zusammenhang zur stochastischen Unabhängigkeit (von Ereignissen — siehe Definition 1.1.6 — und Zufallsvariablen) steht.

Definition 1.4.5. *(Produktverteilung)*
Es seien $P_1, \ldots, P_m$, $m \in \mathbf{N}$, Wahrscheinlichkeitsverteilungen auf der Borel'schen σ–Algebra $\mathcal{B}^1$ mit zugehörigen Verteilungsfunktionen $F_1, \ldots, F_m$. Dann heißt die durch die Verteilungsfunktion

$$F(x_1, \ldots, x_m) = \prod_{i=1}^{m} F_i(x_i), \quad (x_1, \ldots, x_m) \in \mathbf{R}^m, \tag{1.4.30}$$

eindeutig bestimmte Wahrscheinlichkeitsverteilung P auf $\mathcal{B}^m$ die Produktverteilung der Wahrscheinlichkeitsverteilungen $P_1, \ldots, P_m$, in Zeichen:

$$P = \bigotimes_{i=1}^{m} P_i = P_1 \otimes \ldots \otimes P_m. \tag{1.4.31}$$

In der Tat wird durch (1.4.30) eine Δ–monotone Funktion definiert, denn es gilt für $\boldsymbol{x} = (x_1, \ldots, x_m)$, $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbf{R}^m$, $x_i \leq y_i$, $1 \leq i \leq m$:

$$\Delta F_{\boldsymbol{x}}^{\boldsymbol{y}} = \prod_{i=1}^{m} \big(F_i(y_i) - F_i(x_i)\big). \tag{1.4.32}$$

Allgemeiner läßt sich hieraus sogar die Beziehung

$$P\Big(\bigtimes_{i=1}^{m} B_i\Big) = \bigotimes_{i=1}^{m} P_i\Big(\bigtimes_{i=1}^{m} B_i\Big) = \prod_{i=1}^{m} P_i(B_i), \; B_1, \ldots, B_m \in \mathcal{B}^1, \tag{1.4.33}$$

ableiten (für $B_i = (x_i, y_i]$, $1 \leq i \leq m$, fällt nämlich Beziehung (1.4.33) mit (1.4.32) zusammen). Dies bedeutet aber gerade, daß unter dem so definierten Produktmaß $P = \bigotimes_{i=1}^{m} P_i$ Mengen (Ereignisse) $B_1^{(m)}, \ldots, B_m^{(m)}$ der Form

$$B_i^{(m)} = \mathbf{R}^{i-1} \times B_i \times \mathbf{R}^{m-i}, \quad 1 \leq i \leq m,$$

mit $B_i \in \mathcal{B}^1$ stochastisch unabhängig sind, denn es gilt

$$\bigcap_{i=1}^{m} B_i^{(m)} = \bigcap_{i=1}^{m} \mathbf{R}^{i-1} \times B_i \times \mathbf{R}^{m-i} = \bigtimes_{i=1}^{m} B_i,$$

also

$$P\Big(\bigcap_{i=1}^{m} B_i^{(m)}\Big) = P\Big(\bigtimes_{i=1}^{m} B_i\Big) = \prod_{i=1}^{m} P_i(B_i) = \prod_{i=1}^{m} P(B_i^{(m)})$$

wegen $P(B_i^{(m)}) = \prod_{j=1}^{i-1} P_j(\mathbf{R})\, P_i(B_i) \prod_{j=i+1}^{m} P_j(\mathbf{R}) = P_i(B_i)$, $1 \le i \le m$.

Entsprechend erhält man für beliebige Auswahlen $1 \le i_1 < \ldots < i_k \le m$, $1 \le k \le m$,

$$P\Big(\bigcap_{j=1}^{k} B_{i_j}^{(m)}\Big) = \prod_{j=1}^{k} P(B_{i_j}^{(m)}),$$

so daß nach Satz 1.1.2 in der Tat die Ereignisse $B_1^{(m)}, \ldots, B_m^{(m)}$ stochastisch unabhängig sind.

Mit maßtheoretischen Argumenten läßt sich zeigen, daß sogar auf dem Produkt $\big(\bigtimes_{i=1}^{m} \Omega_i, \bigotimes_{i=1}^{m} \mathcal{A}_i\big)$ *beliebiger* Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \le i \le m$, ein Produktmaß $P = \bigotimes_{i=1}^{m} P_i$ vermöge der Beziehung (1.4.33) (mit $B_i \in \mathcal{A}_i$, $1 \le i \le m$) eindeutig definiert werden kann; analog sind dann allgemeiner Ereignisse $B_1^{(m)}, \ldots, B_m^{(m)}$ der Form

$$B_i^{(m)} = \bigtimes_{j=1}^{i-1} \Omega_j \times B_i \times \bigtimes_{j=i+1}^{m} \Omega_j, \quad 1 \le i \le m, \tag{1.4.34}$$

mit $B_i \in \mathcal{A}_i$ stochastisch unabhängig unter P.

Identifiziert man den gegebenen Grundraum $(\Omega, \mathcal{A}, P)$ mit einem *Versuch*, der durch die Wahrscheinlichkeitsverteilung P gesteuert wird, so läßt sich der Produktraum $\big(\bigtimes_{i=1}^{m} \Omega, \bigotimes_{i=1}^{m} \mathcal{A}, \bigotimes_{i=1}^{m} P\big)$ offensichtlich mit einer m–fachen, unabhängigen *Versuchswiederholung* identifizieren. Das Ereignis $B_i^{(m)}$ aus Beziehung (1.4.34) bedeutet dann in diesem Sinne, daß das Ereignis B_i des Grundraums gerade im i–ten Versuch eingetreten ist, und zwar unabhängig von den in den übrigen Versuchen eingetretenden Ereignissen $B_j, j \ne i$. Mit Hilfe von Produkträumen obiger Art läßt sich also in geeigneter Weise die stochastische Unabhängigkeit *endlich* vieler Ereignisse modellieren. Für manche Situationen ist dies jedoch nicht ausreichend, da man häufig auch Folgen unabhängiger Ereignisse in Betracht ziehen muß. Hierzu benötigt man zunächst geeignete σ–Algebren über unendlichen kartesischen Produktmengen.

Definition 1.4.6. *(abzählbare Produktmeßräume)*
Es seien $(\Omega_i, \mathcal{A}_i)$, $i \in \mathbf{N}$, *Meßräume und*

$$\mathcal{E}^{(\infty)} = \Big\{ \bigtimes_{i=1}^{\infty} A_i \mid A_i \in \mathcal{A}_i,\ i \in \mathbf{N} \Big\}. \tag{1.4.35}$$

Dann heißt die durch

$$\bigotimes_{i=1}^{\infty} \mathcal{A}_i = \sigma(\mathcal{E}^{(\infty)}) \tag{1.4.36}$$

definierte σ*–Algebra über der (abzählbaren) kartesischen Produktmenge* $\Omega^{(\infty)} = \bigtimes_{i=1}^{\infty} \Omega_i$ *die (abzählbare) Produkt–*σ*–Algebra der* σ*–Algebren* $\{\mathcal{A}_i\}_{i \in \mathbf{N}}$.

Will man analog zum endlichen Fall auch die stochastische Unabhängigkeit abzählbar vieler Ereignisse — z.B. durch unendliche Versuchswiederholungen —

beschreiben, benötigt man offenbar ein Maß $P^{(\infty)}$ auf $\bigotimes_{i=1}^{\infty} \mathcal{A}$, für das

$$P^{(\infty)}\Big(\bigcap_{i=1}^{m} B_i^{(\infty)}\Big) = P^{(\infty)}\Big(\bigtimes_{i=1}^{m} B_i \times \bigtimes_{i>m} \Omega\Big) = \prod_{i=1}^{m} P^{(\infty)}(B_i^{(\infty)})$$

für alle $m \in \mathbb{N}$ gilt. Ein Vergleich mit (1.4.33) legt dabei nahe, dieses Maß so zu wählen, daß

$$P^{(\infty)}\Big(\bigtimes_{i=1}^{m} A_i \times \bigtimes_{i>m} \Omega\Big) = \bigotimes_{i=1}^{m} P\Big(\bigtimes_{i=1}^{m} A_i\Big) = \prod_{i=1}^{m} P(A_i) \tag{1.4.37}$$

gilt für alle $m \in \mathbb{N}$ und $A_1, \ldots, A_m \in \mathcal{A}$. In der Tat läßt sich mit maßtheoretischen Überlegungen ein solches unendliches Produktmaß (sogar auf Produkten beliebiger Wahrscheinlichkeitsräume) eindeutig konstruieren.

Satz 1.4.3. *(Existenzsatz für allgemeine Produktverteilungen)*
Es seien $(\Omega_i, \mathcal{A}_i, P_i)$, $i \in \mathbb{N}$, Wahrscheinlichkeitsräume. Dann existiert genau eine Wahrscheinlichkeitsverteilung P auf $\bigotimes_{i=1}^{\infty} \mathcal{A}_i$ mit der Eigenschaft

$$P\Big(\bigtimes_{i=1}^{m} A_i \times \bigtimes_{i>m} \Omega_i\Big) = \prod_{i=1}^{m} P_i(A_i) \text{ für alle } m \in \mathbb{N},\ A_i \in \mathcal{A}_i,\ 1 \le i \le m. \tag{1.4.38}$$

P heißt das (abzählbar-unendliche) Produktmaß der Wahrscheinlichkeitsverteilungen $\{P_i\}_{i\in\mathbb{N}}$, in Zeichen:

$$P = \bigotimes_{i=1}^{\infty} P_i. \tag{1.4.39}$$

Einen Beweis dieses und noch allgemeinerer Sätze findet man z.B. in Bauer (1978), §33. Auf die Abzählbarkeit der Indexmenge $\mathbb{N}$ kommt es dabei nicht an, d.h. selbst für *überabzählbar* viele Meßräume mit darauf gegebenen Wahrscheinlichkeitsverteilungen läßt sich ein entsprechender Produktmeßraum mit eindeutig bestimmtem Produktmaß konstruieren.

Mit Hilfe des in Satz 1.4.3 beschriebenen Produktmaßes lassen sich also vermöge (1.4.37) Folgen unabhängiger Ereignisse in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ modellieren.

Damit ist das im Anschluß an Satz 1.1.1 aufgeworfene Problem der Existenz solcher Ereignisfolgen befriedigend gelöst. Durch die Wahl identischer Mengen $B_i = B \in \mathcal{A}$ für alle $i \in \mathbb{N}$ erhält man durch die oben beschriebene Methode insbesondere auch Folgen unabhängiger Ereignisse mit derselben Eintrittswahrscheinlichkeit $p = P(B)$. Ist $0 < p < 1$, so enthält die Grundmenge Ω zwangsläufig mindestens zwei verschiedene Elemente, da die σ–Algebra $\mathcal{A}$ neben $\emptyset$ und Ω (mit den Wahrscheinlichkeiten $P(\emptyset) = 0$ und $P(\Omega) = 1$) noch die Menge $B \neq \emptyset, \Omega$ enthält. Damit besitzt der Grundraum $\bigtimes_{i=1}^{\infty} \Omega = \Omega^{\mathbb{N}}$ aber mindestens dieselbe Mächtigkeit wie die Menge $\{0,1\}^{\mathbb{N}}$, also die Mächtigkeit der reellen Zahlen und ist somit überabzählbar in Übereinstimmung mit Satz 1.1.1.

Im folgenden können und werden wir deshalb stets davon ausgehen, daß der zugrundeliegende Wahrscheinlichkeitsraum "groß" genug ist (im Sinne der Erweiterung über kartesische Produkte), um die Existenz von beliebigen Folgen unabhängiger Ereignisse zu garantieren.

Mit Hilfe des Begriffs des Produktmeßraums lassen sich nun auch auf einfache Weise Zufallsvektoren erklären. Hierzu benötigen wir allerdings noch einen etwas allgemeineren Meßbarkeitsbegriff als in (1.3.1).

Definition 1.4.7. *(Zufallselement)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(\mathcal{X}, \mathcal{B})$ ein beliebiger Meßraum. Eine Abbildung $\boldsymbol{X} : \Omega \longrightarrow \mathcal{X}$ mit der Eigenschaft

$$\{\boldsymbol{X} \in B\} := \boldsymbol{X}^{-1}(B) = \{\omega \mid \boldsymbol{X}(\omega) \in B\} \in \mathcal{A} \text{ für alle } B \in \mathcal{B} \tag{1.4.40}$$

heißt Zufallselement in $(\mathcal{X}, \mathcal{B})$.

Für $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}, \mathcal{B}^1)$ fallen also die Begriffe "Zufallselement" und "Zufallsvariable" zusammen. Die durch (1.4.40) beschriebene Eigenschaft der Meßbarkeit wollen wir auch hier wieder durch die Schreibweise

$$\boldsymbol{X} : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{X}, \mathcal{B})$$

zum Ausdruck bringen.

Ähnlich wie im reellen Fall reicht es auch bei Zufallselementen zum Nachweis der Meßbarkeit aus, Beziehung (1.4.40) für Mengen B eines (beliebigen) Erzeugers $\mathcal{E}$ von $\mathcal{B}$ zu verifizieren.

Zufallsvektoren erhält man nun durch die Spezifizierung des Meßraums $(\mathcal{X}, \mathcal{B})$ als geeigneten Produktmeßraum.

Definition 1.4.8. *(Zufallsvektor)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\boldsymbol{X} : (\Omega, \mathcal{A}) \longrightarrow (\mathbf{R}^m, \mathcal{B}^m)$, $m \in \mathbf{N}$, ein Zufallselement in $(\mathbf{R}^m, \mathcal{B}^m)$. Dann heißt $\boldsymbol{X}$ (m-dimensionaler) Zufallsvektor.

Das folgende Resultat zeigt, daß man sich Zufallsvektoren auch durch komponentenweise Zusammensetzung von Zufallsvariablen entstanden denken kann.

Lemma 1.4.3. *Für $m \in \mathbf{N}$ seien $X_1, \ldots, X_m$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann ist $\boldsymbol{X} = (X_1, \ldots, X_m)$ ein m-dimensionaler Zufallsvektor. Ist umgekehrt $\boldsymbol{X} = (X_1, \ldots, X_m)$ ein m-dimensionaler Zufallsvektor, so bilden die Komponenten $X_1, \ldots, X_m$ gerade Zufallsvariablen auf Ω.*

Beweis. Es sei $\mathcal{F}^{(m)} = \{B \in \mathcal{B}^m \mid \boldsymbol{X}^{-1}(B) \in \mathcal{A}\}$. Man sieht aufgrund von (1.3.3) leicht, daß $\mathcal{F}^{(m)}$ eine σ-Algebra über $\mathcal{B}^m$ bildet, die wegen

$$\boldsymbol{X}^{-1}\Big(\bigtimes_{i=1}^{m} B_i\Big) = \bigcap_{i=1}^{m} \{X_i \in B_i\},\ B_1, \ldots, B_m \in \mathcal{B}^1, \tag{1.4.41}$$

den Erzeuger $\mathcal{E}^{(m)}$ aus (1.4.1) mit $\mathcal{A}_i = \mathcal{B}^1$ enthält. Damit ist aber $\mathcal{B}^m = \sigma(\mathcal{E}^{(m)}) \subseteq \sigma(\mathcal{F}^{(m)}) = \mathcal{F}^{(m)} \subseteq \mathcal{B}^m$, also $\mathcal{F}^{(m)} = \mathcal{B}^m$, was gerade die Meßbarkeit von $\boldsymbol{X}$ bedeutet (vgl. (1.4.40)). Setzt man umgekehrt in (1.4.41) speziell $B_j^{(m)} = \mathbf{R}^{j-1} \times$

$B_j \times \mathbf{R}^{m-j}$, $1 \le j \le m$, so erhält man mit der Meßbarkeit von $\boldsymbol{X} = (X_1, \ldots, X_m)$ gerade

$$\{X_j \in B_j\} = \boldsymbol{X}^{-1}(B_j^{(m)}),\ 1 \le j \le m,$$

also die Meßbarkeit aller Komponenten $X_1, \ldots, X_m$, wie behauptet. ∎

Analog kann man auch eine *Folge* $\{X_n\}_{n\in\mathbf{N}}$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ als Zufallselement in $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^\infty, \mathcal{B}^\infty) = \left(\bigtimes_{i=1}^{\infty} \mathbf{R}, \bigotimes_{i=1}^{\infty} \mathcal{B}\right)$ auffassen. Lemma 1.4.3 gilt dann entsprechend auch für $m = \infty$. Im Zusammenhang mit stochastischen Zähl- und Punktprozessen (etwa im Bereich der Bildverarbeitung oder Mustererkennung) werden wir später sogar noch allgemeinere Bildräume $(\mathcal{X}, \mathcal{B})$ kennenlernen.

Der Begriff der Verteilung einer Zufallsvariablen überträgt sich entsprechend auf den Begriff der Verteilung eines Zufallselements und damit spezieller auf den eines Zufallsvektors oder einer Folge von Zufallsvariablen.

Definition 1.4.9. *(Verteilung eines Zufallselements)*
Es sei $\boldsymbol{X}$ ein Zufallselement in einem Meßraum $(\mathcal{X}, \mathcal{B})$. Dann heißt die durch

$$Q(B) := P(\{\boldsymbol{X} \in B\}) = P(\boldsymbol{X}^{-1}(B)),\ B \in \mathcal{B}, \tag{1.4.42}$$

definierte Wahrscheinlichkeitsverteilung auf $\mathcal{B}$ wieder Verteilung von $\boldsymbol{X}$, in Zeichen: $Q = P^{\boldsymbol{X}}$.

Im Vergleich zu Beziehung (1.3.2) wird also lediglich der spezielle Meßraum $(\mathbf{R}, \mathcal{B}^1)$ durch den Meßraum $(\mathcal{X}, \mathcal{B})$ ersetzt.

Gemäß Satz 1.4.2 ist damit die Verteilung $Q_m = P^{(X_1,\ldots,X_m)}$ eines m–dimensionalen Zufallsvektors $(X_1, \ldots, X_m)$, $m \in \mathbf{N}$, eindeutig durch die zu Q_m gehörige Verteilungsfunktion $F = F_{Q_m}$ bestimmt, d.h. durch alle Wahrscheinlichkeiten

$$\begin{aligned} F(x_1, \ldots, x_m) &= P(X_1 \le x_1, \ldots, X_m \le x_m) \\ &:= P\Big(\bigcap_{i=1}^{m} \{X_i \le x_i\}\Big),\ x_1, \ldots, x_m \in \mathbf{R}. \end{aligned}$$

Um eine entsprechend einfache Beschreibung der Verteilung einer *Folge* von Zufallsvariablen angeben zu können, benötigen wir noch den Begriff der Randverteilung höherdimensionaler Wahrscheinlichkeitsverteilungen.

Definition 1.4.10. *(Randverteilung)*
Es sei $m \in \mathbf{N}$ oder $m = \infty$, P eine Wahrscheinlichkeitsverteilung auf $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^m, \mathcal{B}^m)$ und $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ bzw $\boldsymbol{Y} = \{Y_i\}_{i\in\mathbf{N}}$ mit $Y_i(\omega) = \omega_i$, $\omega \in \mathcal{X}$. Für jedes $n \in \mathbf{N}$ und Zahlen $1 \le i_1 < i_2 < \ldots < i_n$ ($\le m$ für $m < \infty$) heißt dann die Wahrscheinlichkeitsverteilung $P_{(i_1,\ldots,i_n)}$ auf $(\mathbf{R}^n, \mathcal{B}^n)$, welche durch

$$P_{(i_1,\ldots,i_n)} = P^{(Y_{i_1},\ldots,Y_{i_n})} \tag{1.4.43}$$

gegeben ist, die Randverteilung der Ordnung $(i_1, \ldots, i_n)$ zu P.

Die in Definition 1.4.9 auftretenden Zufallsvariablen Y_i, $i \in \mathbf{N}$, heißen auch *Projektion* auf die i–te Komponente; ihre Meßbarkeit folgt aus der Darstellung

$$Y_i^{-1}(B) = \mathbf{R}^{i-1} \times B \times \bigtimes_{j=i+1}^{m} \mathbf{R}, \quad B \in \mathcal{B}^1.$$

Nach Lemma 1.4.3 ist dann auch der Vektor $(Y_{i_1}, \ldots, Y_{i_n})$ meßbar, also ein n–dimensionaler Zufallsvektor (Projektion auf die Komponenten $i_1, \ldots, i_n$).

Im Falle von $m < \infty$ lassen sich Randverteilungen auch direkt durch die zu P gehörige Verteilungsfunktion F beschreiben vermöge

Lemma 1.4.4. *(Verteilungsfunktion der Randverteilung)*
P sei eine Wahrscheinlichkeitsverteilung auf $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^m, \mathcal{B}^m)$, $m \in \mathbf{N}$, und F die zugehörige Verteilungsfunktion. Dann ist die durch

$$F_{(i_1,\ldots,i_n)}(x_{i_1}, \ldots, x_{i_n}) = F(\infty, \ldots, \infty, x_{i_1}, \infty, \ldots, \infty, x_{i_2}, \infty, \ldots, x_{i_n}, \ldots, \infty) \tag{1.4.44}$$
$$:= \lim_{\substack{x_j \to \infty \\ 1 \le j \le m,\ j \notin \{i_1,\ldots,i_n\}}} F(x_1, \ldots, x_m), \quad x_{i_1}, \ldots, x_{i_n} \in \mathbf{R}$$

gegebene Verteilungsfunktion die Verteilungsfunktion der Randverteilung der Ordnung $(i_1, \ldots, i_n)$ zu P.

Beweis. Mit den Bezeichnungen aus Definition 1.4.9 und $\boldsymbol{Y}_m = (Y_1, \ldots, Y_m)$ ist

$$\begin{aligned} F_{(i_1,\ldots,i_n)}(x_{i_1}, \ldots, x_{i_n}) &= P(Y_{i_1} \le x_{i_1}, \ldots, Y_{i_n} \le x_{i_n}) = P\Big(\bigcap_{j=1}^{n} \{Y_{i_j} \le x_{i_j}\}\Big) \\ &= P\Big(\boldsymbol{Y}_m \in \bigcap_{j=1}^{n} \mathbf{R}^{i_j - 1} \times (-\infty, x_{i_j}] \times \mathbf{R}^{m-i_j}\Big) \\ &= P\Big(\boldsymbol{Y}_m \in \mathbf{R}^{i_1 - 1} \times (-\infty, x_{i_1}] \times \mathbf{R}^{i_2 - i_1 - 1} \times (-\infty, x_{i_2}] \times \cdots \\ &\qquad \cdots \times \mathbf{R}^{i_n - i_{n-1} - 1} \times (-\infty, x_{i_n}] \times \mathbf{R}^{m-i_n}\Big), \end{aligned}$$

woraus Beziehung (1.4.44) wegen der Stetigkeit von P von unten folgt. ∎

Besonders einfach lassen sich Randverteilungen von Wahrscheinlichkeitsverteilungen $Q = P^{(X_1,\ldots,X_m)}$, also von Verteilungen von Zufallsvektoren $(X_1, \ldots, X_m)$ mit $m \in \mathbf{N}$ darstellen. Hier gilt nämlich mit den Bezeichnungen aus Definition 1.4.9

$$Q_{(i_1,\ldots,i_n)} = P^{(X_{i_1},\ldots,X_{i_n})},$$

denn setzt man $\boldsymbol{X} = (X_1, \ldots, X_m)$ und $\boldsymbol{Y} = (Y_{i_1}, \ldots, Y_{i_n})$, so ist gerade

$$Q_{(i_1,\ldots,i_n)} = Q^{\boldsymbol{Y}} = \big(P^{\boldsymbol{X}}\big)^{\boldsymbol{Y}} = P^{\boldsymbol{Y} \circ \boldsymbol{X}} = P^{(X_{i_1},\ldots,X_{i_n})},$$

wobei wegen

$$(\boldsymbol{Y} \circ \boldsymbol{X})^{-1} = \boldsymbol{X}^{-1} \circ \boldsymbol{Y}^{-1} \tag{1.4.45}$$

$\boldsymbol{Y} \circ \boldsymbol{X}$ meßbar, also ein n–dimensionaler Zufallsvektor ist ($\circ$ bezeichne hierbei die Hintereinanderausführung von Abbildungen). Hierauf kommen wir in Lemma 2.1.1 noch einmal in allgemeinerem Rahmen zurück.

Ist nun $\{X_i\}_{i\in\mathbf{N}}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$, so sind entsprechend für alle $m \in \mathbf{N}$ die Abbildungen $\boldsymbol{X}_m = (X_1, \ldots, X_m)$ Zufallsvektoren mit der charakteristischen Eigenschaft

$$P^{\boldsymbol{X}_{m+1}}(B_m \times \mathbf{R}) = P^{\boldsymbol{X}_m}(B_m) \text{ für alle } m \in \mathbf{N},\ B_m \in \mathcal{B}^m. \tag{1.4.46}$$

Man sagt, die Verteilungen der Zufallsvektoren $\{\boldsymbol{X}_m\}_{m\in\mathbf{N}}$, $\{P^{\boldsymbol{X}_m}\}_{m\in\mathbf{N}}$, bilden eine *projektive Familie.*

Mit maßtheoretischen Methoden läßt sich wieder zeigen, daß diese Eigenschaft die Verteilung der gesamten Folge (als Zufallselement in $(\mathbf{R}^\infty, \mathcal{B}^\infty)$) eindeutig bestimmt. Genauer gilt:

Satz 1.4.4. *(Existenzsatz für projektive Verteilungsfamilien)*
Es seien Q_m projektive Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^m$, $m \in \mathbf{N}$, d.h. es gelte $Q_{m+1}(B_m \times \mathbf{R}) = Q_m(B_m)$ für alle $m \in \mathbf{N}$, $B_m \in \mathcal{B}^m$. Dann existiert genau eine Wahrscheinlichkeitsverteilung Q auf $\mathcal{B}^\infty$ mit der Eigenschaft:

$$Q\Big(B_m \times \bigtimes_{i>m} \mathbf{R}\Big) = Q_m(B_m) \text{ für alle } m \in \mathbf{N},\ B_m \in \mathcal{B}^m. \tag{1.4.47}$$

Satz 1.4.4 ist ein Spezialfall eines noch wesentlich allgemeineren Satzes, der auf Kolmogoroff zurückgeht; siehe etwa Bauer (1978), Satz 62.3. Er umfaßt insbesondere auch Satz 1.4.3 in der Situation $(\Omega_i, \mathcal{A}_i) = (\mathbf{R}, \mathcal{B}^1)$, $i \in \mathbf{N}$, wenn man $Q_m = \bigotimes_{i=1}^m P_i$, $m \in \mathbf{N}$, setzt.

In der Tat ist damit aufgrund von (1.4.46) die Verteilung einer Folge $\{X_i\}_{i\in\mathbf{N}}$ von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ durch die Verteilungen $P^{(X_1,\ldots,X_m)}$, $m \in \mathbf{N}$, bzw. allgemeiner $P^{(X_{i_1},\ldots,X_{i_n})}$, $1 \le i_1 < \ldots < i_n$, $n \in \mathbf{N}$, eindeutig bestimmt; man sagt auch, die Verteilung von $\{X_i\}_{i\in\mathbf{N}}$, $P^{\{X_i\}_{i\in\mathbf{N}}}$, sei durch die *endlich–dimensionalen Randverteilungen* der Folge bestimmt.

Fragestellungen dieser Art werden uns etwa bei der Behandlung von Markoff–Ketten wiederbegegnen, welche bei der Average–Case–Analyse gewisser Algorithmen und in der Informationstheorie (Kapitel 5) eine wichtige Rolle spielen.

Nunmehr können wir auch die stochastische Unabhängigkeit von Zufallsvariablen bzw. Zufallsvektoren definieren.

Definition 1.4.11. *(Stochastische Unabhängigkeit von Zufallsvariablen)*
Zufallsvariable $X_1, \ldots, X_n$, $n \in \mathbf{N}$, auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißen stochastisch unabhängig, wenn gilt:

$$P^{(X_1,\ldots,X_n)} = \bigotimes_{i=1}^n P^{X_i}. \tag{1.4.48}$$

Eine Folge von Zufallsvariablen $\{X_i\}_{i\in\mathbf{N}}$ auf $(\Omega, \mathcal{A}, P)$ heißt stochastisch unabhängig, wenn für alle $n \in \mathbf{N}$ die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig sind.

Eine Folge $\boldsymbol{X} = \{X_i\}_{i\in\mathbb{N}}$ von Zufallsvariablen ist also nach Satz 1.4.3 genau dann stochastisch unabhängig, wenn

$$P^{\boldsymbol{X}} = \bigotimes_{i=1}^{\infty} P^{X_i} \tag{1.4.49}$$

gilt. Unabhängige Zufallsvariablen sind demnach dadurch charakterisiert, daß ihre gemeinsame Verteilung gerade das Produkt ihrer Randverteilungen ist. Völlig analog läßt sich die stochastische Unabhängigkeit von endlich oder unendlich vielen *Zufallsvektoren* oder sogar *Zufallselementen* $\{\boldsymbol{X}_i\}$ — mit Werten in beliebigen Meßräumen $(\mathcal{X}_i, \mathcal{B}_i)$, $i \in \mathbb{N}$ — definieren; die Beziehungen (1.4.48) und (1.4.49) bleiben dabei wörtlich bestehen, wenn man die Produktbildung der Wahrscheinlichkeitsverteilungen $P^{\boldsymbol{X}_i}$ im Sinne des allgemeinen Satzes 1.4.3 auffaßt.

Definition 1.4.12. *(stochastische Unabhängigkeit von Zufallselementen) Zufallselemente $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $n \in \mathbb{N}$, auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißen stochastisch unabhängig, wenn gilt:*

$$P^{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)} = \bigotimes_{i=1}^{n} P^{\boldsymbol{X}_i}. \tag{1.4.50}$$

Eine Folge von Zufallselementen $\{\boldsymbol{X}_i\}_{i\in\mathbb{N}}$ auf $(\Omega, \mathcal{A}, P)$ heißt stochastisch unabhängig, wenn für alle $n \in \mathbb{N}$ die Zufallselemente $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ stochastisch unabhängig sind.

Stochastische Unabhängigkeit von Zufallsvariablen bzw. allgemeiner Zufallselementen $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $n \in \mathbb{N}$, bedeutet also gerade die stochastische Unabhängigkeit aller Ereignisse $\{\boldsymbol{X}_1 \in B_1\}, \ldots, \{\boldsymbol{X}_n \in B_n\}$ für alle Mengen $B_i \in \mathcal{B}_i$, $1 \le i \le n$. Will man die stochastische Unabhängigkeit von Zufallsvariablen oder allgemeiner Zufallselementen nachprüfen, muß man allerdings nicht alle diese Mengen in Betracht ziehen. Vielmehr gilt:

Lemma 1.4.5. *$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $n \in \mathbb{N}$, seien Zufallselemente auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in Meßräumen $(\mathcal{X}_i, \mathcal{B}_i)$, $1 \le i \le n$; $\mathcal{E}_1, \ldots, \mathcal{E}_n$ seien durchschnittsstabile Erzeuger von $\mathcal{B}_1, \ldots, \mathcal{B}_m$ mit $\mathcal{X}_i \in \mathcal{E}_i$, $1 \le i \le n$. In diesem Fall sind $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ genau dann stochastisch unabhängig, wenn alle Ereignisse $\{\boldsymbol{X}_i \in E_i\}$, $1 \le i \le n$, mit $E_i \in \mathcal{E}_i$ stochastisch unabhängig sind.*

Beweis. Es bleibt nur das Hinreichen der angegebenen Bedingungen für die Unabhängigkeit der Zufallselemente $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ zu zeigen. Unter Heranziehung von Dynkin-Systemen läßt sich nachweisen, daß mit $\{\boldsymbol{X}_i \in E_i\}$, $E_i \in \mathcal{E}_i$, $1 \le i \le n$, auch die Ereignisse $\{\boldsymbol{X}_i \in E_i\}$, $E_i \in \delta(\mathcal{E}_i)$, $1 \le i \le n$, stochastisch unabhängig sind. Wegen der Durchschnittsstabilität kann man dann aber auch noch die Dynkin-Systeme $\delta(\mathcal{E}_i)$ durch die davon erzeugten σ-Algebren $\sigma(\mathcal{E}_i) = \mathcal{B}_i$, $1 \le i \le n$, ersetzen, woraus die Behauptung folgt. ∎

Für einen ausführlichen Beweis sei auf Bauer (1978), Satz 31.2 verwiesen.

Besitzen gewisse der Zufallselemente $\boldsymbol{X}_i$ diskrete Verteilungen mit Wertebereich $\mathcal{W}_i = \{\boldsymbol{x}_{ij}\}_{j\in\mathbb{N}} \subseteq \mathcal{X}_i$, so reicht es nach Lemma 1.4.5 also, sich beim Nachweis

der stochastischen Unabhängigkeit auf die Mengen $\{\boldsymbol{X}_i = \boldsymbol{x}_{ij}\}$, $j \in \mathbb{N}$, zu beschränken, da die Urbilder von Mengen $B \in \mathcal{B}_i$ mit $B \cap \mathcal{W}_i = \emptyset$ ebenfalls leer sind.

Insbesondere sind Zufallselemente $\boldsymbol{X}_i$, die nur *einen* Wert $\boldsymbol{x}_i \in \mathcal{X}_i$ annehmen (d.h. die konstante Abbildungen auf Ω darstellen), stets von anderen Zufallselementen stochastisch unabhängig, da die einzigen Urbilder unter $\boldsymbol{X}_i$ $\emptyset$ und Ω sind. Entsprechendes gilt unter der schwächeren Annahme $P(\boldsymbol{X}_i = \boldsymbol{x}_i) = 1$.

Der in Definition 1.1.6 eingeführte Begriff der stochastischen Unabhängigkeit von Ereignissen läßt sich sogar äquivalent durch die stochastische Unabhängigkeit gewisser Zufallsvariablen, den sogenannten *Indikatorvariablen*, ausdrücken.

Definition 1.4.13. *(Indikatorvariable)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{A}$ ein Ereignis. Die durch

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \notin A \end{cases}$$

definierte Zufallsvariable $\mathbb{1}_A$ heißt Indikatorvariable zu A.

Wegen

$$\mathbb{1}_A^{-1}(B) = \begin{cases} \emptyset & \text{für } B \cap \{0,1\} = \emptyset \\ A & \text{für } 1 \in B, 0 \notin B \\ A^c & \text{für } 0 \in B, 1 \notin B \\ \Omega & \text{für } \{0,1\} \subseteq B, \end{cases} \qquad B \in \mathcal{B}^1$$

ist $\mathbb{1}_A$ tatsächlich meßbar, also eine Zufallsvariable. Sie gibt an, ob das Ereignis A eintritt ($\mathbb{1}_A(\omega) = 1$) oder nicht ($\mathbb{1}_A(\omega) = 0$).

Verknüpfungen von Ereignissen wie Vereinigungs- und Durchschnittsbildung, aber auch Komplementbildung lassen sich über Indikatorvariable entsprechend einfach beschreiben; es gilt z.B. für Ereignisse $A, B \in \mathcal{A}$:

$$\begin{aligned} \mathbb{1}_{A \cup B} &= \max(\mathbb{1}_A, \mathbb{1}_B) \\ \mathbb{1}_{A \cup B} &= \mathbb{1}_A + \mathbb{1}_B, \text{ für } A \cap B = \emptyset \\ \mathbb{1}_{A \cap B} &= \min(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{1}_A \cdot \mathbb{1}_B \\ \mathbb{1}_{A^c} &= 1 - \mathbb{1}_A. \end{aligned}$$

Entsprechende Beziehungen gelten für abzählbar viele solcher Verknüpfungen. Ein Vergleich von Definition 1.1.6 mit Definition 1.4.11 zeigt dann aber gerade, daß endlich oder unendlich viele Ereignisse $\{A_n\}$ in $\mathcal{A}$ genau dann stochastisch unabhängig sind, wenn die zugehörigen Indikatorvariablen $\{\mathbb{1}_{A_n}\}$ stochastisch unabhängig sind. Mit den Bezeichnungen aus (1.1.21) gilt nämlich im letzteren Fall

$$\begin{aligned} P\Big(\bigcap_{i=1}^n B_i\Big) &= P\Big(\bigcap_{i=1}^n \{\mathbb{1}_{B_i} = 1\}\Big) \\ &= \prod_{i=1}^n P(\mathbb{1}_{B_i} = 1) = \prod_{i=1}^n P(B_i), \end{aligned}$$

da mit $\{\mathbb{1}_{A_i}\}$ auch die Indikatorvariablen $\{\mathbb{1}_{B_i}\}$ stochastisch unabhängig sind wegen $\{\mathbb{1}_{A_i} = 1\} = A_i$, $\{\mathbb{1}_{A_i^c} = 1\} = \{\mathbb{1}_{A_i} = 0\} = A_i^c$. Umgekehrt argumentiert man analog, so daß sich mit (1.4.48) und (1.4.33) ein alternativer Beweis des Satzes 1.1.2 ergibt.

Indikatorvariablen — gerade auch unabhängiger Ereignisse — und meßbare Transformationen dieser (und anderer Zufallsvariablen) spielen nicht nur in der Modellierung stochastischer Systeme, sondern auch in der (Lebesgue'schen) Integrationstheorie, auf die wir im Zusammenhang mit Momenten von Verteilungen — z.B. Erwartungswert und Varianz — in Kapitel 1.6 zu sprechen kommen werden, eine große Rolle.

Mit Hilfe von Indikatorvariablen läßt sich auch die stochastische Unabhängigkeit von *Ereignissen* $\{A_i\}_{i\in I} \subseteq \mathcal{A}$ und *Zufallselementen* $\{\boldsymbol{X}_j\}_{j\in J}$ mit (abzählbaren) Indexmengen $I, J \subseteq \mathbb{N}$ formulieren, indem man die zu den Ereignissen $\{A_i\}_{i\in I}$ gehörigen Indikatorvariablen $\{\mathbb{1}_{A_i}\}_{i\in I}$ betrachtet.

Eine besonders einfache Charakterisierung der stochastischen Unabhängigkeit von Zufallsvariablen $X_1, \ldots, X_n$ erhält man im Fall stetiger (Rand-)Verteilungen. Bezeichnen $f_1, \ldots, f_n$ nämlich Dichten zu den Verteilungsfunktionen $F_1, \ldots, F_n$ von $P^{X_1}, \ldots, P^{X_n}$, so ist aufgrund von (1.4.26) und (1.4.30)

$$\prod_{i=1}^{n} f_i(x_i) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} \prod F_i(x_i) \tag{1.4.51}$$

in allen Stetigkeitspunkten $x_i \in \mathbb{R}$ von f_i, $1 \leq i \leq n$. Der Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ besitzt also eine Dichte der Form

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i),\ x_1, \ldots, x_n \in \mathbb{R}, \tag{1.4.52}$$

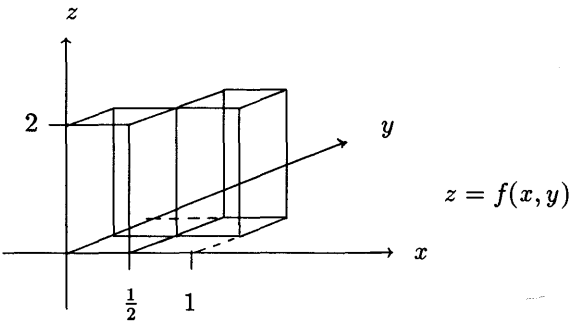
was gerade bedeutet, daß eine gemeinsame Dichte der Zufallsvariablen $X_1, \ldots, X_n$ gegeben ist durch das Produkt der Randdichten $f_1, \ldots, f_n$. Erkennt man umgekehrt, daß die gemeinsame Diche f in einer durch (1.4.52) gegebenen Form mit Dichten $f_1, \ldots, f_n$ darstellbar ist, so sind die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig, und f_i ist eine Dichte der Verteilung P^{X_i}, $1 \leq i \leq n$.

Beziehung (1.4.52) bleibt sogar dann noch gültig, wenn die X_i unabhängige, r_i-dimensionale Zufallsvektoren $\boldsymbol{X}_i$ sind ($r_i \in \mathbb{N}$) und jede der Abbildungen f_i eine r_i-dimensionale Dichte von $P^{\boldsymbol{X}_i}$, $1 \leq i \leq n$ (mit $\boldsymbol{x}_i \in \mathbb{R}^{r_i}$) darstellt. Dichten von Randverteilungen komplexerer Ordnungen erhält man völlig analog durch entsprechende Teilproduktbildungen.

Ist $\boldsymbol{X} = (X_1, \ldots, X_n)$ dagegen ein beliebiger Zufallsvektor mit einer Dichte f (d.h. nicht notwendig unabhängigen Komponenten), so erhält man aufgrund von Lemma 1.4.4 Dichten derartiger Randverteilungen durch Integrieren nach den nicht-relevanten Variablen. Beispielsweise ist durch

$$f(x,y) = \begin{cases} 2 & \text{für } 0 \leq x,y \leq \frac{1}{2} \text{ oder } \frac{1}{2} \leq x,y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte der Gleichverteilung über der Menge $[0, \frac{1}{2}]^2 \cup [\frac{1}{2}, 1]^2$ gegeben.



Ein Zufallsvektor (X,Y) mit dieser Verteilung besitzt die Randdichten $f_1 = \mathbb{1}_{[0,1]} = f_2$, d.h. X und Y sind jeweils auf $[0,1]$ gleichverteilt; dies folgt z.B. aus der Rechnung

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} \int_{-\infty}^{\infty} 0\,dy = 0 & \text{für } x < 0 \text{ oder } x > 1 \\ \int_0^{\frac{1}{2}} 2\,dy = 1 & \text{für } 0 \le x \le \frac{1}{2} \\ \int_{\frac{1}{2}}^1 2\,dy = 1 & \text{für } \frac{1}{2} \le x \le 1; \end{cases}$$

analog für f_2. X und Y sind aber nicht unabhängig, da für das Produkt der Randdichten gilt $f_1(x) \cdot f_2(y) = \mathbb{1}_{[0,1]^2}(x,y)$, $x,y \in \mathbf{R}$, was offensichtlich wesentlich von der gegebenen Dichte f verschieden ist; so ist etwa $P\big((X,Y) \in [0,\frac{1}{2}] \times [\frac{1}{2},1]\big) = 0 \neq \frac{1}{4} = P(X \in [0,\frac{1}{2}]) \cdot P(Y \in [\frac{1}{2},1])$. Für einen auf $[0,1]^2$ gleichverteilten Zufallsvektor (X,Y), dessen Verteilung gerade die Dichte $f(x,y) = \mathbb{1}_{[0,1]^2}(x,y)$, $x,y \in \mathbf{R}$, besitzt, sind allerdings die Komponenten X und Y unabhängig und jeweils auf $[0,1]$ gleichverteilt.

Entsprechendes gilt im Fall von Zähldichten; nach der Bemerkung im Anschluß an Lemma 1.4.5 sind nämlich diskrete Zufallsvariablen $X_1, \ldots, X_n$, $n \in \mathbf{N}$, mit gemeinsamer Zähldichte f und Rand–Zähldichten $f_1, \ldots, f_n$ genau dann stochastisch unabhängig, wenn wieder Beziehung (1.4.52) gilt. Im allgemeinen erhält man im diskreten Fall die Rand–Zähldichten — analog zum stetigen Fall — durch *Summation* über die nicht–relevanten Variablen.

Die durch Beziehung (1.3.10) gegebene Zähldichte des Zufallsvektors (X,Y) aus dem anfangs betrachteten Suchbeispiel ist offenbar nicht das Produkt der Zähldichten von X und Y; die Zufallsvariablen X und Y sind also nicht stochastisch unabhängig, was ja bereits durch Beziehung (1.3.11) angedeutet wurde.

Den funktionalen Zusammenhang dieser Zufallsvariablen kann man neben (1.3.7) auch an der folgenden, auf Indikatorvariablen beruhenden Darstellung ablesen; mit den Bezeichnungen aus (1.1.2) gilt nämlich wegen der paarweisen Disjunktheit der Mengen $A_0, \ldots, A_n$:

$$Y = \sum_{k=0}^{n} k\,\mathbb{1}_{A_k}$$

$$X = \sum_{k=1}^{n} k\,\mathbb{1}_{A_k} + n\,\mathbb{1}_{A_0}.$$

Beide Zufallsvariablen sind also (meßbare) Funktionen derselben Zufallsvariablen $\mathbb{1}_{A_0}, \ldots, \mathbb{1}_{A_n}$.

Wir wollen diesen Abschnitt mit einer Bemerkung zur Existenz beliebiger, stochastisch unabhängiger Folgen von Zufallsvariablen oder allgemeiner Zufallselementen auf einem gemeinsamen Grundwahrscheinlichkeitsraum abschließen; die Existenz solcher Folgen wird nämlich in den meisten stochastischen Modellen stillschweigend angenommen. Ähnlich wie im Fall beliebiger unabhängiger Ereignisse kann man solche Folgen wieder kanonisch mit Hilfe von Produkträumen konstruieren. Sind zunächst $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $n \in \mathbb{N}$, beliebige Zufallselemente auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, so erhält man durch die Definition

$$\boldsymbol{X}_i^*(\omega_1, \ldots, \omega_n) = \boldsymbol{X}_i(\omega_i), \ 1 \leq i \leq n, \tag{1.4.53}$$

unabhängige Zufallselemente $\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_n^*$ auf dem (gemeinsamen) Produktraum $(\bigtimes_{i=1}^n \Omega, \bigotimes_{i=1}^n \mathcal{A}, \bigotimes_{i=1}^n P)$, die dieselbe Verteilung wie die ursprünglichen Zufallselemente besitzen, d.h. es gilt $P^{\boldsymbol{X}_i} = Q^{\boldsymbol{X}_i^*}$, $1 \leq i \leq n$, mit $Q = \bigotimes_{i=1}^n P$. Der Übergang von $\boldsymbol{X}_i$ zu $\boldsymbol{X}_i^*$ geschieht dabei lediglich durch Erweiterung des Arguments ω_i um — redundante — Argumente $\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_n$. Im Fall von Folgen von Zufallselementen, also in der Situation abzählbar-unendlich vieler Argumente $\omega_1, \omega_2, \ldots$ geht man analog vor, indem man den abzählbar-unendlichen Produktraum $(\bigtimes_{i=1}^\infty \Omega, \bigotimes_{i=1}^\infty \mathcal{A}, \bigotimes_{i=1}^\infty P)$ benutzt.

Durch eventuelle Vergrößerung des ursprünglich gegebenen Wahrscheinlichkeitsraumes $(\Omega, \mathcal{A}, P)$ über kartesische Produkte läßt sich also — in obigem Sinne — stets die Existenz sowohl stochastisch unabhängiger Ereignisse als auch unabhängiger Zufallselemente erreichen. Hiervon werden wir im folgenden häufiger Gebrauch machen, ohne jedesmal auf die dahinterstehende Problematik zu verweisen.

1.5. Aufgaben

1.1 Es sei P eine Wahrscheinlichkeitsverteilung auf $\mathcal{B}^1$ mit der Eigenschaft (1.0.6). Zeigen Sie:

$$P\big((a,b]\big) = b - a, \quad 0 \le a < b \le 1.$$

Anleitung: Folgern Sie aus (1.0.6) zunächst, daß wegen $(0,1] = \bigcup_{k=1}^{n} \left(\frac{k-1}{n}, \frac{k}{n}\right]$ und der Additivität von P gilt: $P\Big(\left(\frac{k-1}{n}, \frac{k}{n}\right]\Big) = \frac{1}{n}$ für alle $1 \le k \le n$, $n \in \mathbf{N}$; hieraus ergibt sich die Behauptung für rationale a, b. Benutzen Sie dann die Stetigkeitseigenschaften (1.1.38) und (1.1.39) von P.

1.2 Es sei $\Omega = \{\omega_1, \ldots, \omega_n\}$, $n \in \mathbf{N}$, eine nicht–leere Menge; $\mathcal{P}_{k,n} = Perm_k^n(\Omega; o.W.)$, $1 \le k \le n$, bezeichne die Menge aller (k,n)–Permutationen ohne Wiederholung.
$A_{r,j} = \{\eta = (\eta_1, \ldots, \eta_k) \in \mathcal{P}_{k,n} \mid \eta_j = \omega_r\}$ bezeichne das Ereignis, daß die Permutation η an der j–ten Stelle das Element ω_r enthält ($1 \le r \le n$, $1 \le j \le k$). Zeigen Sie:

$$P(A_{r,j}) = \frac{1}{n}, \quad 1 \le r \le n,\ 1 \le j \le k,$$

wenn alle (k,n)–Permutationen gleichwahrscheinlich sind.

1.3 In Verallgemeinerung von Aufgabe 1.2 sei jetzt eine beliebige Wahrscheinlichkeitsverteilung P über $\mathcal{P}_{k,n}$ gegeben. Zeigen Sie:

$$P(A_{r,j}) = \sum_{\eta \in \mathcal{P}_{k,n} : \eta_j = \omega_r} P(\{\eta\}), \quad 1 \le r \le n,\ 1 \le j \le k.$$

Bestimmen Sie die Wahrscheinlichkeiten $P(A_{r,j})$, $1 \le r \le n$, $1 \le j \le k$ für den Fall $n = 3$, $k = 2$, $P(\{\eta\}) = \frac{\eta_1 + \eta_2}{24}$, $\eta \in \mathcal{P}_{k,n}$, $\Omega = \{1, \ldots, n\}$.

1.4 Zeigen Sie anhand des folgenden Gegenbeispiels, daß in Aufgabe 1.3 selbst im Fall $k = n$ die Verteilung P über $\mathcal{P}_{k,n}$ i.a. *nicht* durch die Wahrscheinlichkeiten $P(A_{r,j})$, $1 \le r \le n$, $1 \le j \le k$ bestimmt ist:
Wählen sie dazu $k = n = 3$ sowie P beliebig mit

$$p_{\min} := \min\{P(\{\eta\}) \mid \eta \in \mathcal{P}_{k,n}\}, \quad p_{\max} := \max\{P(\{\eta\}) \mid \eta \in \mathcal{P}_{k,n}\}.$$

Für $0 \le \alpha \le \min\{p_{\min}, 1 - p_{\max}\}$ sei die Verteilung P_α über $\mathcal{P}_{k,n}$ definiert durch

$$P_\alpha(\{\eta\}) = \begin{cases} P(\{\eta\}) - \alpha & \text{für } \eta \in \{(123),(231),(312)\} \\ P(\{\eta\}) + \alpha & \text{für } \eta \in \{(213),(132),(321)\}. \end{cases}$$

Zeigen Sie: $P_\alpha(A_{r,j}) = P(A_{r,j})$ für alle $1 \le r \le n$, $1 \le j \le k$. Was bedeutet dies im Fall einer Gleichverteilung $P = \mathfrak{L}(\mathcal{P}_{k,n})$?

1.5 (**binary search**) Es sei $n \in \mathbf{N}$ und $\Omega = \{0, 1, \ldots, 2^n - 1\}$ wie in Beispiel 1.1.1. Die Mengen A_k, $0 \le k \le n$ mögen wieder das Ereignis, daß das Schlüsselelement in $k > 0$ Schritten gefunden wird bzw. nicht in dem Feld vorhanden ist, bezeichnen. P sei eine beliebige Wahrscheinlichkeitsverteilung über Ω. Zeigen Sie in Verallgemeinerung von (1.1.2) und (1.1.3):

$$P(A_k) = \sum_{j=1}^{2^{k-1}} P\big(\{(2j-1)2^{n-k}\}\big), \quad 1 \le k \le n.$$

Zeigen Sie speziell für die durch

$$P(\{i\}) = \begin{cases} \dfrac{2i}{4^n} & \text{für } 1 \le i \le 2^n - 1 \\ \dfrac{1}{2^n} & \text{für } i = 0 \end{cases}$$

über Ω gegebene Verteilung:

$$P(A_k) = \frac{2^{k-1}}{2^n}, \quad 1 \le k \le n.$$

Wie ist dieses Ergebnis im Vergleich zu Beziehung (1.1.3) zu interpretieren?

1.6 Ω sei eine nicht-leere Menge sowie $A, B \subset \Omega$, $A \neq B$. Bestimmen Sie den von $\mathcal{E} = \{A, B\}$ erzeugten Ring $\mathcal{R}$ sowie das Dynkin-System $\delta(\mathcal{E})$ und die σ-Algebra $\sigma(\mathcal{E})$. Geben Sie für den Fall $\Omega = \{1,2,3,4\}$, $A = \{1,2,3\}$, $B = \{2,3,4\}$ alle Wahrscheinlichkeitsverteilungen P über Ω mit der Eigenschaft $P(A) = P(B) = \frac{2}{3}$ durch Spezifikation der zugehörigen Elementarwahrscheinlichkeiten an. Wie sieht hier das von $\mathcal{E}$ erzeugte Dynkin-System $\delta(\mathcal{E})$ aus? Warum existiert keine *eindeutige* Fortsetzung von P über $\mathcal{E}$ hinaus?
Gibt es eine Lösung P derart, daß die Ereignisse $\{1,2\}$ und $\{2,3\}$ unter P stochastisch unabhängig sind?
Hinweis: Wählen Sie $p = P(\{2\})$ als Parameter für die Lösungsgesamtheit.

1.7 Es seien $\{(\Omega_n, \mathcal{A}_n)\}_{n \in \mathbf{N}}$ Meßräume mit $\Omega_i \cap \Omega_j = \emptyset$, $i \neq j$, $i, j \in \mathbf{N}$. Zeigen Sie:

$$\mathcal{A} := \left\{ \bigcup_{n=1}^{\infty} A_n \mid A_n \in \mathcal{A}_n,\ n \in \mathbf{N} \right\}$$

ist eine σ-Algebra über der Menge $\Omega := \bigcup_{n=1}^{\infty} \Omega_n$. Gilt eine entsprechende Aussage auch dann, wenn die Systeme $\mathcal{A}_n$ einen Ring (ein Dynkin-System) bilden? Kann man hier auf die paarweise Disjunktheit der Grundmengen Ω_n verzichten (Gegenbeispiele)?

1.8 Es sei $\Omega = \{r+1, r+2, \ldots, r+n\}$, $r, n \in \mathbf{N}$ sowie $1 \le m \le n$. Zeigen Sie, daß durch

$$f(k) = \begin{cases} \dfrac{2(k-r)}{m(n+1)} & \text{für } r+1 \le k \le r+m \\ \dfrac{2(r+n+1-k)}{(n-m+1)(n+1)} & \text{für } r+m \le k \le r+n \end{cases} \qquad (k \in \Omega)$$

eine Wahrscheinlichkeitsverteilung P mit Zähldichte $P(\{k\}) = f(k)$, $k \in \Omega$ gegeben ist. P heißt auch (diskrete) *Dreiecksverteilung* über Ω, i.Z.: $P = \Delta(r; m, n)$. Skizzieren Sie die Zähldichte f sowie die Verteilungsfunktion F zu P.
Zeigen Sie analog, daß für $a < b$, $a, b, r \in \mathbf{R}$, durch

$$f(x) = \begin{cases} \dfrac{2(x-r)}{ab} & \text{für } r \le x \le r+a \\ \dfrac{2(r+b-x)}{(b-a)b} & \text{für } r+a \le x \le r+b \\ 0 & \text{sonst} \end{cases} \qquad (x \in \mathbf{R})$$

die Dichte einer Verteilung P auf $\mathcal{B}^1$ gegeben ist. P heißt entsprechend (stetige) Dreiecksverteilung über $[r, r+b]$, i.Z.: $P = D(r; a, b)$. Skizzieren Sie die Dichte sowie die Verteilungsfunktion zu P.

1.9 Es sei $\Omega = \mathbf{Q} \cap (0,1]$ und $\beta \in (0,1)$. Die Abbildung g sei definiert als

$$g(k,n) = \frac{1}{n} \beta^{n-1}(1-\beta),\ 1 \le k \le n,\ n \in \mathbf{N}.$$

Zeigen Sie: durch

$$P(\{\omega\}) = \sum_{\substack{1 \le k \le n \\ \frac{k}{n} = \omega}} g(k,n), \quad \omega \in \Omega,$$

ist eine diskrete Verteilung über Ω gegeben mit

$$P(\{\omega\}) = \sum_{n=1}^{\infty} \frac{1}{nq} \beta^{nq-1}(1-\beta) = -\frac{1-\beta}{\beta q} \ln\left(1-\beta^q\right),$$

wenn $\omega = \frac{p}{q} \in \Omega$ in teilerfremder Form mit $p, q \in \mathbf{N}$ gegeben ist. Die zugehörige Verteilungsfunktion F läßt sich ausdrücken als

$$F(x) = \sum_{\substack{\omega \in \Omega \\ \omega \leq x}} P(\{\omega\}) = \sum_{\substack{1 \leq k \leq n \\ \frac{k}{n} \leq x}} g(k,n) = (1-\beta) \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{n} \beta^{n-1}$$

für $0 < x \leq 1$.
Bemerkung: Die Verteilung in (1.2.18) ergibt sich durch die Wahl $\beta = \frac{1}{2}$.

1.10 Wie in Aufgabe 1.2 bezeichne $\mathcal{P}_{k,n}$ die Menge aller (k,n)–Permutationen ohne Wiederholung, $1 \leq k \leq n$. Ferner sei $P = \mathfrak{L}\left(\mathcal{P}_{k,n}\right)$ die Gleichverteilung über $\mathcal{P}_{k,n}$, $k \geq 2$. Die Zufallsvariablen $X_1, \ldots, X_k$ seien auf $\mathcal{P}_{k,n}$ definiert durch

$$X_1(\eta) = 1, \quad X_i(\eta) = \begin{cases} 1, & \text{wenn } \eta_i > \max\{\eta_1, \ldots, \eta_{i-1}\} \\ 0, & \text{sonst} \end{cases} \qquad (2 \leq i \leq k,\ \eta \in \mathcal{P}_{k,n}).$$

Zeigen Sie, daß für alle $m \in \{2, \ldots, k\}$ und $x_1, \ldots, x_k \in \{0,1\}$ gilt:

$$\#\left[\bigcap_{i=1}^{m-1} \{X_i = x_i\}\right] = m \cdot \#\left[\bigcap_{i=1}^{m-1} \{X_i = x_i\} \cap \{X_m = 1\}\right].$$

Schließen Sie damit auf

$$P\Big(\bigcap_{i=1}^{m-1} \{X_i = x_i\} \cap \{X_m = 1\}\Big) = \frac{1}{m} P\Big(\bigcap_{i=1}^{m-1} \{X_i = x_i\}\Big)$$

und folgern Sie hieraus, daß die Zufallsvariablen $X_1, \ldots, X_k$ stochastisch unabhängig sind mit

$$P(X_i = 1) = \frac{1}{i}, \quad 1 \leq i \leq k.$$

Anleitung: Zu jeder Permutation $\eta \in \bigcap_{i=1}^{m-1} \{X_i = x_i\} \cap \{X_m = 1\}$ existiert eine Permutation $\sigma = \sigma_\eta$ der Indizes $\{1, \ldots, m-1\}$, welche $\eta_1, \ldots, \eta_m$ der Größe nach ordnet, d.h. $\eta_{\sigma(1)} < \ldots < \eta_{\sigma(m-1)} < \eta_m$, da wegen $X_m = 1$ η_m das größte der Elemente $\eta_1, \ldots, \eta_m$ ist. Die Permutationen der Menge $\bigcap_{i=1}^{m-1} \{X_i = x_i\}$ erhält man nun dadurch, daß man in der Folge $\eta_{\sigma(1)} < \ldots < \eta_{\sigma(m-1)} < \eta_m$ genau ein Element streicht (m Möglichkeiten!), die verbleibenden, geordneten Elemente mit der inversen Permutation σ^{-1} zurückordnet und mit dem gestrichenen Element sowie den restlichen Elementen $\eta_{m+1}, \ldots, \eta_k$ vervollständigt.
Bemerkung: Die Zufallsvariable $S = \sum_{i=1}^{k} X_i$ gibt die Anzahl der (Um–)Speicherungen an, die bei der linearen Maximumsuche in der — zufälligen — Permutation $\eta \in \mathcal{P}_{k,n}$ benötigt werden (vgl. Kemp (1984), §3, oder Knuth (1968), 1.2.10.). Ein allgemeineres Modell ohne Gleichverteilungsannahmen wird in Kapitel 4.1 behandelt.

1.11 Bestimmen Sie eine Dichte f der Gleichverteilung $P = \mathcal{R}(A)$ über dem Quadrat

$$A = \{(x,y) \in \mathbf{R}^2 \mid |x| + |y| \leq 1\}$$

(Skizze!) und geben Sie die zugehörige Verteilungsfunktion in expliziter Form an.
Zeigen Sie, daß die jeweiligen Randverteilungen Dreiecksverteilungen im Sinne von Aufgabe 1.8 sind mit $r = -1, a = 1, b = 2$. Ist P eine Produktverteilung?

2. Transformation und Integration von Zufallsvariablen

Nach der Behandlung eher grundsätzlicher Fragestellungen in den vorangehenden Abschnitten wollen wir uns nun der Untersuchung der Grundlagen einiger wichtiger stochastischer Modelle, die fundamental für die Beschreibung vieler, z.T. erheblich komplizierterer Sachverhalte sind, zuwenden. Die bisherigen Ausführungen haben dabei deutlich gemacht, daß die Beschreibung solcher Modelle — etwa das eingangs aufgeworfene Suchproblem — sowohl in der Sprache von σ–Algebren und Wahrscheinlichkeitsverteilungen als auch durch Zufallsvariable erfolgen kann; beide Zugänge sind sogar in gewisser Weise äquivalent, wie Beziehung (1.3.4) zeigt. Dabei braucht man sich nicht einmal auf Zufalls*variable* zu beschränken; die in (1.3.4) beschriebene Methode funktioniert auch ganz allgemein für Zufallselemente in einem Meßraum $(\mathcal{X}, \mathcal{B})$. Die Verwendung von Zufallsvariablen (oder allgemeiner Zufallselementen) hat allerdings den großen Vorteil, daß damit arithmetische Verknüpfungen zwischen den Zufallsergebnissen und die daraus resultierenden Wahrscheinlichkeitsverteilungen einfacher beschreibbar werden. Eine adäquate Beschreibung stochastischer Modelle kommt damit praktisch ohne die Verwendung von Zufallsvariablen bzw. Zufallselementen und ihren (meßbaren) Transformationen nicht aus.

Besondere Bedeutung kommt dabei der Bildung arithmetischer Mittel von Beobachtungen (d.h. Realisationen von Zufallsvariablen) zu, die dem Informatiker etwa im Bereich der Average–Case–Analyse von Algorithmen begegnet. Interessiert man sich für das Langzeitverhalten solcher Algorithmen, gelangt man über das bereits seit dem 18. Jahrhundert bekannte "Gesetz der großen Zahlen" zwangsläufig zum Begriffs des *Erwartungswerts* einer Zufallsvariablen, d.h. einer Kenngröße von Verteilungen, welche z.B. die Vergleichbarkeit von Algorithmen bezüglich ihrer Effizienz ermöglicht. Eine rigorose Behandlung solcher Sachverhalte erfordert allerdings einen etwas allgemeineren Integrationsbegriff als den des Riemann–Integrals; der hier angemessene Lebesgue'sche Zugang wird deshalb in seinen wesentlichen Grundzügen kurz mitbehandelt.

Die zuletzt angesprochene Problemstellung der Konvergenz von Zufallsvariablen und deren Verteilungen mit ihren Anwendungsmöglichkeiten beispielsweise im CAD (Bézier–Kurven und –Flächen) wird ebenfalls in diesem Kapitel untersucht.

2.1. Spezielle Verteilungen

In diesem Abschnitt wollen wir uns zunächst mit Verknüpfungen bzw. allgemeiner meßbaren Transformationen von Zufallsvariablen beschäftigen, da diese für die Gewinnung einer Reihe spezieller Verteilungen bzw. Verteilungsklassen von grundlegender Bedeutung sind.

Wir beginnen mit einer leichten Verallgemeinerung von Beziehung (1.4.45).

Lemma 2.1.1. *(Meßbarkeit von Transformationen)*
Es seien $\boldsymbol{X} : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{X}, \mathcal{B})$ *und* $\boldsymbol{Y} : (\mathcal{X}, \mathcal{B}) \longrightarrow (\mathcal{Y}, \mathcal{C})$ *Zufallselemente in Meßräumen* $(\mathcal{X}, \mathcal{B})$ *bzw.* $(\mathcal{Y}, \mathcal{C})$. *Dann ist die Komposition* $\boldsymbol{Z} = \boldsymbol{Y} \circ \boldsymbol{X}$ *ein auf* $(\Omega, \mathcal{A})$ *definiertes Zufallselement in* $(\mathcal{Y}, \mathcal{C})$.

Beweis. Für beliebige Mengen $C \in \mathcal{C}$ ist

$$\boldsymbol{Z}^{-1}(C) = (\boldsymbol{Y} \circ \boldsymbol{X})^{-1}(C) = \boldsymbol{X}^{-1}(\underbrace{\boldsymbol{Y}^{-1}(C)}_{\in \mathcal{B}}) \in \mathcal{A}.$$

$\boldsymbol{Z}$ ist also meßbar und somit ein Zufallselement in $(\mathcal{Y}, \mathcal{C})$, wie behauptet. ∎

Besonders wichtig für praktische Anwendungen sind dabei meßbare Transformationen von m–dimensionalen Zufallsvektoren $\boldsymbol{X} = (X_1, \ldots, X_m)$, z.B. durch stetige, stückweise oder auch einseitig stetige Abbildungen $\boldsymbol{Y} : \mathbf{R}^m \longrightarrow \mathbf{R}^n$ mit $m, n \in \mathbf{N}$. Solche Abbildungen sind meßbar, da etwa bei Stetigkeit Urbilder $\boldsymbol{Y}^{-1}(C)$ *offener* Mengen $C \in \mathcal{B}^n$ offene Mengen in $\mathbf{R}^m$ sind, und nach (1.4.13) die offenen Mengen Erzeuger der jeweiligen Borel'schen σ–Algebren bilden.

So sind insbesondere die arithmetischen Verknüfungen Summe, Differenz, Produkt und Quotient stetige Abbildungen von $\mathbf{R}^2$ (bzw. $\mathbf{R} \times (\mathbf{R} \setminus \{0\})$ im letzteren Fall) in $\mathbf{R}$. Damit ergeben endlich viele solcher arithmetischen Verknüpfungen von Zufallsvariablen wieder Zufallsvariablen. Allgemeiner führt auch die abzählbare Supremums–, Infimums– oder Limesbildung von Zufallsvariablen wieder zu Zufallsvariablen (evtl. mit Werten in der größeren Menge $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$; eine geeignete σ–Algebra hierüber ist etwa $\overline{\mathcal{B}^1} = \sigma\left(\mathcal{B}^1 \cup \{\{-\infty\}, \{\infty\}\}\right)$.) Mit einer Folge $\{X_n\}_{n \in \mathbf{N}}$ von Zufallsvariablen sind also in diesem Sinn auch $\sup_{n \in \mathbf{N}} X_n$, $\inf_{n \in \mathbf{N}} X_n$, $\limsup_{n \to \infty} X_n$, $\liminf_{n \to \infty} X_n$ und $\lim_{n \to \infty} X_n$ (falls existent) Zufallsvariablen. Beispielsweise gilt für jede reelle Zahl x

$$\left\{ \sup_{n \in \mathbf{N}} X_n \leq x \right\} = \bigcap_{n \in \mathbf{N}} \{X_n \leq x\} \in \mathcal{A};$$

da das System $\mathcal{E}_5$ aber ein Erzeuger von $\mathcal{B}^1$ ist, folgt somit, daß $\sup_{n \in \mathbf{N}} X_n$ meßbar, also eine Zufallsvariable ist. Ähnlich argumentiert man in den anderen Fällen.

Die anfangs mehrfach erwähnte Gleichverteilung über $\Omega = [0, 1]$ bzw. $\Omega = (0, 1]$ spielt bei Transformationen von Zufallsvariablen eine besondere Rolle, da man jede beliebige Verteilung auf $\mathcal{B}^1$ aus ihr durch eine geeignete Transformation erzeugen kann. Dieses Faktum wird z.B. im Bereich der Simulation (Kapitel 6) ausgenutzt, um Verteilungen auf $\mathcal{B}^1$ zu "simulieren"; man hat sich dazu lediglich eine brauchbare Methode zur Erzeugung gleichverteilter "Zufallszahlen" zu überlegen, die man dann entsprechend transformiert.

Bevor wir das entsprechende Resultat formulieren können, benötigen wir noch den Begriff der Pseudo–Inversen einer schwach monotonen Funktion.

Definition 2.1.1. *(Pseudo–Inverse)*
Es sei $\psi : \mathbf{R} \longrightarrow \mathbf{R}$ *eine schwach monoton wachsende, rechtsseitig stetige Funktion und*

$$I(\psi) = \inf\{\psi(x) \mid x \in \mathbf{R}\}, \quad S(\psi) = \sup\{\psi(x) \mid x \in \mathbf{R}\}. \tag{2.1.1}$$

Dann ist auf dem offenen Intervall $\big(I(\psi), S(\psi)\big)$ *die Pseudo–Inverse* ψ^{-1} *von* ψ *erklärt durch*

$$\psi^{-1}(y) = \inf\{x \in \mathbf{R} \mid \psi(x) \geq y\}, \quad I(\psi) < y < S(\psi). \tag{2.1.2}$$

Aufgrund der rechtsseitigen Stetigkeit von ψ kann dabei das inf in (2.1.1) auch durch min ersetzt werden.

Die im folgenden angegebenen Eigenschaften einer Pseudo–Inversen lassen sich leicht durch geeignete Konvergenzbetrachtungen zeigen.

Lemma 2.1.2. *(Eigenschaften einer Pseudo–Inversen)*
Unter den Voraussetzungen von Definition 2.1.1 gilt:

a) ψ^{-1} *ist auf* $\big(I(\psi), S(\psi)\big)$ *schwach monoton wachsend und linksseitig stetig.*

b) Es ist

$$\psi(\psi^{-1}(y)) \geq y \quad \text{für alle } I(\psi) < y < S(\psi) \tag{2.1.3}$$

mit Gleichheit in (2.1.3), wenn ψ *in* $\psi^{-1}(y)$ *stetig ist.*

c) Es ist

$$\psi^{-1}(\psi(x)) \leq x \quad \text{für alle } I(\psi) < \psi(x) < S(\psi) \tag{2.1.4}$$

mit Gleichheit in (2.1.4), wenn ψ^{-1} *in* $\psi(x)$ *stetig ist.*

d) Es ist

$$y \leq \psi(x) \iff \psi^{-1}(y) \leq x \quad \text{für } I(\psi) < \psi(x), y < S(\psi). \tag{2.1.5}$$

Beweis. Wir wollen hier exemplarisch nur die Aussage d) zeigen.
Unter den Voraussetzungen auf der rechten Seite in (2.1.5) gilt:
Ist $\min\{z \in \mathbf{R} \mid \psi(z) \geq y\} = \psi^{-1}(y) \leq x$, so erhält man für $z^* = \psi^{-1}(y)$ sofort $y \leq \psi(z^*) \leq \psi(x)$ wegen der schwachen Monotonie von ψ, also $y \leq \psi(x)$.
Ist umgekehrt $\psi(x) \geq y$, so ist $\psi^{-1}(y) = \min\{z \in \mathbf{R} \mid \psi(z) \geq y\} \leq x$. ∎

Mit Hilfe von Lemma 2.1.2 läßt sich nun das angekündigte Resultat leicht beweisen.

Satz 2.1.1. *(Erzeugung von Verteilungen)*
Es sei Q *eine Wahrscheinlichkeitsverteilung auf* $\mathcal{B}^1$ *mit Verteilungsfunktion* F *und* U *eine über* $[0,1]$ *gleichverteilte Zufallsvariable auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$. *Dann besitzt die vermöge der Pseudo–Inversen* F^{-1} *durch*

$$X(\omega) = \begin{cases} F^{-1}\big(U(\omega)\big) & \text{für } U(\omega) \in (0,1), \\ 0 & \text{sonst,} \end{cases} \quad \omega \in \Omega, \tag{2.1.6}$$

definierte Zufallsvariable X *die Verteilung* $P^X = Q$. *Ist umgekehrt* X *eine reellwertige Zufallsvariable mit stetiger Verteilungsfunktion* F, *so ist* $F(X)$ *über* $[0,1]$ *gleichverteilt.*

Beweis. Zunächst ist die Meßbarkeit der Abbildung X nachzuweisen. Es ist $I(F) = 0$, $S(F) = 1$; mit der linksseitigen Stetigkeit von F^{-1} gemäß Lemma 2.1.2

a) erhält man für alle Borel–Mengen $B \not\ni 0$

$$X^{-1}(B) = U^{-1}\big(\underbrace{(F^{-1})^{-1}(B)}_{\in(0,1)\cap\mathcal{B}^1}\big) \in \mathcal{A}$$

und speziell für $B = \{0\}$

$$X^{-1}(\{0\}) = U^{-1}\big(\underbrace{(F^{-1})^{-1}(\{0\})}_{\in(0,1)\cap\mathcal{B}^1}\big) \cup U^{-1}\big(\underbrace{(0,1)^c}_{\in\mathcal{B}^1}\big) \in \mathcal{A},$$

also insgesamt die Meßbarkeit von X.
Zum Nachweis der geforderten Verteilungseigenschaft genügt es, die Verteilungsfunktion der Zufallsvariablen X zu betrachten. Mit Lemma 2.1.2 d) folgt nun für $0 < F(x) < 1$

$$P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$$

mit $P(U \notin (0,1)) = 0$. Dies war aber zu zeigen.
Zum Beweis der zweiten Aussage des Satzes können wir ohne Beschränkung der Allgemeinheit annehmen, daß die Zufallsvariable X bereits die durch (2.1.6) gegebene Form besitzt. Wegen der vorausgesetzten Stetigkeit von F gilt dann nach Lemma 2.1.2, Teil b) $F(X) = F\big(F^{-1}(U)\big) = U$, falls $0 < U < 1$. Wegen $P(0 < U < 1) = 1$ ist die Aussage damit bewiesen. ∎

Offensichtlich kommt es in Beziehung (2.1.6) nur auf den oberen Teil der Definition der Zufallsvariablen X an, da der Fall $U(\omega) \notin (0,1)$ nur mit Wahrscheinlichkeit 0 eintritt, die Definition von $X(\omega)$ in dieser Situation also mehr oder weniger willkürlich ist (allerdings unter Beibehaltung der Meßbarkeitseigenschaft). Man sagt auch, daß die Zufallsvariable U *fast sicher* nur Werte im Intervall $(0,1)$ annimmt, bzw. daß die Zufallsvariable X *fast sicher* durch die Beziehung $X = F^{-1}(U)$ definiert ist. Wir werden im folgenden mehrfach von solcher Sprechweise Gebrauch machen, ohne jedesmal explizit die Ausnahmefälle, die lediglich mit Wahrscheinlichkeit 0 eintreten, zu behandeln; es ist dann davon auszugehen, daß stillschweigend eine geeignete Vereinbarung für solche Fälle besteht. Eine Präzisierung dieser Sprechweise enthält die folgende

Definition 2.1.2. *(fast sicher bestehende Eigenschaften)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathfrak{E}$ eine Eigenschaft derart, daß die Menge $A = \{\omega \in \Omega \mid \omega$ besitzt die Eigenschaft $\mathfrak{E}\}$ meßbar, d.h. Element von $\mathcal{A}$ ist. Gilt dann $P(A) = 1$, so sagt man, daß die Eigenschaft $\mathfrak{E}$ fast sicher (abgekürzt: f.s.) besteht.

Im obigen Fall ist etwa $\mathfrak{E}$ die Eigenschaft, durch $F^{-1}(U)$ erklärt zu sein, d.h. es ist $A = \big\{\omega \in \Omega \mid X(\omega) = F^{-1}\big(U(\omega)\big)\big\} = U^{-1}\big((0,1)\big) \in \mathcal{A}$.

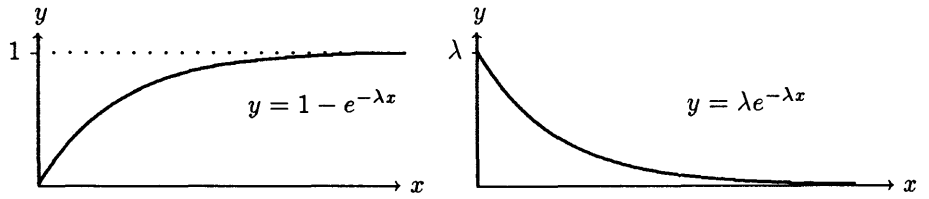
Eine wichtige Klasse von Wahrscheinlichkeitsverteilungen bilden die sogenannten *Exponentialverteilungen* $\{\mathcal{E}(\lambda) \mid \lambda > 0\}$, welche die Verteilungsfunktionen F_λ mit

$$F_\lambda(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \tag{2.1.7}$$

bzw. Dichten f_λ mit

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \tag{2.1.8}$$

besitzen.



Eine Anwendung des Satzes 2.1.1 ergibt hier, daß mit einer auf $[0,1]$ gleichverteilten Zufallsvariablen U die durch

$$X = -\frac{1}{\lambda}\ln(1-U) \quad \text{f.s.} \tag{2.1.9}$$

gegebene Zufallsvariable X $\mathcal{E}(\lambda)$–verteilt ist. Dabei kann die Zufallsvariable $1-U$ in (2.1.9) auch durch die Zufallsvariable U ersetzt werden, da mit U auch $1-U$ gleichverteilt ist wegen

$$P(1-U \le x) = P(U \ge 1-x) = P(U > 1-x) = 1-(1-x) = x$$

für $0 \le x \le 1$.

Exponentialverteilte Zufallsvariablen modellieren häufig Lebensdauern (z.B. von radioaktiven Isotopen) oder Service- und Wartezeiten in Bedienungssystemen (etwa Computer–Netzwerken u.ä.). Man stützt sich dabei auf die für Exponentialverteilungen charakteristische *Gedächtnislosigkeit*, d.h. die Eigenschaft

$$\begin{aligned} P(X > t+h \mid X > t) &= \frac{P(X > t+h)}{P(X > t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P(X > h), \quad t, h > 0, \end{aligned} \tag{2.1.10}$$

wenn X $\mathcal{E}(\lambda)$–verteilt ist. Dies bedeutet in dem ersten Beispiel anschaulich, daß die Wahrscheinlichkeit dafür, daß ein radioaktives Teilchen noch eine Zeitspanne $h > 0$ überlebt, wenn es bis zur Zeit $t > 0$ noch nicht zerfallen ist (ausgedrückt durch das Ereignis $\{X > t\}$), genauso groß wie von Anfang an ist; d.h. der Zerfall geschieht in diesem Sinne zeitlich "rein zufällig". Exponentialverteilte Zufallsvariablen werden also besonders dann zur Modellierung stochastischer Systeme — etwa Netzwerke — eingesetzt, wenn es darum geht, zeitlich rein zufällig eintretende Ereignisse (wie die Beendigung eines im System abgearbeiteten Programms, den Beginn eines neuen Programms usw.) zu beschreiben.

In der Tat legt die Forderung $P(X > t+h \mid X > t) = P(X > h)$, $t, h \ge 0$, bereits die Klasse der Exponentialverteilungen als einzig mögliche Klasse nichttrivialer Verteilungen für P^X fest, d.h. Verteilungen mit $P(X \ne 0) = 1$ bzw. äquivalent $X \ne 0$ f.s. Setzt man nämlich $g(s) = P(X > s)$, $s \ge 0$, so erhält

man zunächst die Funktionalgleichung $\dfrac{g(t+h)}{g(t)} = g(h),\ t,h \geq 0$, mit der Randbedingung $g(0) = P(X > 0) = 1$ und hieraus mit $G = \ln g$:

$$G(t+h) = G(t) + G(h),\ t,h \geq 0,\ G(0) = 0.$$

Die einzig monotonen (oder meßbaren oder stetigen) nicht-konstanten Lösungen dieser Funktionalgleichung sind aber von der Form $G(s) = \lambda s,\ s \geq 0$, mit einem Parameter $\lambda \in \mathbf{R}$, der hier wegen der Monotonie der Verteilungsfunktion positiv sein muß, d.h. es ist

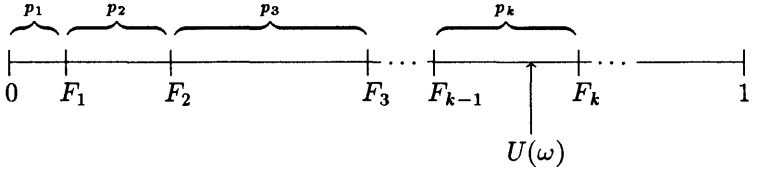
$$P(X \leq x) = 1 - g(x) = 1 - \exp(-G(x)) = 1 - e^{-\lambda x} \text{ für } x \geq 0.$$

Für den Fall, daß F die Verteilungsfunktion einer diskreten Wahrscheinlichkeitsverteilung Q auf $\mathcal{B}^1$ ist, nimmt Beziehung (2.1.6) eine besonders einfache Gestalt an. Wir können dabei ohne Beschränkung der Allgemeinheit annehmen, daß der Träger T der Verteilung gegeben ist durch $T = \{1,2,\ldots,n\},\ n \in \mathbf{N}$, oder $T = \mathbf{N}$ (anderenfalls läßt sich der Träger bijektiv — und wegen der Abzählbarkeit damit auch meßbar — auf eine solche Teilmenge von $\mathbf{N}$ oder $\mathbf{N}$ selbst abbilden). Bezeichnet nun $p_k = Q(\{k\})$ und $F_0 = 0,\ F_k = \sum_{i=1}^{k} p_k = F(k),\ k \in T$, so ist

$$F^{-1}(x) = \min\{k \in T \mid x \in (F_{k-1}, F_k]\},\quad 0 < x < 1. \tag{2.1.11}$$

Wegen der paarweisen Disjunktheit der Intervalle $(F_{k-1}, F_k],\ k \in T$ liegt jedes solche x also in genau einem solchen Intervall $(F_{k-1}, F_k]$. Ist nun wieder U auf $[0,1]$ gleichverteilt, so besagt (2.1.6) gerade

$$X(\omega) = k \quad \text{f.s., falls } U(\omega) \in (F_{k-1}, F_k],\ k \in T.$$



In der Tat läßt sich aus der obigen Skizze unmittelbar ablesen, daß $P(X = k) = P\big(U \in (F_{k-1}, F_k]\big) = p_k,\ k \in T$, gilt, also X die Verteilung $P^X = Q$ besitzt.

Auf dieselbe Weise lassen sich sogar diskrete Verteilungen Q über *beliebigen* Meßräumen $(\mathcal{X}, \mathcal{B})$ erzeugen; besitzt Q nämlich den Träger $T_Q = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\} \subseteq \mathcal{X}$ mit Wahrscheinlichkeiten $p_1 = Q(\{\boldsymbol{x}_1\}),\ p_2 = Q(\{\boldsymbol{x}_2\}), \ldots$, und bezeichnet $\boldsymbol{Y} : T \longrightarrow T_Q : k \mapsto \boldsymbol{x}_k$ mit T wie oben, so besitzt $\boldsymbol{Z} = \boldsymbol{Y} \circ X$ die gewünschte Verteilung.

Beispielsweise läßt sich so die diskrete Gleichverteilung $\mathfrak{L}(\Omega)$ über der Menge $\Omega = \{1,\ldots,n\} \times \{1,\ldots,m\}$ mit $n,m \in \mathbf{N}$ als Verteilung des Zufallsvektors $\boldsymbol{Z}$ aus einer über $[0,1]$ stetig gleichverteilten Zufallsvariablen U erzeugen vermöge

$$\boldsymbol{Z} = (i,j),\quad \text{falls } \frac{(i-1)m + j - 1}{nm} < U \leq \frac{(i-1)m + j}{nm},\quad 1 \leq i \leq n,\ 1 \leq j \leq m. \tag{2.1.12}$$

Wir wollen uns nun allgemeiner mit der Transformation von Zufallsvektoren beschäftigen. Besonders einfach gestaltet sich dies bei diskreten Verteilungen; Lemma 2.1.1 liefert hier unmittelbar

Lemma 2.1.3. *(Transformation von Zufallsvektoren mit diskreten Verteilungen) Es sei $\boldsymbol{X} = (X_1, \ldots, X_m)$ ein diskret verteilter m-dimensionaler Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Zähldichte $f_{\boldsymbol{X}}$ und $G : \mathbf{R}^m \longrightarrow \mathbf{R}^n$ $(m, n \in \mathbf{N})$ eine beliebige meßbare Abbildung. Dann ist $\boldsymbol{Y} = G(\boldsymbol{X})$ ein n-dimensionaler Zufallsvektor, dessen Verteilung gegeben ist durch die Zähldichte $f_{\boldsymbol{Y}}$ mit*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = P(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{\substack{\boldsymbol{x} \in \mathbf{R}^m \\ G(\boldsymbol{x}) = \boldsymbol{y}}} f(\boldsymbol{x}), \quad \boldsymbol{y} \in \mathbf{R}^n. \tag{2.1.13}$$

Beweis. Es ist $P(\boldsymbol{Y} = \boldsymbol{y}) = P\big(G(\boldsymbol{X}) = \boldsymbol{y}\big) = P\big(\boldsymbol{X} \in G^{-1}(\{\boldsymbol{y}\})\big)$, $\boldsymbol{y} \in \mathbf{R}^n$, was mit (2.1.13) zusammenfällt, da nur für höchstens abzählbar viele $\boldsymbol{x} \in \mathbf{R}^m$ die Zähldichte $f(\boldsymbol{x})$ von 0 verschieden ist. ∎

Eine vor allem auch für Anwendungen besonders wichtige Transformation von Zufallsvektoren ist die *Summe* aller Komponenten, die man iterativ als Summe von je zwei Zufallsvariablen darstellen kann. Für diesen Spezialfall (d.h. $m = 2, n = 1$) vereinfacht sich Lemma 2.1.3 zu

Lemma 2.1.4. *(Summe von Zufallsvariablen; Faltungslemma) Es sei $\boldsymbol{X} = (X_1, X_2)$ ein diskret verteilter Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Zähldichte $f_{\boldsymbol{X}}$. Dann ist die Zähldichte der Zufallsvariablen $Y = X_1 + X_2$ gegeben durch*

$$f_Y(y) = \sum_{x \in \mathbf{R}} f_{\boldsymbol{X}}(x, y - x), \quad y \in \mathbf{R}. \tag{2.1.14}$$

Sind speziell die Zufallsvariablen X_1, X_2 stochastisch unabhängig, so ist

$$f_Y(y) = \sum_{x \in \mathbf{R}} f_{X_1}(x) f_{X_2}(y - x), \quad y \in \mathbf{R}, \tag{2.1.15}$$

wobei f_{X_1}, f_{X_2} die Zähldichten von X_1, X_2 bezeichnen.

Man beachte wieder, daß in den Beziehungen (2.1.14) und (2.1.15) höchstens abzählbar viele Summanden von Null verschieden sind, die dort betrachteten Ausdrücke also wohldefiniert sind.

Im Falle der Unabhängigkeit bezeichnet man die Verteilung Q von $X_1 + X_2$ — auch bei nicht-diskreten Verteilungen — als *Faltung* von $P_1 = P^{X_1}$ und $P_2 = P^{X_2}$, in Zeichen:

$$Q = P_1 * P_2. \tag{2.1.16}$$

Sind alle Komponenten $X_1, \ldots, X_m$ des Zufallsvektors $\boldsymbol{X}$ unabhängig, so heißt die Verteilung Q von $\sum_{i=1}^m X_i$ ebenfalls Faltung der Verteilungen $P_i = P^{X_i}$, $1 \le i \le m$, in Zeichen:

$$Q = \mathop{*}_{i=1}^{m} P_i = \mathop{*}_{i=1}^{k} P_i * \mathop{*}_{i=k+1}^{m} P_i, \tag{2.1.17}$$

d.h. die Faltungsoperation $*$ ist *assoziativ* (wegen der Assoziativität der gewöhnlichen Addition reller Zahlen); mit ähnlicher Begründung ist sie darüberhinaus *kommutativ*.

Im Falle von Indikatorvariablen $X_i = \mathbb{1}_{A_i}$, $1 \leq i \leq m$, mit unabhängigen Ereignissen $A_1, \ldots, A_m \in \mathcal{A}$, und gleichen Eintrittswahrscheinlichkeiten $p = P(A_i)$, $1 \leq i \leq m$, ergibt sich über Faltungsbildung die Klasse der *Binomialverteilungen* $\mathfrak{B}(m,p)$, d.h. es ist

$$P\Big(\sum_{i=1}^{m} X_i = k\Big) = \mathfrak{B}(m,p)\big(\{k\}\big) = \binom{m}{k} p^k (1-p)^{m-k}, \quad k = 0,1,\ldots,m. \tag{2.1.18}$$

Dies läßt sich leicht induktiv einsehen:
Für $m = 1$ und $k = 1$ ist gerade

$$P(X_1 = 1) = \mathfrak{B}(1,p)\big(\{1\}\big) = p = \binom{1}{1} p^1 (1-p)^0;$$

analog für $k = 0$. Mit dem Faltungslemma 2.1.4 erhält man dann unter der Voraussetzung, daß (2.1.18) für ein $m \in \mathbb{N}$ richtig ist:

$$\begin{aligned} P\Big(\sum_{i=1}^{m+1} X_i = k\Big) &= \sum_{j=0}^{m} P\Big(\sum_{i=1}^{m} X_i = j\Big) P(X_{m+1} = k-j) \\ &= \sum_{j=k-1}^{k} P\Big(\sum_{i=1}^{m} X_i = j\Big) P(X_{m+1} = k-j) \\ &= \binom{m}{k-1} p^{k-1} (1-p)^{m-k+1} \cdot p + \binom{m}{k} p^k (1-p)^{m-k} \cdot (1-p) \\ &= \left\{ \binom{m}{k-1} + \binom{m}{k} \right\} p^k (1-p)^{m+1-k} \\ &= \binom{m+1}{k} p^k (1-p)^{m+1-k} \end{aligned}$$

für $1 \leq k \leq m$; analog bei $k \in \{0, m+1\}$.

Die Gültigkeit von (2.1.18) läßt sich aber auch durch kombinatorische Überlegungen herleiten:
Es ist ja $\sum_{i=1}^{m} X_i = \sum_{i=1}^{m} \mathbb{1}_{A_i}$, also das Ereignis $\Big\{ \sum_{i=1}^{m} X_i = k \Big\}$, $0 \leq k \leq m$, identisch mit dem Ereignis

$$\bigcup_{1 \leq i_1 < i_2 < \ldots < i_k \leq m} \Big\{ \bigcap_{j=1}^{k} A_{i_j} \cap \bigcap_{l \notin \{i_1,\ldots,i_k\}} A_l^c \Big\}$$

(genau k der Ereignisse $A_1, \ldots, A_m$ treten ein). Da es gerade $\binom{m}{k}$ solcher Auswahlen gibt und wegen der stochastischen Unabhängigkeit

$$P\Big(\bigcap_{j=1}^{k} A_{i_j} \cap \bigcap_{l \notin \{i_1,\ldots,i_k\}} A_l^c \Big) = p^k (1-p)^{m-k}$$

für jede solche Auswahl gilt, folgt ebenfalls (2.1.18).

Binomialverteilungen $\mathfrak{B}(m,p)$ beschreiben also die Verteilung der Anzahl der Erfolge in einer unabhängigen Versuchsserie (Folge unabhängiger Ereignisse) vom Umfang m mit derselben Eintrittswahrscheinlichkeit $p \in [0,1]$.

Beispielsweise ist die Anzahl der Einsen im (8 Bit–) ASCII–Code eines nach einer diskreten Gleichverteilung ausgewählten Zeichens $\mathfrak{B}(8,\frac{1}{2})$–verteilt, da bei der ASCII–Codierung gerade alle $2^8 = 256$ Kombinationen von 0 und 1 (= alle Binärdarstellungen der Zahlen $0,1,\ldots,255$) mit derselben Wahrscheinlichkeit auftreten.

Umgekehrt kann man in einer unendlichen Versuchsserie gleichartiger, unabhängiger Ereignisse $\{A_n\}_{n\in\mathbb{N}}$ mit derselben Eintrittswahrscheinlichkeit $p \in (0,1]$ nach der Verteilung des *Eintrittszeitpunktes* X des ersten Erfolges fragen, d.h. formal

$$X = \inf\{n \in \mathbb{N} \mid \mathbb{1}_{A_n} = 1\}. \tag{2.1.19}$$

X besitzt dann eine sogenannte *geometrische Verteilung* $\mathfrak{G}(p)$, die gegeben ist durch

$$P(X=k) = \mathfrak{G}(p)(\{k\}) = p(1-p)^{k-1}, \quad k \in \mathbb{N}. \tag{2.1.20}$$

Das Ereignis $\{X=k\}$, $k \in \mathbb{N}$ läßt sich nämlich darstellen als

$$\{X=k\} = \bigcap_{i=1}^{k-1} A_i^c \cap A_k, \quad k \in \mathbb{N}, \tag{2.1.21}$$

mit $P\Big(\bigcap_{i=1}^{k-1} A_i^c \cap A_k\Big) = (1-p)^{k-1}p$, $k \in \mathbb{N}$, aufgrund der Unabhängigkeit der Ereignisse $\{A_n\}_{n\in\mathbb{N}}$. Insbesondere folgt aus (2.1.20) $P(X<\infty) = 1$ durch Summation über alle $k \in \mathbb{N}$; d.h. X ist f.s. endlich, was bedeutet, daß in einer unabhängigen Versuchsserie der beschriebenen Art wenigstens ein Erfolg nach endlicher Beobachtungszeit f.s. eintritt.

Strenggenommen muß allerdings für den Fall der Nichtexistenz des Infimums in (2.1.19) das Ereignis $\{X=\infty\} = \bigcap_{n=1}^{\infty} A_n^c = \bigcap_{n=1}^{\infty}\{\mathbb{1}_{A_n} = 0\}$ miterfaßt werden, wenn es auch lediglich mit Wahrscheinlichkeit Null eintritt.

Es sei hier noch einmal daran erinnert, daß zur Modellierung dieses Wartezeitproblems mit $p \in (0,1)$ ein *überabzählbarer* Grundraum Ω benötigt wird, obwohl die Verteilung von X diskret ist (vgl. Satz 1.1.1).

Gelegentlich ist man statt an dem Eintrittszeitpunkt X des ersten Erfolges in der Versuchsserie an der Zahl der vorausgehenden *Mißerfolge*, also an $X-1$, interessiert. Die sich hieraus ergebende Verteilung

$$P(X-1=k) = P(X=k+1) = p(1-p)^k, \quad k \in \mathbb{N}_0, \tag{2.1.22}$$

über den nicht–negativen ganzen Zahlen $\mathbb{N}_0$ heißt ebenfalls geometrische Verteilung (mit Träger $\mathbb{N}_0$).

Zufallsvariable X der Art (2.1.19) heißen auch *Stoppzeiten* bezüglich der Folge $\{X_n\}_{n\in\mathbb{N}}$. Ihre charakteristische Form läßt sich allgemeiner wie folgt fassen.

Definition 2.1.3. *(Stoppzeit)*
Es sei $\{\boldsymbol{X}_n\}$ eine Folge von Zufallselementen in einem Meßraum $(\mathcal{X}, \mathcal{B})$. Eine Zufallsvariable S mit Werten in $\mathbb{N}$ heißt Stoppzeit bezgl. $\{X_n\}$, wenn es zu jedem $n \in \mathbb{N}$ eine Menge $B_n \in \mathcal{B}^{(n)} = \bigotimes_{i=1}^{n} \mathcal{B}$ gibt derart, daß

$$\{S = n\} = \{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in B_n\} \tag{2.1.23}$$

gilt.

Die Bedingung (2.1.23) besagt also in anderer Form, daß das Ereignis $\{S = n\}$ nur von den Zufallselementen $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ (meßbar) abhängt. Insbesondere ist das durch

$$(\boldsymbol{X}_S)(\omega) = \boldsymbol{X}_{S(\omega)}(\omega), \quad \omega \in \Omega, \tag{2.1.24}$$

definierte Zufallselement meßbar wegen

$$\{\boldsymbol{X}_S \in C\} = \bigcup_{n=1}^{\infty} \{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in B_n, \boldsymbol{X}_n \in C\} \quad (C \in \mathcal{B}) \tag{2.1.25}$$

mit den Mengen B_n aus Definition 2.1.3. $\boldsymbol{X}_S$ beschreibt dabei die im Augenblick des Stoppens vorliegende Situation.

Will man die Zufallselemente $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{S-1}$ vor dem Stoppzeitpunkt S bzw. das Zufallselement $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_S)$ in einem stochastischen Modell miterfassen, so muß man beachten, daß $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_S)$ ein Zufallselement der *zufälligen* Länge S ist; als Bildraum für $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_S)$ wählt man dann zweckmäßigerweise den Meßraum $(\mathcal{Y}, \mathcal{C})$ mit $\mathcal{Y} = \bigcup_{m=1}^{\infty} \mathcal{X}^m$ und $\mathcal{C} = \{\bigcup_{m=1}^{\infty} B_m \mid B_m \in \mathcal{B}^{(m)},\ m \in \mathbb{N}\}$ (vgl. auch Aufgabe 1.7).

Stoppzeiten werden uns insbesondere bei der Behandlung von Markoff–Ketten im Zusammenhang mit Modellen zur Average–Case–Analyse gewisser Algorithmen in Kapitel 4 begegnen.

Das folgende Resultat zeigt, daß bei unabhängigen Folgen von Zufallselementen die Unabhängigkeit der Restfolge durch Stoppen erhalten bleibt.

Satz 2.1.2. *(Eigenschaften von Stoppzeiten)*
Es sei unter den Voraussetzungen von Definition 2.1.3 $\{\boldsymbol{X}_n\}$ eine unabhängige Folge mit Verteilung $P^{\boldsymbol{X}_n} = Q$ für alle $n \in \mathbb{N}$. Dann gilt:
a) $\{S = n\}$ und $\{\boldsymbol{X}_{n+1}, \boldsymbol{X}_{n+2}, \ldots\}$ sind unabhängig für alle $n \in \mathbb{N}$
b) $\{\boldsymbol{X}_{S+n}\}_{n \in \mathbb{N}}$ ist unabhängig und identisch verteilt mit Verteilung Q
c) $(S, \boldsymbol{X}_S)$ und $\{\boldsymbol{X}_{S+n}\}_{n \in \mathbb{N}}$ sind unabhängig.
Die letzte Aussage bleibt auch gültig, wenn $(S, \boldsymbol{X}_S)$ durch $(S, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_S)$ ersetzt wird.

Beweis. Für $m, n \in \mathbb{N}$ und $C_0, \ldots, C_m \in \mathcal{B}$ gilt

$$\begin{aligned} P\Big(\{S = n\} \cap \bigcap_{k=1}^{m} \{\boldsymbol{X}_{n+k} \in C_k\}\Big) &= P\big((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n+m}) \in B_n \times \bigtimes_{k=1}^{m} C_k\big) \\ &= P\big((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in B_n\big) P\big((\boldsymbol{X}_{n+1}, \ldots, \boldsymbol{X}_{n+m}) \in \bigtimes_{k=1}^{m} C_k\big) \\ &= P(S = n) \prod_{k=1}^{m} Q(C_k), \end{aligned} \tag{2.1.26}$$

wobei wieder B_n die die Stoppeigenschaft definierende Menge aus (2.1.23) sei. Hieraus folgt a). Ferner ist

$$\begin{aligned} &P\big(\{S=n, \boldsymbol{X}_S \in C_0\} \cap \bigcap_{k=1}^{m} \{\boldsymbol{X}_{S+k} \in C_k\}\big) \\ &\quad = P\big((\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{n+m}) \in (B_n \cap (\mathcal{X}^{n-1} \times C_0)) \times \bigtimes_{k=1}^{m} C_k\big) \\ &\quad = P\big((\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \in B_n \cap (\mathcal{X}^{n-1} \times C_0)\big) \prod_{k=1}^{m} Q(C_k) \\ &\quad = P(S=n, \boldsymbol{X}_S \in C_0) \prod_{k=1}^{m} Q(C_k), \end{aligned} \tag{2.1.27}$$

was c) und b) ergibt (letzteres mit $C_0 = \mathcal{X}$ und Summation über alle $n \in \mathbf{N}$). Die noch fehlende Aussage ergibt sich analog, wenn in (2.1.27) "$\boldsymbol{X}_S \in C_0$" durch "$(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_S) \in C_0^{(n)}$" ersetzt wird mit einer Menge $C_0^{(n)} \in \mathcal{B}^{(n)}$; anstatt $\mathcal{X}^{n-1} \times C_0$ ist entsprechend $C_0^{(n)}$ zu wählen. ∎

Manchmal kann es erforderlich sein, Stoppzeiten S auch mit Wert ∞ zu betrachten (etwa wie im Beispiel der geometrischen Verteilung), d.h.

$$\{S=\infty\} = \{(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots) \in B_\infty\} \tag{2.1.28}$$

mit

$$B_\infty = \Big(\bigcup_{n=1}^{\infty} B_n \times \bigtimes_{k>n} \mathcal{X}\Big)^c. \tag{2.1.29}$$

B_∞ beschreibt das Ereignis, daß im Verlauf der durch die Folge $\{\boldsymbol{X}_n\}_{n\in\mathbf{N}}$ beschriebenen Beobachtungen *keinmal* die gewünschte Stoppsituation eintritt.

Für $S = \infty$ muß Beziehung (2.1.24) entsprechend erweitert werden, indem man formal ein weiteres Zufallselement $\boldsymbol{X}_\infty$ definiert, dessen Wertebereich außerhalb des Wertebereichs der ursprünglichen Folge liegt, z.B. im Fall von *Zufallsvariablen* durch $X_\infty := \infty$.

Zur Behandlung von Zufallsvariablen mit Werten $\pm\infty$ muß man dann allerdings wieder die größere Borel-σ-Algebra

$$\overline{\mathcal{B}^1} = \sigma\left(\mathcal{B}^1 \cup \{\{-\infty\},\{\infty\}\}\right) = \left\{B \cup C \mid B \in \mathcal{B}^1,\ C \subseteq \{-\infty,\infty\}\right\} \tag{2.1.30}$$

über $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty,\infty\}$ zugrundelegen.

Der Fall der geometrischen Verteilung läßt sich nun unter diesen allgemeineren Zugang subsumieren durch die Wahl der Stoppmengen

$$B_n = \begin{cases} \bigtimes_{k=1}^{n-1} \{0\} \times \{1\} & \text{für } n \in \mathbf{N} \\ \bigtimes_{k=1}^{\infty} \{0\} & \text{für } n = \infty \end{cases} \tag{2.1.31}$$

mit den Zufallsvariablen $X_n = \mathbb{1}_{A_n}$, $n \in \mathbb{N}$. Speziell ist dann wieder

$$\begin{aligned}\{S = n\} = \{(X_1, \ldots, X_n) \in B_n\} &= \bigcap_{k=1}^{n-1} \{X_k = 0\} \cap \{X_n = 1\} \\ &= \bigcap_{k=1}^{n-1} A_k^c \cap A_n, \quad n \in \mathbb{N},\end{aligned} \tag{2.1.32}$$

wie in Beziehung (2.1.21), sowie

$$\{S = \infty\} = \{(X_1, \ldots, X_n) \in B_\infty\} = \bigcap_{k=1}^{\infty} \{X_k = 0\} = \bigcap_{k=1}^{\infty} A_k^c. \tag{2.1.33}$$

Ist allgemeiner $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallselementen in $(\mathcal{X}, \mathcal{B})$ und $B \in \mathcal{B}$ ein Ereignis mit $0 < Q(B) < 1$, so ist analog

$$S = \inf\{n \in \mathbb{N} \mid \boldsymbol{X}_n \in B\} \tag{2.1.34}$$

ebenfalls eine Stoppzeit; insbesondere ist S geometrisch verteilt mit

$$P(S = n) = p(1-p)^{n-1}, \; p = Q(B), \; n \in \mathbb{N}, \tag{2.1.35}$$

wenn die Folge unabhängig ist mit derselben Verteilung $Q = P^{\boldsymbol{X}_n}$, $n \in \mathbb{N}$. Dies ergibt sich entweder wie in (2.1.31) bis (2.1.33) mit

$$B_n = \begin{cases} \bigtimes_{k=1}^{n-1} B^c \times B & \text{für } n \in \mathbb{N} \\ \bigtimes_{k=1}^{\infty} B^c & \text{für } n = \infty \end{cases} \tag{2.1.36}$$

oder auf übliche Weise mit der speziellen Wahl $A_n = \{\boldsymbol{X}_n \in B\}$ $(\in \mathcal{A})$, $n \in \mathbb{N}$, und der äquivalenten Darstellung

$$S = \inf\{n \in \mathbb{N} \mid \mathbb{1}_{A_n} = 1\}. \tag{2.1.37}$$

S heißt auch *erste Eintrittszeit* der Folge in die Menge B. Die Folge $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ wird also durch S gestoppt, wenn erstmalig eine Beobachtung in der Menge B auftritt. Die Verteilung von $\boldsymbol{X}_S$, also der Beobachtung im Augenblick des Stoppens, läßt sich dabei folgendermaßen berechnen.

Lemma 2.1.5. *(Eigenschaften der ersten Eintrittszeit)*
Es sei $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ eine unabhängige Folge von Zufallselementen in einem Meßraum $(\mathcal{X}, \mathcal{B})$ mit Verteilung $Q = P^{\boldsymbol{X}_n}$, $n \in \mathbb{N}$, und $B \in \mathcal{B}$ ein Ereignis mit $0 < Q(B) < 1$. Bezeichnet S die erste Eintrittszeit der Folge in die Menge B gemäß (2.1.34), so besitzt die gestoppte Beobachtung $\boldsymbol{X}_S$ die elementare bedingte Verteilung unter (der Hypothese) B; genauer:

$$P(\boldsymbol{X}_S \in A) = \frac{Q(A \cap B)}{Q(B)} = Q(A \mid B), \quad A \in \mathcal{B}. \tag{2.1.38}$$

Ferner sind S und $\boldsymbol{X}_S$ stochastisch unabhängig, und S ist $\mathfrak{G}(p)$-verteilt mit $p = Q(B)$. Die letzte Aussage bleibt gültig, wenn $\boldsymbol{X}_S$ durch $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_S)$ ersetzt wird.

Beweis. Die geometrische Verteilung von S ist nach den obigen Ausführungen klar; insbesondere ist S f.s. endlich, d.h. $\boldsymbol{X}_S$ nimmt f.s. nur Werte im Wertebereich der ursprünglichen Folge an. Durch direktes Nachrechnen erhält man nun

$$\begin{aligned} P(S = n, \boldsymbol{X}_S \in A) &= P\big((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in \bigtimes_{k=1}^{n-1} B^c \times B, \boldsymbol{X}_n \in A\big) \\ &= P\big((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1}) \in \bigtimes_{k=1}^{n-1} B^c, \boldsymbol{X}_n \in B \cap A\big) \\ &= Q^{n-1}(B^c)Q(A \cap B) = (1-p)^{n-1}p \cdot \frac{Q(A \cap B)}{Q(B)} \\ &= \mathfrak{G}(p)\big(\{n\}\big) \cdot Q(A \mid B) \\ &= P(S = n)Q(A \mid B),\ n \in \mathbb{N},\ A \in \mathcal{B}. \end{aligned} \tag{2.1.39}$$

Durch Summation über n folgt

$$\begin{aligned} P(\boldsymbol{X}_S \in A) &= \sum_{n=1}^{\infty} P(S = n, \boldsymbol{X}_S \in A) \\ &= \sum_{n=1}^{\infty} P(S = n)Q(A \mid B) = Q(A \mid B),\ A \in \mathcal{B}, \end{aligned} \tag{2.1.40}$$

also der erste Teil der Behauptung. Ein Vergleich mit (2.1.39) zeigt somit, daß

$$P(S = n, \boldsymbol{X}_S \in A) = P(S = n)P(\boldsymbol{X}_S \in A) \tag{2.1.41}$$

gilt für alle $n \in \mathbb{N}$, $A \in \mathcal{B}$, d.h. S und $\boldsymbol{X}_S$ sind stochastisch unabhängig, wie behauptet. Der Beweis der restlichen Aussage verläuft ähnlich dem Beweis des zweiten Teils der Aussage c) in Satz 2.1.2. ∎

Das Anfangsbeispiel der binären Suche kann ebenfalls als (endliches) Stopp-Problem formuliert werden. Dazu hat man lediglich mit den Bezeichnungen aus (1.1.2) die Mengen $C_k = A_k$, $1 \le k \le n-1$, $C_n = A_n \cup A_0$ und $C_k = \emptyset$, $k \ge n+1$, zu wählen; die Anzahl der Schritte bis zum Abbruch des Algorithmus läßt sich dann durch die Stoppzeit $S = \min\{k \in \mathbb{N} \mid \mathbb{1}_{C_k} = 1\}$ beschreiben mit Stoppmengen analog (2.1.31). Wegen $\Omega = \bigcup_{k=1}^{n} C_k$ und $C_k = \emptyset$ für $k > n$ ist dabei $S \le n$, also S insbesondere endlich. Allerdings sind hier die Ereignisse $\{C_k\}_{k\in\mathbb{N}}$ (und damit auch die Zufallsvariablen $\{\mathbb{1}_{C_k}\}_{k\in\mathbb{N}}$) weder stochastisch unabhängig, wie schon in Abschnitt 1.3 gezeigt wurde, noch besitzen sie dieselbe Eintrittswahrscheinlichkeit. Die Verteilung von S ist daher auch nicht geometrisch, sondern gegeben durch die Beziehung (1.3.6).

Ein Vergleich mit der Beziehung (1.3.5) und den Ausführungen am Ende von Abschnitt 1.4 zeigt, daß es durchaus möglich ist, dieselbe konkrete Situation auf verschiedene Weisen in einem stochastischen Modell zu erfassen. Welches Modell

man wählt, hängt dabei davon ab, welche Art von Aussagen man über die gegebene Situation treffen bzw. welche Aspekte man besonders hervorheben will. Hierauf kommen wir später im Rahmen der Average–Case–Analyse von Algorithmen noch einmal zurück.

Beispiel 2.1.1. (Verwerfungsmethode)
Lemma 2.1.5 ist ein wesentliches Hilfsmittel zur Erzeugung von Verteilungen aus gleichverteilten Zufallsvariablen durch die sogenannte *Verwerfungsmethode*, die in Kapitel 5 ausführlich behandelt wird. Ist beispielsweise $\{(X_{1n}, X_{2n})\}_{n\in\mathbb{N}}$ eine Folge unabhängiger, über dem Quadrat $[-1,1] \times [-1,1]$ gleichverteilter Zufallsvektoren (d.h. insbesondere sind die Komponenten X_{1n} und X_{2n} stochastisch unabhängig und jeweils über $[-1,1]$ gleichverteilt mit Dichte $\frac{1}{2}\mathbb{1}_{[-1,1]}(\cdot)$), und setzt man

$$S = \inf\{n \in \mathbb{N} \mid X_{1n}^2 + X_{2n}^2 \leq 1\}, \tag{2.1.42}$$

so ist S geometrisch verteilt mit

$$p = \frac{\displaystyle\int_{-1}^{1}\int_{-1}^{1} \mathbb{1}_K \, dx_1 \, dx_2}{\displaystyle\int_{-1}^{1}\int_{-1}^{1} dx_1 \, dx_2} = \frac{\pi}{4} \tag{2.1.43}$$

(hierbei bezeichne $K = K^a(0,0;1)$ den Einheitskreis), und $(X_{1,S}, X_{2,S})$ ist über K gleichverteilt. Man verwirft also in der Folge $\{(X_{1n}, X_{2n})\}_{n\in\mathbb{N}}$ alle Beobachtungen, die *nicht* in den Einheitskreis fallen; die erste Beobachtung, die in den Einheitskreis fällt — d.h. $(X_{1,S}, X_{2,S})$ — ist dann dort ebenfalls gleichverteilt. ∎

In der Situation von Lemma 2.1.5 zeigt das Borel–Cantelli–Lemma (Satz 1.1.3), daß nicht nur f.s. wenigstens eines, sondern sogar f.s. *unendlich viele* der Ereignisse $\{\boldsymbol{X}_n \in B\}$ eintreten, da für die Reihe $\sum_{n=1}^{\infty} P(\boldsymbol{X}_n \in B) = \sum_{n=1}^{\infty} p = \infty$ und damit $P\big(\limsup_{n\to\infty}\{\boldsymbol{X}_n \in B\}\big) = 1$ gilt. Es kommen also im Verlauf der Beobachtungsfolge $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ f.s. immer wieder Beobachtungen vor, die in der Menge B liegen. Satz 2.1.2 lehrt dabei, daß sich die einer solchen, in B liegenden Beobachtung anschließende Folge wieder wie die ursprüngliche Folge verhält, also insbesondere stochastisch unabhängig und identisch verteilt ist. Durch wiederholtes Stoppen erhält man so eine Folge von Beobachtungen, die die elementare bedingte Verteilung $Q(\cdot \mid B)$ besitzen, wobei die zugehörigen Stoppzeiten rekursiv definiert sind vermöge

$$\begin{aligned} S_1 &= \inf\{n \in \mathbb{N} \mid \boldsymbol{X}_n \in B\} \\ S_{k+1} &= \inf\{n > S_k \mid \boldsymbol{X}_n \in B\}, \quad k \in \mathbb{N}. \end{aligned} \tag{2.1.44}$$

Wir wollen uns im folgenden kurz mit der Verteilung dieser Stoppzeiten beschäftigen. Dazu zeigen wir zunächst, daß alle S_k, $k \in \mathbb{N}$, tatsächlich die Stoppeigenschaft besitzen. Für $n, k \in \mathbb{N}$ ist nämlich

$$\{S_k = n\} = \Big\{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in \bigcup_{\substack{0\leq i_1,\ldots,i_k\leq n-k \\ i_1+\ldots+i_k=n-k}} (B^c)^{i_1} \times B \times (B^c)^{i_2} \times \cdots \times (B^c)^{i_k} \times B\Big\}; \tag{2.1.45}$$

diese Menge beschreibt also das Ereignis, daß die n–te Beobachtung $\boldsymbol{X}_n$ genau zum k–ten Mal in der Menge B liegt.

Mit Satz 2.1.2 läßt sich nun zeigen, daß die *Wartezeiten*

$$\Delta_k = \begin{cases} S_k - S_{k-1} & \text{für } k > 1 \\ S_1 & \text{für } k = 1 \end{cases} \tag{2.1.46}$$

zwischen dem Auftreten von Beobachtungen in der Menge B stochastisch unabhängig und identisch geometrisch verteilt sind. Hierzu benötigen wir allerdings noch die folgenden Hilfsresultate, die auch sonst von Bedeutung sind.

Lemma 2.1.6. *(Blöcke unabhängiger Zufallselemente)*
Es sei $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallselemente in Meßräumen $(\mathcal{X}_n, \mathcal{B}_n)$. $I_1, I_2 \subseteq \mathbb{N}$, $I_1 \cap I_2 = \emptyset$ seien disjunkte Indexmengen. Dann sind auch

$$\boldsymbol{X}_{I_1} = \{\boldsymbol{X}_i \mid i \in I_1\} \quad \text{und} \quad \boldsymbol{X}_{I_2} = \{\boldsymbol{X}_i \mid i \in I_2\}$$

stochastisch unabhängig.

Beweis. Für Mengen $B_i \in \mathcal{B}_i$, $i \in I_1$, $B_j \in \mathcal{B}_j$, $j \in I_2$ ist

$$\begin{aligned} P\Big(\boldsymbol{X}_{I_1} \in \bigtimes_{i\in I_1} B_i, \boldsymbol{X}_{I_2} \in \bigtimes_{j\in I_2} B_j\Big) &= P\Big(\bigcap_{i\in I_1}\{\boldsymbol{X}_i \in B_i\} \cap \bigcap_{j\in I_2}\{\boldsymbol{X}_j \in B_j\}\Big) \\ &= \prod_{i\in I_1} P(\boldsymbol{X}_i \in B_i) \prod_{j\in I_2} P(\boldsymbol{X}_j \in B_j) = P\Big(\boldsymbol{X}_{I_1} \in \bigtimes_{i\in I_1} B_i\Big) P\Big(\boldsymbol{X}_{I_2} \in \bigtimes_{j\in I_2} B_j\Big). \end{aligned}$$

Hieraus folgt die Behauptung. ∎

Lemma 2.1.6 läßt sich induktiv sofort auf den Fall endlich und sogar abzählbar–unendlich vieler paarweise disjunkter Indexmengen $I_1, \ldots, I_m \subseteq \mathbb{N}$, $m \in \mathbb{N}$, erweitern, d.h. allgemeiner sind in der Situation von Lemma 2.1.6 die Zufallselemente $\boldsymbol{X}_{I_1}, \ldots, \boldsymbol{X}_{I_m}$ stochastisch unabhängig. Die stochastische Unabhängigkeit von Zufallselementen bleibt also erhalten, wenn diese über paarweise disjunkte Indexbereiche gruppiert werden. Analog zeigt man

Lemma 2.1.7. *(Blocktransformation unabhängiger Zufallselemente)*
Es seien $\boldsymbol{X}, \boldsymbol{Y}$ Zufallselemente in einem Meßraum $(\mathcal{X}, \mathcal{B})$ und $g : (\mathcal{X}, \mathcal{B}) \longrightarrow (\mathcal{Y}, \mathcal{C})$, $h : (\mathcal{X}, \mathcal{B}) \longrightarrow (\mathcal{Z}, \mathcal{D})$ meßbare Abbildungen in weitere Meßräume $(\mathcal{Y}, \mathcal{C})$, $(\mathcal{Z}, \mathcal{D})$. Dann ist $(g \circ \boldsymbol{X}, h \circ \boldsymbol{Y})$ ein Zufallselement in dem Meßraum $(\mathcal{Y} \times \mathcal{Z}, \mathcal{C} \otimes \mathcal{D})$, und die Zufallselemente $g \circ \boldsymbol{X}$ und $h \circ \boldsymbol{Y}$ sind stochastisch unabhängig.

Beweis. Für $C \in \mathcal{C}$, $D \in \mathcal{D}$ ist

$$\begin{aligned} P(g \circ \boldsymbol{X} \in C, h \circ \boldsymbol{Y} \in D) &= P(\boldsymbol{X} \in g^{-1}(C), \boldsymbol{Y} \in h^{-1}(D)) \\ &= P(\boldsymbol{X} \in g^{-1}(C)) P(\boldsymbol{Y} \in h^{-1}(D)) \\ &= P(g \circ \boldsymbol{X} \in C) P(h \circ \boldsymbol{Y} \in D). \end{aligned}$$

Hieraus folgt die Behauptung. ∎

Auch dieses Lemma läßt sich sofort auf den Fall endlich bzw. abzählbar–unendlich vieler Koordinatentransformationen verallgemeinern.

Lemma 2.1.6 und Lemma 2.1.7 besagen also zusammen anschaulich, daß bei einer stochastisch unabhängigen Folge von Zufallselementen (meßbare) Funktionen von disjunkten Teilabschnitten dieser Folge stochastisch unabhängig bleiben. Dieser Sachverhalt spielt bei sehr vielen stochastischen Modellbildungen eine zentrale Rolle, da man häufig komplizierte Strukturen durch solche Art von Transformationen auf Strukturen mit unabhängigen Folgen zurückführen kann. Beispielsweise lassen sich die in Kapitel 3 behandelten Markoff-Ketten durch entsprechende Überlegungen kanonisch mit unabhängigen Folgen beschreiben.

Das folgende Lemma behandelt schließlich noch ein einfaches Kriterium für die stochastische Unabhängigkeit von Folgen von Zufallselementen.

Lemma 2.1.8. *(Stochastische Unabhängigkeit bei Zufallsfolgen)*
Es sei $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallselementen in Meßräumen $(\mathcal{X}_n, \mathcal{B}_n)$. Die Folge ist genau dann stochastisch unabhängig, wenn für alle $n \in \mathbb{N}$ die Zufallselemente $\boldsymbol{X}_{n+1}$ und $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ stochastisch unabhängig sind.

Beweis. Wir zeigen induktiv, daß die gemeinsame Verteilung von $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, $n \in \mathbb{N}$ durch das Produktmaß $\bigotimes_{i=1}^{n} P^{\boldsymbol{X}_i}$ (im Sinne von Satz 1.4.3) gegeben ist, wenn die im Lemma angegebene Bedingung erfüllt ist.

Für $n = 1$ ist nichts zu zeigen. Unter der Annahme, daß die Gültigkeit der letzteren Aussage für $n \in \mathbb{N}$ nachgewiesen ist, folgt

$$P^{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{n+1})} = P^{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)} \otimes P^{\boldsymbol{X}_{n+1}} = \bigotimes_{i=1}^{n} P^{\boldsymbol{X}_i} \otimes P^{\boldsymbol{X}_{n+1}} = \bigotimes_{i=1}^{n+1} P^{\boldsymbol{X}_i}$$

und damit die stochastische Unabhängigkeit der Folge.

Die umgekehrte Aussage folgt aus Lemma 2.1.6. ∎

Nunmehr können wir das angekündigte Resultat bezüglich der in (2.1.44) eingeführten Stoppzeiten beweisen.

Satz 2.1.3. *(Unabhängigkeit bei ersten Eintrittszeiten)*
Es sei $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ eine unabhängige Folge von Zufallselementen in einem Meßraum $(\mathcal{X}, \mathcal{B})$ mit Verteilung $Q = P^{\boldsymbol{X}_n}$, $n \in \mathbb{N}$, und $B \in \mathcal{B}$ ein Ereignis mit $0 < Q(B) < 1$. Bezeichnet $\{S_k\}_{k\in\mathbb{N}}$ die in (2.1.44) rekursiv definierte Folge von Eintrittszeiten in die Menge B, so gilt:

Die in (2.1.46) definierte Folge von Wartezeiten $\{\Delta_k\}_{k\in\mathbb{N}}$ ist stochastisch unabhängig und identisch verteilt mit

$$P^{\Delta_k} = \mathfrak{G}(p),\ p = Q(B);\ k \in \mathbb{N}. \tag{2.1.47}$$

Beweis. Nach Lemma 2.1.8 und der allgemeineren Fassung der Lemmata 2.1.6 und 2.1.7 reicht es zu zeigen, daß für alle $k \in \mathbb{N}$ die Zufallsvektoren $(S_1, \ldots, S_k)$ und Δ_{k+1} stochastisch unabhängig sind, da $S_j = \sum_{i=1}^{j} \Delta_i$, $1 \le j \le k$, gilt. Nun ist aber

$$\Delta_{k+1} = \inf\{n \in \mathbb{N} \mid \boldsymbol{X}_{S_k+n} \in B\} \tag{2.1.48}$$

eine meßbare Funktion der Folge $\{\boldsymbol{X}_{S_k+n}\}_{n\in\mathbb{N}}$, welche nach Satz 2.1.2 von dem Zufallselement $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{S_k})$ stochastisch unabhängig ist. Die Stoppzeiten $S_1,\dots,S_j$, $1\le j\le k$, sind wegen $S_j\le S_k$, $1\le j\le k$, ihrerseits meßbare Funktionen von $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{S_k})$, so daß insgesamt Δ_{k+1} von $(S_1,\dots,S_k)$ und damit auch von $(\Delta_1,\dots,\Delta_k)$ stochastisch unabhängig ist.
Da jede der Folgen $\{\boldsymbol{X}_{S_j+n}\}_{n\in\mathbb{N}}$, $j\in\mathbb{N}$ nach Teil b) des Satzes 2.1.2 unabhängig und identisch wie die ursprüngliche Folge verteilt ist, besitzt nach (2.1.48) auch jedes Δ_{k+1}, $k\in\mathbb{N}$, dieselbe Verteilung wie $\Delta_1=S_1$, die nach Lemma 2.1.5 durch $\mathfrak{G}(p)$ mit $p=Q(B)$ gegeben ist. ∎

Die Unabhängigkeit und geometrische Verteilung der Folge $\{\Delta_k\}_{k\in\mathbb{N}}$ kann auch direkt nachgewiesen werden, da für $m\in\mathbb{N}$, $n_1,\dots,n_m\in\mathbb{N}$, gilt:

$$\bigcap_{k=1}^{m}\{\Delta_k=n_k\}=\bigcap_{k=1}^{m}\{\boldsymbol{X}_{n_k}\in B\}\cap\bigcap_{\substack{1\le i\le\sum_{j=1}^{m}n_j\\ i\notin\{n_1,n_1+n_2,\dots,\sum_{j=1}^{m}n_j\}}}\{\boldsymbol{X}_i\notin B\},$$

d.h. wegen der Unabhängigkeit der Folge $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ ist

$$\begin{aligned}P\Big(\bigcap_{k=1}^{m}\{\Delta_k=n_k\}\Big)&=\prod_{k=1}^{m}P(\boldsymbol{X}_{n_k}\in B)\prod_{\substack{1\le i\le\sum_{j=1}^{m}n_j\\ i\notin\{n_1,n_1+n_2,\dots,\sum_{j=1}^{m}n_j\}}}P(\boldsymbol{X}_i\notin B)\\&=\big(Q(B)\big)^m\big(1-Q(B)\big)^{\sum_{j=1}^{m}(n_j-1)}\\&=\prod_{k=1}^{m}\big(p(1-p)^{n_k-1}\big),\ p=Q(B).\end{aligned}\tag{2.1.49}$$

Hieraus folgt $P^{(\Delta_1,\dots,\Delta_m)}=\bigotimes_{k=1}^{m}P^{\Delta_k}$ für alle $m\in\mathbb{N}$ und somit die stochastische Unabhängigkeit der Folge $\{\Delta_k\}_{k\in\mathbb{N}}$. Die geometrische Verteilung der Folge ergibt sich aus der letzten Gleichheit in (2.1.49).

Satz 2.1.3 besagt insbesondere, daß sich die Stoppzeiten S_m, $m\in\mathbb{N}$, f.s. als Summe von m unabhängigen, jeweils $\mathfrak{G}(p)$–verteilten Zufallsvariablen darstellen lassen, so daß

$$P^{S_m}=\mathop{*}_{i=1}^{m}\mathfrak{G}(p),\ p=Q(B),\ m\in\mathbb{N},\tag{2.1.50}$$

gilt. Diese Verteilung heißt *negative Binomialverteilung* mit Parametern m und p, in Zeichen: $\overline{\mathfrak{B}}(m,p)$ und ist in allgemeiner Form gegeben durch

$$\overline{\mathfrak{B}}(m,p)(\{k\})=\binom{k-1}{m-1}p^m(1-p)^{k-m},\ k\in\mathbb{N},\ k\ge m,\tag{2.1.51}$$

mit $p\in(0,1]$.

Dies läßt sich ähnlich wie im Fall der Binomialverteilungen wieder leicht induktiv nachweisen. Dazu sei $\{X_n\}_{n\in\mathbb{N}}$ eine unabhängige Folge $\mathfrak{G}(p)$–verteilter Zufallsvariablen mit $p \in (0,1]$. Wir zeigen:

$$P\Big(\sum_{i=1}^{m} X_i = k\Big) = \overline{\mathfrak{B}}(m,p)(\{k\}) = \binom{k-1}{m-1} p^m (1-p)^{k-m},\ k \geq m. \qquad (2.1.52)$$

Für $m = 1$ ist nichts zu zeigen, da dann (2.1.52) mit (2.1.20) zusammenfällt. Mit dem Faltungslemma 2.1.4 erhält man unter der Voraussetzung, daß (2.1.52) für ein $m \in \mathbb{N}$ richtig ist:

$$\begin{aligned} P\Big(\sum_{i=1}^{m+1} X_i = k\Big) &= \sum_{j=m}^{k} P\Big(\sum_{i=1}^{m} X_i = j\Big) P(X_{m+1} = k-j) \\ &= \sum_{j=m}^{k} \binom{j-1}{m-1} p^m (1-p)^{j-m} \cdot p(1-p)^{k-j-1} \\ &= p^{m+1}(1-p)^{k-(m+1)} \sum_{j=m}^{k} \binom{j-1}{m-1} \\ &= \binom{k-1}{m} p^{m+1}(1-p)^{k-(m+1)},\ k \geq m+1. \end{aligned}$$

Auch hier läßt sich die Beziehung (2.1.52) wieder kombinatorisch deuten: Nimmt man wie im Fall der Binomialverteilungen an, daß $A_1, \ldots, A_m \in \mathcal{A}$, $m \in \mathbb{N}$, unabhängige Ereignisse mit derselben Eintrittswahrscheinlichkeit p sind und X_i die Wartezeit zwischen dem Eintreten des $(i-1)$–ten und i–ten Ereignisses bezeichnet, so ist das Ereignis $\left\{\sum_{i=1}^{m} X_i = k\right\}$, $k \geq m$, identisch mit dem Ereignis

$$\bigcup_{1\leq i_1<i_2<\ldots<i_{m-1}<i_m=k} \Big\{\bigcap_{j=1}^{m} A_{i_j} \cap \bigcap_{l\notin\{i_1,\ldots,i_m\}} A_l^c\Big\}$$

(das m–te Ereignis tritt genau im k–ten Versuch ein). Da es gerade $\binom{k-1}{m-1}$ solcher Auswahlen gibt und wegen der stochastischen Unabhängigkeit

$$P\Big(\bigcap_{j=1}^{m} A_{i_j} \cap \bigcap_{l\notin\{i_1,\ldots,i_m\}} A_l^c\Big) = p^m(1-p)^{k-m}$$

für jede solche Auswahl gilt, folgt ebenfalls (2.1.52).

Zählt man wieder nur die dem m–ten Erfolg im k–ten Versuch vorausgehenden *Mißerfolge*, d.h. betrachtet man statt der Zufallsvariablen $\sum_{i=1}^{m} X_i$ die Zufallsva-

riable $\sum_{i=1}^m (X_i - 1) = \sum_{i=1}^m X_i - m$, so heißt die hieraus resultierende Verteilung

$$\begin{aligned} P\Big(\sum_{i=1}^m X_i - m = k\Big) &= P\Big(\sum_{i=1}^m X_i = k + m\Big) \\ &= \binom{k+m-1}{m-1} p^m (1-p)^k \\ &= \binom{k+m-1}{k} p^m (1-p)^k,\ m \in \mathbb{N},\ k \in \mathbb{N}_0, \end{aligned} \tag{2.1.53}$$

ebenfalls negative Binomialverteilung (über $\mathbb{N}_0$), i.Z.: $\overline{\mathfrak{B}}^+(m,p)$; analog $\mathfrak{G}^+(p) = \overline{\mathfrak{B}}^+(1,p)$.

Der Name *negative* Binomialverteilung gründet sich auf der alternativen Darstellung

$$\binom{k+m-1}{k} p^m (1-p)^k = \binom{-m}{k} p^m (p-1)^k,\ m \in \mathbb{N},\ k \in \mathbb{N}_0,$$

wenn man formal

$$\begin{aligned} \binom{-m}{k} &= \frac{(-m)_k}{k!} = \frac{(-m-k+1)(-m-k+2)\cdots(-m)}{k!} \\ &= (-1)^k \frac{(m+k-1)_k}{k!} = (-1)^k \binom{k+m-1}{k},\ m \in \mathbb{N},\ k \in \mathbb{N}_0, \end{aligned}$$

setzt.

Für kleine Parameter $p \ll 1$ bzw. $1-p \ll 1$ lassen sich die Binomial– bzw. negative Binomialverteilung $\mathfrak{B}(m,p)$ bzw. $\overline{\mathfrak{B}}^+(m,p)$ durch die sogenannte *Poisson–Verteilung* $\mathfrak{P}(\lambda)$ approximieren, die erklärt ist durch

$$\mathfrak{P}(\lambda)(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!},\ k \in \mathbb{N}_0;\ \lambda > 0. \tag{2.1.54}$$

Genauer gilt:

Satz 2.1.4. *(Poisson (1837))*
Es sei $\{p_n\}_{n\in\mathbb{N}} \subseteq (0,1)$ *eine Folge reeller Zahlen mit* $\lim_{n\to\infty} np_n = \lambda > 0$ *bzw.* $\lim_{n\to\infty} n(1-p_n) = \mu > 0$. *Dann gilt*

$$\lim_{n\to\infty} \mathfrak{B}(n,p_n)(\{k\}) = \mathfrak{P}(\lambda)(\{k\})$$

bzw.

$$\lim_{n\to\infty} \overline{\mathfrak{B}}^+(n,p_n)(\{k\}) = \mathfrak{P}(\mu)(\{k\}),\ k \in \mathbb{N}_0.$$

Beweis. Es ist für $0 \le k \le n$

$$\begin{aligned} \mathfrak{B}(n,p_n)(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n!}{(n-k)!\big(n(1-p_n)\big)^k} \Big(1 - \frac{\lambda}{n} \cdot \frac{np_n}{\lambda}\Big)^n \frac{(np_n)^k}{k!}. \end{aligned} \tag{2.1.55}$$

Nach Voraussetzung strebt aber $\left(1-\frac{\lambda}{n}\cdot\frac{np_n}{\lambda}\right)^n$ gegen $e^{-\lambda}$, $(np_n)^k$ gegen λ^k, sowie der dritte Faktor in (2.1.55) gegen 1. Damit ergibt sich die erste Behauptung. Im zweiten Fall gilt für $n\in\mathbb{N}$, $k\in\mathbb{N}_0$,

$$\begin{aligned}\overline{\mathfrak{B}}^+(n,p_n)(\{k\}) &= \binom{n+k-1}{k}p_n^n(1-p_n)^k \\ &= \frac{(n+k-1)!}{(n-1)!n^k}\left(1+\frac{n(1-p_n)}{np_n}\right)^{-n}\frac{\left(n(1-p_n)\right)^k}{k!}.\end{aligned} \tag{2.1.56}$$

Nach Voraussetzung strebt hier $\left(1+\frac{n(1-p_n)}{np_n}\right)^{-n}$ gegen $e^{-\mu}$, $\left(n(1-p_n)\right)^k$ gegen μ^k sowie der dritte Faktor in (2.1.56) wieder gegen 1. Damit ist der Satz bewiesen. ∎

Beispiel 1.1.4 zeigt, daß Poisson–Approximationen auch in allgemeineren Situationen möglich sind: bezeichnet nämlich dort die Zufallsvariable X die Anzahl der bereits richtig sortierten Feldelemente, so besitzt X für große n aufgrund von (1.1.50) approximativ eine $\mathfrak{P}(1)$–Verteilung. X ist aber *nicht* exakt binomialverteilt; allerdings ist dies der Fall, wenn man Wiederholungen von Feldelementen zuläßt (vgl. die Bemerkung im Anschluß an Beispiel 1.1.4), da dann mit den dortigen Bezeichnungen $X=\sum_{i=1}^n \mathbb{1}_{A_i}$ mit $P(A_i)=\frac{1}{n}$, $1\le i\le n$, gilt, was das vorherige Ergebnis zumindest anschaulich erklärt.

Eine wesentliche Eigenschaft der Poisson–Verteilung ist ihre Faltungsstabilität, die im folgenden beschrieben wird.

Lemma 2.1.9. *(Faltungsstabilität der Poisson–Verteilung)*
Für Poisson–Verteilungen $\mathfrak{P}(\lambda)$ und $\mathfrak{P}(\mu)$ mit $\lambda,\mu>0$ gilt

$$\mathfrak{P}(\lambda)*\mathfrak{P}(\mu)=\mathfrak{P}(\lambda+\mu), \tag{2.1.57}$$

d.h. sind X und Y stochastisch unabhängige Zufallsvariable mit $P^X=\mathfrak{P}(\lambda)$, $P^Y=\mathfrak{P}(\mu)$, so ist auch die Summe $X+Y$ Poisson–verteilt mit $P^{X+Y}=\mathfrak{P}(\lambda+\mu)$.

Beweis. Es ist für $k\in\mathbb{N}_0$

$$\begin{aligned}P(X+Y=k) &= \sum_{i=1}^k P(X=i)P(Y=k-i) = e^{-(\lambda+\mu)}\sum_{i=1}^k\frac{\lambda^i\mu^{k-i}}{i!(k-i)!} \\ &= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{i=1}^k\binom{k}{i}\lambda^i\mu^{k-i} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}.\end{aligned}$$

Hieraus folgt die Behauptung. ∎

Will man das obige Resultat in der Praxis anwenden, stellt sich häufig die Frage nach der Güte der Approximation in Form einer Abschätzung des Approximationsfehlers. Hierfür wollen wir das folgende Abstandsmaß zwischen Wahrscheinlichkeitsverteilungen betrachten.

Definition 2.1.4. *(Kolmogoroff–Abstand)*
Es bezeichne $\mathcal{P}$ die Menge aller Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^1$. F_P bezeichne die zu der Wahrscheinlichkeitsverteilung $P \in \mathcal{P}$ gehörige Verteilungsfunktion. Dann heißt die durch

$$\rho(P,Q) = \sup_{x \in \mathbf{R}} \{|F_P(x) - F_Q(x)|\}, \; P,Q \in \mathcal{P}, \tag{2.1.58}$$

auf $\mathcal{P} \times \mathcal{P}$ erklärte Metrik ρ Kolmogoroff–Abstand zwischen den Verteilungen P und Q.

ρ ist in der Tat eine Metrik auf $\mathcal{P} \times \mathcal{P}$, denn es gelten die drei für Metriken charakteristischen Eigenschaften

(1) $\rho(P,Q) = 0 \iff P = Q, \; P,Q \in \mathcal{P}$
(2) $\rho(P,Q) = \rho(Q,P), \; P,Q \in \mathcal{P}$ *(Symmetrie)*
(3) $\rho(P,W) \leq \rho(P,Q) + \rho(Q,W), \; P,Q,W \in \mathcal{P}$ *(Dreiecks – Ungleichung).*

Beziehung (1) ist dabei eine Konsequenz des Eindeutigkeitssatzes 1.2.3; Beziehung (3) ergibt sich aus der entsprechenden Dreiecks–Ungleichung für reelle Zahlen.

Eine für den Fall der Poisson–Approximation wichtige Eigenschaft von ρ ist die sogenannte *Subadditivität*, die im folgenden beschrieben wird.

Satz 2.1.5. *(Subadditivität des Kolmogoroff–Abstands)*
Es seien $P_1, \ldots, P_n$, $Q_1, \ldots, Q_n$, $n \in \mathbf{N}$, Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^1$. Dann gilt für den Kolmogoroff–Abstand ρ:

$$\rho\Big(\mathop{*}_{i=1}^{n} P_i, \mathop{*}_{i=1}^{n} Q_i \Big) \leq \sum_{i=1}^{n} \rho(P_i, Q_i). \tag{2.1.59}$$

Beweis. Wir beweisen den Satz hier zunächst nur für den Fall diskreter Verteilungen mit Trägern, die Teilmengen von $\mathbf{N}_0$ sind, was für das betrachtete Approximationsproblem ausreicht.
Seien dazu X, Y, Z Zufallsvariable mit Werten in $\mathbf{N}_0$. Wir zeigen als ersten Schritt:

$$\rho\big(P^{X+Z}, P^{Y+Z}\big) \leq \rho\big(P^X, P^Y\big). \tag{2.1.60}$$

Für beliebige $m \in \mathbf{N}$ gilt nämlich

$$\begin{aligned}
&\Big| \sum_{k=0}^{m} P(X+Z=k) - P(Y+Z=k) \Big| \\
&\quad = \Big| \sum_{k=0}^{m} \sum_{i=0}^{k} \Big(P(X=k-i) - P(Y=k-i) \Big) P(Z=i) \Big| \\
&\quad = \Big| \sum_{0 \leq i \leq k \leq m} \sum \Big(P(X=k-i) - P(Y=k-i) \Big) P(Z=i) \Big| \\
&\quad \leq \sum_{i=0}^{m} \Big| \sum_{j=0}^{m-i} P(X=j) - P(Y=j) \Big| P(Z=i) \\
&\quad \leq \sum_{i=0}^{\infty} \rho\big(P^Y, P^Y\big) P(Z=i) = \rho\big(P^X, P^Y\big).
\end{aligned}$$

Hieraus folgt aber (2.1.60), da sich die zugehörigen Verteilungsfunktionen höchstens an den ganzzahligen Stellen $m \in \mathbb{N}_0$ ändern.
Sind nun $X_1, \ldots, X_n$ bzw. $Y_1, \ldots, Y_n$, $n \in \mathbb{N}$, jeweils unabhängige Zufallsvariable mit Verteilungen $P^{X_k} = P_k$, $P^{Y_k} = Q_k$, $1 \leq k \leq n$, so läßt sich unter Heranziehung von (2.1.60) induktiv zeigen:

$$\rho\Big(P^{\sum_{k=1}^n X_k}, P^{\sum_{k=1}^n Y_k}\Big) \leq \sum_{k=1}^n \rho\Big(P^{X_k}, P^{Y_k}\Big). \tag{2.1.61}$$

Für $n = 1$ ist nichts zu zeigen.
Unter der Annahme, daß (2.1.61) für ein $n \in \mathbb{N}$ gültig ist, folgt mit der Eigenschaft (3) von ρ und (2.1.60)

$$\begin{aligned}
\rho\Big(P^{\sum_{k=1}^{n+1} X_k}, P^{\sum_{k=1}^{n+1} Y_k}\Big) &\leq \rho\Big(P^{\sum_{k=1}^{n+1} X_k}, P^{\sum_{k=1}^n Y_k + X_{n+1}}\Big) \\
&\quad + \rho\Big(P^{\sum_{k=1}^n Y_k + X_{n+1}}, P^{\sum_{k=1}^n Y_k + Y_{n+1}}\Big) \\
&\leq \rho\Big(P^{\sum_{k=1}^n X_k}, P^{\sum_{k=1}^n Y_k}\Big) + \rho\Big(P^{X_{n+1}}, P^{Y_{n+1}}\Big) \\
&\leq \sum_{k=1}^n \rho\Big(P^{X_k}, P^{Y_k}\Big) + \rho\Big(P^{X_{n+1}}, P^{Y_{n+1}}\Big) = \sum_{k=1}^{n+1} \rho\Big(P^{X_k}, P^{Y_k}\Big),
\end{aligned}$$

d.h. (2.1.61) gilt auch für $n + 1$.
Damit ist aber die Aussage des Satzes bewiesen. ∎

In Bezug auf Satz 2.1.4 ist es auch sinnvoll, für Verteilungen $P, Q \in \mathcal{P}$ mit Trägern in $\mathbb{N}_0$ den Abstand

$$\rho_0(P, Q) = \sup_{k \in \mathbb{N}_0} \big|P(\{k\}) - Q(\{k\})\big| \tag{2.1.62}$$

zu betrachten. Wegen $P(\{k\}) = F_P(k) - F_P(k-1)$, $k \in \mathbb{N}_0$, $P \in \mathcal{P}$, gilt jedenfalls

$$\rho_0(P, Q) \leq 2\rho(P, Q), \; P, Q \in \mathcal{P}. \tag{2.1.63}$$

Die Subadditivität des Kolmogoroff-Abstandes ρ (vgl. Satz 2.1.5) überträgt sich völlig analog auch auf ρ_0, da die für den Beweis von Satz 2.1.5 wesentliche Beziehung (2.1.60) entsprechend gültig bleibt: es ist ja unter den dortigen Voraussetzungen für $k \in \mathbb{N}_0$

$$\begin{aligned}
\rho_0\big(P^{X+Z}, P^{Y+Z}\big) &= \sup_{k \in \mathbb{N}_0} \big|P(X + Z = k) - P(Y + Z = k)\big| \\
&\leq \sup_{k \in \mathbb{N}_0} \sum_{i=0}^k \Big|P(X = k - i) - P(Y = k - i)\Big| P(Z = i) \\
&\leq \sum_{i=0}^\infty \rho_0\big(P^X, P^Y\big) P(Z = i) = \rho_0\big(P^X, P^Y\big).
\end{aligned} \tag{2.1.64}$$

Das folgende Lemma gestattet nun Abschätzungen bezüglich der Konvergenz der Binomial- gegen die Poissonverteilung in Satz 2.1.4, was für praktische Anwendungen wichtig ist.

Lemma 2.1.10. *Es seien $p \in (0,1)$, $\lambda, \mu > 0$. Dann gilt:*

$$\rho\big(\mathfrak{B}(p), \mathfrak{P}(\lambda)\big) \le \frac{1}{2}\lambda^2 + |p-\lambda|, \quad \rho_0\big(\mathfrak{B}(p), \mathfrak{P}(\lambda)\big) \le \lambda^2 + |p-\lambda| \tag{2.1.65}$$

$$\rho\big(\mathfrak{P}(\lambda), \mathfrak{P}(\mu)\big) \le |\lambda-\mu|, \quad \rho_0\big(\mathfrak{P}(\lambda), \mathfrak{P}(\mu)\big) \le |\lambda-\mu|. \tag{2.1.66}$$

Beweis. Es ist

$$\begin{aligned}\rho\big(\mathfrak{B}(p), \mathfrak{P}(\lambda)\big) &= \max\Big\{\big|1-p-e^{-\lambda}\big|,\ \big|1-(1+\lambda)e^{-\lambda}\big|,\ \sup_{k\ge 2}\Big|1-\sum_{i=0}^{k} e^{-\lambda}\frac{\lambda^i}{i!}\Big|\Big\} \\ &\le \max\big\{\big|1-\lambda-e^{-\lambda}\big| + |p-\lambda|,\ \big|1-(1+\lambda)e^{-\lambda}\big|\big\} \le \frac{1}{2}\lambda^2 + |p-\lambda|\end{aligned}$$

und

$$\begin{aligned}\rho_0\big(\mathfrak{B}(p), \mathfrak{P}(\lambda)\big) &= \max\Big\{\big|1-p-e^{-\lambda}\big|,\ \big|p-\lambda e^{-\lambda}\big|,\ \sup_{k\ge 2}\big|e^{-\lambda}\frac{\lambda^k}{k!}\big|\Big\} \\ &\le \max\Big\{\big|1-\lambda-e^{-\lambda}\big| + |p-\lambda|,\ \big|\lambda-\lambda e^{-\lambda}\big| + |p-\lambda|,\ \sup_{k\ge 2}\big|e^{-\lambda}\frac{\lambda^k}{k!}\big|\Big\} \\ &\le \lambda^2 + |p-\lambda|.\end{aligned}$$

Damit ist (2.1.65) bewiesen. Zum Beweis von (2.1.66) können wir aus Symmetriegründen $\lambda > \mu$ annehmen. Aufgrund von (2.1.60) und Lemma 2.1.9 reicht es zu zeigen, daß für Zufallsvariable $X \equiv 0$ und Y mit $P^Y = \mathfrak{P}(\lambda-\mu)$ gilt:

$$\rho\big(P^X, P^Y\big) \le |\lambda-\mu|,\ \rho_0\big(P^X, P^Y\big) \le |\lambda-\mu|.$$

Setzen wir zur Abkürzung $\delta = \lambda - \mu > 0$, so ergibt sich

$$\begin{aligned}\rho\big(P^X, P^Y\big) &= \sup_{k\in\mathbb{N}_0}\big|P(X\le k) - P(Y\le k)\big| \\ &= \sup_{k\in\mathbb{N}_0}\big|1-P(Y\le k)\big| = \sup_{k\in\mathbb{N}_0} P(Y>k) \\ &= P(Y>0) = 1-e^{-\delta} \le \delta = |\lambda-\mu|.\end{aligned}$$

Ist nun Z eine Zufallsvariable mit $P^Z = \mathfrak{P}(\mu)$, unabhängig von X und Y, so folgt

$$\rho\big(\mathfrak{P}(\lambda), \mathfrak{P}(\mu)\big) = \rho\big(P^{X+Z}, P^{Y+Z}\big) \le \rho\big(P^X, P^Y\big) \le |\lambda-\mu|.$$

Entsprechend erhält man

$$\rho_0\big(P^X, P^Y\big) = \max\big\{|1-P(Y=0)|,\ \sup_{k\in\mathbb{N}} P(Y=k)\big\} \le \max\big\{\delta,\ \sup_{k\in\mathbb{N}} e^{-\delta}\frac{\delta^k}{k!}\big\}.$$

Nun nimmt aber für $k \in \mathbb{N}$ die Abbildung $\delta \mapsto e^{-\delta}\frac{\delta^{k-1}}{k!}$ ihr Maximum bei $\delta = k-1$ mit Wert $e^{-(k-1)}\frac{(k-1)^{k-1}}{k!}$ an; die letztere Folge ist aber monoton fallend und durch

1 beschränkt (bei Wahl von $k = 1$). Demnach ist $e^{-\delta}\frac{\delta^k}{k!} \leq \delta$ für alle $k \in \mathbf{N}$ und somit

$$\rho_0(P^X, P^Y) \leq \delta = |\lambda - \mu|,$$

woraus wie oben die zweite Aussage in (2.1.66) folgt.
Man beachte, daß im obigen Beweis mehrfach von der Ungleichung

$$1 + x \leq e^x \leq 1 + x + \frac{x^2}{2}e^{|x|}, \quad x \in \mathbf{R},$$

Gebrauch gemacht wurde, die z.B. leicht aus der Taylor–Entwicklung der Exponentialfunktion hergeleitet werden kann.
Damit ist das Lemma bewiesen. ∎

Mit Hilfe von Lemma 2.1.10 lassen sich jetzt die Konvergenzaussagen in Beziehung (2.1.67) wie folgt präzisieren:

Lemma 2.1.11. *(Konvergenzabschätzungen bei Poisson–Approximation) Unter den Voraussetzungen von Satz 2.1.4 gilt:*

$$\begin{aligned} \rho\big(\mathfrak{B}(n,p_n), \mathfrak{P}(\lambda)\big) &\leq \frac{\lambda^2}{2n} + |np_n - \lambda| \\ \rho_0\big(\mathfrak{B}(n,p_n), \mathfrak{P}(\lambda)\big) &\leq \frac{\lambda^2}{n} + |np_n - \lambda|. \end{aligned} \tag{2.1.67}$$

Beweis. Dies folgt aus Satz 2.1.5 und der Subadditivität von ρ_0, wenn man in Beziehung (2.1.65) p durch p_n und λ durch $\frac{\lambda}{n}$ ersetzt. ∎

Man beachte, daß Beziehung (2.1.67) auch dann noch gültig bleibt, wenn auf der rechten Seite der Term λ^2 durch den Term $(np_n)^2$ ersetzt wird; dies folgt aus der Dreiecksungleichung für Metriken sowie (2.1.66) vermöge der Beziehung

$$\rho\big(\mathfrak{B}(n,p_n), \mathfrak{P}(\lambda)\big) \leq \rho\big(\mathfrak{B}(n,p_n), \mathfrak{P}(np_n)\big) + \rho\big(\mathfrak{P}(np_n), \mathfrak{P}(\lambda)\big),$$

analog für ρ_0.

Binomialverteilungen spielen in verschiedenen stochastischen Modellen innerhalb der Informatik eine große Rolle, etwa beim Sortieren reeller Zahlen durch Verteilen mittels `hybridsort` (Mehlhorn (1986), Abschnitt II.2.2.). Wir wollen dieses Verfahren hier kurz beschreiben und dabei den Zusammenhang mit der Poisson–Approximation aufzeigen.

Beispiel 2.1.2. (`hybridsort`)
Gegeben seien $n \in \mathbf{N}$ reelle Zahlen $x_1, \ldots, x_n \in (0,1]$, die der Größe nach zu sortieren sind. Für eine feste Zahl $\alpha > 0$ sei $k = \lfloor \alpha n \rfloor$. `hybridsort` operiert dann in zwei Stufen:

a) Schaffe k leere Körbe; wirf die i–te Zahl x_i in den Korb $\lceil kx_i \rceil$, $1 \leq i \leq n$.
b) Wende auf jeden Korb ein (gewöhnliches) Sortierverfahren an und konkateniere die Körbe, d.h. reihe sie hintereinander auf (dabei bezeichne $\lceil x \rceil$ die kleinste ganze Zahl oberhalb von x).

Für eine stochastische Analyse des Verfahrens nehmen wir an, daß die Zahlen $x_1, \ldots, x_n$ Realisationen unabhängiger, jeweils $\mathcal{R}((0,1])$–verteilter Zufallsvariablen $X_1, \ldots, X_n$ sind. Dann ist die Anzahl N_j der in Korb j befindlichen Zahlen $\mathfrak{B}\left(n, \frac{1}{k}\right)$–verteilt für $1 \leq j \leq k$, da die Erfolgswahrscheinlichkeit p_j dafür, daß die Zahl X_i in Korb j geworfen wird, gegeben ist durch

$$\begin{aligned} p_j &= P(j-1 < kX_i \leq j) = P\Big(\frac{j-1}{k} < X_i \leq \frac{j}{k}\Big) \\ &= \frac{1}{k}, \quad 1 \leq j \leq k,\ 1 \leq i \leq n. \end{aligned} \tag{2.1.68}$$

Diese Feststellung läßt sich dann z.B. dazu verwenden, Aussagen über die mittlere Laufzeit des Verfahrens zu gewinnen; hierauf werden wir später in Kapitel 4 noch einmal gesondert zurückkommen.
Ist die Anzahl n der zu sortierenden Zahlen groß, so wächst entsprechend die Zahl k der Körbe; für die Erfolgswahrscheinlichkeiten $p_j = \frac{1}{k}$, $1 \leq j \leq k$, bedeutet dies:

$$np_j = \frac{n}{k} = \frac{n}{\lfloor \alpha n \rfloor} \longrightarrow \frac{1}{\alpha} := \lambda \quad (n \to \infty).$$

Für große n ist also jede der Anzahlen N_j, $1 \leq j \leq k$, näherungsweise $\mathfrak{P}(\lambda)$–verteilt. Mit Lemma 2.1.11 erhält man in diesem Fall z.B. die Fehlerabschätzung

$$\left|P(N_j = m) - \mathfrak{P}(\lambda)(\{m\})\right| \leq \frac{2}{\alpha \lfloor \alpha n \rfloor} = \frac{2}{\alpha k} = \frac{2\lambda}{k} \tag{2.1.69}$$

für $1 \leq j \leq k$, $m \in \mathbb{N}_0$. ∎

Man beachte jedoch, daß die Zufallsvariablen $N_1, \ldots, N_k$ z.B. aufgrund der funktionalen Beziehung $\sum_{i=1}^{k} N_i = n$ nicht stochastisch unabhängig sind; vielmehr ist die gemeinsame Verteilung von $N_1, \ldots, N_k$ durch eine sogenannte *Multinomial–Verteilung* gegeben, die man allgemeiner wie folgt erhält:

a) Es seien $A_1, \ldots, A_k, k \in \mathbb{N}$, disjunkte Ereignisse in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit $\bigcup_{i=1}^{k} A_i = \Omega$ und $P(A_i) > 0$, $1 \leq i \leq k$. Dann ist der Vektor der zugehörigen Indikatorfunktionen $\boldsymbol{I} = (I_1, \ldots, I_k) = (\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k})$ *multinomialverteilt* mit Parametern $p_i = P(A_i)$, $1 \leq i \leq k$, in Zeichen: $P^{\boldsymbol{I}} = \mathfrak{M}(1; p_1, \ldots, p_k)$, oder explizit:

$$P(I_1 = i_1, \ldots, I_k = i_k) = \begin{cases} p_1^{i_1} \cdot \ldots \cdot p_k^{i_k} & \text{für } i_1, \ldots, i_k \in \{0,1\}, \\ & \sum_{j=1}^{k} i_j = 1 \\ & \text{(d.h. genau für ein } j \text{ ist } i_j = 1) \\ 0 & \text{sonst.} \end{cases} \tag{2.1.70}$$

Insbesondere ist jede der Komponenten $I_j = \mathbb{1}_{A_j}$, $1 \leq j \leq k$, selbst $\mathfrak{B}(1, p_j)$–verteilt, jedoch sind $I_1, \ldots, I_k$ nicht stochastisch unabhängig (beispielsweise ist $P(I_i = I_j = 1) = P(A_i \cap A_j) = P(\emptyset) = 0 < p_i p_j = P(I_i = 1)P(I_j = 1)$ für $i \neq j$).

b) Es seien $\boldsymbol{J}_1, \ldots, \boldsymbol{J}_n$ stochastisch unabhängige, jeweils $\mathfrak{M}(1; p_1, \ldots, p_k)$–verteilte Zufallsvektoren ($n, k \in \mathbb{N}$). Dann ist der (k–dimensionale) Zufallsvektor $\boldsymbol{S}_n = \sum_{j=1}^{n} \boldsymbol{J}_j$ ebenfalls *multinomialverteilt* mit Parametern $p_i = P(A_i)$, $1 \leq i \leq k$, in Zeichen: $P^{\boldsymbol{S}_n} = \mathfrak{M}(n; p_1, \ldots, p_k)$, oder explizit:

$$P(\boldsymbol{S}_n = (i_1, \ldots, i_k)) = \begin{cases} \binom{n}{i_1, \ldots, i_k} p_1^{i_1} \cdot \ldots \cdot p_k^{i_k} & \text{für } i_1, \ldots, i_k \in \mathbb{N}_0, \\ & \sum_{j=1}^{k} i_j = n \\ 0 & \text{sonst,} \end{cases} \tag{2.1.71}$$

wobei $\binom{n}{i_1, \ldots, i_k} = \frac{n!}{i_1! i_2! \cdots i_k!}$, $i_1, \ldots, i_k \in \mathbb{N}_0$, die *Multinomialkoeffizienten* bezeichne. Nach Lemma 2.1.6 sind für festes $j \in \{1, \ldots, k\}$ die Projektionen J_{ij}, $1 \leq i \leq n$, stochastisch unabhängige, jeweils $\mathfrak{B}(1, p_j)$–verteilte Zufallsvariablen; die Komponenten S_{nj} des Zufallsvektors $\boldsymbol{S}_n$ sind daher jeweils $\mathfrak{B}(n, p_j)$–verteilt (aber nicht stochastisch unabhängig).

Anschaulich entstehen Multinomialverteilungen $\mathfrak{M}(n; p_1, \ldots, p_k)$ also durch n unabhängige Versuchswiederholungen, bei denen im Einzelversuch genau eines von k möglichen, paarweise disjunkten (und die Grundmenge Ω insgesamt ausschöpfenden) Ereignissen eintritt, wenn man die Häufigkeiten des Eintretens der verschiedenen Ereignisse zählt.

Entsprechend den Ausführungen zur Binomialverteilung läßt sich die Beziehung (2.1.71) auch wieder kombinatorisch deuten, da es gerade

$$\binom{n}{i_1, \ldots, i_k} = \binom{n}{i_1} \binom{n - i_1}{i_2} \cdots \binom{n - i_1 - \ldots - i_{k-1}}{i_k}$$

Möglichkeiten gibt, n Objekte in k disjunkte Gruppen mit den Umfängen $i_1, \ldots, i_k$ einzuteilen.

Aus dem gerade Gesagten folgt also, daß die bei `hybridsort` auftretenden Anzahlen $(N_1, \ldots, N_k)$ $\mathfrak{M}(n; \frac{1}{k}, \ldots, \frac{1}{k})$–multinomialverteilt sind.

Will man in einem Modell zur stochastischen Analyse von `hybridsort` auch die Möglichkeit einer nicht–gleichmäßigen Verteilung der zu sortierenden Elemente über $(0, 1]$ miterfassen, so läßt sich dies durch die Annahme einer beliebigen Verteilung Q über $(0, 1]$ mit Verteilungsfunktion F für die unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ bewerkstelligen. Die Anzahlen N_j der nun in Korb j befindlichen Elemente sind dann wieder $\mathfrak{B}(n, p_j)$–verteilt, allerdings hier mit Parameter $p_j = F(\frac{j}{k}) - F(\frac{j-1}{k})$, $1 \leq j \leq k$, wie sich unmittelbar aus Beziehung (2.1.68) ablesen läßt. Entsprechend ist der Zufallsvektor $(N_1, \ldots, N_k)$ jetzt $\mathfrak{M}(n; p_1, \ldots, p_k)$–verteilt.

Will man ein günstiges "durchschnittliches" Verhalten bezüglich der Laufzeit des Algorithmus erreichen, ist es vorteilhaft, die Verteilung der Anzahlen der Elemente in den einzelnen Körben möglichst gleichartig zu gestalten. Diese Prämisse ist in dem zuletzt diskutierten Fall — bei beliebiger Verteilung Q — offenbar nicht mehr gegeben. Man kann aber bei Kenntnis von Q im Falle der Stetigkeit der Verteilungsfunktion F durch die Anwendung von F auf die zu sortierenden Elemente diese vorteilhafte Situation leicht erreichen, da in diesem Fall nach Satz 2.1.1 die Zufallsvariablen $F(X_1), \ldots, F(X_n)$ jeweils $\mathcal{R}((0, 1])$–verteilt sind. Man beachte,

daß dies sogar in der Situation einer *beliebigen* Verteilung Q (nicht nur über $(0,1]$) mit stetiger Verteilungsfunktion F gilt. Die anfangs getroffene Annahme über die Verteilung der Elemente bedeutet also praktisch keine Einschränkung der Allgemeinheit.

In Verallgemeinerung von Satz 2.1.4 läßt sich mit völlig analogen Methoden zeigen, daß auch die Multinomialverteilung unter bestimmten Bedingungen an die Parameter eine Grenzverteilung besitzt, und zwar eine Produktverteilung, deren Faktoren jeweils Poisson–Verteilungen bilden. Genauer gilt:

Satz 2.1.6. *(Grenzverteilung von Multinomialverteilungen)*
Es sei für festes $k \in \mathbb{N}$ und $1 \le j \le k-1$ $\{p_{jn}\}_{n\in\mathbb{N}} \subseteq (0,1)$ je eine Folge reeller Zahlen mit der Eigenschaft $\lim_{n\to\infty} np_{jn} = \lambda_j > 0$, $1 \le j \le k-1$. Dann gilt

$$\begin{aligned}\lim_{n\to\infty} \mathfrak{M}(n; p_{1n}, \ldots, p_{kn})\big(\{i_1, \ldots, i_{k-1}, n - \sum_{j=1}^{k-1} i_j\}\big) \\ = \mathfrak{P}(\lambda_1)(\{i_1\}) \cdots \mathfrak{P}(\lambda_{k-1})(\{i_{k-1}\}) \\ = \mathfrak{P}(\lambda_1) \otimes \cdots \otimes \mathfrak{P}(\lambda_{k-1})\big(\{i_1, \ldots, i_{k-1}\}\big)\end{aligned} \tag{2.1.72}$$

für alle $i_1, \ldots, i_{k-1} \ge 0$, $\sum_{j=1}^{k-1} i_j \le n$.

Da der Beweis analog zum Beweis von Satz 2.1.4 verläuft, wollen wir ihn hier nicht explizit führen.

Man beachte, daß wegen $\sum_{j=1}^{k} p_{jn} = 1$ für alle $n \in \mathbb{N}$ aus den angegebenen Bedingungen auch folgt: $\lim_{n\to\infty} p_{kn} = 1$.

Beziehung (2.1.72) impliziert also insbesondere, daß unter den angegebenen asymptotischen Bedingungen die ersten $k-1$ Komponenten entsprechend multinomialverteilter Zufallsvektoren asymptotisch unabhängig und asymptotisch Poissonverteilt sind.

Auch für Satz 2.1.6 lassen sich — ähnlich wie in Lemma 2.1.11 — Untersuchungen der Approximationsgüte anstellen, die allerdings technisch aufwendiger sind; so erhält man z.B. mit den Bezeichnungen in Beziehung (2.1.72)

$$\begin{aligned}\Big|\mathfrak{M}(n; p_{1n}, \ldots, p_{kn})\big(\{i_1, \ldots, i_{k-1}, n - \sum_{j=1}^{k-1} i_j\}\big) - \mathfrak{P}(\lambda_1)(\{i_1\}) \cdots \mathfrak{P}(\lambda_{k-1})(\{i_{k-1}\})\Big| \\ \le \frac{1}{n}\Big(\sum_{j=1}^{k-1} \lambda_j\Big)^2 + \sum_{1\le j\le k-1} |np_{jn} - \lambda_j|.\end{aligned} \tag{2.1.73}$$

Die Aussage bleibt gültig, wenn in (2.1.73) der Term $\Big(\sum_{j=1}^{k-1} \lambda_j\Big)^2$ durch den Term $\Big(\sum_{j=1}^{k-1} np_{jn}\Big)^2$ ersetzt wird.

Man beachte, daß sich hieraus bei der Wahl $k = 1$ das Ergebnis des Lemmas 2.1.10 ergibt.

Für tiefergehende Resultate sei hier etwa auf die Arbeit von Deheuvels & Pfeifer (1988b) verwiesen.

Für das Beispiel von `hybridsort` bedeutet dies, daß jeweils $m \in \mathbb{N}$ feste Anzahlen $(N_{i_1}, \ldots, N_{i_m})$ der in den Körben $i_1, \ldots, i_m$ befindlichen Elemente näherungsweise — bei großem n — die Produktverteilung $\mathfrak{P}(\frac{1}{\alpha}) \otimes \cdots \otimes \mathfrak{P}(\frac{1}{\alpha})$ (m Faktoren) besitzen. Dies ergibt sich unmittelbar aus Satz 2.1.6 bzw. der Abschätzung (2.1.73) mit $k = m + 1$, wenn man die restlichen $n - m$ Körbe zu einem einzigen Korb zusammenfaßt. Analog zu (2.1.69) ergibt die Fehlerabschätzung (2.1.73) hier eine obere Schranke von $m(m+1)\frac{\lambda}{k}$. Dies bedeutet, daß eine sinnvolle Poisson–Approximation auch noch für den Fall gegeben ist, daß die Anzahl $m = m_n$ der betrachteten Körbe von n abhängt, solange $\lim_{n\to\infty} \frac{m_n}{\sqrt{n}} = 0$ gilt.

Stärker noch als in den bisher betrachteten Situationen erweisen sich Poisson–Approximationen besonders dann als nützlich, wenn man an der Verteilung einer Summe unabhängiger, aber mit verschiedenen Parametern binomialverteilter Zufallsvariablen interessiert ist, die z.B. bei der Average–Case–Analyse gewisser Suchalgorithmen auftreten. Solche sogenannten *Poisson–Binomialverteilungen* lassen sich nämlich i.a. nicht mehr geschlossen darstellen. Mit Hilfe des allgemeinen Satzes 2.1.5 und Lemma 2.1.10 kann man diese Situation aber ebenfalls einfach behandeln. Wir wollen dazu die Poisson–Binomialverteilung, die sich aus der Summe unabhängiger, mit Parametern $p_1, \ldots, p_n$ binomialverteilter Zufallsvariablen ergibt, mit $\mathfrak{PB}(n; p_1, \ldots, p_n)$ bezeichnen.

Satz 2.1.7. *(Poisson–Approximation für Poisson–Binomialverteilungen)*
Es seien $X_1, \ldots, X_n$, $n \in \mathbb{N}$ unabhängige, binomialverteilte Zufallsvariablen mit Parametern $p_i = P(X_i = 1) \in [0,1]$, $1 \le i \le n$. Dann gilt:

$$\begin{aligned}
\rho\Big(P^{\sum_{i=1}^n X_i}, \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) &= \rho\Big(\mathfrak{PB}(n; p_1, \ldots, p_n), \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) \le \frac{1}{2}\sum_{i=1}^n p_i^2 \\
\rho_0\Big(P^{\sum_{i=1}^n X_i}, \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) &= \rho_0\Big(\mathfrak{PB}(n; p_1, \ldots, p_n), \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) \le \sum_{i=1}^n p_i^2
\end{aligned} \tag{2.1.74}$$

Beweis. Dies folgt sofort aus (2.1.59), (2.1.64) und (2.1.67), wenn man in der letzteren Beziehung jeweils $n = 1$ und $p = p_i$, $1 \le i \le n$ wählt. ∎

Eine in einigen Fällen günstigere Abschätzung wurde von Barbour & Hall (1984) entwickelt; aus den dort angestellten Untersuchungen ergeben sich z.B. die Ungleichungen

$$\begin{aligned}
\rho\Big(\mathfrak{PB}(n; p_1, \ldots, p_n), \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) &\le \frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n p_i} \\
\rho_0\Big(\mathfrak{PB}(n; p_1, \ldots, p_n), \mathfrak{P}\Big(\sum_{i=1}^n p_i\Big)\Big) &\le \frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n p_i},
\end{aligned} \tag{2.1.75}$$

die insbesondere für große Werte von $\sum_{i=1}^n p_i$ zu besseren Ergebnissen führen.

Mit Hilfe der Abschätzungen (2.1.74) oder (2.1.75) läßt sich auch eine Konvergenzaussage analog zu Satz 2.1.4 für Poisson–Binomialverteilungen beweisen.

Satz 2.1.8. *(Konvergenz von Poisson–Binomialverteilungen)*
Es sei $\{p_{in} \mid 1 \le i \le n,\ n \in \mathbb{N}\} \subseteq [0,1]$ *eine Doppelfolge reeller Zahlen mit der Eigenschaft*

$$\lim_{n\to\infty} \sum_{i=1}^{n} p_{in} = \lambda > 0, \quad \lim_{n\to\infty} \max_{1\le i\le n} \{p_{in}\} = 0. \tag{2.1.76}$$

Dann gilt

$$\lim_{n\to\infty} \mathfrak{PB}(n; p_{1n}, \ldots, p_{nn})(\{k\}) = \mathfrak{P}(\lambda)(\{k\}),\ k \in \mathbb{N}_0. \tag{2.1.77}$$

Beweis. Die vorausgesetzten Limesbeziehungen in (2.1.76) implizieren

$$\lim_{n\to\infty} \sum_{i=1}^{n} p_{in}^2 \le \lim_{n\to\infty} \max_{1\le i\le n} \{p_{in}\} \lim_{n\to\infty} \sum_{i=1}^{n} p_{in} = 0.$$

Die Aussage des Satzes ergibt sich damit aus der Beziehung (2.1.74) oder (2.1.75) für ρ_0 mittels der Dreiecksungleichung für Metriken und der Beziehung (2.1.66). ∎

Ähnlich wie bei der Multinomialverteilung bleibt eine Poisson–Approximation für Poisson–Binomialverteilungen auch dann noch sinnvoll, wenn $\sum_{i=1}^{n} p_{in}$ nicht beschränkt bleibt, aber der Quotient $\dfrac{\sum_{i=1}^{n} p_{in}^2}{\sum_{i=1}^{n} p_{in}}$ noch gegen Null strebt für $n \to \infty$, wie man an den in (2.1.75) angegebenen Abschätzungen sehen kann; dabei braucht man nicht einmal die zweite Limesbeziehung aus (2.1.76) vorauszusetzen. Für diesen Fall sind die angegebenen Ungleichungen sogar recht scharf; asymptotisch gilt nämlich hier bei $n \to \infty$

$$\begin{aligned} \rho\Big(\mathfrak{PB}(n; p_{1n}, \ldots, p_{nn}), \mathfrak{P}\Big(\sum_{i=1}^{n} p_{in}\Big)\Big) &\sim \frac{1}{2\sqrt{2\pi e}} \frac{\sum_{i=1}^{n} p_{in}^2}{\sum_{i=1}^{n} p_{in}} \\ &\approx 0{,}121 \frac{\sum_{i=1}^{n} p_{in}^2}{\sum_{i=1}^{n} p_{in}}. \end{aligned} \tag{2.1.78}$$

Der interessierte Leser sei hier verwiesen auf die Arbeit von Deheuvels & Pfeifer (1988a).

Zum Abschluß dieses Abschnitts wollen wir uns noch mit der Transformation stetig verteilter Zufallsvariablen bzw. –vektoren beschäftigen. Wir beginnen mit einem Analogon zu Lemma 2.1.3.

Satz 2.1.9. *(Transformationssatz für stetige Verteilungen)*
Es sei $\boldsymbol{X} = (X_1, \ldots, X_m)$, $m \in \mathbb{N}$, *ein* m*–dimensionaler Zufallsvektor auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *mit Dichte* $f_{\boldsymbol{X}}$ *und* $G : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ *eine meßbare Abbildung mit folgenden Eigenschaften:*

G ist injektiv und stetig differenzierbar auf $\mathbb{R}^m$;

die Funktionaldeterminante $\det\Big(\dfrac{\partial G_i}{\partial x_j}\Big)_{\{1\le i,j\le m\}}$ von G

ist entweder *positiv* oder *negativ* auf $\mathbb{R}^m$.

Dann besitzt der Zufallsvektor $\boldsymbol{Y} = G(\boldsymbol{X})$ *ebenfalls eine Dichte* $f_{\boldsymbol{Y}}$*; eine geeignete Wahl hierfür ist*

$$f_{\boldsymbol{Y}}(y_1,\ldots,y_m) = \begin{cases} \dfrac{f_{\boldsymbol{X}}(G^{-1}((y_1,\ldots,y_m))}{\left|\det\left(\frac{\partial G_i}{\partial x_j}(G^{-1}(y_1,\ldots,y_m))\right)\right|} & \text{für } (y_1,\ldots,y_m) \in G(\mathbf{R}^m) \\ 0 & \text{sonst.} \end{cases} \tag{2.1.79}$$

Beweis. Dies folgt aus der für das Riemann–Integral gültigen Substitutionsregel; vgl. Heuser (1988), 205.2. Einen maßtheoretischer Beweis findet man bei Floret (1981), §17; insbesondere ist hier $G(\mathbf{R}^m) \in \mathcal{B}^m$. ∎

Man beachte, daß im Gegensatz zu Lemma 2.1.3 in Satz 2.1.9 der Bildbereich von G dieselbe Dimension m wie der Definitionsbereich besitzten muß, da sonst die Funktionaldeterminante von G nicht existiert. Will man auch die allgemeinere Situation eines n–dimensionalen Bildbereichs für G mit z.B. $n < m$ miterfassen, so kann man dies durch eine geeignete Erweiterung erreichen, indem man etwa

$$G^*(y_1,\ldots,y_m) = \big(y_1,\ldots,y_{m-n},G(y_1,\ldots,y_m)\big), \quad y_1,\ldots,y_m \in \mathbf{R},$$

setzt und nur die letzten n Komponenten des hieraus resultierenden Zufallsvektors $\boldsymbol{Y}^* = G^*(\boldsymbol{X})$ betrachtet. Diese Vorgehensweise bietet sich z.B. dann an, wenn man — analog zu Lemma 2.1.4 — an der Verteilung der Summe von stetig verteilten Zufallsvariablen interessiert ist.

Lemma 2.1.12. *(Summe stetig verteilter Zufallsvariablen; Faltungslemma) Es sei* $\boldsymbol{X} = (X_1, X_2)$ *ein stetig verteilter Zufallsvektor mit Dichte* $f_{\boldsymbol{X}}$*. Dann besitzt die Zufallsvariable* $Y = X_1 + X_2$ *eine Dichte* f_Y *der Form*

$$f_Y(y) = \int f_{\boldsymbol{X}}(x, y-x)\,dx, \quad y \in \mathbf{R}. \tag{2.1.80}$$

Sind speziell die Zufallsvariablen X_1, X_2 *stochastisch unabhängig, so ist*

$$f_Y(y) = \int f_{X_1}(x) f_{X_2}(y-x)\,dx, \quad y \in \mathbf{R}. \tag{2.1.81}$$

Beweis. Wir betrachten die Abbildung $G(x_1,x_2) = (x_1, x_1+x_2)$, $x_1, x_2 \in \mathbf{R}$. Diese ist offensichtlich injektiv und auf $\mathbf{R}^2$ stetig differenzierbar mit

$$\det\left(\frac{\partial G_i}{\partial x_j}\right)_{\{1\le i,j\le 2\}} = \det\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = 1.$$

Ferner ist $G^{-1}(y_1,y_2) = (y_1, y_2 - y_1)$, $y_1, y_2 \in \mathbf{R}$, so daß nach Satz 2.1.9 der Zufallsvektor $\boldsymbol{Z} = G(\boldsymbol{X}) = (X_1, X_1 + X_2)$ eine Dichte der Form

$$f_{\boldsymbol{Z}}(y_1,y_2) = f_{\boldsymbol{X}}(y_1, y_2 - y_1), \quad y_1, y_2 \in \mathbf{R},$$

besitzt. Die zweite Randverteilung von $\boldsymbol{Z}$ ist nun die gewünschte Verteilung der Summe $X_1 + X_2$, welche man nach Lemma 1.4.4 durch Integration von $f_{\boldsymbol{Z}}$ nach der

Variablen y_1 erhält. Dies ergibt Beziehung (2.1.80). Die andere Aussage folgt aus der Tatsache, daß bei stochastischer Unabhängigkeit die Dichte $f_{\boldsymbol{X}}$ gerade durch das Produkt der Randdichten f_{X_1} und f_{X_2} gegeben ist. ∎

Sind z.B. $X_1, \ldots, X_n$, $n \in \mathbb{N}$, unabhängige, $\mathcal{E}(\lambda)$–exponentialverteilte Zufallsvariablen, so erhält man für die Summenvariable $Y_n = \sum_{k=1}^{n} X_k$ eine sogenannte *Erlang–Verteilung* $\mathcal{E}(n, \lambda)$ mit Dichte

$$f_{Y_n}(y) = \begin{cases} \dfrac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y} & \text{für } y > 0 \\ 0 & \text{sonst.} \end{cases} \tag{2.1.82}$$

Dies läßt sich mit vollständiger Induktion leicht nachprüfen:
Für $n = 1$ ergibt sich offenbar sofort Beziehung (2.1.8), d.h. es ist $\mathcal{E}(1, \lambda) = \mathcal{E}(\lambda)$. Angenommen, Beziehung (2.1.82) sei gültig für ein $n \in \mathbb{N}$. Nach dem Faltungslemma 2.1.12 ist dann für $y > 0$

$$\begin{aligned} f_{Y_{n+1}}(y) &= \int_0^y f_{Y_n}(x) f_{X_{n+1}}(y-x)\, dx = \int_0^y \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \lambda e^{-\lambda(y-x)}\, dx \\ &= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda y} \int_0^y x^{n-1}\, dx = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda y} \left. \frac{x^n}{n} \right|_0^y = \frac{\lambda^{n+1}}{n!} e^{-\lambda y}, \end{aligned}$$

was mit (2.1.82) für $n+1$ übereinstimmt.

Erlang–Verteilungen sind also ebenfalls faltungsstabil, d.h. es gilt

$$\mathcal{E}(n, \lambda) * \mathcal{E}(m, \lambda) = \mathcal{E}(n+m, \lambda), \quad n, m \in \mathbb{N}. \tag{2.1.83}$$

Wegen $\int f_{Y_{n+1}}(y)\, dy = 1$ ergibt sich mit $\lambda = 1$ also speziell auch die Beziehung

$$n! = \int_0^\infty x^n e^{-x}\, dx = \Gamma(n+1)$$

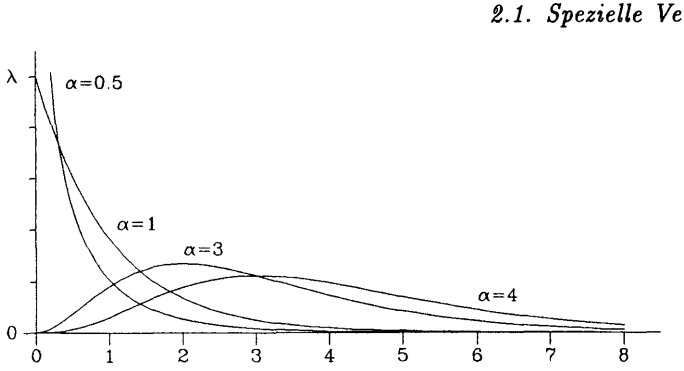
(Euler'sche Γ–Funktion an der Stelle n). Erlang–Verteilungen stellen sich daher als Spezialfälle der allgemeineren Γ–*Verteilungen* $\Gamma(\alpha, \lambda)$ mit Parameter $\alpha > 0$ heraus, deren Dichten $f_{\alpha,\lambda}$ gegeben sind durch

$$f_{\alpha,\lambda}(y) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} & \text{für } y > 0 \\ 0 & \text{sonst,} \end{cases} \tag{2.1.84}$$

d.h. es ist $\mathcal{E}(n, \lambda) = \Gamma(n+1, \lambda)$, $n \in \mathbb{N}$. Ähnlich wie die Erlang–Verteilungen sind auch die Γ–Verteilungen faltungsstabil, d.h. es gilt

$$\Gamma(\alpha, \lambda) * \Gamma(\beta, \lambda) = \Gamma(\alpha + \beta, \lambda), \quad \alpha, \beta, \lambda > 0. \tag{2.1.85}$$

Im folgenden sind einige der Dichten $f_{\alpha,\lambda}$ für verschiedene Parameterwerte von α skizziert (man beachte, daß der Parameter λ lediglich eine Skalenänderung der Achsen bewirkt).



Für große Werte von α werden die Dichten der Γ–Verteilung immer symmetrischer. Wir wollen jetzt untersuchen, inwieweit sich diese Dichten — bei geeigneter Zentrierung — einer Grenzdichte annähern. Dazu ist es vorteilhaft, die Dichten linear transformierter, stetig verteilter Zufallsvariablen berechnen zu können.

Lemma 2.1.13. *(Dichten linear transformierter Zufallsvariablen)*
Es sei X eine Zufallsvariable mit Dichte f_X und $Y = aX + b$ mit reellen Zahlen a, b, $a \neq 0$. Dann besitzt Y eine Dichte der Form

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \quad y \in \mathbf{R}. \tag{2.1.86}$$

Beweis. Wir können den Transformationssatz 2.1.8 mit $m = 1$ direkt anwenden auf die injektive, stetig differenzierbare Abbildung $G(x) = ax + b$, $x \in \mathbf{R}$. Es ist nämlich $G'(x) \equiv a$ mit $G^{-1}(y) = \frac{y-b}{a}$, so daß (2.1.86) unmittelbar aus (2.1.79) folgt. ∎

Sei nun X_α eine $\Gamma(\alpha, 1)$–verteilte Zufallsvariable. Die linear transformierte Zufallsvariable $Y_\alpha = \frac{1}{\sqrt{\alpha}} X_\alpha - \sqrt{\alpha}$ besitzt dann nach Lemma 2.1.13 die Dichte

$$f_{Y_\alpha}(y) = \sqrt{\alpha} f_{\alpha,1}(y\sqrt{\alpha} + \alpha), \quad y \in \mathbf{R}.$$

Benutzt man jetzt die *Stirling'sche Formel* zur Approximation der Γ–Funktion

$$\lim_{z \to \infty} \frac{\Gamma(z)}{\sqrt{2\pi} e^{-z} z^{z-\frac{1}{2}}} = 1 \tag{2.1.87}$$

(Abramowitz & Stegun (1984), 6.1.37 oder Heuser (1988), Abschnitt 96), so ergibt sich durch Einsetzen in (2.1.84) für $y \in \mathbf{R}$ und genügend große α

$$\begin{aligned} f_{Y_\alpha}(y) &\sim \frac{1}{\sqrt{2\pi}} \Big(1 + \frac{y}{\sqrt{\alpha}}\Big)^{\alpha - 1} e^{-y\sqrt{\alpha}} \sim \frac{1}{\sqrt{2\pi}} \Big(1 + \frac{y}{\sqrt{\alpha}}\Big)^{\alpha} e^{-y\sqrt{\alpha}} \\ &= \frac{1}{\sqrt{2\pi}} \exp\Big(- y\sqrt{\alpha} + \alpha \ln\big(1 + \frac{y}{\sqrt{\alpha}}\big)\Big) \\ &\sim \frac{1}{\sqrt{2\pi}} \exp\Big(- y\sqrt{\alpha} + \alpha\big[\frac{y}{\sqrt{\alpha}} - \frac{y^2}{2\alpha}\big]\Big) \\ &= \frac{1}{\sqrt{2\pi}} \exp\big(- \frac{y^2}{2}\big) =: f(y) \quad (\alpha \to \infty). \end{aligned} \tag{2.1.88}$$

Die Grenzfunktion $f(y)$ bildet dabei die Dichte der sogenannten (Standard–)*Normalverteilung*, welche mit $\mathcal{N}(0,1)$ bezeichnet wird. Besitzt eine Zufallsvariable X eine solche Verteilung, so heißt die Verteilung der positiv–lineartransformierten Zufallsvariablen $Y = aX + b$, $a > 0, b \in \mathbf{R}$, ebenfalls Normalverteilung, in Zeichen: $\mathcal{N}(b, a^2)$. Die Bedeutung der Parameter a und b sowie der Bezeichnungsweise wird allerdings erst im nächsten Abschnitt klar, wenn die Begriffe Erwartungswert und Varianz von Verteilungen eingeführt sind.

Wählt man den Parameter $\alpha = n+1$, $n \in \mathbf{N}$, ganzzahlig, d.h. betrachtet man Erlang–Verteilungen $\mathcal{E}(n,1)$, so besagt das gerade hergeleitete Ergebnis, daß für unabhängige Zufallsvariablen $X_1, \ldots, X_n$, die jeweils $\mathcal{E}(1)$–exponentialverteilt sind, die zentrierte Summe $\dfrac{\sum_{k=1}^n X_k - n}{\sqrt{n}}$ für große n asymptotisch $\mathcal{N}(0,1)$–normalverteilt ist. Es wird sich später zeigen, daß dieses Resultat im wesentlichen (d.h. bis auf die Wahl der zentrierenden Konstanten n und $\sqrt{n}$) allgemeingültig ist, also insbesondere nicht von der Art der zugrundeliegenden Verteilung der Zufallsvariablen $X_1, \ldots, X_n$ abhängt. Diese Beobachtung rechtfertigt die große Bedeutung, welche der Normalverteilung in stochastischen Modellen zukommt.

Auch die Normalverteilung ist faltungsstabil, wie man durch Nachrechnen mit Hilfe von Lemma 2.1.12 feststellen kann; genauer gilt:

$$\mathcal{N}(b, a^2) * \mathcal{N}(d, c^2) = \mathcal{N}(b+d, a^2+c^2), \quad a, c > 0, b, d \in \mathbf{R}. \tag{2.1.89}$$

Erlang–Verteilungen spielen vor allem bei der Modellierung von Bediensystemen (z.B. Ein–Prozessor–Systemen) eine zentrale Rolle, wenn man davon ausgeht, daß die individuellen Bedienzeiten (Verweildauern von Programmen im Rechner) $X_1, \ldots, X_n$, $n \in \mathbf{N}$, voneinander unabhängig und — etwa begründet durch die Gedächtnislosigkeit — $\mathcal{E}(\lambda)$–exponentialverteilt sind. Die $\mathcal{E}(n,\lambda)$–Erlangverteilten Summenvariablen $S_n = \sum_{k=1}^n X_k$ geben dann den Zeitpunkt an, zu dem der $n+1$–te Kunde bedient (das $n+1$–te Programm gestartet) wird. Für eine äquivalente, aber mehr "zeitbezogene" Modellierung des Systems kann man auch den durch

$$N_t = \#\{n \in \mathbf{N} \mid S_n \le t\}, \quad t > 0, \tag{2.1.90}$$

definierten *Zählprozeß* betrachten, der angibt, wieviele Kunden (Programme) bis zum Zeitpunkt $t > 0$ abgefertigt (beendet) wurden. Setzt man noch $S_0 := 0$, so läßt sich wegen

$$\begin{aligned} \{N_t = n\} &= \{S_n \le t < S_{n+1}\} \\ &= \{S_{n+1} > t\} \setminus \{S_n > t\}, \quad n \in \mathbf{N}_0, t > 0, \end{aligned} \tag{2.1.91}$$

und der Subtraktivität von P leicht die Verteilung von N_t für alle $t > 0$ bestimmen durch die Rechnung (partielle Integration)

$$\begin{aligned} P(N_t = n) &= P(S_{n+1} > t) - P(S_n > t) \\ &= \int_t^\infty \underbrace{\frac{\lambda^n x^n}{n!}}_{u(x)} \underbrace{\lambda e^{-\lambda x}}_{-v'(x)}\, dx - \int_t^\infty \underbrace{\frac{\lambda^n x^{n-1}}{(n-1)!}}_{u'(x)} \underbrace{e^{-\lambda x}}_{v(x)}\, dx \\ &= -\int_t^\infty \Big(u(x)v'(x) + u'(x)v(x)\Big)\, dx = -u(x)v(x)\Big|_t^\infty \\ &= u(t)v(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbf{N}_0, \end{aligned} \tag{2.1.92}$$

d.h. N_t ist $\mathfrak{P}(\lambda t)$-verteilt für jeden Zeitpunkt $t > 0$. Der Prozeß $\{N_t \mid t > 0\}$ heißt daher auch *Poisson-Prozeß* mit Parameter λ. Hierauf und auf weitere charakteristische Eigenschaften des Poisson-Prozesses werden wir noch einmal in Kapitel 3 zurückkommen.

Bedienungssysteme, die in der Informatik eine besondere Rolle spielen, werden in Ross (1983) sowie ausführlicher auch in Pflug (1986) und Bolch (1989) behandelt.

Zum Abschluß dieses Abschnitts wollen wir noch eine Erweiterung von Satz 2.1.9 behandeln, die vor allem dann benötigt wird, wenn die dort betrachtete Transformation G nicht auf ganz $\mathbf{R}^m$, sondern nur auf einer geeigneten Teilmenge $T \subset \mathbf{R}^m$ definiert ist, etwa $G(x) = \sqrt{x}$ mit $T = [0, \infty) \subset \mathbf{R}^1$.

Satz 2.1.10. *(Allgemeiner Transformationssatz für stetige Verteilungen)*
Es sei $\boldsymbol{X} = (X_1, \ldots, X_m)$, $m \in \mathbf{N}$, ein m-dimensionaler Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Dichte $f_{\boldsymbol{X}}$ und $T \subseteq \mathbf{R}^m$ eine Jordan-Menge mit $P(\boldsymbol{X} \in T) = 1$[1], d.h. $f_{\boldsymbol{X}}(x_1, \ldots, x_m) = 0$ für $(x_1, \ldots, x_m) \in T^c$. Ferner sei $G : T \longrightarrow \mathbf{R}^m$ eine meßbare Abbildung mit den Eigenschaften:

> G ist injektiv und stetig differenzierbar
> auf dem Inneren T° von T;
> die Funktionaldeterminante $\det\left(\frac{\partial G_i}{\partial x_j}\right)_{\{1 \le i,j \le m\}}$ von G
> ist entweder *positiv* oder *negativ* auf T°.

Dann besitzt für jede meßbare Fortsetzung G^* von G auf ganz $\mathbf{R}^m$ der Zufallsvektor $\boldsymbol{Y} = G^*(\boldsymbol{X})$ ebenfalls eine Dichte $f_{\boldsymbol{Y}}$; eine geeignete Wahl hierfür ist

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_m) = \begin{cases} \dfrac{f_{\boldsymbol{X}}(G^{-1}((y_1, \ldots, y_m))}{\left|\det\left(\frac{\partial G_i}{\partial x_j}(G^{-1}(y_1, \ldots, y_m))\right)\right|} & \text{für } (y_1, \ldots, y_m) \in G(T^\circ) \\ 0 & \text{sonst.} \end{cases} \tag{2.1.93}$$

Auch der Beweis dieses allgemeineren Satzes läßt sich mit der Substitutionsregel für mehrfache Riemann-Integrale führen, wobei wieder $G(T^\circ) \in \mathcal{B}^m$ ist.

Man beachte, daß man formal in Satz 2.1.10 eine Fortsetzung G^* von G betrachten muß, da der Wertebereich von X nach unserer Konvention ganz $\mathbf{R}^m$ ist, die Wahl dieser Fortsetzung aber grundsätzlich keine Rolle spielt, da Werte außerhalb von T° nur mit Wahrscheinlichkeitkeit 0 angenommen werden.

Wir wollen die Wirksamkeit des Satzes 2.1.10 an einem abschließenden Beispiel illustrieren. Dazu betrachten wir einen auf dem Einheitsquadrat $T = [0,1] \times [0,1]$ $\mathcal{R}(T)$-gleichverteilten Zufallsvektor $\boldsymbol{X} = (X_1, X_2)$ mit Dichte $f_{\boldsymbol{X}} = \mathbb{1}_T$. Die Abbildung G sei auf T definiert vermöge

$$G(x_1, x_2) = \left(\sqrt{x_1}\cos(2\pi x_2), \sqrt{x_1}\sin(2\pi x_2)\right), \ (x_1, x_2) \in T. \tag{2.1.94}$$

---

[1] vgl. die Fußnote auf Seite 44

Für G^* wählen wir die (meßbare) Fortsetzung

$$G^*(x_1,x_2)=\begin{cases} G(x_1,x_2) & \text{für } (x_1,x_2)\in T \\ 0 & \text{sonst.}\end{cases} \tag{2.1.95}$$

Dann sind die Voraussetzungen von Satz 2.1.10 erfüllt, und $\boldsymbol{Y}=G^*(\boldsymbol{X})$ ist auf dem Einheitskreis $K=K^a(0,0;1)$ $\mathcal{R}(K)$–gleichverteilt. Es ist nämlich

$$\det\Big(\frac{\partial G_i}{\partial x_j}\Big)_{\{1\le i,j\le 2\}}=\begin{vmatrix} \frac{1}{2\sqrt{x_1}}\cos(2\pi x_2) & -2\pi\sqrt{x_1}\sin(2\pi x_2) \\ \frac{1}{2\sqrt{x_1}}\sin(2\pi x_2) & 2\pi\sqrt{x_1}\cos(2\pi x_2)\end{vmatrix}=\pi$$

auf dem Inneren $T^\circ=(0,1)\times(0,1)$ von T und somit

$$f_{\boldsymbol{Y}}(y_1,y_2)=\begin{cases}\frac{1}{\pi} & \text{für } (y_1,y_2)\in G(T^\circ) \\ 0 & \text{sonst.}\end{cases} \tag{2.1.96}$$

Man beachte hierbei, daß das Bild $G(T^\circ)$ allerdings nicht den ganzen offenen Kreis $K^\circ(0,0;1)$ darstellt, sondern nur den "geschlitzten" Kreis $K^\circ(0,0;1)\setminus[0,1)\times\{0\}$. Jedoch besitzt die Menge $M=[0,1)\times\{0\}$ die Fläche $\int\int_M dy_1dy_2=0$, so daß durch (2.1.96) tatsächlich eine Dichte der Gleichverteilung über dem Einheitskreis gegeben ist (vgl. Beziehung (1.4.28)).

Satz 2.1.10 läßt sich auch noch auf den Fall ausdehnen, daß die Menge T *unbeschränkt* ist. Hierzu hat man z.B. lediglich zu fordern, daß die Durchschnitte von T mit allen abgeschlossenen, beschränkten (m–dimensionalen) Intervallen jeweils Jordan–Mengen bilden. Durch Abbildungen $g:[0,\infty)\longrightarrow\mathbf{R}$, die auf $(0,\infty)$ injektiv und stetig differenzierbar sind mit entweder $g'\cdot g>0$ oder $g'\cdot g<0$ erhält man dann allgemein *rotationssymmetrische* Verteilungen in $\mathbf{R}^2$ vermöge der Abbildung

$$G(x_1,x_2)=\big(g(x_1)\cos(2\pi x_2),g(x_1)\sin(2\pi x_2)\big) \tag{2.1.97}$$

für $(x_1,x_2)\in T=[0,\infty)\times[0,1]$ mit

$$\begin{aligned}\det\Big(\frac{\partial G_i}{\partial x_j}\Big)_{\{1\le i,j\le 2\}}&=\begin{vmatrix} g'(x_1)\cos(2\pi x_2) & -2\pi g(x_1)\sin(2\pi x_2) \\ g'(x_1)\sin(2\pi x_2) & 2\pi g(x_1)\cos(2\pi x_2)\end{vmatrix} \\ &=2\pi g'(x_1)g(x_1)\end{aligned}$$

für $(x_1,x_2)\in T^\circ=(0,\infty)\times(0,1)$. Besitzt nun X_1 eine Dichte f mit $f(x)=0$, $x<0$ und $f(x)>0$, $x\ge 0$, und ist X_2 unabhängig von X_1 und über $[0,1]$ gleichverteilt, so ist mit der Fortsetzung G^* wie in (2.1.95) der Zufallsvektor $\boldsymbol{Y}=G^*(\boldsymbol{X})$ rotationssysmmetrisch verteilt mit Dichte

$$f_{\boldsymbol{Y}}(y_1,y_2)=\frac{f\Big(g^{-1}\big(\sqrt{y_1^2+y_2^2}\big)\Big)}{2\pi\sqrt{y_1^2+y_2^2}\,\big|g'\big(g^{-1}\big(\sqrt{y_1^2+y_2^2}\big)\big)\big|},\ (y_1,y_2)\in G(T^\circ). \tag{2.1.98}$$

Im nächsten Abschnitt wollen wir uns nun u.a. mit den Begriffsbildungen Erwartungswert und Varianz von Verteilungen bzw. Zufallsvariablen beschäftigen, da diese Parameter nicht nur häufig Verteilungen charakterisieren, sondern auch etwa bei der Average–Case–Analyse von Algorithmen, im CAD, bei der Bildverarbeitung und anderen Bereichen der Informatik von zentraler Bedeutung sind.

2.2. Erwartungswert und Varianz

Betrachtet man ein Experiment, dessen Ausgang eine zufällige, reellwertige Größe mit nur endlich vielen möglichen Werten $\{x_1, \ldots, x_m\} \subset \mathbf{R}$ ist, und kennt man die Wahrscheinlichkeiten p_i für das Auftreten der x_i, $i = 1, \ldots, m$, so ist intuitiv klar, daß die Zahl $\sum_{i=1}^{m} x_i \cdot p_i$ das "mittlere" oder "erwartete Ergebnis" repräsentiert. Wir wollen das an der Laufzeitanalyse eines einfachen Sortieralgorithmus verdeutlichen.

Beispiel 2.2.1. (Sortieren durch Einfügen - `straightinsertion`)
Der Algorithmus `straightinsertion` vergleicht bei der Sortierung eines Feldes `a: ARRAY[1..n] OF REAL` der Länge $n \in \mathbf{N}$ in der k–ten Iteration, $k = 2, \ldots, n$, das k–te Element `a[k]` mit den vorhergehenden `a[j]`, $j = k-1, \ldots, 1$. Diese werden vom $(k-1)$–ten beginnend um einen Platz nach rechts geschoben, bis das k–te Element an der richtigen Stelle eingefügt werden kann. Wie beim Ordnen eines Blatts beim Kartenspiel werden die Elemente sukzessive durch Verschieben der bereits sortierten an ihre richtige Stelle plaziert. Nach $n-1$ Iterationen ist das resultierende Feld geordnet.
Bei einem Feld der Länge 3 gibt es $6 = 3!$ Anordnungen, die durch den oben beschriebenen Sortieralgorithmus wie folgt behandelt werden. Die Zahlen $1, 2, 3$ repräsentieren lediglich die Größe der einzelnen Elemente.

Ausgangsanord.	Anord. nach Umspeicherungen	Anz. Umspeich.
(123)		0
(132)	$\curvearrowright$ (123)	1
(213)	$\curvearrowright$ (123)	1
(231)	$\curvearrowright$ (213) $\curvearrowright$ (123)	2
(312)	$\curvearrowright$ (132) $\curvearrowright$ (123)	2
(321)	$\curvearrowright$ (231) $\curvearrowright$ (213) $\curvearrowright$ (123)	3

Unter der Annahme, daß jede Anordnung mit gleicher Wahrscheinlichkeit $1/6$ als Ausgangsanordnung auftritt, beträgt die "mittlere" oder "erwartete" Anzahl der Umspeicherungen

$$E = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} = \frac{9}{6} = 1.5 \tag{2.2.1}$$

Weiß man dagegen, daß die Anordnungen, bei denen genau zwei Elemente an der falschen Stelle stehen, mit doppelter Wahrscheinlichkeit auftreten wie die anderen, so erhalten (132), (213) und (321) je die Wahrscheinlichkeit $\frac{2}{9}$, die verbleibenden Anordnungen je die Wahrscheinlichkeit $\frac{1}{9}$. "Im Mittel" beträgt die Anzahl der Umspeicherungen dann

$$E = 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 1 \cdot \frac{2}{9} + 2 \cdot \frac{1}{9} + 2 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} = \frac{14}{9} \approx 1.556$$

In beiden Fällen wurde die mit den zugehörigen Wahrscheinlichkeiten gewichtete Summe der möglichen Werte gebildet. ∎

Obiges Beispiel wird nun vollständig in ein wahrscheinlichkeitstheoretisches Modell eingebettet. Wir wählen

$$\begin{aligned}\Omega &= \{(123),(132),(213),(231),(312),(321)\}\\ &= Perm_n^n(\{1,2,3\}; o.W.)\end{aligned}$$

in der Schreibweise von Definition 1.1.3 und als zugehörige σ–Algebra $\mathcal{A} = \mathfrak{P}(\Omega)$. Die Abbildung $X : \Omega \to \mathbf{R}$ ordne jeder Anfangsanordnung die Anzahl der nötigen Umspeicherungen unter `straightinsertion` bis zum Erreichen einer aufsteigenden Sortierung zu. Sie ist explizit in der Tabelle angegeben mit den Urbildern $\omega \in \Omega$ in der linken Spalte und Werten $x \in \mathbf{R}$ in der rechten.

Die Verteilung P^X der Zufallsvariablen X besitzt den Träger $T = \{0,1,2,3\}$, und es gilt

$$\begin{aligned}P^X(\{0\}) &= P(X=0) = P(\{(123)\})\\ P^X(\{1\}) &= P(X=1) = P(\{(132),(213)\})\\ P^X(\{2\}) &= P(X=2) = P(\{(231),(312)\})\\ P^X(\{3\}) &= P(X=3) = P(\{(321)\})\end{aligned}$$

Unter der Annahme einer Gleichverteilung auf der Menge der möglichen Anfangsverteilungen ergeben sich konkret die Werte $P^X(\{0\}) = 1/6$, $P^X(\{1\}) = 2/6$, $P^X(\{2\}) = 2/6$, $P^X(\{3\}) = 1/6$, womit die Verteilung P^X vollständig festgelegt ist.

Für eine beliebige endlich–diskrete Wahrscheinlichkeitsverteilung P auf $(\Omega, \mathcal{A})$ berechnen wir wie oben die mittlere oder erwartete Anzahl E der Umspeicherungen wieder als das gewichtete Mittel aller möglichen Ergebnisse, wobei die Gewichte gerade die zugehörigen Wahrscheinlichkeiten sind, daß ein Ergebnis von der Zufallsvariablen X realisiert wird. Mit $x_0 = 0$, $x_1 = 1$, $x_2 = 2$ und $x_3 = 3$ folgt

$$\begin{aligned}E &= \sum_{i=0}^{3} x_i \cdot P(X = x_i)\\ &= 0 \cdot P(X = x_0) + 1 \cdot P(X = x_1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3)\end{aligned} \tag{2.2.2}$$

Im Fall einer Gleichverteilung auf der Menge der möglichen Anfangsanordnungen erhält man speziell

$$E = 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = 1.5\ ,$$

also genau die in (2.2.1) ausgewertete Zahl.

Offensichtlich hängt die Zahl E nur von der Verteilung der Zufallsvariablen X ab. Die bisherigen Überlegungen werden deshalb allgemein in der folgenden Definition zusammengefaßt.

Definition 2.2.1. *(Erwartungswert für diskrete Zufallsvariable)*
Sei X eine diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbf{R}$ und endlichem Träger $T = \{x_1, \ldots, x_m\} \subset \mathbf{R}$, $m \in \mathbf{N}$. Dann heißt

$$E(X) = \sum_{i=1}^{m} x_i \cdot P(X = x_i) \tag{2.2.3}$$

der Erwartungswert von X.

Beispiel 2.2.2. (`binary search`)
Wir betrachten die Situation des binären Suchens aus Beispiel 1.1.1 bzw. Beispiel 1.3.1, d.h. die Beschreibung der Anzahl der Schritte bis zum Abbruch des Verfahrens durch die Zufallsvariable X aus (1.3.5) mit der in (1.3.6) gegebenen Verteilung. Hier ist $T = \{1, \ldots, n\}$ mit

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} i \cdot P(X = i) = n2^{-n} + \sum_{i=1}^{n} i \cdot 2^{i-1-n} \\ &= n2^{-n} + \sum_{1 \le j \le i \le n} \sum 2^{i-1-n} = n2^{-n} + \sum_{j=1}^{n} 2^{-n-1} \sum_{i=j}^{n} 2^i \\ &= n2^{-n} + \sum_{j=1}^{n} 2^{-n-1}\left(2^{n+1} - 2^j\right) = n2^{-n} + \sum_{j=1}^{n} \left(1 - 2^{j-1-n}\right) \\ &= n2^{-n} + n - \left(1 - 2^{-n}\right) = n - 1 + \frac{n+1}{2^n}, \end{aligned}$$

d.h. die erwartete Schrittzahl bis zum Abbruch des Algorithmus ist — für große n — praktisch nur um 1 besser als im schlechtesten Fall. ∎

Gelegentlich ist es notwendig, Erwartungswerte auch unter der Bedingung zu berechnen, daß eine bestimmtes Ereignis $B \in \mathcal{A}$ mit $P(B) > 0$ eintritt, wie die Ausführungen zum binären Suchen am Ende des Abschnitts 1.3 zeigen (Anzahl der Schritte bis zum Auffinden des Schlüsselelements unter der Bedingung, daß dieses in dem Feld tatsächlich vorhanden ist). Die entsprechende Begriffsbildung des bedingten Erwartungswerts läßt sich dann völlig analog mit der bedingten Wahrscheinlichkeit formulieren.

Definition 2.2.2. *(elementarer bedingter Erwartungswert)*
Es sei $B \in \mathcal{A}$ ein Ereignis mit $P(B) > 0$. Dann heißt unter den Voraussetzungen von Definition 2.2.1 die Zahl

$$E(X \mid B) = \sum_{i=1}^{m} x_i \cdot P(X = x_i \mid B) \tag{2.2.4}$$

der (elementare) bedingte Erwartungswert von X unter (der Hypothese) B.

Der elementare bedingte Erwartungswert ist also nichts anderes als der Erwartungswert bezüglich der bedingten Verteilung $P(X \in \cdot \mid B)$.

Eine analoge Rechnung wie in Beispiel 2.2.1 zeigt nun, daß die erwartete Anzahl der Schritte bis zum Auffinden des gesuchten Elements unter der Bedingung, daß dieses in dem Feld vorhanden ist, mit dem Bezeichnungen aus (1.3.7) bis (1.3.9) ausgedrückt werden kann als

$$E(Y \mid 1 \le Y \le n) = \sum_{i=1}^{n} i \frac{2^{i-1}}{2^n - 1} = n - 1 + \frac{n}{2^n - 1} \tag{2.2.5}$$

(vgl. auch Knuth (1973), S. 410ff, wo dieses Problem graphentheoretisch behandelt wird, oder Aigner (1988), Kapitel 1).

Auf allgemeinere bedingte Erwartungswerte werden wir später in Kapitel 3 noch einmal ausführlicher eingehen.

Natürlich möchte man den Begriff des Erwartungswerts nicht nur für endlich diskrete Zufallsvariablen zur Verfügung haben. Eine direkte Verallgemeinerung von (2.2.3) bietet sich bei diskreten Zufallsvariablen mit abzählbarem Träger $T = \{x_1, x_2, \dots\}$ an. Man könnte hier einfach $E(X) = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i)$ setzen. Allerdings bringt auch das schon gewisse Schwierigkeiten mit sich, da die unendliche Reihe nicht konvergent oder absolut konvergent sein muß und der Grenzwert, falls existent, von der Summationsreihenfolge abhängen kann.

Wir betrachten zum Beispiel für festes $0 < p < 1$

$$T = \left\{ x_i = \frac{(-1)^i}{ip(1-p)^{i-1}} \mid i \in \mathbf{N} \right\} \quad \text{und} \quad P(X = x_i) = p(1-p)^{i-1},\ i \in \mathbf{N}. \tag{2.2.6}$$

Dann ist $\sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} \frac{(-1)^i}{i}$ konvergent aber nicht absolut konvergent, da die harmonische Reihe $\sum_{i=1}^{\infty} 1/i$ bekanntlich divergiert.

Zur Einführung eines allgemeinen Erwartungswertbegriffs gehen wir nun in den folgenden zwei Schritten vor. Die Beweise einiger technischer Sachverhalte werden wir hierbei auslassen.

1. Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbf{R}$, $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ in der Schreibweise von Definition 1.3.1. Ferner gelte $X(\omega) \geq 0$ für alle $\omega \in \Omega$. Dann existiert eine Folge von diskreten Zufallsvariablen $\{X_n\}_{n \in \mathbf{N}}$ mit endlichem Träger, $X_n : (\Omega, \mathcal{A}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$, mit $X_n(\omega) \uparrow X(\omega)$ für $n \to \infty$. Wähle etwa

$$\begin{aligned} X_n(\omega) &= \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathbb{1}_{A_{i,n}}(\omega) + n \mathbb{1}_{\{X(\omega) > n\}}(\omega), \\ &\text{wobei } A_{i,n} = \left\{ \frac{i-1}{2^n} < X(\omega) \leq \frac{i}{2^n} \right\} \in \mathcal{A},\ 1 \leq i \leq n2^n. \end{aligned} \tag{2.2.7}$$

$X_n(\omega)$ ist lediglich der Werte $\{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, n\}$ fähig und approximiert $X(\omega)$ punktweise von unten.

Wir setzen

$$E(X) = \lim_{n \to \infty} E(X_n), \tag{2.2.8}$$

wobei $E(X_n)$ auf der rechten Seite durch Definition 2.2.1 bestimmt ist. Der Träger von X_n ist dabei gegeben durch diejenigen Zahlen $\frac{i}{2^n}, 0 \leq i \leq n2^n$, für die $P(A_{i+1,n}) > 0$ ist (mit $A_{n2^n+1,n} = \{X > n\}$). Es läßt sich zeigen, daß die Folge $\{E(X_n)\}_{n \in \mathbf{N}}$ monoton wächst und damit (eventuell uneigentlich mit Wert ∞) konvergiert. Ferner ist der Grenzwert in (2.2.8) unabhängig von der speziellen Gestalt der approximierenden Folge $\{X_n\}$. Ist nämlich $\{Y_n\}$

eine weitere Folge von diskreten Zufallsvariablen mit endlichem Träger und gilt $Y_n \uparrow X$ $(n \to \infty)$, so ist $\lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} E(Y_n)$.

Die Darstellung (2.2.7) etwa führt zu

$$E(X) = \lim_{n\to\infty} \Big\{ \sum_{i=1}^{n2^n} \frac{i-1}{2^n} P\Big(\frac{i-1}{2^n} < X \le \frac{i}{2^n}\Big) + nP(X > n) \Big\} \tag{2.2.9}$$

$$= \lim_{n\to\infty} \Big\{ \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \left(F\left(\frac{i}{2^n}\right) - F\left(\frac{i-1}{2^n}\right)\right) + n\,(1 - F(n)) \Big\},$$

wobei F die Verteilungsfunktion von X bezeichne.

2. X sei eine (vorzeichenunbeschränkte) Zufallsvariable, $X : (\Omega, \mathcal{A}) \to (\mathbf{R}, \mathcal{B}^1)$. Dann sind

$$X^+ = \max\{X, 0\} \quad \text{und} \quad X^- = \max\{-X, 0\}$$

wieder Zufallsvariable. X^+ (X^-) heißt Positiv- (Negativ-)teil der Zufallsvariablen X. Er entsteht, indem man die Abbildung X (das Negative von X) "nach unten bei Null abschneidet". Die Zufallsvariable X läßt sich durch $X = X^+ - X^-$ wieder einfach zusammensetzen.
X^+ und X^- sind beide nicht-negativ und ihre Erwartungswerte $E(X^+)$ und $E(X^-)$ existieren nach (2.2.8), sind jedoch unter Umständen nicht endlich.

Definition 2.2.3. *(Erwartungswert von Zufallsvariablen)*
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ eine Zufallsvariable, $X = X^+ - X^-$. Gilt $E(X^+) < \infty$ und $E(X^-) < \infty$, so heißt X integrierbar bezüglich P oder kurz: P-integrierbar. In diesem Fall heißt $E(X) = E(X^+) - E(X^-)$ der Erwartungswert oder das Lebesgue-Integral von X bezüglich P.

Wir bezeichnen im folgenden synonym

$$E(X) = E_P(X) = \int X\, dP = \int X(\omega)\, dP(\omega), \tag{2.2.10}$$

je nachdem, welche Schreibweise die angemessene ist.

Aus der Darstellung $|X| = X^+ + X^-$ und der Definition des Integrals folgt, daß X genau dann integrierbar ist, wenn $|X|$ diese Eigenschaft besitzt.

Ist nun genau eine der Zahlen $E(X^+)$ oder $E(X^-)$ unendlich, so ist $E(X) = E(X^+) - E(X^-)$ immer noch mit Wert $+\infty$ oder $-\infty$ definiert. In diesem Fall nennen wir X quasi-integrierbar. Im Fall, daß beide Integrale $E(X^+)$ und $E(X^-)$ unendlich sind, kann man der Zufallsvariablen X keinen Erwartungswert zuordnen. Ein Beispiel hierfür wurde in (2.2.6) vorgestellt.

Definition 2.2.3 ist zwar sehr allgemein, jedoch über den Approximationsprozeß (2.2.7) mit endlich-diskreten Zufallsvariablen für Positiv- und Negativteil nur schwer zur expliziten Bestimmung von Erwartungswerten auszuwerten. Der folgende Satz zeigt, wie man im allgemeinen für beliebige diskrete bzw. absolutstetige Verteilungen Erwartungswerte einfach ausrechnen kann.

Satz 2.2.1. *(Berechnung von Erwartungswerten)*
Es sei X eine reelle Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

a) *Ist X diskret verteilt mit Zähldichte f, so ist X genau dann P–integrierbar, wenn*

$$\sum_{x \in \mathbf{R}} |x| f(x) < \infty \tag{2.2.11}$$

gilt. In diesem Fall ist

$$E(X) = \sum_{x \in \mathbf{R}} x f(x). \tag{2.2.12}$$

Ist darüberhinaus P selbst eine diskrete Verteilung, so gilt auch

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}). \tag{2.2.13}$$

b) *Ist X absolut–stetig verteilt mit einer Dichte f, so ist X genau dann P–integrierbar, wenn*

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty \tag{2.2.14}$$

gilt. In diesem Fall ist

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx. \tag{2.2.15}$$

c) *Bezeichnet F die Verteilungsfunktion von X und ist X P–integrierbar, so gilt in jedem Fall auch die Darstellung*

$$E(X) = -\int_{-\infty}^{0} F(x)\, dx + \int_{0}^{\infty} (1 - F(x))\, dx. \tag{2.2.16}$$

Existiert umgekehrt jedes der beiden letzteren Integrale, so ist X P–integrierbar, wobei der Erwartungswert $E(X)$ wiederum durch die Beziehung (2.2.16) gegeben ist.
Gilt $P(X \in \mathbf{Z}) = 1$, d.h. nimmt X fast sicher nur ganzzahlige Werte an, so läßt sich (2.2.16) auch einfacher schreiben als

$$E(X) = -\sum_{k=1}^{\infty} P(X \leq -k) + \sum_{k=0}^{\infty} P(X > k). \tag{2.2.17}$$

Beweis. a) Es sei $T^+ = \{x_1, x_2, \ldots\}$ der (abzählbare) Träger von X^+ und $T^- = \{y_1, y_2, \ldots\}$ der (abzählbare) Träger von X^-. Wir setzen

$$X_n^+ = \sum_{k=1}^{n} x_k \mathbb{1}_{\{X^+ = x_k\}}, \quad X_n^- = \sum_{k=1}^{n} y_k \mathbb{1}_{\{X^- = y_k\}}, \quad n \in \mathbf{N},$$

d.h. auf den Mengen $\{X^+ \in \{x_1, \ldots, x_n\}\}$ bzw. $\{X^- \in \{y_1, \ldots, y_n\}\}$ stimmen X_n^+ und X^+ bzw. X_n^- und X^- überein; ferner gilt jeweils

$$X_n^+ \uparrow X^+, \quad X_n^- \uparrow X^- \quad (n \to \infty).$$

Gemäß Definition 2.2.1 berechnen sich die entsprechenden Erwartungswerte zu

$$E(X_n^+) = \sum_{k=1}^{n} x_k f(x_k), \quad E(X_n^-) = \sum_{k=1}^{n} y_k f(y_k);$$

existiert nun $E(X)$, so konvergieren nach Voraussetzung beide Reihen mit

$$\lim_{n\to\infty} E(X_n^+) = \sum_{x\geq 0} x f(x) = E(X^+), \quad \lim_{n\to\infty} E(X_n^-) = -\sum_{x\leq 0} x f(x) = E(X^-).$$

Daher konvergiert auch $\sum_{x\in\mathbf{R}} |x| f(x)$ mit Wert $E(|X|) = E(X^+) + E(X^-)$.
Ist umgekehrt die letztere Reihe konvergent, so konvergieren jeweils auch die Reihen $\sum_{x\geq 0} x f(x)$ und $\sum_{x\leq 0} x f(x)$, d.h. es existieren jeweils $E(X^+)$ und $E(X^-)$ und damit auch $E(X)$.
Ist nun P diskret, so läßt sich wegen der absoluten Konvergenz der Reihe in (2.2.12) diese auch folgendermaßen umordnen:

$$\begin{aligned} E(X) &= \sum_{x\in\mathbf{R}} x P(X=x) = \sum_{x\in\mathbf{R}} x \sum_{\omega: X(\omega)=x} P(\{\omega\}) \\ &= \sum_{x\in\mathbf{R}} \sum_{\omega: X(\omega)=x} X(\omega) P(\{\omega\}) = \sum_{\omega\in\Omega} X(\omega) P(\{\omega\}); \end{aligned}$$

dies ist gerade (2.2.13).
b) Für die Zufallsvariable X^+ sei X_n^+ in der zu (2.2.7) analogen Darstellung gegeben. Wie in (2.2.9) ergibt sich dann

$$\begin{aligned} E(X_n^+) &= \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(x)\,dx \;+\; n\int_n^\infty f(x)\,dx \\ &\leq \sum_{i=1}^{n2^n} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} x f(x)\,dx \;+\; \int_n^\infty x f(x)\,dx = \int_0^\infty x f(x)\,dx. \end{aligned}$$

Aus der für alle $n \in \mathbf{N}$ gültigen Abschätzung

$$\begin{aligned} E(X_n^+) &\geq \sum_{i=1}^{n2^n} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} x f(x)\,dx - \sum_{i=1}^{n2^n} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} \left(x - \frac{i-1}{2^n}\right) f(x)\,dx \\ &\geq \int_0^n x f(x)\,dx - \frac{1}{2^n} \sum_{i=1}^{n2^n} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(x)\,dx \\ &= \int_0^n x f(x)\,dx - \frac{1}{2^n} \int_0^n f(x)\,dx \geq \int_0^n x f(x)\,dx - \frac{1}{2^n} \end{aligned}$$

folgt ferner

$$\lim_{n\to\infty} E(X_n^+) = \int_0^\infty x f(x)\,dx = E(X^+).$$

Eine analoge Rechnung für X^- ergibt die Darstellung

$$E(X^-) = -\int_{-\infty}^{0} x f(x)\,dx;$$

durch entsprechende Argumentation wie im Fall a) folgt hieraus die Behauptung.
c) Wir zeigen zunächst:

$$E(X^+) = \int_0^\infty (1 - F(x))\,dx, \quad E(X^-) = \int_{-\infty}^{0} F(x)\,dx. \tag{2.2.18}$$

Durch Umordnung der Reihe in (2.2.9) erhält man für X^+ aufgrund der Monotonie von F

$$\begin{aligned} E(X^+) &= \lim_{n\to\infty} \Big\{ -\sum_{i=1}^{n2^n} \frac{i-1}{2^n} \left(1 - F\left(\frac{i}{2^n}\right)\right) \\ &\quad + \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \left(1 - F\left(\frac{i-1}{2^n}\right)\right) + n(1 - F(n)) \Big\} \\ &= \lim_{n\to\infty} \frac{1}{2^n} \sum_{i=1}^{n2^n} \left(1 - F\left(\frac{i}{2^n}\right)\right) \le \lim_{n\to\infty} \int_0^n (1 - F(x))\,dx \\ &= \int_0^\infty (1 - F(x))\,dx \end{aligned}$$

sowie entsprechend

$$E(X^+) \ge \lim_{n\to\infty} \int_{\frac{1}{2^n}}^{n+\frac{1}{2^n}} (1 - F(x))\,dx = \int_0^\infty (1 - F(x))\,dx.$$

Dies ergibt die erste Beziehung in (2.2.18); die entsprechende Aussage für X^- erhält man völlig analog. Die ersten beiden Aussagen des Teils c) erhält man nun mit derselben Argumentation wie in a) oder b). (2.2.17) folgt aus (2.2.16) durch Summation und der Konstanz von F auf $(k, k+1)$, $k \in \mathbf{Z}$, d.h.

$$\int_{-(k+1)}^{-k} F(x)\,dx = F(-k) = P(X \le -k)$$

sowie

$$\int_k^{k+1} (1 - F(x))\,dx = 1 - F(k+1) = P(X > k), \quad k \in \mathbf{N}_0,$$

womit der Satz bewiesen ist. ∎

Beziehung (2.2.13) motiviert auch die Schreibweise (2.2.10) für den Erwartungswert als abstraktes Integral, welches ja üblicherweise in der Analysis als stilisiertes Summenzeichen aufgefaßt wird (man denke etwa an den Riemann'schen Integrationszugang über Ober- und Untersummen).

Für den Fall, daß die Verteilungsfunktion F überall differenzierbar (d.h. die Dichte f stetig) ist, kann man Beziehung (2.2.18) und damit Teil c) des Satzes 2.2.1 auch direkt durch partielle Integration aus Beziehung (2.2.15) erhalten: Es ist nämlich bei Existenz von $E(X)$

$$\begin{aligned} E(X^+) &= \int_0^\infty x f(x)\,dx = -\int_0^\infty x \frac{d}{dx}(1-F(x))\,dx \\ &= -x(1-F(x))\Big|_0^\infty + \int_0^\infty (1-F(x))\,dx \\ &= \int_0^\infty (1-F(x))\,dx \end{aligned}$$

wegen $0 \le \lim_{x\to\infty} x(1-F(x)) \le \lim_{x\to\infty} \int_x^\infty y f(y)\,dy = 0$ sowie analog

$$\begin{aligned} E(X^-) &= -\int_{-\infty}^0 x f(x)\,dx = -\int_{-\infty}^0 x \frac{d}{dx} F(x)\,dx \\ &= -xF(x)\Big|_{-\infty}^0 + \int_{-\infty}^0 F(x)dx \\ &= \int_{-\infty}^0 F(x)\,dx \end{aligned}$$

wegen $0 \ge \lim_{x\to-\infty} xF(x) \ge \lim_{x\to-\infty} \int_{-\infty}^x y f(y)\,dy = 0$.

Aus den obigen Ausführungen ergibt sich unmittelbar, daß der Erwartungswert $E(X)$ einer reellen Zufallsvariablen X nicht von ihrer speziellen Definition als meßbare Abbildung auf $(\Omega, \mathcal{A}, P)$ abhängt, sondern lediglich von ihrer Verteilung P^X. Man sagt daher auch gelegentlich, die *Verteilung* P^X habe den Erwartungswert $E(X)$. Insbesondere ist die Erwartungswertbildung invariant gegenüber (meßbaren) Abänderungen der Zufallsvariablen X auf P-Nullmengen, d.h. gilt $X = Y$ f.s. für eine weitere Zufallsvariable Y auf $(\Omega, \mathcal{A}, P)$, so ist im Falle der Existenz $E(X) = E(Y)$.

Die in (1.2.18) beschriebene Verteilung über $\mathbb{Q} \cap (0,1]$ zeigt, daß es manchmal schwierig, wenn nicht unmöglich ist, den Erwartungswert selbst einer diskreten Verteilung direkt über die Gleichung (2.2.12) (bzw. (2.2.15) im Fall stetiger Verteilungen) zu bestimmen. In dem angesprochenen Beispiel läßt sich dies aber recht einfach vermöge Teil c) des letzten Satzes bzw. Beziehung (2.2.18) bewerkstelligen, denn es gilt

$$\int_0^1 \lfloor nx \rfloor\,dx = \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \lfloor nx \rfloor\,dx = \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} (k-1)\,dx = \sum_{k=1}^n \frac{k-1}{n} = \frac{n-1}{2},$$

so daß sich für den Erwartungswert nach (1.2.18) ergibt

$$\begin{aligned} \int_0^1 (1-F(x))\,dx &= 1 - \int_0^1 \sum_{n=1}^\infty \frac{\lfloor nx \rfloor}{n2^n}\,dx = 1 - \sum_{n=1}^\infty \int_0^1 \frac{\lfloor nx \rfloor}{n2^n}\,dx \\ &= 1 - \frac{1}{2}\sum_{n=1}^\infty \frac{n-1}{n2^n} = \frac{1}{2} + \frac{1}{2}\sum_{n=1}^\infty \frac{1}{n2^n} = \frac{1}{2} + \frac{1}{2}\ln 2. \end{aligned}$$

Weitere charakteristische Eigenschaften des Erwartungswerts sind in dem folgenden Satz zusammengefaßt.

Satz 2.2.2. *(Eigenschaften des Erwartungswerts)*
$X, Y, \{X_n\}_{n\in\mathbf{N}}$ seien Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Es gilt:

a) Ist X P–integrierbar, so läßt sich $E(X)$ auch darstellen als

$$E(X) = \int_0^1 F^{-1}(x)\,dx, \tag{2.2.19}$$

wobei F^{-1} wieder die Pseudo–Inverse von F bezeichne.

b) Ist X P–integrierbar, so gilt

$$|E(X)| \leq E(|X|). \tag{2.2.20}$$

c) Sind jeweils X und Y P–integrierbar, so ist

$$E(aX + bY) = aE(X) + bE(Y) \text{ für alle } a, b \in \mathbf{R} \quad (\textit{Linearität}). \tag{2.2.21}$$

d) Unter den Voraussetzungen von c) gilt:

$$X \leq Y \text{ f.s.} \implies E(X) \leq E(Y) \quad (\textit{Monotonie}). \tag{2.2.22}$$

Wird lediglich Y als P–integrierbar vorausgesetzt und ist $0 \leq X \leq Y$ f.s, so ist auch X P–integrierbar, und es gilt wieder (2.2.22).

e) Sind alle X_n nicht–negativ, so gilt für $n \to \infty$:

$$X_n \uparrow X \text{ f.s.} \implies E(X_n) \uparrow E(X) \quad (\textit{monotone Konvergenz}). \tag{2.2.23}$$

f) Ist Y P–integrierbar und gilt $|X_n| \leq Y$ für alle $n \in \mathbf{N}$, so gilt für $n \to \infty$:

$$X_n \longrightarrow X \text{ f.s.} \implies X \text{ integrierbar und } E\big(|X_n - X|\big) \longrightarrow 0 \tag{2.2.24}$$

(majorisierte Konvergenz); insbesondere gilt dann auch $E(X_n) \longrightarrow E(X)$.

g) Ist $X = \mathbb{1}_A$ für eine Menge $A \in \mathcal{A}$, so gilt

$$E(X) = E(\mathbb{1}_A) = P(A). \tag{2.2.25}$$

h) Ist X P–integrierbar, so gilt für beliebige $c > 0$:

$$P\big(|X| > c\big) \leq \frac{E(|X|)}{c} \quad (\textit{Markoff - Ungleichung}). \tag{2.2.26}$$

i) Sind X und Y P–integrierbar, so gilt:

$$X \text{ und } Y \text{ stochastisch unabhängig} \implies E(XY) = E(X)E(Y). \tag{2.2.27}$$

j) Ist X P–integrierbar, so gilt:

$$X \geq 0 \text{ und } E(X) = 0 \implies X = 0 \text{ f.s.} \tag{2.2.28}$$

Beweis. a) Sei zunächst $X \geq 0$. Dann ist auch $F^{-1} \geq 0$, und man erhält

$$\int_0^1 F^{-1}(x)\,dx = \int_0^1 \int_0^{F^{-1}(x)} dy\,dx = \iint\limits_{\substack{0\leq y\leq F^{-1}(x)\\ 0<x<1}} dy\,dx$$

$$= \iint\limits_{\substack{F(y)\leq x<1\\ y\geq 0}} dx\,dy = \int_0^\infty (1-F(x))\,dx = E(X).$$

Ist $X \leq 0$, so ergibt sich analog

$$\int_0^1 F^{-1}(x)\,dx = -\int_0^1 \int_{F^{-1}(x)}^0 dy\,dx = -\iint\limits_{\substack{F^{-1}(x)\leq y\leq 0\\ 0<x<1}} dy\,dx$$

$$= -\iint\limits_{\substack{0<x\leq F(y)\\ y\leq 0}} dx\,dy = -\int_{-\infty}^0 F(x)\,dx = E(X).$$

Für beliebige X erhält man nun das Ergebnis durch die Zerlegung $X = X^+ - X^-$ mittels Satz 2.2.1 c) bzw. den Beziehungen (2.2.20) und (2.2.21).
b) Es ist

$$|E(X)| = \left|E(X^+) - E(X^-)\right| \leq E(X^+) + E(X^-) = E\left(|X|\right).$$

c) Seien zunächst $a = b = 1$ und X und Y endlich-diskrete Zufallsvariablen mit einer gemeinsamen Zähldichte f. Dann ist auch $X + Y$ eine endlich-diskrete Zufallsvariable. Mit Lemma 2.1.4 und der Definition 2.2.1 des Erwartungswerts erhält man in diesem Fall

$$\begin{aligned} E(X+Y) &= \sum_{z\in\mathbf{R}} z \sum_{x\in\mathbf{R}} f(x,z-x) = \sum_{x\in\mathbf{R}}\sum_{z\in\mathbf{R}} z f(x,z-x) \\ &= \sum_{x\in\mathbf{R}}\sum_{y\in\mathbf{R}} (x+y) f(x,y) = \sum_{x\in\mathbf{R}} x \sum_{y\in\mathbf{R}} f(x,y) + \sum_{y\in\mathbf{R}} y \sum_{x\in\mathbf{R}} f(x,y) \\ &= \sum_{x\in\mathbf{R}} x P(X=x) + \sum_{y\in\mathbf{R}} y P(Y=y) = E(X) + E(Y) \end{aligned}$$

(man beachte, daß in allen Summen nur *endlich viele* von Null verschiedene Summanden enthalten sind). Sind nun X und Y beliebige, nicht-negative Zufallsvariablen und $\{X_n\}_{n\in\mathbf{N}}$ bzw. $\{Y_n\}_{n\in\mathbf{N}}$ X bzw. Y monoton approximierende Folgen endlich-diskreter Zufallsvariablen, so gilt allgemeiner ebenfalls

$$E(X+Y) = \lim_{n\to\infty} E(X_n + Y_n) = \lim_{n\to\infty} E(X_n) + \lim_{n\to\infty} E(Y_n) = E(X) + E(Y).$$

Sind Xund Y schließlich beliebige P–integrierbare Zufallsvariablen, so erhält man hieraus durch die Zerlegung in Positiv– und Negativteil wieder das gewünschte Ergebnis.
Für beliebige a, b argumentiert man analog; ist etwa X wieder endlich–diskret verteilt, erhält man unmittelbar aus Definition 2.2.1

$$E(aX) = \sum_{x \in \mathbf{R}} axP(X = x) = a \sum_{x \in \mathbf{R}} xP(X = x) = aE(X).$$

Durch monotone Approximation und Zerlegung in Positiv– und Negativteil läßt sich dann die letztere Gleichheit auch für beliebige P–integrierbare Zufallsvariablen X zeigen.
d) Bezeichnen F_X und F_Y die Verteilungsfunktionen von X und Y, so folgt aus $X \leq Y$ f.s. sofort auch $\{Y \leq x\} \subseteq \{X \leq x\}$ für alle $x \in \mathbf{R}$ und somit $F_Y(x) = P(Y \leq x) \leq P(X \leq x) = F_X(x)$, $x \in \mathbf{R}$. Die erste Behauptung ergibt sich dann aus Satz 2.2.1 c).
Die zweite Behauptung sieht man wie folgt ein: Ist $\{X_n\}_{n \in \mathbf{N}}$ eine monotone Folge nicht–negativer endlich–diskreter Zufallsvariablen mit $X_n \uparrow X$ $(n \to \infty)$, so gilt jedenfalls auch $E(X_n) \uparrow E(X)$ $(n \to \infty)$. Wegen $X_n \leq X \leq Y$ f.s., $n \in \mathbf{N}$, ist nach der ersten Behauptung aber $E(X_n) \leq E(Y)$ für alle $n \in \mathbf{N}$ und daher auch $E(X) \leq E(Y)$, also insbesondere $E(X)$ endlich und somit X P–integrierbar.
e) Zu jeder Zufallsvariablen X_n, $n \in \mathbf{N}$, exisistiert eine nicht–negative monotone Folge $\{Y_{nm}\}_{m \in \mathbf{N}}$ endlich–diskreter Zufallsvariablen mit $Y_{nm} \uparrow X_n$ $(m \to \infty)$. Wählt man $Z_m := \max(Y_{1m}, \ldots, Y_{mm})$, $m \in \mathbf{N}$, so bildet $\{Z_m\}_{m \in \mathbf{N}}$ ebenfalls eine nicht–negative monotone Folge endlich–diskreter Zufallsvariablen, für die man leicht auch $Z_m \uparrow X$ $(m \to \infty)$ nachweist. Somit gilt jedenfalls $E(Z_m) \uparrow E(X)$ $(m \to \infty)$. Wegen $Z_m \leq X_m$ und der Monotonie des Erwartungswerts (Teil d) dieses Satzes) folgt damit die Behauptung.
f) Es bezeichne $D_n := |X_n - X|$, $n \in \mathbf{N}$. Dann ist nach Voraussetzung $D_n \leq Z := Y + |X| \leq 2Y$ f.s. für alle $n \in \mathbf{N}$, also nach Teil d) dieses Satzes jedes D_n P–integrierbar mit $E(D_n) \leq E(Z) \leq 2E(Y)$. Setzt man ferner $I_n := \inf_{m \geq n}(Z - D_m)$, $n \in \mathbf{N}$, so ist $\{I_n\}_{n \in \mathbf{N}}$ monoton wachsend mit $I_n \uparrow Z$ $(n \to \infty)$ f.s., da nach Voraussetzung D_n f.s. gegen 0 strebt. Es folgt $E(Z) - E(D_n) \geq E(I_n) \uparrow E(Z)$, $n \to \infty$; d.h. es ist $\lim_{n \to \infty} E(D_n) = 0$, wie behauptet. Die zweite Behauptung ergibt sich hieraus sofort vermöge der Abschätzung $|E(X_n) - E(X)| \leq E(|X_n - X|) = E(D_n)$, $n \in \mathbf{N}$ (Teil b)).
g) Dies folgt direkt aus Definition 2.2.1 wegen $\{X = 1\} = A$, $\{X = 0\} = A^c$, also $P(X = 1) = P(A)$, $P(X = 0) = P(A^c)$.
h) Wir setzen

$$Y = c \cdot \mathbb{1}_{\{|X| > c\}} = \begin{cases} c, & |X| > c \\ 0, & |X| \leq c. \end{cases}$$

Dann ist $Y \leq |X|$, also nach Teil d) dieses Satzes auch $E(Y) \leq E(|X|)$. Nach dem gerade gezeigten Teil g) sowie aufgrund der Linearität (Teil c)) ist aber $E(Y) = cP(Y = 1) = cP(|X| > c)$, woraus die Behauptung folgt.
i) Seien zunächst wieder X und Y endlich–diskrete Zufallsvariablen mit Zähldichten f_X und f_Y. Dann ist auch XY eine endlich–diskrete Zufallsvariable, und

gemäß Definition 2.2.1 ergibt sich

$$E(XY) = \sum_{x\in\mathbf{R}}\sum_{y\in\mathbf{R}} xyf_X(x)f_Y(y)$$
$$= \sum_{x\in\mathbf{R}} xf_X(x)\sum_{y\in\mathbf{R}} yf_Y(y) = E(X)E(Y).$$

Durch monotone Approximation mit endlich–diskreten Zufallsvariablen und Zerlegung in Positiv– und Negativteil läßt sich dann das gewünschte Ergebnis auch allgemein zeigen.

j) Nach Teil h) folgt für alle $c > 0$: $P(|X| > c) \leq \frac{E(X)}{c} = 0$, also $P(|X| > c) = 0$ für alle $c > 0$ und somit $P(X = 0) = 1$, wie behauptet. ∎

Teil a) des Satzes 2.2.2 läßt sich auch mit Hilfe der Substitutionsregel für Integrale beweisen, wenn die Zufallsvariable X eine stetige, positive Dichte f besitzt; substituiert man nämlich in dem Integral in (2.2.19) $y = F^{-1}(x)$, also $F(y) = x$, so geht dieses über in $\int_{-\infty}^{\infty} yf(y)\,dy$, was ja nach (2.2.15) mit $E(X)$ übereinstimmt.

Die Aussage des Teils f) des letzten Satzes ist auch als *Satz von Lebesgue* bekannt; eine etwas allgemeinere Formulierung findet man etwa in Bauer (1978), Satz 15.4. Auf die einschränkenden Voraussetzungen an die Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ kann dabei i.a. nicht verzichtet werden, wie das folgende Beispiel zeigt:
Man wähle z.B. den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P) = \big((0,1],\ (0,1]\cap\mathcal{B}^1,\ P\big)$ mit der Gleichverteilung P über $\Omega = (0,1]$ und als Zufallsvariablen die Abbildungen $X_n = n \cdot \mathbb{1}_{(0,1/n]}$, $n \in \mathbf{N}$, d.h. $\frac{X_n}{n}$ ist $\mathfrak{B}(1, \frac{1}{n})$–verteilt. Dann konvergiert für jedes $\omega \in (0,1]$ die Folge $\{X_n(\omega)\}$ gegen 0, d.h. es gilt $X_n \longrightarrow X \equiv 0$ $(n \to \infty)$. Jedoch ist $E(X_n) = n \cdot \frac{1}{n} = 1$ für alle $n \in \mathbf{N}$, so daß $E(X_n) \longrightarrow 1 \neq 0 = E(X)$ $(n \to \infty)$ gilt. Die Folge $\{X_n\}$ läßt sich also nicht durch eine P–integrierbare Zufallsvariable Y majorisieren; die kleinste solche Zufallsvariable Y ist nämlich gegeben durch

$$Y = \sup_{n\in\mathbf{N}} X_n = \sum_{k=1}^{\infty} k\mathbb{1}_{(\frac{1}{k+1},\frac{1}{k}]}$$

mit

$$E(Y) = \sum_{k=1}^{\infty} kP\left(\left(\frac{1}{k+1}, \frac{1}{k}\right]\right) = \sum_{k=1}^{\infty} k\frac{1}{k(k+1)} = \sum_{k=2}^{\infty}\frac{1}{k} = \infty,$$

d.h. Y ist nicht P–integrierbar.

Die in Abschnitt 1.5 betrachteten Transformationen G von Zufallsvariablen und Zufallsvektoren bzw. allgemeiner Zufallselementen $\boldsymbol{X}$ werfen nun die Frage auf, ob man die Erwartungswerte von $G(\boldsymbol{X})$ statt über die Verteilung von $G(\boldsymbol{X})$, die ja u.U. recht kompliziert sein kann, nicht auch direkt aus der Verteilung von $\boldsymbol{X}$ berechnen kann. Der folgende Satz gibt hierauf eine erste Antwort.

Satz 2.2.3. *(Transformationssatz für Erwartungswerte)*
Es sei $\boldsymbol{X}$ ein Zufallselement auf $(\Omega, \mathcal{A}, P)$ mit Werten in einem Meßraum $(\mathcal{X}, \mathcal{B})$ und $G : (\mathcal{X}, \mathcal{B}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ eine meßbare Abbildung derart, daß $G \geq 0$ oder $G(\boldsymbol{X})$ P–integrierbar ist. Dann gilt

$$E(G(\boldsymbol{X})) = \int G(\boldsymbol{X})\,dP = \int G\,dP^{\boldsymbol{X}}. \tag{2.2.29}$$

Ist umgekehrt G $P^{\boldsymbol{X}}$–integrierbar, so ist auch $G(\boldsymbol{X})$ P–integrierbar, und es gilt abermals (2.2.29).

Beweis. Wir zeigen die Aussage zunächst für Abbildungen G der Form $G = \mathbb{1}_B$ mit $B \in \mathcal{B}$. Dann ist nämlich für alle $\omega \in \Omega$

$$G\big(\boldsymbol{X}(\omega)\big) = \mathbb{1}_B\big(\boldsymbol{X}(\omega)\big) = \begin{cases} 1 & \text{für } \boldsymbol{X}(\omega) \in B \\ 0 & \text{sonst} \end{cases} = \mathbb{1}_{\boldsymbol{X}^{-1}(B)}(\omega),$$

also $G(\boldsymbol{X}) = \mathbb{1}_{\boldsymbol{X}^{-1}(B)}$ und somit nach (2.2.25)

$$\begin{aligned} E_P\big(G(\boldsymbol{X})\big) &= E_P\big(\mathbb{1}_{\boldsymbol{X}^{-1}(B)}\big) = P\big(\boldsymbol{X}^{-1}(B)\big) = P^{\boldsymbol{X}}(B) \\ &= E_{P^{\boldsymbol{X}}}\big(\mathbb{1}_B\big) = E_{P^{\boldsymbol{X}}}(G), \end{aligned} \tag{2.2.30}$$

was eine andere Schreibweise für den in (2.2.29) ausgedrückten Sachverhalt ist. Nimmt nun G endlich viele Werte $\alpha_1, \ldots, \alpha_n$, $n \in \mathbb{N}$, an, so besitzt G die Form $G = \sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}$ mit $B_i = \{G = \alpha_i\}$, $1 \le i \le n$; aufgrund der Linearität des Erwartungswerts (Beziehung (2.2.21)) gilt dann mit (2.2.30) wieder

$$\begin{aligned} E_P\big(G(\boldsymbol{X})\big) &= E_P\Big(\sum_{i=1}^n \alpha_i \mathbb{1}_{\boldsymbol{X}^{-1}(B_i)}\Big) = \sum_{i=1}^n \alpha_i E_P\big(\mathbb{1}_{\boldsymbol{X}^{-1}(B_i)}\big) \\ &= \sum_{i=1}^n \alpha_i E_{P^{\boldsymbol{X}}}\big(\mathbb{1}_{B_i}\big) = E_{P^{\boldsymbol{X}}}\Big(\sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}\Big) = E_{P^{\boldsymbol{X}}}(G). \end{aligned} \tag{2.2.31}$$

Ist allgemeiner $G \ge 0$, so existiert eine isotone Folge $\{G_n\}_{n\in\mathbb{N}}$ nicht–negativer meßbarer Funktionen auf $\mathcal{X}$, die je endlich viele Werte annehmen und für die $G = \sup_{n\in\mathbb{N}} G_n$, also auch $G(\boldsymbol{X}) = \sup_{n\in\mathbb{N}} G_n(\boldsymbol{X})$ gilt. Mit (2.2.23) ergibt sich wieder

$$\begin{aligned} E_P\big(G(\boldsymbol{X})\big) &= E_P\big(\sup_{n\in\mathbb{N}} G_n(\boldsymbol{X})\big) = \sup_{n\in\mathbb{N}} E_P\big(G_n(\boldsymbol{X})\big) \\ &= \sup_{n\in\mathbb{N}} E_{P^{\boldsymbol{X}}}(G_n) = E_{P^{\boldsymbol{X}}}\big(\sup_{n\in\mathbb{N}} G_n\big) = E_{P^{\boldsymbol{X}}}(G). \end{aligned} \tag{2.2.32}$$

Durch Zerlegung von G in Positiv– und Negativteil folgt schließlich die erste Behauptung.
Die zweite Behauptung zeigen wir durch Kontraposition, d.h.: Ist $G(\boldsymbol{X})$ nicht P–integrierbar, so ist G auch nicht $P^{\boldsymbol{X}}$–integrierbar. Es ist dann nämlich wenigstens einer der Ausdrücke $E_P\big(G^+(\boldsymbol{X})\big)$ oder $E_P\big(G^-(\boldsymbol{X})\big)$ unendlich, also nach dem ersten Teil des Satzes auch wenigstens einer der Ausdrücke $E_{P^{\boldsymbol{X}}}(G^+)$ oder $E_{P^{\boldsymbol{X}}}(G^-)$. Damit ist aber G nicht $P^{\boldsymbol{X}}$–integrierbar. ∎

Im letzten Satz wurde von einer Beweismethode Gebrauch gemacht, die unter dem Namen *algebraische Induktion* bekannt ist: Man zeigt die gewünschte Aussage zunächst für Indikatorfunktionen und benutzt dann die Linearität des Erwartungswerts, so daß die Aussage damit auch für Funktionen mit endlich vielen Werten gilt; durch monotone Approximation und Zerlegung in Positiv– und Negativteil erhält man dann die Aussage allgemein für meßbare Abbildungen. Wir werden

im folgenden einige weitere Aussagen mit diesem Prinzip beweisen, dabei aber nur den ersten — wesentlichen — Schritt explizit ausführen.

Satz 2.2.3 gestattet formal auch die folgende Darstellung des Erwartungswerts:

$$E(X) = \int X(\omega)\, dP(\omega) = \int x\, dP^X(x), \tag{2.2.33}$$

indem man für die Abbildung G auf $\mathbf{R}$ speziell die Identität wählt, d.h. $G(x) = x$, $x \in \mathbf{R}$. Dies zeigt noch einmal, daß der Erwartungswert in der Tat nicht von der speziellen Wahl der Zufallsvariablen X, sondern nur von deren Verteilung P^X abhängt.

Besitzt die Verteilung P^X eine Dichte f und vergleicht man die Beziehungen (2.2.33) und (2.2.15) miteinander, kann man also formal das Integrationssymbol "$dP^X(x)$" mit dem Ausdruck "$f(x)\,dx$" identifizieren. Man schreibt daher gelegentlich auch für die Dichte f den formalen "Quotienten" $f(x) = \dfrac{dP^X(x)}{dx}$.

Für den Fall, daß P^X eine Dichte f besitzt, erhebt sich daher zwangsläufig allgemeiner die Frage, für welche Abbildungen G der Erwartungswert $E\big(G(X)\big)$ als *Riemann*–Integral darstellbar ist. Daß dies nicht für jede (meßbare) Abbildung G der Fall zu sein braucht, zeigt das folgende Beispiel:
Man wähle etwa $G = \mathbb{1}_B$ mit $B = \mathbf{Q} \cap [0,1]$ und für P^X die stetige Gleichverteilung über $[0,1]$. Dann ist G nicht Riemann–integrierbar über $[0,1]$, wohl aber Lebesgue–integrierbar bezüglich P^X mit $E_{P^X}(G) = P(X \in B) = 0$!

Diese Problematik behandelt allgemeiner der nachfolgende Satz.

Satz 2.2.4. *(Erwartungswert und Riemann–Integral)*
Es sei X eine stetig verteilte Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Dichte f und $G : (\mathbf{R}, \mathcal{B}^1) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ eine meßbare Abbildung. Ferner sei U_G die Menge aller Unstetigkeitsstellen von G.

a) *Es ist $U_G \in \mathcal{B}^1$, d.h. die Menge der Unstetigkeitsstellen von G ist Borel'sch. Ist ferner $G \cdot f$ (evtl. uneigentlich) Riemann–integrierbar, so ist notwendig $P^X(U_G) = 0$. Diese Bedingung ist auch hinreichend für die Riemann–Integrierbarkeit von $G \cdot f$ über ein Intervall $[a,b]$ mit $a < b$, $a, b \in \mathbf{R}$, wenn $G \cdot f$ dort beschränkt ist.*

b) *Ist $|G| \cdot f$ (evtl. uneigentlich) Riemann–integrierbar, so ist auch $G(X)$ P–integrierbar, und es gilt*

$$E\big(G(X)\big) = \int_{-\infty}^{\infty} G(x) f(x)\, dx. \tag{2.2.34}$$

Beweis. Wir wollen hier nur Teil b) des Satzes beweisen, da der Beweis des ersten Teils eine aufwendige maßtheoretische Argumentation erfordert (vgl. etwa Benedetto (1976), Abschnitt 3.4 und speziell Theorem 3.13).
Hierzu beschränken wir uns zunächst auf die Situation, daß ein endliches Intervall $[a,b]$ mit $a < b$ existiert, so daß $P(X \in [a,b]) = 1$ gilt und G dort beschränkt ist. (Der allgemeine Fall kann durch Grenzwertbetrachtungen im wesentlichen hierauf zurückgeführt werden.) Für $n \in \mathbf{N}$ sei $a = x_{n0} < x_{n1} < \ldots < x_{nn} = b$ eine Zerlegung von $[a,b]$ mit

$$\lim_{n\to\infty} \max\{x_{nk} - x_{n,k-1} \mid 1 \le k \le n\} = 0. \tag{2.2.35}$$

Nach dem erweiterten Mittelwertsatz für Integrale (vgl. Heuser (1980), 85.6) existiert dann für jedes Zerlegungsintervall $[x_{n,k-1}, x_{nk}]$ eine Zahl g_{nk} mit der Eigenschaft

$$\inf_{x_{n,k-1}\le x\le x_{nk}} G(x) \le g_{nk} \le \sup_{x_{n,k-1}\le x\le x_{nk}} G(x) \quad \text{und}$$

$$\int_{x_{n,k-1}}^{x_{nk}} G(x)f(x)\,dx = g_{nk}\int_{x_{n,k-1}}^{x_{nk}} f(x)\,dx = g_{nk}P^X\big((x_{n,k-1}, x_{nk}]\big)$$

für $1 \le k \le n$, $n \in \mathbb{N}$. Definiere nun meßbare Abbildungen

$$G_n = \sum_{k=1}^{n} g_{nk}\mathbb{1}_{(x_{n,k-1},x_{nk}]}, \quad n \in \mathbb{N}.$$

Es gilt dann

$$\begin{aligned}\int_a^b G(x)f(x)\,dx &= \sum_{k=1}^{n}\int_{x_{n,k-1}}^{x_{nk}} G(x)f(x)\,dx = \sum_{k=1}^{n} g_{nk}\int_{x_{n,k-1}}^{x_{nk}} f(x)\,dx \\ &= \sum_{k=1}^{n} g_{nk}P^X\big((x_{n,k-1}, x_{nk}]\big) = E_{P^X}(G_n)\end{aligned} \tag{2.2.36}$$

für alle $n \in \mathbb{N}$. Da nach Voraussetzung G auf $[a,b]$ beschränkt ist — etwa $|G(x)| \le M$ für alle $x \in [a,b]$ — folgt auch $|G_n| \le Y := M\mathbb{1}_{[a,b]}$ für alle $n \in \mathbb{N}$, d.h. alle G_n sind durch die P^X–integrierbare Zufallsvariable Y mit $E_{P^X}(Y) = M$ majorisiert. Wir wollen jetzt noch zeigen, daß die Folge $\{G_n \mid n \in \mathbb{N}\}$ P^X–f.s. gegen G konvergiert, so daß damit die Voraussetzungen des Satzes von der majorisierten Konvergenz (Satz 2.2.2 Teil f)) erfüllt sind. Sei dazu $x \in (a,b]$ ein Stetigkeitspunkt von G, d.h. $x \in U_G^c$. Für jede der betrachteten Zerlegungen von $[a,b]$ existiert dann ein Index $k_n = k_n(x) \in \{1,2,\dots,n\}$ mit $x \in (x_{n\,k_n-1}, x_{n\,k_n}]$, $n \in \mathbb{N}$. Wegen (2.2.35) konvergiert also $g_{n\,k_n} = G_n(x)$ gegen $G(x)$ mit $n \to \infty$. Nach Teil a) des Satzes gilt aber $P^X(U_G) = 0$, was mit dem gerade gezeigten $\lim_{n\to\infty} G_n = G$ P^X–f.s. bedeutet. Mit Satz 2.2.2 Teil f) ergibt sich somit: G ist P^X–integrierbar (d.h. mit Satz 2.2.3 auch: $G(X)$ ist P–integrierbar), und es gilt nach (2.2.36)

$$E_P\big(G(X)\big) = E_{P^X}(G) = \lim_{n\to\infty} E_{P^X}(G_n) = \int_a^b G(x)f(x)\,dx = \int_{-\infty}^{\infty} G(x)f(x)\,dx,$$

wie behauptet. ∎

Der letzte Satz zeigt damit noch einmal aus einer anderen Sicht, warum in dem obigen Beispiel mit $G = \mathbb{1}_B$, $B = \mathbb{Q}\cap[0,1]$ und $P^X = \mathcal{R}\big([0,1]\big)$ der Erwartungswert $E\big(G(X)\big)$ nicht als Riemann–Integral existiert: es ist ja $U_G = [0,1]$, also $P^X(U_G) = P^X\big([0,1]\big) = 1 \ne 0$ und damit G nicht P^X–f.s. stetig.

Definiert man allgemeiner für eine Dichte f einer Verteilung P^X und eine Borel–Menge $B \in \mathcal{B}^1$ das Integral

$$\int_B f(x)\,dx := \int \mathbb{1}_B(x)\,f(x)\,dx = E_P\big(\mathbb{1}_B(X)\big) = P(X \in B) \tag{2.2.37}$$

(vgl. Beziehung (1.2.25)), so besagt Satz 2.2.3 also, daß dieses (Lebesgue–)Integral genau dann als Riemann–Integral darstellbar ist, wenn die Abbildung $G = \mathbb{1}_B$ P^X–f.s. stetig ist. Man macht sich leicht klar, daß hier $U_G = \partial B$ ist, wobei ∂B den Rand von B, also die abgeschlossene Hülle von B ohne die inneren Punkte von B, $\overline{B} \setminus B^\circ$, bezeichnet. Das Integral in (2.2.37) existiert also genau dann im Riemann'schen Sinne, wenn $P^X(\partial B) = 0$ gilt.

Ist nun allgemeiner X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $A \in \mathcal{A}$, so bezeichne analog

$$\int_A X \, dP = \int \mathbb{1}_A X \, dP. \tag{2.2.38}$$

Beziehung (2.2.37) ist hierin für den Spezialfall $(\Omega, \mathcal{A}, P) = (\mathbf{R}, \mathcal{B}^1)$ enthalten, wenn X durch die Identitätsabbildung und P durch P^X ersetzt wird.

Durch (2.2.37) bzw. (2.2.38) ist also eine echte Erweiterung des Riemann–Integrals gegeben. Wir werden in Zukunft häufiger von dieser Schreibweise Gebrauch machen, ohne jedesmal zu unterscheiden, ob das Integral im Riemann'schen oder Lebesgue'schen Sinn zu verstehen ist.

Für beliebige (meßbare) Abbildungen G ist z.B. dann $P^X(U_G) = 0$, wenn U_G abzählbar ist, also etwa für den Fall, daß G stückweise schwach monoton ist. Satz 2.2.3 liefert damit auch einen weiteren Beweis für die durch (2.2.19) gegebene Darstellung des Erwartungswerts $E(X)$, wenn X eine beliebige, P–integrierbare Zufallsvariable mit Verteilungsfunktion F ist: Für eine $\mathcal{R}\big((0,1)\big)$–verteilte Zufallsvariable U besitzt ja nach Satz 2.1.1 die Zufallsvariable $F^{-1}(U)$ die Verteilung P^X, d.h. nach (2.2.29) ist

$$E_P(X) = E_P\big(F^{-1}(U)\big) = E_{P^U}(F^{-1}) = \int_0^1 F^{-1}(u) \, du.$$

Entsprechende Sachverhalte gelten auch für m–dimensionale Zufallsvektoren mit m–dimensionalen Dichten ($m \in \mathbf{N}$), die wir hier allerdings ohne Beweis formulieren.

Satz 2.2.5. *(Transformationssatz für Zufallsvektoren)*
Es sei $\boldsymbol{X}$ ein m–dimensionaler Zufallsvektor auf $(\Omega, \mathcal{A}, P)$ ($m \in \mathbf{N}$) und $G : (\mathbf{R}^m, \mathcal{B}^m) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ eine meßbare Abbildung.

a) Ist $\boldsymbol{X}$ diskret verteilt mit Zähldichte f und konvergiert $\sum_{\boldsymbol{x} \in \mathbf{R}^m} |G(\boldsymbol{x})| f(\boldsymbol{x})$ absolut, so existiert $E\big(G(\boldsymbol{X})\big)$ mit

$$E\big(G(\boldsymbol{X})\big) = \sum_{\boldsymbol{x} \in \mathbf{R}^m} G(\boldsymbol{x}) f(\boldsymbol{x}). \tag{2.2.39}$$

b) Ist $\boldsymbol{X}$ stetig verteilt mit Dichte f und ist $|G| \cdot f$ (evtl. uneigentlich) m–dimensional Riemann–integrierbar, so existiert $E\big(G(\boldsymbol{X})\big)$ mit

$$E\big(G(\boldsymbol{X})\big) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} G(x_1, \dots, x_m) f(x_1, \dots, x_m) \, dx_1 \dots dx_m. \tag{2.2.40}$$

Man beachte, daß die Aussage a) des letzten Satzes eine unmittelbare Konsequenz des Transformationssatzes 2.2.3 sowie der Beziehung (2.2.13) ist (wobei hier für Ω der Wertebereich von $\boldsymbol{X}$ zu wählen ist).

Die Ausführungen zu Satz 2.2.4 gelten entsprechend auch im höherdimensionalen Fall[1)].

Beispiel 2.2.2. (Gleichverteilung auf dem Kreis)
Es sei (X,Y) auf dem Einheitskreis stetig gleichverteilt (vgl. (1.4.37)).

a) Wir betrachten den Abstand $R = G(X,Y) = \sqrt{X^2+Y^2}$ des Zufallsvektors (X,Y) vom Nullpunkt. Die Verteilung von R ist dann gegeben durch

$$P(R \le x) = P(X^2+Y^2 \le x^2) = \frac{1}{\pi} \iint\limits_{K^a(0,0;x)} du\,dv = x^2 \tag{2.2.41}$$

für $0 \le x \le 1$. R besitzt somit eine Dichte

$$f_R(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0 & \text{sonst,} \end{cases} \tag{2.2.42}$$

woraus sich der Erwartungswert von R berechnen läßt zu

$$E(R) = \int_0^1 2x^2\,dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3}. \tag{2.2.43}$$

Eine direkte Anwendung des Transformationssatzes 2.2.5 ergibt im Vergleich dazu das folgende — kompliziertere — Integral:

$$E(R) = \frac{1}{\pi} \iint\limits_{K^a(0,0;1)} \sqrt{x^2+y^2}\,dx\,dy = \frac{1}{\pi} \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sqrt{x^2+y^2}\,dx\,dy,$$

welches sich nicht unmittelbar leicht auswerten läßt. Unter Verwendung von Beziehung (2.1.94) können wir allerdings auch folgendermaßen vorgehen:
Seien U, V unabhängige, je $\mathcal{R}\big((0,1)\big)$–verteilte Zufallsvariablen und X und Y definiert vermöge

$$X = \sqrt{U}\cos(2\pi V), \quad Y = \sqrt{U}\sin(2\pi V). \tag{2.2.44}$$

Dann besitzt der Zufallsvektor (X,Y) die gewünschte Gleichverteilung über $K^a(0,0;1)$, und es ist $R = \sqrt{X^2+Y^2} = G^*(U,V) = \sqrt{U}$, unabhängig von V. Eine Anwendung von Satz 2.2.4 ergibt dann entsprechend

$$E(R) = E\big(G^*(U,V)\big) = E(\sqrt{U}) = \int_0^1 \sqrt{u}\,du$$
$$= \frac{2}{3}u^{\frac{2}{3}}\Big|_0^1 = \frac{2}{3}.$$

[1)] vgl. hierzu auch die Fußnote auf S. 44

b) Für den Fall $G(X,Y) = X \cdot Y$ läßt sich — im Gegensatz zu a) — die Verteilung von $G(X,Y)$ nicht auf einfache Weise berechnen. Satz 2.2.5 liefert dagegen

$$\begin{aligned} E(XY) &= \iint\limits_{K^a(0,0;1)} xy\,dx\,dy = \int_{-1}^{1} y \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\,dx\,dy \\ &= \int_{-1}^{1} y \frac{x^2}{2}\Big|_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dy = 0 \end{aligned} \tag{2.2.45}$$

sowie

$$\begin{aligned} E(X) = E(Y) &= \iint\limits_{K^a(0,0;1)} x\,dx\,dy = \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\,dx\,dy \\ &= \int_{-1}^{1} \frac{x^2}{2}\Big|_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dy = 0, \end{aligned} \tag{2.2.46}$$

d.h. es gilt $E(XY) = E(X)E(Y) = 0$. ∎

Beispiel 2.2.2 zeigt, daß die explizite Berechnung von Erwartungswerten mitunter nicht oder nur schwer mit elementaren Mitteln zu bewerkstelligen ist. Man ist daher gelegentlich auch an einfachen Abschätzungen für Erwartungswerte interessiert. Eine solche Abschätzung liefert das folgende

Lemma 2.2.1. *(Jensen'sche Ungleichung)*
Es sei $\boldsymbol{X} = (X_1, \ldots, X_m)$ ein m-dimensionaler Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in einer konvexen Menge $\mathcal{M} \in \mathcal{B}^m$ ($m \in \mathbb{N}$), und $G : \mathcal{M} \longrightarrow \mathbb{R}$ eine meßbare, konvexe (konkave) Abbildung. $G(\boldsymbol{X})$ sowie die Komponenten $X_1, \ldots, X_m$ von $\boldsymbol{X}$ seien sämtlich P-integrierbar. Dann ist $\big(E(X_1), \ldots, E(X_m)\big) \in \mathcal{M}$, und es gilt:

$$E\big(G(\boldsymbol{X})\big) \begin{cases} \geq G\big(E(X_1), \ldots, E(X_m)\big), & \text{falls } G \text{ konvex} \\ \leq G\big(E(X_1), \ldots, E(X_m)\big), & \text{falls } G \text{ konkav.} \end{cases} \tag{2.2.47}$$

Ist G nicht-negativ und konkav, braucht dabei lediglich die Integrierbarkeit aller Komponenten von $\boldsymbol{X}$ gefordert zu werden.

Beweis. Die erste Aussage des Lemmas wollen wir hier nur für den Fall beweisen, daß $\mathcal{M}$ kompakt ist (einen vollständigen Beweis findet man z.B. in Gänssler & Stute (1977), Satz 5.4.1.).
Es bezeichne $\mu = (\mu_1, \ldots, \mu_m) = \big(E(X_1), \ldots, E(X_m)\big)$. Angenommen, μ liegt außerhalb von $\mathcal{M}$. Dann existiert eine μ von $\mathcal{M}$ strikt trennende Hyperebene, d.h. es existieren Zahlen $a_1, \ldots, a_m \in \mathbb{R}$ mit

$$\sum_{i=1}^{m} a_i(x_i - \mu_i) > 0 \text{ für alle } (x_1, \ldots, x_m) \in \mathcal{M}$$

(vgl. etwa Valentine (1968)). Insbesondere ist also $S = \sum_{i=1}^{m} a_i(X_i - \mu_i) > 0$ und damit auch $E(S) > 0$, denn anderenfalls ergäbe sich aus Monotoniegründen $E(S) = 0$ und daher mit (2.2.28) $S = 0$ P–f.s. im Widerspruch zu $S > 0$. Damit folgt

$$0 < E(S) = \sum_{i=1}^{m} a_i\big(E(X_i) - \mu_i\big) = 0,$$

also ein erneuter Widerspruch, womit die Gültigkeit von $\mu \in \mathcal{M}$ bewiesen ist.
Zum Beweis von (2.2.47) können wir annehmen, daß G konvex ist (sonst Übergang von G zu $-G$). Es existiert dann eine Stützhyperebene, die den Graphen von G im Punkt μ berührt, d.h. es existieren Zahlen $b_1, \ldots, b_m \in \mathbf{R}$ mit

$$G(x_1, \ldots, x_m) \geq \sum_{i=1}^{m} b_i(x_i - \mu_i) + G(\mu_1, \ldots, \mu_m), \quad (x_1, \ldots, x_m) \in \mathcal{M}.$$

(Im Falle der Differenzierbarkeit von G in μ ist etwa $b_i = \frac{\partial G}{\partial x_i}(\mu)$, $1 \leq i \leq m$.)
Mit der Monotonie und Linearität des Erwartungswerts folgt daher

$$E\big(G(X_1, \ldots, X_m)\big) \geq \sum_{i=1}^{m} b_i\big(E(X_i) - \mu_i\big) + G(\mu_1, \ldots, \mu_m) = G(\mu_1, \ldots, \mu_m),$$

wie behauptet.
Die restliche Behauptung ergibt sich nun mit Satz 2.2.2 d). ∎

Beispiel 2.2.3. (Entropie einer Verteilung)
Es sei $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \in \mathbf{N}$, eine endliche Menge und P eine Wahrscheinlichkeitsverteilung mit Träger $\mathcal{X}$. Für $1 \leq i \leq m$ bezeichne $p_i = P(\{x_i\})$. Insbesondere sind dann alle $p_i > 0$. Die *Entropie* von P ist definiert durch

$$H = -\sum_{i=1}^{m} p_i \cdot \log p_i, \tag{2.2.48}$$

wobei als Basis für den Logarithmus eine beliebige, aber feste Zahl > 1 gewählt werden kann. Die Entropie charakterisiert die "Unbestimmtheit", welche durch die Verteilung P gegeben ist; der Untersuchung ihrer spezifischen Eigenschaften ist der spätere Abschnitt 5.1 gewidmet. Wir wollen an dieser Stelle nur zeigen, wie man die Jensen'sche Ungleichung verwenden kann, um eine — scharfe — obere Schranke für die Entropie herzuleiten.
Hierzu definieren wir auf $\mathcal{X}$ die Zufallsvariable Z durch $Z(x_i) = \frac{1}{p_i}$, $1 \leq i \leq m$. Es ist dann nach (2.2.13)

$$E(Z) = \sum_{i=1}^{m} Z(x_i)P(\{x_i\}) = \sum_{i=1}^{m} \frac{1}{p_i} \cdot p_i = m, \tag{2.2.49}$$

also mit der Jensen'schen Ungleichung (2.2.47) aufgrund der Konkavität des Logarithmus

$$\begin{aligned} E(\log Z) &= \sum_{i=1}^{m} P(\{x_i\}) \log Z(x_i) = \sum_{i=1}^{m} p_i \log\Big(\frac{1}{p_i}\Big) \\ &= H \leq \log E(Z) = \log m \end{aligned} \tag{2.2.50}$$

mit Gleichheit für den Fall einer diskreten Gleichverteilung, d.h. $P(\{x_i\}) = \frac{1}{m}$, $1 \leq i \leq m$, wie man unmittelbar nachrechnet. ∎

Eine diskrete Gleichverteilung über einer endlichen Menge beinhaltet also in diesem Sinn das größte Maß an "Unbestimmtheit", weshalb diese Verteilung häufig — gerade auch in der Informatik — zur Modellierung zufälliger Vorgänge herangezogen wird, für die keine Vorkenntnisse zur Verfügung stehen, etwa die (zufällige) Reihenfolge von Elementen eines Feldes, aus denen ein Schlüsselelement ausgesucht werden soll — man vergleiche etwa das bereits mehrfach erwähnte Beispiel der binären Suche (Beispiele 1.1.1 und 2.2.2) oder auch das am Anfang dieses Abschnitts betrachtete Sortierverfahren `straightinsertion` (Beispiel 2.2.1).[1)]

Von besonderer Bedeutung sind speziell Erwartungswerte *polynomialer* Transformationen G von Zufallsvariablen und Zufallsvektoren, die im folgenden gesondert behandelt werden.

Definition 2.2.4. *(Varianz, Kovarianz und höhere Momente)*
Es seien X und Y Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ sowie $k \in \mathbb{N}$.

a) Ist X^2 P–integrierbar, so auch X; in diesem Fall heißt

$$\mathrm{Var}(X) = E\big((X - E(X))^2\big) \tag{2.2.51}$$

die Varianz von X bzw. von P^X.

b) Sind X, Y und XY jeweils P–integrierbar, so heißt

$$\mathrm{Kov}(X,Y) = E\Big[\big(X - E(X)\big)\big(Y - E(Y)\big)\Big] \tag{2.2.52}$$

die Kovarianz von X und Y. Ist $\mathrm{Kov}(X,Y) = 0$, so heißen X und Y unkorreliert.

c) Ist $|X|^k$ P–integrierbar, so auch X; in diesem Fall heißt $E\big(|X|^k\big)$ das k–te absolute Moment[2)] von X, $E\big(X^k\big)$ das k–te Moment von X, $E\big(|X - E(X)|^k\big)$ das k–te absolute zentrale Moment von X sowie $E\big((X - E(X))^k\big)$ das k–te zentrale Moment von X (bzw. P^X).

Man beachte, daß die Zufallsvariable $|X|^k + 1$ für jedes $k \in \mathbb{N}$ eine P–integrierbare Majorante zu $|X|$ ist, wenn $|X|^k$ P–integrierbar ist (vgl. Satz 2.2.2 d)).

Das folgende Lemma behandelt einige einfache Eigenschaften von Momenten von Verteilungen.

[1)] Dies bedeutet allerdings nicht, daß die in einer bestimmten Situation vorliegende Verteilung tatsächlich eine Gleichverteilung ist; die Unkenntnis über das "wahre" Verteilungsmodell führt daher nicht *zwingend* zur Annahme einer Gleichverteilung!

[2)] Stellt man sich die reelle Achse gemäß der Verteilung P^X mit Masse belegt vor, so entspricht $E(X)$ gerade dem *Massenschwerpunkt*; $E\big(X^2\big)$ ist dann das zugehörige *Trägheitsmoment*, wenn diese mit Masse behaftete Achse um den Nullpunkt mit der y–Achse als Drehachse rotiert. $Var(X)$ ist das entsprechende Trägheitsmoment, wenn die x–Achse um den Schwerpunkt $E(X)$ rotiert (vgl. hierzu auch Heuser (1989), Abschnitt 90).

Lemma 2.2.2. *(Eigenschaften von Momenten)*
Es seien X und Y Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann gilt:

a) Sind X^2, Y und XY P–integrierbar, so ist auch

$$\begin{aligned} \operatorname{Var}(X) &= \operatorname{Kov}(X,X) = E(X^2) - \big(E(X)\big)^2 \\ \operatorname{Kov}(X,Y) &= E(XY) - E(X)E(Y). \end{aligned} \tag{2.2.53}$$

Für beliebige $c \in \mathbf{R}$ gilt darüberhinaus:

$$E\big((X-c)^2\big) = \operatorname{Var}(X) + \big(E(X) - c\big)^2, \tag{2.2.54}$$

d.h. $E\big((X-c)^2\big)$ ist minimal für $c = E(X)$.[1)]
Ferner gilt:

$$\operatorname{Var}(X) = 0 \quad \Longleftrightarrow \quad X = const\ P - \text{f.s.} \tag{2.2.55}$$

b) Unter den Voraussetzungen von a) gilt:

$$\begin{aligned} \operatorname{Var}(aX+b) &= a^2 \operatorname{Var}(X) \\ \operatorname{Kov}(aX+b, cY+d) &= ac \operatorname{Kov}(X,Y) \end{aligned} \tag{2.2.56}$$

für alle $a, b, c, d \in \mathbf{R}$. Ferner ist

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Kov}(X,Y). \tag{2.2.57}$$

Sind speziell X und Y unkorreliert, so gilt

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y). \tag{2.2.58}$$

c) Ist X^2 P–integrierbar, so gilt

$$P(|X - E(X)| > c) \leq \frac{\operatorname{Var}(X)}{c^2} \quad (Tschebyscheff - Ungleichung) \tag{2.2.59}$$

für alle $c > 0$.

d) Existiert für ein $k \in \mathbf{N}$ das k–te absolute Moment von X, so existieren auch alle m–ten Momente von X für $m \leq k$ (d.h. absolute und nicht–absolute, zentrale und nicht–zentrale), und es gilt

$$\begin{aligned} E\big((X-c)^m\big) \leq E\big(|X-c|^m\big) &\leq \Big\{E\big(|X-c|^k\big)\Big\}^{\frac{m}{k}} \\ &\leq 2^m \Big(E\big(|X|^k\big) + |c|^k\Big)^{\frac{m}{k}} \end{aligned} \tag{2.2.60}$$

[1)] Dies spiegelt die physikalisch intuitive Tatsache wieder, daß das Trägheitsmoment bei Drehung um den *Schwerpunkt* am kleinsten ist.

für alle $c \in \mathbf{R}$. Bezeichnet F die Verteilungsfunktion von P^X, so gilt auch die Darstellung

$$\begin{aligned} E(X^k) &= (-1)^k \int_{-\infty}^{0} kx^{k-1}F(x)\,dx + \int_0^{\infty} kx^{k-1}(1-F(x))\,dx \\ &= \int_0^1 \left(F^{-1}(x)\right)^k dx \\ E(|X|^k) &= \int_{-\infty}^{0} kx^{k-1}F(x)\,dx + \int_0^{\infty} kx^{k-1}(1-F(x))\,dx \\ &= \int_0^1 |F^{-1}(x)|^k\,dx. \end{aligned} \tag{2.2.61}$$

Gilt wieder $P(X \in \mathbf{Z}) = 1$, so läßt sich das k-te Moment von X auch schreiben als

$$\begin{aligned} E(X^k) = (-1)^k \sum_{m=1}^{\infty} &\left[(-m+1)^k - (-m)^k\right] P(X \le -m) \\ &+ \sum_{m=0}^{\infty} \left[(m+1)^k - m^k\right] P(X > m). \end{aligned} \tag{2.2.62}$$

e) Sind $|X|^p$ und $|Y|^q$ P-integrierbar für $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$, so ist auch $|XY|$ P-integrierbar, und es gilt

$$E(|XY|) \le \left\{E(|X|^p)\right\}^{\frac{1}{p}} \cdot \left\{E(|Y|^q)\right\}^{\frac{1}{q}} \quad (H\ddot{o}lder - Ungleichung). \tag{2.2.63}$$

Ist speziell $p = q = 2$, so gilt auch

$$|\operatorname{Kov}(X,Y)| \le \sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)} \tag{2.2.64}$$

mit Gleichheit in (2.2.64) genau dann, wenn X und Y P-f.s. linear voneinander abhängen, d.h. wenn Zahlen $a, b, c \in \mathbf{R}$, a, b nicht beide 0, existieren mit $aX + bY = c$ P-f.s.

Beweis. a) Beziehung (2.2.53) ergibt sich unmittelbar aus (2.2.51) und (2.2.52) durch Auflösen der Klammern und der Linearität des Erwartungswerts, entsprechend bei (2.2.54): es ist mit $\mu = E(X)$

$$\begin{aligned} E\left((X-c)^2\right) &= E(X^2 - 2cX + c^2) \\ &= E(X^2 - 2\mu X + \mu^2) + 2(\mu - c)E(X) + c^2 - \mu^2 \\ &= \operatorname{Var}(X) + 2(\mu - c)\mu + c^2 - \mu^2 = \operatorname{Var}(X) + c^2 - 2c\mu + \mu^2 \\ &= \operatorname{Var}(X) + \left(E(X) - c\right)^2. \end{aligned}$$

b) Beziehung (2.2.56) ergibt sich durch Rechnung wie in a). Weiter ist dann mit $\mu = E(X), \nu = E(Y)$ nach Teil a)

$$\begin{aligned}\operatorname{Var}(X+Y) &= E\big((X+Y)^2\big) - (\mu+\nu)^2 \\ &= E\big(X^2 + 2XY + Y^2\big) - (\mu^2 + 2\mu\nu + \nu^2) \\ &= E\big(X^2\big) - \mu^2 + E\big(Y^2\big) - \nu^2 + 2\big[E(XY) - \mu\nu\big] \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Kov}(X,Y)\end{aligned}$$

wie behauptet. Sind X und Y zusätzlich unkorreliert, so ist $\operatorname{Kov}(X,Y) = 0$, womit (2.2.58) folgt.
Gilt $X = c$ P–f.s. für ein $c \in \mathbf{R}$, so ist auch $E(X) = c$, $E\big(X^2\big) = c^2$ und somit $\operatorname{Var}(X) = 0$. Ist umgekehrt $\operatorname{Var}(X) = E\big(X - E(X))^2\big) = 0$, so auch $X - E(X) = 0$ P–f.s. nach (2.2.28).
c) Dies ist ein Spezialfall der Markoff–Ungleichung (2.2.26), wenn man dort X durch $(X - E(X))^2$ und c durch c^2 ersetzt.
d) Es reicht aufgrund von Satz 2.2.2 d), Beziehung (2.2.60) nachzuweisen. Diese folgt aber für $m < k$ aus der Hölder–Ungleichung (2.2.63) für die Wahl $p = \frac{k}{m}$, $q = \frac{k}{k-m}$ und $Y \equiv 1$, wenn dort X durch $|X|^m$ ersetzt wird sowie aus der für alle reellen Zahlen a, b gültigen Abschätzung $|a-b|^k \le \big(|a| + |b|\big)^k \le \big(2\max(|a|,|b|)\big)^k \le 2^k\big(|a|^k + |b|^k\big)$.
Beziehung (2.2.61) ergibt sich mit Hilfe von Satz 2.2.4 analog zu Satz 2.2.1 c) und der Argumentation zu Beziehung (2.2.47). Die letzte Aussage folgt schließlich analog Beziehung (2.2.17).
e) Wir benutzen die Jensen'sche Ungleichung (2.2.47) für die konkave Abbildung $G(u,v) = u^{\frac{1}{p}} v^{\frac{1}{q}}$, $u, v > 0$. (Die Konkavität ergibt sich z.B. aufgrund von

$$\frac{\partial^2 G}{\partial u^2}(u,v) = -\frac{p-1}{p^2} u^{\frac{1}{p}-2} v^{\frac{1}{q}} < 0, \quad \frac{\partial^2 G}{\partial v^2}(u,v) = -\frac{q-1}{q^2} u^{\frac{1}{p}} v^{\frac{1}{q}-2} < 0 \quad \text{und}$$

$$\frac{\partial^2 G}{\partial u \partial v} = \frac{1}{pq} u^{\frac{1}{p}-1} v^{\frac{1}{q}-1},$$

d.h. die Matrix $\begin{pmatrix} \frac{\partial^2 G}{\partial u^2} & \frac{\partial^2 G}{\partial u\,\partial v} \\ \frac{\partial^2 G}{\partial u\,\partial v} & \frac{\partial^2 G}{\partial v^2} \end{pmatrix}$ ist negativ–semidefinit auf der konvexen Menge $\mathcal{M} = (0,\infty) \times (0,\infty)$ (vgl. Kall (1976), Satz 2.17).) Nach Lemma 2.2.1 folgt also

$$\begin{aligned}E\big(|XY|\big) &= E\Big(G\big(|X|^p |Y|^q\big)\Big) \\ &\le G\Big(E\big(|X|^p\big), E\big(|Y|^q\big)\Big) = \Big\{E\big(|X|^p\big)\Big\}^{\frac{1}{p}} \cdot \Big\{E\big(|Y|^q\big)\Big\}^{\frac{1}{q}}\end{aligned}$$

wie behauptet. Beziehung (2.2.64) ergibt sich nun unter Verwendung von (2.2.20) aus (2.2.63), wenn man $p = q = 2$ wählt und X durch $X - E(X)$ sowie Y durch

$Y - E(Y)$ ersetzt.
Seien nun $a, b, c \in \mathbf{R}$, a, b nicht beide 0; etwa $a \neq 0$. Dann gilt nach (2.2.56)

$$\begin{aligned} |a| \cdot |\operatorname{Kov}(X,Y)| &= |\operatorname{Kov}(aX,Y)| = |\operatorname{Kov}(c-bY,Y)| = |b| \operatorname{Var}(Y) \\ &= \sqrt{b^2 \operatorname{Var}(Y)}\sqrt{\operatorname{Var}(Y)} = \sqrt{\operatorname{Var}(aX) \operatorname{Var}(Y)} \\ &= |a|\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}, \end{aligned}$$

also Gleichheit in (2.2.64). Ist umgekehrt $\operatorname{Kov}(X,Y) = \pm\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}$, so ergibt sich für alle $a, b \in \mathbf{R}$ die Gleichung

$$\begin{aligned} \operatorname{Var}(aX+bY) &= a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Kov}(X,Y) \\ &= a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) \pm 2ab\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)} \\ &= \left(a\sqrt{\operatorname{Var}(X)} \pm b\sqrt{\operatorname{Var}(Y)}\right)^2, \end{aligned}$$

also speziell $\operatorname{Var}(aX+bY) = 0$ für die Wahl $a = \sqrt{\operatorname{Var}(Y)}$, $b = \mp\sqrt{\operatorname{Var}(X)}$. Nach (2.2.55) ist dann aber $aX + bY = const$ P–f.s., wie behauptet. (Für $\sqrt{\operatorname{Var}(X)} = \sqrt{\operatorname{Var}(Y)} = 0$ sind jeweils X und Y P–f.s. konstant, so daß sich in diesem Fall a und b stets so wählen lassen, daß nicht $a = b = 0$ gilt.) ∎

Teil e) des letzten Satzes gibt unmittelbar Anlaß zu der folgenden

Definition 2.2.5. *(Korrelation)*
Es seien X und Y Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit $0 < \operatorname{Var}(X)$, $\operatorname{Var}(Y) < \infty$. Dann heißt

$$\operatorname{Korr}(X,Y) = \frac{\operatorname{Kov}(X,Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}} \tag{2.2.65}$$

die Korrelation zwischen X und Y.

Die Korrelation ist also ein Maß für die *lineare* Abhängigkeit zwischen X und Y; sie nimmt nach obigem Werte zwischen -1 und 1 an mit $\operatorname{Korr}(X,Y) = \pm 1$ genau dann, wenn Zahlen $a, b, c \in \mathbf{R}$, a, b nicht beide 0, existieren mit $aX + bY = c$ P–f.s. Ist $\operatorname{Korr}(X,Y) > 0$, so sagt man auch, X und Y seien *positiv* korreliert; ist $\operatorname{Korr}(X,Y) < 0$, heißen X und Y *negativ* korreliert. Stochastisch unabhängige Zufallsvariablen sind nach (1.6.32) stets auch unkorreliert; die Umkehrung hiervon ist i.a. aber nicht richtig, wie Beispiel 2.2.2 b) zeigt: dort sind wegen $X^2 + Y^2 \leq 1$ X und Y nicht stochastisch unabhängig, wohl aber unkorreliert nach (2.2.45).

Die Bedeutung der Varianz als Maß der Streuung einer Verteilung kommt insbesondere auch in der folgenden Form der Tschebyscheff–Ungleichung (Teil c) des letzten Satzes) zum Ausdruck: bezeichnet nämlich $\mu = E(X)$, $\sigma = \sqrt{\operatorname{Var}(X)}$, so gilt stets

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2} \tag{2.2.66}$$

für alle $k \in \mathbf{N}$. Dies folgt unmittelbar aus (2.2.59), wenn dort $c = k\sigma$ gewählt wird. Das Intervall $[\mu - k\sigma, \mu + k\sigma]$ heißt auch $k\sigma$–Bereich der Verteilung P^X; in

in ihm liegen also — für *jede* Verteilung P^X — stets mindestens $\left(1-\frac{1}{k^2}\right)\cdot 100$ % aller Beobachtungen von X. Die Tschebyscheff-Ungleichung ist dabei scharf: Betrachtet man z.B. die durch

$$F(x)=\begin{cases}0 & \text{falls } x<-B\\ \frac{A^2}{2x^2} & \text{falls } -B\le x\le -A\\ \frac{1}{2} & \text{falls } -A\le x\le A\\ 1-\frac{A^2}{2x^2} & \text{falls } A\le x<B\\ 1 & \text{falls } x\ge B\end{cases} \tag{2.2.67}$$

mit reellen Konstanten $A>0$, $B=A\cdot e^{\varepsilon/2}$ $(\varepsilon\ge 0)$ gegebene Verteilungsfunktion, so ergibt sich aus Symmetriegründen $E(X)=0$ sowie

$$\begin{aligned}E(X^2)&=\int_0^\infty P(X^2>x)\,dx=\int_0^{A^2}dx+\int_{A^2}^{B^2}\frac{A^2}{x}\,dx\\ &=A^2+2A^2\ln\left(\frac{B}{A}\right)=A^2(1+\varepsilon)\end{aligned}$$

und somit

$$P(|X-E(X)|>c)=\frac{A^2}{c^2}=\frac{\operatorname{Var}(X)}{(1+\varepsilon)c^2}\le\frac{\operatorname{Var}(X)}{c^2}$$

für $A\le c<B$. Für $\varepsilon\downarrow 0$ erhält man hier asymptotisch Gleichheit.

Die Tschebyscheff-Ungleichung spielt auch eine wesentliche Rolle bei dem sogenannten *Gesetz der großen Zahlen*, welches in der einfachsten Form schon von J. Bernoulli im 18. Jahrhundert formuliert wurde und auf empirische Weise den Erwartungswertbegriff durch Mittelbildung von Beobachtungen veranschaulicht.

Satz 2.2.6. *(schwaches Gesetz der großen Zahlen)*
Es sei $\{X_n\}_{n\in\mathbf{N}}$ eine Folge paarweise unkorrelierter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$ mit $E(X_n)=\mu\in\mathbf{R}$ und $\operatorname{Var}(X_n)=\sigma^2>0$ für alle $n\in\mathbf{N}$. Dann gilt für die arithmetischen Mittel $S_n=\frac{1}{n}\sum_{i=1}^n X_i$ der durch die Folge $\{X_n\}_{n\in\mathbf{N}}$ beschriebenen Beobachtungen:

$$P(|S_n-\mu|>\varepsilon)\le\frac{\sigma^2}{n\varepsilon^2}\quad\text{für alle }\varepsilon>0\text{ und }n\in\mathbf{N}, \tag{2.2.68}$$

d.h. die Wahrscheinlichkeit für positive Abweichungen der arithmetischen Mittel $S_n=\frac{1}{n}\sum_{i=1}^n X_i$ vom Erwartungswert μ der Folge wird mit wachsendem n beliebig klein.

Beweis. Zunächst ist $E(S_n)=\frac{1}{n}\sum_{i=1}^n E(X_i)=\mu$ für alle $n\in\mathbf{N}$ sowie aufgrund von (2.2.56) und mehrfacher Anwendung von (2.2.58)

$$\operatorname{Var}(S_n)=\frac{1}{n^2}\sum_{i=1}^n\operatorname{Var}(X_i)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

Die Aussage ergibt sich somit direkt aus der Tschebyscheff–Ungleichung (2.2.59) mit $X = S_n$ und $c = \varepsilon$. ∎

Man beachte, daß die Voraussetzungen des Satzes 2.2.6 insbesondere dann erfüllt sind, wenn die Folge $\{X_n\}_{n\in\mathbb{N}}$ *stochastisch unabhängig* ist. Für diesen Fall lassen sich noch stärkere Aussagen bzgl. P–fast sicherer Konvergenz herleiten, was u.a. Gegenstand des Abschnitts 2.3 dieses Kapitels ist.

Im Rahmen der Average–Case–Analyse von Algorithmen etwa läßt sich das Gesetz der großen Zahlen so interpretieren: Ist X_i die (zufällige) Laufzeit des Algorithmus beim i–ten Aufruf (vgl. Beispiel 2.2.1), und sind die jeweiligen Ausgangskonstellationen (wie etwa die Anordnung der Feldelemente beim binären Suchen) unabhängig voneinander, so nähert sich die mittlere Laufzeit $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ nach n Aufrufen mit wachsendem n immer mehr dem Erwartungswert μ an. Der Erwartungswert ist also eine geeignete Größe, um durchschnittliche Laufzeiten von Algorithmen auf einer theoretischen Ebene miteinander zu vergleichen.

Wählt man in Satz 2.2.6 speziell $X_n = \mathbb{1}_{A_n}$ für Ereignisse $A_n \in \mathcal{A}$ mit den Eigenschaften

$$P(A_n \cap A_m) = P(A_n)P(A_m) \quad \text{für alle } n, m \in \mathbb{N},\ n \neq m$$
$$P(A_n) = p \in [0,1] \quad \text{für alle } n \in \mathbb{N},$$

d.h. die Ereignisse A_n sind paarweise stochastisch unabhängig mit derselben Eintrittswahrscheinlichkeit p, so besagt das Gesetz der großen Zahlen in diesem Fall, daß die relativen Eintrittshäufigkeiten $S_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_n}$ für große n nahe bei p liegen, genauer:

$$P\big(|S_n - p| > \varepsilon\big) \leq \frac{1}{4n\varepsilon^2} \quad \text{für alle } \varepsilon > 0 \text{ und } n \in \mathbb{N}. \tag{2.2.69}$$

Es ist nämlich dann $E(X_n) = E(X_n^2) = p$, also $\mathrm{Var}(X_n) = p - p^2 = p(1-p) \leq \frac{1}{4}$ für alle $n \in \mathbb{N}$ sowie $E(X_nX_m) = E\big(\mathbb{1}_{A_n\cap A_m}\big) = P(A_n \cap A_m) = P(A_n)P(A_m) = E\big(\mathbb{1}_{A_n}\big)E\big(\mathbb{1}_{A_m}\big) = E(X_n)E(X_m)$, also $\mathrm{Kov}(X_n, X_m) = 0$ für $n \neq m$, womit (2.2.69) aus (2.2.68) folgt.

Beziehung (2.2.69) ist fundamental für die *mathematische Statistik* oder auch die in Kapitel 6 behandelten Simulationsverfahren: einerseits lassen sich damit Wahrscheinlichkeiten von Ereignissen empirisch durch relative Häufigkeiten des Eintretens solcher Ereignisse aus einer genügend großen Zahl von Wiederholungen gleichartiger Versuche gewinnen; andererseits ermöglicht eine im Mittel "ausgewogene" Verteilung des Eintretens bzw. Nichteintretens von Ereignissen in einer Serie von Experimenten — die in der Regel durch geeignete *Algorithmen* (und damit letztlich deterministisch) erzeugt werden — die Simulation zufälliger Vorgänge etwa durch Einsatz von Rechnern.

Wir wollen uns zum Abschluß dieses Abschnitts noch kurz mit erzeugenden Funktionen beschäftigen, die die Berechnung von Momenten (insbesondere Erwartungswerten, Varianzen und Kovarianzen) i.a. erheblich vereinfachen und die für diskrete Verteilungen auch in der Kombinatorik von Bedeutung sind.

Definition 2.2.6. *(erzeugende Funktionen)*
Es sei X eine reelle Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $I \subseteq \mathbf{R}$ derart, daß

$$\Psi_X(t) = E\big(e^{tX}\big),\ t \in I \tag{2.2.70}$$

für alle $t \in I$ endlich ist. Dann heißt die auf I definierte Abbildung Ψ_X die momenterzeugende Funktion zu X bzw. zu P^X.
Existiert die momenterzeugende Funktion $\Psi_X(t)$ für $t \in I$, so heißt die durch

$$\psi_X(s) = \Psi_X(\ln s) = E\big(s^X\big), \quad s \in J := \{e^u \mid u \in I\} \tag{2.2.71}$$

gegebene Funktion die wahrscheinlichkeitserzeugende (oder auch kürzer: erzeugende) Funktion zu X bzw. P^X.

Die momenterzeugende Funktion Ψ_X charakterisiert die Verteilung P^X eindeutig, wenn die Menge I ein Intervall $[-\delta, \delta]$ für ein $\delta > 0$ enthält. Einen Beweis dieser Aussage, der üblicherweise Methoden der Fourier–Analyse benötigt, können wir hier nicht führen; der interessierte Leser sei etwa auf Billingsley (1986), Theorem 30.1 verwiesen.

Die Bedeutung erzeugender Funktionen wird in dem nachfolgenden Lemma deutlich.

Lemma 2.2.3. *(Eigenschaften erzeugender Funktionen)*
Es sei X eine reelle Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und Ψ_X bzw. ψ_X die moment– bzw. wahrscheinlichkeitserzeugende Funktion zu X. Dann gilt:

a) Es ist stets $\Psi_X(0) = \psi_X(1) = 1$. Existiert ferner $\Psi_X(t)$ für ein $t^* > 0$ ($t_* < 0$), so auch für alle $t \in [0, t^*]$ ($t \in [t_*, 0]$). Bezeichnet speziell

$$t_+ = \sup\{t \in \mathbf{R} \mid \Psi_X(t) < \infty\},\ t_- = \inf\{t \in \mathbf{R} \mid \Psi_X(t) < \infty\}, \tag{2.2.72}$$

so existiert $\Psi_X(t)$ für alle $t \in (t_-, t_+)$, und $\psi_X(s)$ existiert für alle $s \in (e^{t_-}, e^{t_+})$ (mit der Konvention $e^{-\infty} = 0$, $e^{\infty} = \infty$).
Ist speziell $X \geq 0$ P–f.s. ($X \leq 0$ P–f.s.), so ist $t_- = -\infty$ ($t_+ = \infty$).

b) Mit den Bezeichnungen aus (2.2.72) sei $0 < \delta < \min(t_+, -t_-)$. Dann existieren sämtliche Momente $E\big(|X|^k\big)$, $k \in \mathbf{N}$, Ψ_X ist im Nullpunkt beliebig oft differenzierbar, und es gilt

$$\begin{aligned} \Psi_X^{(k)}(0) &= E\big(X^k\big) \quad \text{für alle } k \in \mathbf{N} \\ \Psi_X(t) &= \sum_{k=0}^{\infty} \frac{E\big(X^k\big)}{k!} t^k \quad (|t| \leq \delta). \end{aligned} \tag{2.2.73}$$

Insbesondere ist

$$E(X) = \Psi_X'(0), \quad \operatorname{Var}(X) = \Psi_X''(0) - \big(\Psi_X'(0)\big)^2. \tag{2.2.74}$$

Ferner ist dann auch $\psi_X(s)$ für $s = 1$ differenzierbar, und es gilt

$$\begin{aligned} \psi_X^{(k)}(1) &= E\Big(\prod_{i=0}^{k-1}(X - i)\Big) \quad \text{für alle } k \in \mathbf{N} \\ \psi_X(s) &= \sum_{k=0}^{\infty} \frac{E\big(\prod_{i=0}^{k-1}(X - i)\big)}{k!}(s-1)^k \quad (|s-1| \leq 1 - e^{-\delta}) \end{aligned} \tag{2.2.75}$$

Insbesondere ist

$$E(X) = \psi_X'(1), \quad \mathrm{Var}(X) = \psi_X''(1) + \psi_X'(1)\big(1 - \psi_X'(1)\big). \tag{2.2.76}$$

c) *Gilt* $P(X \in \mathbf{N}_0) = 1$, *d.h. nimmt* X *f.s. nur die Werte* $\{0, 1, 2, \ldots\}$ *an, so läßt sich* ψ_X *fortsetzen, d.h.* $\psi_X(s) = E\big(s^X\big)$ *existiert auch für alle* $|s| \leq 1$, *und es gilt*

$$\begin{aligned} \psi_X^{(k)}(0) &= P(X = k) \quad \text{für alle } k \in \mathbf{N} \\ \psi_X(s) &= \sum_{k=0}^{\infty} P(X = k)\, s^k \quad (|s| \leq 1). \end{aligned} \tag{2.2.77}$$

d) *Sind* X *und* Y *stochastisch unabhängige Zufallsvariablen mit momenterzeugenden Funktionen* $\Psi_X(t)$ *und* $\Psi_Y(t)$, *die beide für* $t \in I \subseteq \mathbf{R}$ *existieren, so besitzt dort auch* $X + Y$ *eine momenterzeugende Funktion, und es gilt*

$$\Psi_{X+Y}(t) = \Psi_X(t) \cdot \Psi_Y(t), \quad t \in I \tag{2.2.78}$$

bzw. auch

$$\psi_{X+Y}(s) = \psi_X(s) \cdot \psi_Y(s), \quad s \in J = \{e^u \mid u \in I\} \tag{2.2.79}$$

(Multiplikationsregel).

Beweis. a) Die erste Aussage ergibt sich direkt aus der Definition (moment-)erzeugender Funktionen. Mit der Hölder–Ungleichung (2.2.63) erhält man ferner für $0 \leq t \leq t^*$

$$\Psi_X(t) = E\big(e^{tX}\big) \leq \Big\{E\big(e^{t^* X}\big)\Big\}^{t/t^*} = \big(\Psi_X(t^*)\big)^{t/t^*}$$

mit $p = \frac{t^*}{t}$ und $Y \equiv 1$, wenn man dort X durch e^{tX} ersetzt. Mit analoger Argumentation folgt für $t_* \leq t \leq 0$

$$\begin{aligned} \Psi_X(t) &= E\big(e^{tX}\big) = E\big(e^{(-t)(-X)}\big) \\ &\leq \Big\{E\big(e^{(-t^*)(-X)}\big)\Big\}^{t/t^*} = \big(\Psi_X(t_*)\big)^{t/t_*}. \end{aligned}$$

Somit existiert $\Psi_X(t)$ für alle $t \in (t_-, t_+)$, also definitionsgemäß auch $\psi_X(s)$ für $s \in (e^{t_-}, e^{t_+})$. Für $X \geq 0$ ist stets $e^{tX} \leq 1$, $t \leq 0$, d.h. es folgt $t_- = -\infty$ (analog bei $X \leq 0$).

b) Für $|t| \leq \delta$ seien die meßbaren Abbildungen (Zufallsvariablen) $G_n(t; \cdot)$ auf $(\mathbf{R}, \mathcal{B}^1)$ definiert durch

$$G_n(t; x) = \sum_{k=0}^{n} \frac{x^k}{k!} t^k \quad (x \in \mathbf{R},\ n \in \mathbf{N}).$$

Es ist dann für alle $n \in \mathbf{N}$ und $|t| \leq \delta$

$$\left|G_n(t;X)\right| \leq e^{|tX|} \leq e^{\delta X} + e^{-\delta X} =: Z,$$

also nach Voraussetzung Z P-integrierbar mit $E(Z) = \Psi_X(\delta) + \Psi_X(-\delta)$. Ferner ist

$$|X|^n \leq \frac{n!}{\delta^n} G_n(\delta, |X|) \leq \frac{n!}{\delta^n} Z$$

für alle $n \in \mathbf{N}$, d.h. es existieren sämtliche Momente $E\big(|X|^n\big)$, $n \in \mathbf{N}$. Wegen

$$e^{tX} = \lim_{n\to\infty} G_n(t;X) \quad (|t| \leq \delta)$$

sind also die Voraussetzungen des Satzes von der majorisierten Konvergenz erfüllt (Satz 2.2.2 f)), d.h. es gilt

$$\begin{aligned}\Psi_X(t) = E\big(e^{tX}\big) &= \lim_{n\to\infty} E\big(G_n(t;X)\big) = \lim_{n\to\infty} E\Big(\sum_{k=0}^{n} \frac{X^k}{k!} t^k\Big) \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} \frac{E(X^k)}{k!} t^k = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k,\end{aligned}$$

wie behauptet. Die Differenzierbarkeit von Ψ_X im Nullpunkt ergibt sich nunmehr durch gliedweise Differentiation dieser Reihe. Beziehung (2.2.74) folgt dabei direkt aus (2.2.53).
Zum Beweis von (2.2.75) und (2.2.76) geht man analog vor unter Beachtung der Reihenentwicklung

$$s^x = \sum_{k=0}^{\infty} \frac{\prod_{i=0}^{k-1}(x-i)}{k!} (s-1)^k \quad (|s-1| < 1,\ x \in \mathbf{R}).$$

c) Unter der angegebenen Bedingung läßt sich ψ_X vermöge (2.2.39) darstellen als

$$\psi_X(s) = E\big(s^X\big) = \sum_{k=0}^{\infty} s^k P(X = k),$$

wobei die Reihe wegen $\sum_{k=0}^{\infty} P(X = k) = 1$ für $|s| \leq 1$ absolut konvergiert. Hieraus folgt die Behauptung.
d) Nach Lemma 2.1.7 sind mit X und Y auch e^{tX} und e^{tY}, $t \in I$, stochastisch unabhängig, woraus sofort

$$\begin{aligned}\Psi_{X+Y}(t) &= E\big(e^{t(X+Y)}\big) = E\big(e^{tX} \cdot e^{tY}\big) \\ &= E\big(e^{tX}\big) E\big(e^{tY}\big) = \Psi_X(t)\Psi_Y(t), \quad t \in I\end{aligned}$$

folgt. Die Aussage für ψ_{X+Y} ergibt sich völlig analog. ∎

Teil b) des letzten Satzes besagt also insbesondere auch, daß die Verteilung P^X eindeutig durch die Momente $E(X^k)$, $k \in \mathbf{N}$, bestimmt ist, wenn die Reihe in (2.2.73) einen positiven Konvergenzradius besitzt, d.h. wenn

$$\limsup_{k\to\infty} \sqrt[k]{\frac{E(X^k)}{k!}} < \infty \tag{2.2.80}$$

gilt, da in diesem Fall die momenterzeugende Funktion Ψ_X in einem Intervall der Form $[-\delta, \delta]$ für ein $\delta > 0$ existiert und durch die Reihe in (2.2.73) dargestellt wird. Allerdings gibt es auch Verteilungen, für die dies nicht zutrifft; eine solche wird z.B. in Aufgabe 2.14 behandelt.

Wir wollen die Nützlichkeit insbesondere der Beziehungen (2.2.73), (2.2.74) und (2.2.76) an einigen Verteilungen beispielhaft verdeutlichen.

Ist z.B. X $\mathcal{E}(\lambda)$–verteilt mit $\lambda > 0$, so ist für $t < \lambda = t_+$

$$\Psi_X(t) = \lambda \int_0^\infty e^{(t-\lambda)x}\, dx = \frac{\lambda}{\lambda - t} \tag{2.2.81}$$

mit der Entwicklung (geometrische Reihe)

$$\Psi_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - \frac{t}{\lambda}} = \sum_{k=0}^{\infty} \left(\frac{t}{\lambda}\right)^k, \quad |t| < \lambda, \tag{2.2.82}$$

d.h. es ist

$$E(X^k) = \frac{k!}{\lambda^k}, \quad k \in \mathbf{N}$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Unter Verwendung der Multiplikationsregel (2.2.78) erhält man hieraus sofort auch die momenterzeugende Funktion Ψ_Y einer $\mathcal{E}(n, \lambda)$–Erlang–verteilten Zufallsvariablen Y ($n \in \mathbf{N}$) als

$$\Psi_Y(t) = \Big(\Psi_X(t)\Big)^n = \Big(\frac{\lambda}{\lambda - t}\Big)^n, \quad t < \lambda \tag{2.2.83}$$

(vgl. auch (2.1.82)).

Ist X $\mathcal{N}(0,1)$–normalverteilt, so erhält man entsprechend für alle $t \in \mathbf{R}$

$$\begin{aligned} \Psi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} e^{t^2/2}\, dx \\ &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx = e^{t^2/2}. \end{aligned} \tag{2.2.84}$$

Die Zufallsvariable $Y = \sigma X + \mu$ mit $\sigma > 0$, $\mu \in \mathbf{R}$, ist dann nach den Ausführungen im Anschluß an (2.1.88) $\mathcal{N}(\mu, \sigma^2)$–verteilt; ihre momenterzeugende Funktion ist also gegeben durch

$$\begin{aligned} \Psi_Y(t) &= E\big(e^{tY}\big) = E\big(e^{\mu t} e^{\sigma t X}\big) \\ &= e^{\mu t} \Psi_X(\sigma t) = \exp\Big(\frac{\sigma^2 t^2}{2} + \mu t\Big), \quad t \in \mathbf{R}, \end{aligned} \tag{2.2.85}$$

woraus $\Psi'_Y(t) = (\sigma^2 t + \mu)\Psi_Y(t)$, $\Psi''_Y(t) = \big(\sigma^2 + (\sigma^2 t + \mu)^2\big)\Psi_Y(t)$, $t \in \mathbf{R}$, also

$$E(Y) = \Psi'_Y(0) = \mu$$
$$\mathrm{Var}\,(Y) = \Psi''_Y(0) - \big(\Psi'_Y(0)\big)^2 = \sigma^2$$

folgt, wodurch die Bedeutung der Parameter μ und σ^2 als Erwartungswert und Varianz der $\mathcal{N}(\mu, \sigma^2)$–Verteilung deutlich wird. Insbesondere liefert die Multiplikationsregel (2.2.78) auch einen Beweis der Beziehung (2.1.89): ist nämlich Z eine weitere $\mathcal{N}(\nu, \tau^2)$– verteilte Zufallsvariable, unabhängig von Y mit $\nu \in \mathbf{R}$, $\tau > 0$, so gilt

$$\Psi_{Y+Z}(t) = \Psi_Y(t)\Psi_Z(t) = \exp\Big(\frac{\sigma^2 t^2}{2} + \mu t + \frac{\tau^2 t^2}{2} + \nu t\Big)$$
$$= \exp\Big(\frac{(\sigma^2 + \tau^2)t^2}{2} + (\mu + \nu)t\Big), \quad t \in \mathbf{R},$$

d.h. Ψ_{Y+Z} ist die momenterzeugende Funktion einer $\mathcal{N}(\mu+\nu, \sigma^2+\tau^2)$–Verteilung.

Für die geometrische Verteilung $\mathfrak{G}^+(p)$, $0 < p < 1$, erhält man beispielsweise mit der Abkürzung $q = 1 - p$

$$\psi_X(s) = \sum_{k=0}^{\infty} pq^k s^k = \frac{p}{1 - qs}, \quad |s| < \frac{1}{q}, \tag{2.2.86}$$

also $\psi'_X(s) = \dfrac{pq}{(1-qs)^2}$, $\psi''_X(s) = \dfrac{2pq^2}{(1-qs)^3}$ und damit gemäß (2.2.76)

$$E(X) = \psi'_X(1) = \frac{q}{p}$$
$$\mathrm{Var}\,(X) = \psi''_X(1) + \psi'_X(1)\big(1 - \psi'_X(1)\big) = \frac{q}{p^2}.$$

Die Multiplikationsregel führt hier unmittelbar zu

$$\psi_Y(s) = \Big(\frac{p}{1 - qs}\Big)^n, \quad n \in \mathbf{N},\ |s| < \frac{1}{q} \tag{2.2.87}$$

für eine $\overline{\mathfrak{B}}^+(n, p)$–verteilte Zufallsvariable Y und daher mit analoger Rechnung zu

$$E(Y) = \psi'_Y(1) = \frac{nq}{p}$$
$$\mathrm{Var}\,(Y) = \psi''_Y(1) + \psi'_Y(1)\big(1 - \psi'_Y(1)\big) = \frac{nq}{p^2}.$$

Das letzte Ergebnis folgt übrigens auch direkt aus (2.2.87) unter Verwendung der Linearität des Erwartungswerts und der Additivitätseigenschaft (2.2.58) der Varianz bei stochastischer Unabhängigkeit der Summanden.

Ist X schließlich $\mathfrak{B}(n,p)$–verteilt ($n \in \mathbf{N}$, $0 < p < 1$), so ergibt sich durch analoge Rechnung mit denselben Bezeichnungen

$$\psi_X(s) = (q+ps)^n, \quad s \in \mathbf{R}, \tag{2.2.88}$$

also

$$E(X) = \psi_X'(1) = np$$
$$\mathrm{Var}(X) = \psi_X''(1) + \psi_X'(1)\big(1 - \psi_X'(1)\big) = npq.$$

Für eine $\mathfrak{P}(\lambda)$–Poisson–verteilte Zufallsvariable X mit $\lambda > 0$ gilt entsprechend

$$\psi_X(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} s^k = e^{\lambda(s-1)}, \qquad s \in \mathbf{R}, \tag{2.2.89}$$

mit

$$E(X) = \psi_X'(1) = \lambda$$
$$\mathrm{Var}(X) = \psi_X''(1) + \psi_X'(1)\big(1 - \psi_X'(1)\big) = \lambda.$$

Die folgenden Tabellen enthalten für einige wichtige Verteilungen jeweils Erwartungswert, Varianz sowie die erzeugende bzw. momenterzeugende Funktion (vgl. hierzu auch Aufgabe 2.15).

P^X		$E(X)$	$\mathrm{Var}(X)$	$\psi_X(s)$	
$\mathfrak{L}\big(\{1,2,\ldots,n\}\big)$	$(n\in\mathbf{N})$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{s^{n+1}-s}{n(s-1)}$	$(s\neq 1)$
$\mathfrak{B}(n,p)$	$(n\in\mathbf{N}, 0\le p\le 1)$	np	$np(1-p)$	$(1-p+ps)^n$	
$\mathfrak{PB}(n;p_1,\ldots,p_n)$	$(0\le p_i\le 1)$	$\sum_{k=1}^{n} p_i$	$\sum_{k=1}^{n} p_i(1-p_i)$	$\prod_{k=1}^{n}(1-p_i+p_is)$	
$\overline{\mathfrak{B}}^+(n,p)$	$(n\in\mathbf{N}, 0<p\le 1)$	$n\frac{1-p}{p}$	$n\frac{1-p}{p^2}$	$\Big(\frac{p}{1-(1-p)s}\Big)^n$	$\big(\lvert s\rvert<\frac{1}{1-p}\big)$
$\mathfrak{P}(\lambda)$	$(\lambda>0)$	λ	λ	$e^{\lambda(s-1)}$	

P^X		$E(X)$	$\mathrm{Var}(X)$	$\Psi_X(t)$	
$\mathcal{R}\big([a,b]\big)$	$(a<b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{t(b-a)}$	$(t\neq 0)$
$\mathcal{E}(\lambda)$	$(\lambda>0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{\lambda-t}$	$(t<\lambda)$
$\Gamma(\alpha,\lambda)$	$(\alpha,\lambda>0)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\Big(\frac{1}{\lambda-t}\Big)^{\alpha}$	$(t<\lambda)$
$\mathcal{N}(\mu,\sigma^2)$	$(\mu\in\mathbf{R}, \sigma^2>0)$	μ	σ^2	$\exp\Big(\frac{\sigma^2t^2}{2}+\mu t\Big)$	

Der Begriff der (moment–)erzeugenden Funktion für Zufallsvariablen X läßt sich auch noch in natürlicher Weise auf Zufallsvektoren $\boldsymbol{X} = (X_1, \ldots, X_m)$, $m \in \mathbb{N}$, übertragen. Existieren beispielsweise alle momenterzeugenden Funktionen $\Psi_{X_i}(t_i)$, $1 \le i \le m$ für $|t_i| < \delta_i$ mit Zahlen $\delta_i > 0$, $1 \le i \le m$, so auch das Produkt $E\big(\exp \sum_{k=1}^m t_i X_i\big)$, $|t_i| < \delta_i$, $1 \le i \le m$, wie man sich analog zu Lemma 2.2.3 a) mit Hilfe der Hölder–Ungleichung (2.2.63) leicht überlegen kann. Die momenterzeugende Funktion $\Psi_{\boldsymbol{X}}(t_1, \ldots, t_m)$ ist dann definiert durch

$$\Psi_{\boldsymbol{X}}(t_1, \ldots, t_m) = E\big(\exp \sum_{k=1}^m t_i X_i\big), \qquad |t_i| < \delta_i,\ 1 \le i \le m. \tag{2.2.90}$$

Analog erklärt man die wahrscheinlichkeitserzeugende Funktion $\psi_{\boldsymbol{X}}(s_1, \ldots, s_m)$ durch

$$\psi_{\boldsymbol{X}}(s_1, \ldots, s_m) = E\big(\prod_{k=1}^m s_i^{X_i}\big), \qquad |s_i - 1| < 1 - e^{-\delta_i},\ 1 \le i \le m. \tag{2.2.91}$$

Ähnlich wie im Fall von Zufallsvariablen kann man dann auch Momente der Komponenten von $\boldsymbol{X}$ durch Differentiation erhalten, etwa

$$\begin{aligned} E(X_j) &= \frac{\partial}{\partial t_j} \Psi_{\boldsymbol{X}}(\boldsymbol{0}) \\ E(X_j X_k) &= \frac{\partial^2}{\partial t_j\, \partial t_k} \Psi_{\boldsymbol{X}}(\boldsymbol{0}) \end{aligned} \qquad (1 \le j, k \le m) \tag{2.2.92}$$

oder allgemeiner

$$E\Big(\prod_{j=1}^k X_{i_j}\Big) = \frac{\partial^k}{\partial t_{i_1} \ldots \partial t_{i_k}} \Psi_{\boldsymbol{X}}(\boldsymbol{0}) \qquad (1 \le i_1 \le \ldots \le i_k \le m), \tag{2.2.93}$$

wobei $\boldsymbol{0} = (0, \ldots, 0)$ den Nullvektor bezeichne. Für die wahrscheinlichkeitserzeugende Funktion ergeben sich entsprechende Formeln, z.B.

$$\begin{aligned} E(X_j) &= \frac{\partial}{\partial t_j} \psi_{\boldsymbol{X}}(\boldsymbol{1}) \\ E(X_j X_k) &= \frac{\partial^2}{\partial t_j\, \partial t_k} \psi_{\boldsymbol{X}}(\boldsymbol{1}) \end{aligned} \qquad (1 \le j, k \le m,\ j \ne k) \tag{2.2.94}$$

mit der Bezeichnung $\boldsymbol{1} = (1, \ldots, 1)$.

Nehmen die Zufallsvariablen $X_1, \ldots, X_m$ wieder (P–f.s.) nur Werte in $\mathbb{N}_0$ an, gilt analog zu (2.2.77) auch

$$\psi_{\boldsymbol{X}}(s_1, \ldots, s_m) = \sum_{k_1=0}^{\infty} \ldots \sum_{k_m=0}^{\infty} P(X_1 = k_1, \ldots, X_m = k_m) s_1^{k_1} \cdot \ldots \cdot s_m^{k_m} \tag{2.2.95}$$

für $|s_1|, \ldots |s_m| \leq 1$.

Die letzten beiden Beziehungen sind geeignet, die Kovarianzen der Komponenten multinomialverteilter Zufallsvektoren zu berechnen, die etwa in Beispiel (2.1.1) (`hybridsort`) auftreten. Die erzeugende Funktion eines $\mathfrak{M}(n; p_1, \ldots, p_m)$-multinomialverteilten Zufallsvektors $\boldsymbol{X}$ ($n \in \mathbf{N}$, $0 \leq p_j \leq 1$, $1 \leq j \leq m$) ist nämlich gegeben durch

$$\begin{aligned} \psi_{\boldsymbol{X}}(s_1, \ldots, s_m) = E\Big(\prod_{k=1}^{m} s_i^{X_i}\Big) &= \sum_{\substack{0 \leq i_1, \ldots, i_m \leq n \\ \sum_{j=1}^{m} i_j = n}} \binom{n}{i_1, \ldots, i_m} \prod_{j=1}^{m} (p_j s_j)^{i_j} \\ &= \Big(\sum_{j=1}^{m} p_j s_j\Big)^n, \qquad s_1, \ldots, s_m \in \mathbf{R}. \end{aligned} \tag{2.2.96}$$

Gemäß (2.2.94) erhält man dann durch einfache Rechnung

$$\begin{aligned} E(X_j) &= np_j \\ E(X_j X_k) &= n(n-1)p_j p_k \qquad (1 \leq j, k \leq m,\ j \neq k). \\ \operatorname{Kov}(X_j X_k) &= E(X_j X_k) - E(X_j)E(X_k) = -np_j p_k \end{aligned}$$

Die Komponenten $X_1, \ldots, X_m$ sind also *negativ* korreliert mit

$$\operatorname{Korr}(X_j X_k) = -\sqrt{\frac{p_j p_k}{(1-p_j)(1-p_k)}}, \qquad 1 \leq j, k \leq m,\ j \neq k,$$

was wegen $\sum_{k=1}^{m} X_k = n$ auch zu erwarten war.

Zum Abschluß dieses Kapitels greifen wir noch einmal Fragestellungen zur Konvergenz von Verteilungen und Zufallsvariablen auf, wobei in diesem Zusammenhang auch die im CAD eingesetzten Bézier-Kurven und -Flächen besprochen werden.

2.3. Grenzwertsätze

Ausgangspunkt unserer Überlegungen ist das in Satz 2.2.6 formulierte (schwache) Gesetz der großen Zahlen, welches zu folgendem allgemeinen Konvergenzbegriff für eine Folge von Zufallsvariablen Anlaß gibt.

Definition 2.3.1. *(stochastische Konvergenz)*
Es seien X_n, $n \in \mathbb{N}$, und X Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Die Folge $\{X_n\}_{n\in\mathbb{N}}$ heißt stochastisch konvergent gegen X, wenn gilt:

$$\lim_{n\to\infty} P\big(|X_n - X| > \varepsilon\big) = 0 \quad \text{für alle } \varepsilon > 0, \tag{2.3.1}$$

in Zeichen: $X_n \xrightarrow{P} X$ $(n \to \infty)$.

Man macht sich leicht klar, daß im Falle stochastischer Konvergenz der Grenzwert X P–f.s. eindeutig bestimmt ist, d.h. ist Y eine weitere Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit $X_n \xrightarrow{P} Y$ $(n \to \infty)$, so ist $P(X = Y) = 1$. Für beliebiges $\varepsilon > 0$ und $n \in \mathbb{N}$ ist nämlich

$$\begin{aligned} P(|X - Y| > \varepsilon) &= P\big(\big|(X - X_n) - (Y - X_n)\big| > \varepsilon\big) \\ &\le P\big(\{|X - X_n| > \tfrac{\varepsilon}{2}\} \cup \{|Y - X_n| > \tfrac{\varepsilon}{2}\}\big) \\ &\le P\big(\{|X - X_n| > \tfrac{\varepsilon}{2}\}\big) + P\big(\{|Y - X_n| > \tfrac{\varepsilon}{2}\}\big); \end{aligned}$$

die rechte Seite in (2.3.2) strebt aber nach Voraussetzung gegen 0, so daß $P(|X - Y| > \varepsilon) = 0$ gilt für alle $\varepsilon > 0$, d.h. es ist $P(X = Y) = 1$.

Das Gesetz der großen Zahlen in der Form des Satzes 2.2.6 besagt also gerade, daß — mit den dortigen Bezeichnungen — die arithmetischen Mittel S_n stochastisch gegen den Erwartungswert μ konvergieren.

Einen Zusammenhang zwischen stochastischer und fast sicherer Konvergenz stellt das folgende Resultat her.

Satz 2.3.1. *Unter den Voraussetzungen von Definition 2.3.1 gilt:*

a) Konvergiert die Folge $\{X_n\}_{n\in\mathbb{N}}$ P–f.s. gegen X, so auch stochastisch, d.h. es gilt

$$X_n \longrightarrow X \ P-\text{f.s.} \quad \Longrightarrow \quad X_n \xrightarrow{P} X \quad (n \to \infty). \tag{2.3.2}$$

b) Konvergiert die Folge $\{\sup_{m\ge n} |X_m - X|\}_{n\in\mathbb{N}}$ stochastisch gegen 0, so konvergiert die Folge $\{X_n\}_{n\in\mathbb{N}}$ P–fast sicher gegen X, d.h. es gilt

$$\sup_{m\ge n} |X_m - X| \xrightarrow{P} 0 \quad \Longrightarrow \quad X_n \longrightarrow X \ P-\text{f.s.} \quad (n \to \infty). \tag{2.3.3}$$

Beweis. a) Für jedes $\varepsilon > 0$ ergibt sich die folgende Ungleichungskette

$$\begin{aligned} 0 &\le \lim_{n\to\infty} P(|X_n - X| > \varepsilon) \le \lim_{n\to\infty} P\Big(\bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\Big) \\ &= P\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\Big) = P\big(\limsup_{n\to\infty} \{|X_n - X| > \varepsilon\}\big) \\ &= P\big(|X_n - X| > \varepsilon \text{ für } \infty \text{ viele } n\big) \le P\big(\{\lim_{n\to\infty} X_n = X\}^c\big) = 0, \end{aligned} \tag{2.3.4}$$

was Teil a) des Satzes beweist.
b) Es bezeichne $D_n = \sup_{m \geq n} |X_m - X|$, $n \in \mathbf{N}$. Dann ist die Folge $\{D_n\}_{n \in \mathbf{N}}$ schwach monoton fallend und nach unten durch 0 beschränkt, also konvergent gegen eine Zufallsvariable $D \geq 0$, die wegen der vorausgesetzten stochastischen Konvergenz sogar P–fast sicher endlich ist. Mit dem gerade bewiesenen Teil a) des Satzes folgt aber auch $D_n \xrightarrow{P} D$ $(n \to \infty)$, also wegen der P–fast sicheren Eindeutigkeit des stochastischen Limes $D = 0$ P–f.s., da nach Voraussetzung ja $\{D_n\}_{n \in \mathbf{N}}$ stochastisch gegen 0 strebt. Es konvergiert also D_n f.s. gegen 0 und damit X_n f.s. gegen X $(n \to \infty)$, was zu zeigen war. ∎

Die beiden Aussagen des Satzes 2.3.1 lassen sich offensichtlich auch wie folgt zusammenfassen:

$$X_n \longrightarrow X \; P-\text{f.s.} \qquad \Longleftrightarrow \qquad \sup_{m \geq n} |X_m - X| \xrightarrow{P} 0 \quad (n \to \infty). \tag{2.3.5}$$

Das folgende Beispiel zeigt, daß die Aussage a) des letzten Satzes nicht direkt umkehrbar ist, d.h. stochastische Konvergenz impliziert i.a. nicht fast–sichere Konvergenz:
Man wähle hierzu den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P) = \big([0,1],\ [0,1] \cap \mathcal{B}^1,\ \mathcal{R}([0,1])\big)$ sowie die durch

$$X_{n^2+k} = \mathbb{1}_{\left[\frac{k-1}{2n+1}, \frac{k}{2n+1}\right]}, \quad 1 \leq k \leq 2n+1,\ n \in \mathbf{N}_0$$

definierten Zufallsvariablen $\{X_n\}_{n \in \mathbf{N}}$. Dann ist für alle $\varepsilon > 0$

$$P\big(|X_n| > \varepsilon\big) = P\big(|X_n| = 1\big) \leq \frac{1}{2\sqrt{n} - 1}, \quad n \in \mathbf{N},$$

d.h. es gilt $X_n \xrightarrow{P} 0$ $(n \to \infty)$. Die Folge $\{X_n\}_{n \in \mathbf{N}}$ konvergiert aber nicht P–f.s. gegen 0, da $\sup_{m \geq n} |X_m| = 1$ gilt für alle $n \in \mathbf{N}$, d.h. es ist $\limsup_{n \to \infty} |X_n| = 1$.
Für die nachfolgenden Untersuchungen ist es nützlich, eine etwas allgemeinere Version des schwachen Gesetzes der großen Zahlen zur Verfügung zu haben, als es in Satz 2.2.6 formuliert wurde.

Satz 2.3.2. *(allgemeines schwaches Gesetz der großen Zahlen)*
Es sei $\{X_n\}_{n \in \mathbf{N}}$ eine Folge paarweise unkorrelierter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$E(X_n) = \mu \in \mathbf{R} \quad \text{und} \quad \lim_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^{n} \operatorname{Var}(X_k) = 0. \tag{2.3.6}$$

Dann gilt für die arithmetischen Mittel $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ der Folge $\{X_n\}_{n \in \mathbf{N}}$:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu \qquad (n \to \infty). \tag{2.3.7}$$

Beweis. Es ist $E(S_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$ für alle $n \in \mathbf{N}$ sowie aufgrund von (2.2.56) und mehrfacher Anwendung von (2.2.58) $\mathrm{Var}(S_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i)$. Die Aussage ergibt sich also wieder aus der Tschebyscheff–Ungleichung (2.2.59) mit $X = S_n$ und $c = \varepsilon$. ∎

Will man in dem schwachen Gesetz der großen Zahlen in der Form des letzten Satzes P–fast sichere Konvergenz der arithmetischen Mittel S_n gegen μ erhalten, muß man dafür Sorge tragen, daß $\sup_{m \geq n} |S_m - \mu|$ mit $n \to \infty$ stochastisch gegen 0 strebt. Der folgende Satz liefert hierzu eine Möglichkeit.

Satz 2.3.3. *(allgemeines starkes Gesetz der großen Zahlen)*
Es sei $\{X_n\}_{n\in\mathbf{N}}$ eine Folge stochastisch unabhängiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$E(X_n) = \mu \in \mathbf{R},\ n \in \mathbf{N} \quad \text{und} \quad \sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^2} < \infty. \tag{2.3.8}$$

Dann gilt für die arithmetischen Mittel $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ der Folge $\{X_n\}_{n\in\mathbf{N}}$:

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \mu \quad P-\text{f.s.} \tag{2.3.9}$$

Beweis. Wir beweisen den Satz unter der stärkeren Annahme

$$M := \sup_{n\in\mathbf{N}} E\big(X_n^4\big) < \infty. \tag{2.3.10}$$

(Einen vollständigen Beweis findet man etwa in Bauer (1978), Satz 37.1.) Aufgrund der Jensen'schen Ungleichung (2.2.47) sind dann auch alle Varianzen $\mathrm{Var}(X_n)$ beschränkt mit

$$\mathrm{Var}(X_n) \leq E\big(X_n^2\big) \leq \sqrt{E\big(X_n^4\big)} \leq \sqrt{M}, \quad n \in \mathbf{N}.$$

Die Konvergenzbedingung in (2.3.8) ist also unter (2.3.10) erfüllt. Mit der Bezeichnung $Y_n = X_n - \mu$, $n \in \mathbf{N}$, folgt dann aus (2.3.10) unter Heranziehen von (2.2.60) zunächst

$$\sup_{n\in\mathbf{N}} E\big(Y_n^4\big) \leq 16 \sup_{n\in\mathbf{N}} \Big(E\big(X_n^4\big) + \mu^4\Big) \leq 16\big(M + \mu^4\big) =: M^*,$$

wobei jetzt $E(Y_n) = 0$, $\mathrm{Var}(Y_n) = \mathrm{Var}(X_n)$ gilt für alle $n \in \mathbf{N}$. Wir entwickeln nun den Ausdruck $E\Big(\Big(\sum_{k=1}^{n} Y_k\Big)^4\Big)$, $n \in \mathbf{N}$, wie folgt:

$$\begin{aligned}
E\Big(\Big(\sum_{k=1}^{n} Y_k\Big)^4\Big) &= E\Big(\sum_{1\leq k_1,k_2,k_3,k_4\leq n} \sum\sum\sum\sum Y_{k_1} Y_{k_2} Y_{k_3} Y_{k_4}\Big) \\
&= \sum_{k=1}^{n} E\big(Y_k^4\big) + 2 \sum_{1\leq j<k\leq n} \sum\sum E\big(Y_j^2\big) E\big(Y_k^2\big) \\
&\leq \sum_{k=1}^{n} E\big(Y_k^4\big) + \Big(\sum_{k=1}^{n} E\big(Y_k^2\big)\Big)^2 \\
&\leq nM^* + n^2 M \leq 2n^2 M^*, \quad n \in \mathbf{N}.
\end{aligned}$$

Hierbei ist zu beachten, daß wegen (2.2.27) alle Terme der Form $E(Y_j \cdot Y_k^3)$, $1 \leq j,k \leq n$, $j \neq k$, verschwinden. Für die arithmetischen Mittel $S_n^* = \frac{1}{n}\sum_{i=1}^n Y_i = S_n - \mu$ der Folge $\{Y_n\}_{n\in\mathbf{N}}$ erhält man hieraus also die Ungleichung

$$E(|S_n - \mu|^4) \leq \frac{2n^2 M^*}{n^4} = \frac{2M^*}{n^2}, \quad n \in \mathbf{N},$$

so daß die Reihe $\sum_{n=1}^\infty E(|S_n - \mu|^4)$ konvergiert. Man erhält damit unter Verwendung der Markoff–Ungleichung (2.2.26) für jedes $\varepsilon > 0$ die Abschätzung

$$\begin{aligned} P(\sup_{m\geq n} |S_m - \mu| > \varepsilon) &= P(\bigcup_{m=n}^\infty \{|S_m - \mu| > \varepsilon\}) \\ &\leq \sum_{m=n}^\infty P(|S_m - \mu| > \varepsilon) \leq \sum_{m=n}^\infty \frac{E(|S_m - \mu|^4)}{\varepsilon^4} \\ &\leq \frac{2M^*}{\varepsilon^4} \sum_{m=n}^\infty \frac{1}{m^2} \leq \frac{2M^*}{\varepsilon^4} \cdot \frac{1}{n-1}, \quad n \geq 2. \end{aligned}$$

Damit gilt aber: $\sup_{m\geq n} |S_m - \mu| \xrightarrow{P} 0$, $n \to \infty$, also auch $S_n \longrightarrow \mu$ $P-$f.s., $n \to \infty$, nach Satz 2.3.1 b). ∎

Nimmt man spezieller an, daß die Folge $\{X_n\}_{n\in\mathbf{N}}$ in Satz 2.3.3 *identisch* verteilt ist, kann man auf die Varianz–Bedingung in (2.3.8) sogar gänzlich verzichten.

Satz 2.3.4. *(Gesetz der großen Zahlen)*
Es sei $\{X_n\}_{n\in\mathbf{N}}$ eine Folge stochastisch unabhängiger und identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$E(X_n) = \mu \in \mathbf{R}, \quad n \in \mathbf{N}. \tag{2.3.11}$$

Dann gilt für die arithmetischen Mittel $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ der Folge $\{X_n\}_{n\in\mathbf{N}}$ wieder:

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \mu \quad P-\text{f.s.} \tag{2.3.12}$$

Der Beweis dieses Satzes ist technisch aufwendiger, weshalb wir hier auf ihn verzichten; vgl. etwa Billingsley (1986), Theorem 22.1 oder Bauer (1978), Satz 37.2.

In Bezug auf die Average–Case–Analyse von Algorithmen lassen sich die beiden letzten Resultate dahingehend interpretieren, daß die mittlere Laufzeit $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ nach n Aufrufen (vgl. die Ausführungen im Anschluß an Satz 2.2.6) nicht nur stochastisch, sondern sogar P–fast sicher gegen den Erwartungswert μ konvergiert; praktisch bedeutet dies, daß "fast jede" Beobachtungsfolge, die aus stochastisch unabhängigen und identisch verteilten Zufallsvariablen gewonnen wird, durch entsprechende Mittelbildung langfristig zum Erwartungswert führt.

Die Beziehungen (2.3.9) und (2.3.12) lassen sich dabei auch so formulieren:

$$\lim_{n\to\infty} S_n - \mu = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) = 0 \qquad P - \text{f.s.} \tag{2.3.13}$$

Tatsächlich zeigt der Beweis des Satzes 2.3.3 aber, daß hier sogar eine schärfere Aussage möglich ist, nämlich

$$\lim_{n\to\infty} n^{\alpha}(S_n - \mu) = \lim_{n\to\infty} \frac{1}{n^{1-\alpha}}\sum_{i=1}^{n}(X_i - \mu) = 0 \qquad P - \text{f.s.} \tag{2.3.14}$$

für $0 < \alpha < \frac{1}{4}$; Beziehung (2.3.17) ergibt nämlich auch

$$E\big((n^{\alpha}|S_n - \mu|)^4\big) \le \frac{2M^*}{n^{2-4\alpha}}, \quad n \in \mathbb{N}, \tag{2.3.15}$$

so daß die Reihe $\sum_{n=1}^{\infty} E\big((n^{\alpha}|S_n - \mu|)^4\big)$ immer noch konvergiert. Neben der reinen Konvergenzaussage lassen sich also auch Aussagen zur *Konvergenzgeschwindigkeit* im Gesetz der großen Zahlen gewinnen. Der folgende Satz zeigt, daß in der Situation des Satzes 2.3.4 diese Konvergenzgeschwindigkeit exakt bestimmt werden kann, wenn man die Endlichkeit der Varianz der Verteilungen voraussetzt.

Satz 2.3.5. *(Gesetz vom iterierten Logarithmus)*
Es sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge stochastisch unabhängiger, identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$E(X_n) = \mu \in \mathbb{R} \quad \text{und} \quad 0 < \operatorname{Var}(X_n) = \sigma^2 < \infty \tag{2.3.16}$$

für alle $n \in \mathbb{N}$. Dann gilt für die arithmetischen Mittel $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ der Folge $\{X_n\}_{n\in\mathbb{N}}$:

$$\begin{aligned} \limsup_{n\to\infty} \sqrt{\frac{n}{2\ln\ln n}}(S_n - \mu) &= \limsup_{n\to\infty} \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{2n\ln\ln n}} = \sigma \quad P - \text{f.s.} \\ \liminf_{n\to\infty} \sqrt{\frac{n}{2\ln\ln n}}(S_n - \mu) &= \liminf_{n\to\infty} \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{2n\ln\ln n}} = -\sigma \quad P - \text{f.s.} \end{aligned} \tag{2.3.17}$$

Auch der Beweis dieses Satzes ist zu umfangreich, um hier darauf eingehen zu können; vgl. etwa Billingsley (1986), Theorem 9.5 oder Gänssler & Stute (1977), Satz 4.3.20.

Das Wachstum des Terms $\ln\ln n$ in (2.3.17) (von dem der Satz seinen Namen hat) ist äußerst langsam; so gilt etwa $\ln\ln n \le 3$ für $n \le 5 \times 10^8$. Von Interesse ist daher das Grenzverhalten von $\sqrt{n}(S_n - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu)$ für $n \to \infty$. Nach Satz 2.3.5 kann man natürlich unter den dortigen Voraussetzungen keine fast sichere Konvergenz mehr erwarten; entsprechendes gilt für stochastische Konvergenz. Andererseits zeigt die Rechnung in (2.1.88), daß sich

für den Fall von $\mathcal{E}(1)$–exponentialverteilten Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ die Verteilung von $\sqrt{n}(S_n-\mu)=\dfrac{\sum_{i=1}^n X_i - n}{\sqrt{n}}$ für große n einer $\mathcal{N}(0,1)$–Normalverteilung "annähert", wobei allerdings zu klären bleibt, in welchem Sinne diese Approximation zu verstehen ist. Wir werden damit zunächst auf das allgemeinere Problem geführt, was wir unter der Konvergenz einer Folge von *Verteilungen* $\{Q_n\}_{n\in\mathbf{N}}$ gegen eine Verteilung Q auf einem Meßraum $(\mathcal{X},\mathcal{B})$ zu verstehen haben. Es ist dabei i.a. nicht sinnvoll, diese Konvergenz "punktweise" zu verstehen, d.h. zu verlangen, daß $Q_n(B)\longrightarrow Q(B)$ $(n\to\infty)$ für *jede* Menge $B\in\mathcal{B}$ Gültigkeit haben soll. Das folgende einfache Beispiel verdeutlicht dies: man wähle etwa mit $(\mathcal{X},\mathcal{B})=(\mathbf{R},\mathcal{B}^1)$ die (konstanten) Zufallsvariablen $X_n=a_n$ $(n\in\mathbf{N})$, wobei $\{a_n\}_{n\in\mathbf{N}}$ eine beliebige, gegen $a\in\mathbf{R}$ konvergente Zahlenfolge sei, und $Q_n=P^{X_n}$, $n\in\mathbf{N}$; es konvergiert dann die Folge $\{X_n\}_{n\in\mathbf{N}}$ überall gegen die Zufallsvariable $X=a$, jedoch *keine* der Folgen $\{Q_n(B)\}_{n\in\mathbf{N}}$ konvergiert, wenn z.B. $a\in\mathcal{B}\in\mathcal{B}^1$ gilt und $B\cap\{a_1,a_2,\ldots\}$ endlich ist. Der punktweise Konvergenzbegriff ist also für die Konvergenz von Verteilungen zu eng; wir betrachten deshalb im folgenden einen etwas schwächeren Konvergenzbegriff, der sich für stochastische Fragestellungen als besonders geeignet erweist.

Definition 2.3.2. *(schwache Konvergenz)*
Es seien $\{Q_n\}_{n\in\mathbf{N}}$, Q Verteilungen auf dem Meßraum $(\mathcal{X},\mathcal{B})=(\mathbf{R},\mathcal{B}^1)$. Die Folge $\{Q_n\}_{n\in\mathbf{N}}$ heißt schwach (oder auch verteilungs–)konvergent gegen Q, wenn

$$\int g\,dQ_n\longrightarrow\int g\,dQ\qquad(n\to\infty)\tag{2.3.18}$$

gilt für jede auf $\mathbf{R}$ stetige, beschränkte Funktion g, in Zeichen:

$$Q_n\xrightarrow{w}Q\qquad(n\to\infty).\tag{2.3.19}$$

Man beachte, daß die in Definition 2.3.2 auftretenden Funktionen g meßbar, also Zufallsvariable sind, und damit — wegen der Beschränktheit — auch Q_n– bzw. Q–integrierbar mit endlichem Wert für alle $n\in\mathbf{N}$. Das Symbol $\xrightarrow{w}$ leitet sich dabei aus dem englischen "weak convergence" ab.

Wir wollen zunächst zeigen, daß die Stetigkeitsbedingung an die Funktion g in Definition 2.3.2 zu einer gleichmäßigen Stetigkeitsbedingung äquivalent ist, was einige nachfolgende Beweisgänge vereinfacht.

Lemma 2.3.1. *Es seien $\{Q_n\}_{n\in\mathbf{N}}$, Q Verteilungen auf dem Meßraum $(\mathcal{X},\mathcal{B})=(\mathbf{R},\mathcal{B}^1)$. Die Folge $\{Q_n\}_{n\in\mathbf{N}}$ ist genau dann schwach konvergent gegen Q, wenn*

$$\int g\,dQ_n\longrightarrow\int g\,dQ\qquad(n\to\infty)\tag{2.3.20}$$

gilt für jede auf $\mathbf{R}$ gleichmäßig stetige, beschränkte Funktion g.

Beweis. Da jede gleichmäßig stetige Funktion g auch stetig ist, reicht es, das Hinreichen der angegebenen Bedingung zu zeigen. Sei dazu g stetig und beschränkt,

etwa $|g| \leq M \in \mathbf{R}$, sowie $\varepsilon > 0$. Wähle eine Zahl $a \in \mathbf{R}$ so, daß $\int_{|x|\geq a} dQ(x) \leq \frac{\varepsilon}{2M}$ gilt sowie stetige Funktionen g_*, g^* mit den Eigenschaften

$$\begin{aligned} &g_*(x) = g^*(x) = g(x), && |x| \leq a \\ &-M \leq g_*(x) \leq g(x), && a \leq |x| \leq a+1 \\ &g(x) \leq g^*(x) \leq M, && a \leq |x| \leq a+1 \\ &g_*(x) = -M, && |x| \geq a+1 \\ &g^*(x) = M, && |x| \geq a+1. \end{aligned} \tag{2.3.21}$$

Dann sind g_*, g^* gleichmäßig stetig, da die Einschränkung der stetigen Funktionen g_* und g^* auf das kompakte Intervall $[-a-1, a+1]$ jedenfalls gleichmäßig stetig ist (vgl. Heuser (1988), Satz 111.10) und g_* und g^* außerhalb dieses Intervalls konstant sind. Ferner ist

$$\int (g^* - g_*)\, dQ \leq 2M \int_{|x|\geq a} dQ(x) \leq \varepsilon.$$

Es folgt

$$\begin{aligned} \liminf_{n\to\infty} \int g\, dQ_n &\geq \liminf_{n\to\infty} \int g_*\, dQ_n = \int g_*\, dQ \\ &= \int g\, dQ - \int (g - g_*)\, dQ \\ &\geq \int g\, dQ - \int (g^* - g_*)\, dQ \geq \int g\, dQ - \varepsilon \end{aligned}$$

sowie analog

$$\limsup_{n\to\infty} \int g\, dQ_n \leq \int g\, dQ + \varepsilon.$$

Da ε beliebig war, folgt also $\lim_{n\to\infty} \int g\, dQ_n = \int g\, dQ$, was zu zeigen war. ∎

Sind die Verteilungen Q_n und Q speziell Verteilungen von Zufallsvariablen, d.h. gilt $Q_n = P^{X_n}$, $Q = P^X$ für Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$, X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, so läßt sich aufgrund des Transformationssatzes 2.2.3 Beziehung (2.3.17) auch schreiben als

$$E(g(X_n)) \longrightarrow E(g(X)) \qquad (n \to \infty). \tag{2.3.22}$$

Für das oben angegebene Beispiel erweist sich der Begriff der schwachen Konvergenz unmittelbar als geeignet: für stetige, beschränkte Funktionen g gilt nämlich definitionsgemäß

$$E(g(X_n)) = g(a_n) \longrightarrow g(a) = E(g(X)) \qquad (n \to \infty), \tag{2.3.23}$$

so daß hier die (fast sichere) Konvergenz der Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ die schwache Konvergenz der Verteilungen $\{P^{X_n}\}_{n\in\mathbf{N}}$ nach sich zieht; eine Eigenschaft, die sich als allgemeingültig erweist, da sogar die stochastische Konvergenz schwache Konvergenz impliziert.

Lemma 2.3.2. *Es seien $\{X_n\}_{n\in\mathbf{N}}$, X Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann gilt:*

$$X_n \xrightarrow{P} X \quad \Longrightarrow \quad P^{X_n} \xrightarrow{w} P^X \qquad (n \to \infty). \tag{2.3.24}$$

Ist X P–fast sicher konstant, gilt hiervon auch die Umkehrung:

$$P^{X_n} \xrightarrow{w} P^X \quad \Longrightarrow \quad X_n \xrightarrow{P} X \qquad (n \to \infty). \tag{2.3.25}$$

Beweis. Gemäß Lemma 2.3.1 reicht es, gleichmäßig stetige, beschränkte Funktionen g auf $\mathbf{R}$ zu betrachten; etwa $|g| \le M \in \mathbf{R}$. Wegen der gleichmäßigen Stetigkeit gibt es dann zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit der Eigenschaft

$$|x - y| \le \delta \quad \Longrightarrow \quad |g(x) - g(y)| \le \varepsilon, \quad x, y \in \mathbf{R}.$$

Für $n \in \mathbf{N}$ bezeichne $A_n = \{|X_n - X| \le \delta\}$. Es ist dann für alle $n \in \mathbf{N}$

$$\begin{aligned} \big|E\big(g(X_n)\big) - E\big(g(X)\big)\big| &\le \int_{A_n} |X_n - X|\, dP + \int_{A_n^c} |X_n - X|\, dP \\ &\le \varepsilon \int_{A_n} dP + 2M \int_{A_n^c} dP \\ &\le \varepsilon + 2M\, P\big(A_n^c\big) = \varepsilon + 2M\, P(|X_n - X| > \delta), \end{aligned} \tag{2.3.26}$$

also nach Voraussetzung der stochastischen Konvergenz

$$\lim_{n\to\infty} \big|E\big(g(X_n)\big) - E\big(g(X)\big)\big| \le \varepsilon$$

für jedes $\varepsilon > 0$. Der letztere Limes ist daher Null, was die erste Behauptung beweist.
Zum Beweis der Umkehrung nehmen wir an, daß $X = a$ P–f.s. gilt für ein $a \in \mathbf{R}$. Mit der Wahl der stetigen, beschränkten Funktion

$$g(x) = \max\big\{1, \frac{1}{\varepsilon}|x - a|\big\}, \qquad x \in \mathbf{R}$$

erhält man dann für jedes $\varepsilon > 0$

$$\begin{aligned} P\big(|X_n - X| > \varepsilon\big) &= E\big(\mathbb{1}_{\{|X_n - a| > \varepsilon\}}\big) \\ &\le E\big(g(X_n)\big) \longrightarrow E\big(g(X)\big) = g(a) = 0 \end{aligned}$$

für $n \to \infty$, d.h. es gilt $X_n \xrightarrow{P} X$, $n \to \infty$. ∎

Auch die Aussage in (2.3.24) kann nicht ohne weiteres umgekehrt werden: wählt man wieder den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P) = \big([0,1],\ [0,1] \cap \mathcal{B}^1,\ \mathcal{R}([0,1])\big)$ sowie die durch

$$X_n = \begin{cases} \mathbb{1}_{[0,\frac{1}{2})} & \text{falls } n \text{ gerade} \\ \mathbb{1}_{[\frac{1}{2},1]} & \text{falls } n \text{ ungerade,} \end{cases} \qquad n \in \mathbf{N},$$

definierten Zufallsvariablen, so ist $P^{X_n} = \mathfrak{B}\left(1, \frac{1}{2}\right)$ für alle $n \in \mathbf{N}$, d.h. es gilt auch $P^{X_n} \xrightarrow{w} Q = \mathfrak{B}\left(1, \frac{1}{2}\right)$ $(n \to \infty)$; $\{X_n\}_{n\in\mathbf{N}}$ konvergiert aber nicht stochastisch.

Das nachfolgende Lemma zeigt die Bedeutung der schwachen Konvergenz noch einmal aus anderer Sicht.

Lemma 2.3.3. *(Eigenschaften der schwachen Konvergenz)*
Es seien $\{Q_n\}_{n\in\mathbf{N}}$, Q *Verteilungen über* $(\mathbf{R},\mathcal{B}^1)$. F_n, $n\in\mathbf{N}$, F *bezeichne die Verteilungsfunktionen zu* Q_n, $n\in\mathbf{N}$, *bzw.* Q. *Die folgenden vier Aussagen sind äquivalent:*

a) $Q_n \xrightarrow{w} Q \quad (n\to\infty)$ (2.3.27)

b) $F_n(x) \longrightarrow F(x)$ für alle Stetigkeitspunkte x von $F \quad (n\to\infty)$ (2.3.28)

c) $F_n^{-1}(y) \longrightarrow F^{-1}(y)$ für alle Stetigkeitspunkte $y\in(0,1)$ von $F^{-1} \quad (n\to\infty)$ (2.3.29)

d) $Q_n(B) \longrightarrow Q(B)$ für alle $B\in\mathcal{B}^1$ mit $Q(\partial B)=0 \quad (n\to\infty)$, (2.3.30)

wobei wieder ∂B *den Rand der Menge* B *bezeichne.*
Ferner gilt:

e) *Ist* U *eine* $\mathcal{R}((0,1))$*–verteilte Zufallsvariable und konvergiert für* $n\to\infty$ *die Folge* $\{Q_n\}_{n\in\mathbf{N}}$ *schwach gegen* Q, *so konvergiert die Folge* $\{X_n\}_{n\in\mathbf{N}}$ *mit*

$$X_n = F_n^{-1}(U), \quad n\in\mathbf{N}, \tag{2.3.31}$$

fast sicher gegen $X = F^{-1}(U)$. *Dabei ist* $P^X = Q$ *und* $P^{X_n} = Q_n$ *für alle* $n\in\mathbf{N}$.

f) *Konvergiert die Folge* $\{Q_n\}_{n\in\mathbf{N}}$ *in der Metrik* ρ, *so auch schwach, d.h.*

$$\rho(Q_n,Q)\longrightarrow 0 \quad\Longrightarrow\quad Q_n \xrightarrow{w} Q \quad (n\to\infty). \tag{2.3.32}$$

Beweis. Wir zeigen den ersten Teil des Lemmas nach dem Ringschluß–Schema

a) $\to$ b)
$\uparrow \quad \swarrow \quad \uparrow$, wobei wir mit dem oberen Teil beginnen.
c) $\to$ d)

Wir zeigen zunächst, daß a) b) impliziert. Sei dazu x ein Stetigkeitspunkt von F und $\varepsilon>0$. Dann existiert ein $\delta>0$ mit der Eigenschaft

$$|x-y|\le\delta \quad\Longrightarrow\quad |F(x)-F(y)|\le\varepsilon$$

für alle $y\in\mathbf{R}$. Wähle nun stetige Funktionen $0\le g_*, g^*\le 1$ mit

$$g_*(y)=\begin{cases}1, & \text{falls } y\le x-\delta\\ 0, & \text{falls } y\le x\end{cases} \quad\text{und}\quad g^*(y)=\begin{cases}1, & \text{falls } y\le x\\ 0, & \text{falls } y\le x+\delta.\end{cases} \tag{2.3.33}$$

Es folgt

$$\begin{aligned}\liminf_{n\to\infty} F_n(x) &= \liminf_{n\to\infty}\int \mathbb{1}_{(-\infty,x]}\,dQ_n\\ &\ge \liminf_{n\to\infty}\int g_*(y)\,dQ_n(y) = \int g_*(y)\,dQ(y)\\ &\ge \int \mathbb{1}_{(-\infty,x-\delta]}\,dQ = F(x-\delta)\ge F(x)-\varepsilon.\end{aligned} \tag{2.3.34}$$

Analog erhält man

$$\limsup_{n\to\infty} F_n(x) \le F(x) + \varepsilon; \tag{2.3.35}$$

da ε beliebig war, gilt also $\lim_{n\to\infty} F_n(x) = F(x)$, wie behauptet.
Wir zeigen nun, daß b) c) impliziert. Hierzu bemerken wir zunächst, daß aufgrund der schwachen Monotonie von F und F^{-1} die Menge der Unstetigkeitsstellen U_F bzw. $U_{F^{-1}}$ jeweils höchstens abzählbar ist, die Menge der Stetigkeitspunkte also jeweils dicht in $\mathbf{R}$ bzw. in $(0,1)$ liegt.
Sei nun $\varepsilon > 0$ und $y \in (0,1)$ Stetigkeitspunkt von F^{-1}; dann existieren Stetigkeitspunkte x und z von F mit

$$x < F^{-1}(y) < z \qquad \text{und} \qquad z - x < \varepsilon. \tag{2.3.36}$$

Nach der Eigenschaft (2.1.5) der Pseudo–Inversen muß dann auch $F(x) < y$ gelten; gemäß (2.1.3) ist ferner $y \le F\big(F^{-1}(y)\big) \le F(z)$. Dann ist aber $y < F(z)$; anderenfalls wäre nämlich $y = F(z)$ und damit für jedes $0 < \delta < 1 - y$

$$F^{-1}(y+\delta) = \inf\{u \in \mathbf{R} \mid F(u) \ge y + \delta\} > z,$$

zugleich aber $F^{-1}(y) < z$ im Widerspruch zur Stetigkeit von F^{-1} in y. Damit erhält man insgesamt

$$F(x) < y < F(z).$$

Gemäß b) existiert nun eine Zahl $n(\varepsilon) \in \mathbf{N}$ mit

$$F_n(x) < y < F_n(z) \qquad \text{für alle } n \ge n(\varepsilon),$$

was wiederum nach (2.1.5)

$$x \le F_n^{-1}(y) \le z \qquad \text{für alle } n \ge n(\varepsilon)$$

nach sich zieht. Aufgrund von (2.3.36) ist also

$$\big|F_n^{-1}(y) - F^{-1}(y)\big| \le z - x \le \varepsilon \qquad \text{für alle } n \ge n(\varepsilon),$$

womit c) bewiesen ist.
Aus c) ergibt sich unmittelbar auch Teil e) des Lemmas: nach Satz 2.1.1 gilt nämlich $P^{X_n} = Q_n$, $n \in \mathbf{N}$, sowie $P^X = Q$; wegen der Abzählbarkeit von $U_{F^{-1}}$ ist also $P(U_{F^{-1}}) = 0$ (wobei hier $P = \mathcal{R}\big((0,1)\big)$ zu setzen ist), so daß aus der Konvergenz $F_n^{-1}(y) \longrightarrow F^{-1}(y)$, $y \in U_{F^{-1}}^c$ $(n \to \infty)$ die gewünschte Konvergenz $X_n = F_n^{-1}(U) \longrightarrow F^{-1}(U) = X$ $(n \to \infty)$ P–fast sicher folgt.
Damit ist zugleich gezeigt, daß c) wieder a) impliziert: man wähle dazu nur die zuletzt durchgeführte Konstruktion der Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$, X; die fast sichere Konvergenz von $\{X_n\}_{n\in\mathbf{N}}$ gegen X liefert dann nach Lemma 2.3.2 die schwache Konvergenz von $\{Q_n\}_{n\in\mathbf{N}}$ gegen Q.
Dieselbe Konstruktion zeigt auch, daß c) d) nach sich zieht: ist nämlich $B \in \mathcal{B}^1$ mit $Q(\partial B) = P(X \in \partial B) = 0$, so zeigen die Überlegungen im Anschluß an Beziehung (2.2.37), daß die Indikatorfunktion $\mathbb{1}_B$ Q–fast sicher stetig ist, da hier

wieder $U_{\mathbb{1}_B} = \partial B$ gilt; d.h. ist $\{y_n\}_{n\in\mathbf{N}}$ eine Folge mit $\lim_{n\to\infty} y_n = y \notin \partial B$, so gilt auch $\lim_{n\to\infty} \mathbb{1}_B(y_n) = \mathbb{1}_B(y)$. Für alle $\omega \in \big(X^{-1}(\partial B)\big)^c$ erhält man somit auch $\lim_{n\to\infty} \mathbb{1}_B\big(X_n(\omega)\big) = \mathbb{1}_B\big(X(\omega)\big)$, d.h. $Z_n = \mathbb{1}_B(X_n) \longrightarrow Z = \mathbb{1}_B(X)$ $(n \to \infty)$ P–fast sicher. Da die Zufallsvariablen $\{Z_n\}_{n\in\mathbf{N}}$ alle durch die integrierbare Konstante 1 absolut beschränkt sind, liefert eine Anwendung des Satzes von der majorisierten Konvergenz (Beziehung (2.2.24)) das gewünschte Ergebnis, nämlich

$$\begin{aligned} Q_n(B) = P(X_n \in B) = E\big(\mathbb{1}_B(X_n)\big) = E(Z_n) \longrightarrow E(Z) \\ = E\big(\mathbb{1}_B(X)\big) = P(X \in B) = Q(B), \quad n \to \infty. \end{aligned}$$

Die Aussage b) folgt nun trivialerweise aus d), denn der Rand eines jeden Intervalls $B = (-\infty, x]$, $x \in \mathbf{R}$, besteht gerade aus der Menge $\{x\}$; ist x ein Stetigkeitspunkt von F, so ist nach (1.2.4) und (1.2.5) also $Q(\{x\}) = 0$, d.h. es folgt mit dieser Bezeichnung

$$F_n(x) = Q_n(B) \longrightarrow Q(B) = F(x), \quad n \to \infty.$$

Teil f) des Lemmas ist eine unmittelbare Konsequenz aus b). Das Lemma ist damit vollständig bewiesen. ∎

Teil d) des letzten Lemmas zeigt also, daß schwache Konvergenz von Verteilungen in einem gewissen Sinn doch als punktweise Konvergenz verstanden werden kann, wenn man sich auf solche Mengen beschränkt, deren Rand das (Grenz–)Maß Null besitzen. In dem vor Definition 2.3.2 betrachteten Beispiel ist aber gerade diese Bedingung verletzt: die Grenzverteilung Q ist nämlich die Verteilung der Zufallsvariablen $X \equiv a$, so daß $Q(\partial B) = 0$ genau dann gilt, wenn a kein Element von ∂B ist. In der betrachteten Situation ist aber $a \in \partial B$: a kann nämlich kein innerer Punkt von B sein, da die Folge $\{a_n\}_{n\in\mathbf{N}}$ gegen a konvergiert, also ansonsten schließlich alle Folgenglieder in B liegen müßten im Wiederspruch zur Voraussetzung, daß B höchstens endlich viele Folgenglieder enthält. Es ist somit notwendig $a \in B \setminus B^\circ \subseteq \partial B$, wie behauptet.

Teil e) des Lemmas besagt, daß man sich bei Betrachtungen der schwachen Konvergenz sogar auf fast–sichere Konvergenz zurückziehen kann, wenn lediglich die *Verteilungen* der beteiligten Zufallsvariablen von Interesse sind, nicht aber ihre explizite Definition als meßbare Abbildungen. Allerdings bedeutet dies nicht, daß man in *jedem* Fall schwache durch fast–sichere Konvergenz ersetzen darf, wie ja bereits das in (2.3.40) vorgestellte Beispiel zeigt.

Eine Umkehrung des Teils f) des Lemmas ist i.a. falsch, wie das Beispiel $\mathcal{E}(\lambda) \xrightarrow{w} Q$ mit $Q(\{0\}) = 1$ für $\lambda \to 0$ lehrt: hier ist nämlich $\rho\big(\mathcal{E}(\lambda), Q\big) \equiv 1$ für alle $\lambda > 0$, d.h. es liegt keine Konvergenz in der Metrik ρ vor.

Der Beweis des Teils b) des Lemmas zeigt schließlich, daß man sich bei der Prüfung auf schwache Konvergenz von Verteilungen $\{Q_n\}_{n\in\mathbf{N}}$ gegen Q im Sinne von (2.3.18) sogar auf echte Teilklassen der Menge der (gleichmäßig) stetigen beschränkten Funktionen zurückziehen kann, da die Hilfsfunktionen g_* und g^* in (2.3.33) stets auch so gewählt werden können, daß z.B. Ableitungen beliebiger Ordnung exisitieren und diese auch noch sämtlich gleichmäßig stetig und beschränkt sind. Mit einer Argumentation ähnlich wie in (2.3.34) und (2.3.35) kann man sich entsprechend auf die Menge aller stetigen, beschränkten Funktionen h

zurückziehen, die im Nullpunkt beliebig häufig differenzierbar sind und die für alle reellen x durch ihre Taylorreihe dargestellt werden, für die also

$$h(x) = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} x^k \quad \text{für alle } x \in \mathbf{R} \tag{2.3.37}$$

absolut konvergiert. Wählt man nämlich beispielsweise

$$h(x) = \frac{1}{2} - \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{k!(2k+1)} = \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du, \quad x \in \mathbf{R},$$

so gilt für $a \in \mathbf{R}$, $b > 0$

$$h(b(x-a)) = h(bx) + \frac{1}{\sqrt{\pi}} \int_0^{ab} e^{-u^2}\, du, \quad x \in \mathbf{R},$$

d.h. $h(b(x-a))$ besitzt ebenfalls eine überall konvergente Taylorreihe, und $h(b(x-a))$ approximiert für große b die Indikatorfunktion $\mathbb{1}_{(-\infty,a]}(x)$ in jedem Bereich $(-\infty, a-\varepsilon) \cup (a+\varepsilon, \infty)$ $(\varepsilon > 0)$ gleichmäßig, symmetrisch und einseitig monoton, d.h. $h(b(x-a))$ ist bezgl. b monoton wachsend für $x < a$ und monoton fallend für $x > a$. Das gerade gewählte Beispiel zeigt, daß man sich bei den Funktionen h sogar noch auf solche einschränken kann, deren Majorante

$$h^*(x) = \sum_{k=0}^{\infty} \frac{|h^{(k)}(0)|}{k!} |x|^k, \quad x \in \mathbf{R} \tag{2.3.38}$$

ebenfalls beschränkt ist.

Die zuletzt getroffene Feststellung können wir z.B. dazu benutzen, einen Zusammenhang zwischen der Konvergenz momenterzeugender Funktionen und schwacher Konvergenz herzuleiten.

Satz 2.3.6. *(momenterzeugende Funktionen und schwache Konvergenz)*
Es seien $\{X_n\}_{n\in\mathbf{N}}$ und X Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit momenterzeugenden Funktionen $\{\Psi_n\}_{n\in\mathbf{N}}$, Ψ, die sämtlich in einem Intervall $[-\delta, \delta]$ mit $\delta > 0$ existieren mögen. Dann gilt:

$$\Psi_n(t) \longrightarrow \Psi(t), \ |t| \le \delta \quad \Longrightarrow \quad P^{X_n} \xrightarrow{w} P^X \quad (n \to \infty), \tag{2.3.39}$$

und es konvergieren sämtliche Momente

$$E(X_n^k) \longrightarrow E(X^k) \quad \text{für alle } k \in \mathbf{N} \quad (n \to \infty). \tag{2.3.40}$$

Beweis. Gemäß (2.2.74) lassen sich unter den angegebenen Voraussetzungen alle momenterzeugenden Funktionen als Taylor–Reihen darstellen; es gilt dann für $|t| \le \delta$:

$$\begin{aligned} \Psi_n(t) = E(e^{tX_n}) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X_n^k) \\ &\longrightarrow \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) = E(e^{tX}) = \Psi(t) \quad (n \to \infty), \end{aligned} \tag{2.3.41}$$

woraus sich unmittelbar die Konvergenz aller Momente

$$E(X_n^k) \longrightarrow E(X^k) \quad \text{für alle } k \in \mathbb{N} \quad (n \to \infty) \tag{2.3.42}$$

ergibt. Ähnlich folgt, daß sogar alle absoluten Momente konvergieren (vgl. Lemma 2.2.3 b)). Sei nun h eine stetige, beschränkte Funktion mit der Darstellung (2.3.37) und der Eigenschaft (2.3.38). Mit der dortigen Bezeichnung sind also alle Majoranten $h^*(X_n)$, $n \in \mathbb{N}$ und $h^*(X)$ beschränkt (etwa durch $0 < M < \infty$) und daher mit endlichem Wert integrierbar. Nach dem Satz von der majorisierten Konvergenz (Satz 2.2.2 f)) erhält man somit die Darstellungen

$$E\big(h(X_n)\big) = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} E(X_n^k), \quad n \in \mathbb{N}$$

$$E\big(h(X)\big) = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} E(X^k),$$

wobei alle Reihen sogar absolut konvergieren und diese durch dieselbe Konstante M beschränkt sind. Setzt man zur Abkürzung

$$s_n(k) = \sup_{m \geq n} \big|E(X_m^k) - E(X^k)\big|, \quad n, k \in \mathbb{N},$$

so folgt aus dem gerade gesagten zunächst

$$\sum_{k=0}^{\infty} \frac{|h^{(k)}(0)|}{k!} s_1(k) \leq 2M < \infty.$$

Sei nun $\varepsilon > 0$. Dann existiert eine Zahl $N(\varepsilon) \in \mathbb{N}$ mit

$$\sum_{k > N(\varepsilon)} \frac{|h^{(k)}(0)|}{k!} s_1(k) \leq \varepsilon.$$

Aufgrund der Konvergenz der Momente nach (2.3.42) ergibt sich dann aber $\lim_{n\to\infty} s_n(k) = 0$ für alle $k = 1, 2, \ldots, N(\varepsilon)$, also

$$\begin{aligned} \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{|h^{(k)}(0)|}{k!} s_n(k) &\leq \lim_{n\to\infty} \sum_{k > N(\varepsilon)}^{\infty} \frac{|h^{(k)}(0)|}{k!} s_n(k) \\ &\leq \sum_{k > N(\varepsilon)}^{\infty} \frac{|h^{(k)}(0)|}{k!} s_1(k) \leq \varepsilon. \end{aligned}$$

Da ε beliebig war, folgt also

$$\lim_{n\to\infty} \big|E\big(h(X_n)\big) - E\big(h(X)\big)\big| \leq \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{|h^{(k)}(0)|}{k!} s_n(k) = 0$$

und somit $E\big(h(X_n)\big) \longrightarrow E\big(h(X)\big)$ $(n \to \infty)$, was zum Nachweis der schwachen Konvergenz in (2.3.39) genügt. ∎

Mit Hilfe des letzten Resultats können wir nun in Ergänzung des Gesetzes der großen Zahlen (Satz 2.3.4) und des Gesetzes vom iterierten Logarithmus (Satz 2.3.5) die noch offenen Fragestellung der schwachen Konvergenz abschließend behandeln.

Satz 2.3.7. *(Zentraler Grenzwertsatz)*
Es sei $\{X_n\}_{n\in\mathbf{N}}$ eine Folge stochastisch unabhängiger, identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$E(X_n) = \mu \in \mathbf{R} \quad \text{und} \quad 0 < \operatorname{Var}(X_n) = \sigma^2 < \infty \tag{2.3.43}$$

für alle $n \in \mathbf{N}$. Dann gilt für die arithmetischen Mittel $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ der Folge $\{X_n\}_{n\in\mathbf{N}}$:

$$P^{\frac{\sqrt{n}}{\sigma}(S_n-\mu)} \xrightarrow{w} \mathcal{N}(0,1). \tag{2.3.44}$$

Beweis. Wir können die Aussage hier nur unter der stärkeren Annahme zeigen, daß die momenterzeugende Funktion $\Psi(t)$ der Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ für $|t| \le \delta$ $(\delta > 0)$ existiert. Dazu bezeichne $Y_n = \dfrac{X_n - \mu}{\sigma}$, $n \in \mathbf{N}$. Dann ist $E(Y_n) = 0$, $\operatorname{Var}(Y_n) = 1$, und es existieren alle Momente von X_n und Y_n sowie die momenterzeugende Funktion Ψ^* der Y_n, $n \in \mathbf{N}$ für $|t| \le \delta$ mit $\Psi^*(t) = \Psi\left(\frac{t}{\sigma}\right) \cdot e^{-\frac{\mu t}{\sigma}}$. Bezeichnet ferner Ψ_n die momenterzeugende Funktion der Zufallsvariablen $\frac{Y_i}{\sqrt{n}}$, $i, n \in \mathbf{N}$, so erhält man weiter für $|t| \le \delta$ und alle $i \in \mathbf{N}$

$$\begin{aligned}\Psi_n(t) = \Psi^*\Big(\frac{t}{\sqrt{n}}\Big) &= \sum_{k=0}^{\infty} \frac{t^k}{k!\sqrt{n}^k} E(Y_i^k) \\ &= 1 + \frac{t^2}{2n} + \sum_{k=3}^{\infty} \frac{t^k}{k!\sqrt{n}^k} E(Y_i^k) \\ &= 1 + \frac{t^2}{2n} + R_n(t)\end{aligned}$$

mit

$$R_n(t) \le \frac{1}{n\sqrt{n}} \sum_{k=3}^{\infty} \frac{t^k}{k!} E\big(|Y_i^k|\big) \le \frac{1}{n\sqrt{n}}\Big(\Psi^*(\delta) + \Psi^*(-\delta)\Big)$$

(vgl. den Beweis zu Lemma 2.2.3 b)). Setzt man $T_n = \frac{\sqrt{n}}{\sigma}(S_n - \mu)$, $n \in \mathbf{N}$, so erhält man für die momenterzeugende Funktion von T_n für $|t| \le \delta$:

$$\Psi_{T_n}(t) = \Big(\Psi_n(t)\Big)^n = \Big(1 + \frac{t^2}{2n} + R_n(t)\Big)^n \longrightarrow \exp\Big(\frac{t^2}{2}\Big) \ (n \to \infty). \tag{2.3.45}$$

Die Grenzfunktion auf der rechten Seite von (2.3.45) ist aber gerade die momenterzeugende Funktion der $\mathcal{N}(0,1)$–Verteilung, so daß mit Satz 2.3.6 folgt: $P^{T_n} \xrightarrow{w} \mathcal{N}(0,1)$ $(n \to \infty)$, was zu zeigen war. ∎

Bemerkenswert an diesem Resultat ist vor allem die Tatsache, daß die Konvergenz in (2.3.44) *unabhängig* davon ist, welche spezielle Gestalt die ursprüngliche Verteilung der Folge $\{X_n\}_{n\in\mathbf{N}}$ besitzt, und diese lediglich über die beiden ersten Momente Erwartungswert und Varianz als zentrierende Größen in die Konvergenz eingeht. Dies erklärt, warum man häufig Normalverteilungsannahmen in solchen stochastischen Modellen findet, in denen Summen unabhängiger Zufallsvariablen eine Rolle spielen.

Allgemeinere Formulierungen des Zentralen Grenzwertsatzes findet man z.B. in Bauer (1978), §51, oder Billingsley (1986), Section 27.

Damit erklärt sich auch der durch Beziehung (2.1.88) gegebene Sachverhalt noch einmal aus anderer Sicht: sind die Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ nämlich sämtlich $\mathcal{E}(1)$–verteilt, so ist ja gerade $E(X_n) = 1$, $\mathrm{Var}(X_n) = 1$, $n \in \mathbf{N}$, d.h. nach dem Zentralen Grenzwertsatz ist $\dfrac{\sum_{k=1}^n X_k - n}{\sqrt{n}}$ asymptotisch $\mathcal{N}(0,1)$–verteilt, wie behauptet.

Ähnlich wie bei der Poisson–Approximation läßt sich auch im Zentralen Grenzwertsatz die Güte der Approximation abschätzen, wenn beispielsweise das dritte absolute zentrale Moment

$$E\big(|X_n - \mu|^3\big) =: M < \infty \tag{2.3.46}$$

existiert.

Satz 2.3.8. *(Berry–Esséen; van Beek)*
Unter den Voraussetzungen von Satz 2.3.7 sowie (2.3.46) gilt mit den Bezeichungen aus Definition 2.1.4:

$$\rho\Big(P^{\frac{\sqrt{n}}{\sigma}(S_n-\mu)}, \mathcal{N}(0,1)\Big) = \sup_{x\in\mathbf{R}} \left| P\left(\frac{\sum_{k=1}^n (X_k - \mu)}{\sigma\sqrt{n}} \le x \right) - \Phi(x) \right| \le \frac{0.8 \cdot M}{\sigma^3\sqrt{n}}, \tag{2.3.47}$$

wobei

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du, \quad x \in \mathbf{R}, \tag{2.3.48}$$

die Verteilungsfunktion der $\mathcal{N}(0,1)$–Verteilung bezeichne.

Einen Beweis dieser Aussage findet man z.B. bei Gänssler & Stute (1977), Satz 4.2.10 und Korollar 4.2.15.

Die in (2.3.47) angegebene Konvergenzgeschwindigkeit kann dabei i.a. nicht verbessert werden. Wählt man etwa die unabhängige Folge $\{Y_n\}_{n\in\mathbf{N}}$ als $\mathfrak{B}(1, \frac{1}{2})$–verteilt und setzt $X_n = 2Y_n - 1$, $n \in \mathbf{N}$, so ist $E(X_n) = 0$, $\mathrm{Var}(X_n) = 1 = M$ für alle $n \in \mathbf{N}$. Einfache kombinatorische Überlegungen ergeben

$$P\Big(\sum_{k=1}^n X_k = 0\Big) = \binom{n}{n/2} \frac{1}{2^n} \sim \frac{2}{\sqrt{2\pi n}} \quad (n \in \mathbf{N},\ n \text{ gerade}),$$

wie man etwa mit Hilfe der Stirling'schen Formel (2.1.87) einsehen kann (es ist $n! = \Gamma(n+1)$, $n \in \mathbf{Z}^+$). Wegen

$$\rho\left(P^{\frac{\sqrt{n}}{\sigma}(S_n-\mu)}, \mathcal{N}(0,1)\right) \ge \frac{1}{2} P\Big(\sum_{k=1}^n X_k = 0\Big) \sim \frac{1}{\sqrt{2\pi n}}, \quad n \in \mathbf{N},$$

kann also die durch den Term $\frac{1}{\sqrt{n}}$ gegebene Konvergenzgeschwindigkeit in (2.3.47) nicht verbessert werden (siehe auch Gänssler & Stute (1977), Beispiel 4.2.16).

Mit Hilfe von Satz 2.3.6 läßt sich darüberhinaus auch die schwache Konvergenz bei Poisson–Approximation (Sätze 2.1.4 und 2.1.7) direkt nachweisen. Wir wollen dies hier exemplarisch nur am Beispiel des letztgenannten Satzes vorführen, d.h. wir zeigen in Analogie zu Beziehung (2.1.77) mit den dortigen Bezeichnungen:

$$\mathfrak{PB}(n; p_{1n}, \ldots, p_{nn}) \xrightarrow{w} \mathfrak{P}(\lambda) \quad (n \to \infty). \tag{2.3.49}$$

Nach der Multiplikationsregel (2.2.79) besitzt die momenterzeugende Funktion Ψ_n einer $\mathfrak{PB}(n; p_{1n}, \ldots, p_{nn})$–verteilten Zufallsvariablen die Form (vgl. auch die Tabelle S. 135)

$$\begin{aligned}\Psi_n(t) &= \prod_{k=1}^{n}(1 - p_{kn} + p_{kn}e^t) = \prod_{k=1}^{n}(1 + p_{kn}(e^t - 1)) \\ &= \exp\Big(\sum_{k=1}^{n}\ln(1 + p_{kn}(e^t - 1))\Big) \sim \exp\Big(\sum_{k=1}^{n} p_{kn}(e^t - 1)\Big) \\ &\sim \exp\big(\lambda(e^t - 1)\big) \quad (t \in \mathbf{R},\ n \to \infty)\end{aligned}$$

unter den in (2.1.76) spezifizierten Bedingungen; der letztere Ausdruck stellt aber gerade die momenterzeugende Funktion einer $\mathfrak{P}(\lambda)$–verteilten Zufallsvariablen dar, womit die Aussage bewiesen ist.

Der Begriff der schwachen Konvergenz läßt sich in natürlicher Weise auch auf höherdimensionale Verteilungen bzw. Verteilungen von Zufallselementen mit Werten in einem geeigneten metrischen Meßraum $(\mathcal{X}, \mathcal{B})$[1)] übertragen, indem man in (2.3.18) Funktionen g betrachtet, die auf $\mathcal{X}$ (gleichmäßig) stetig, reellwertig und beschränkt sind. Die in Lemma 2.3.3 d) gegebene Charakterisierung der schwachen Konvergenz bleibt dabei sinngemäß gültig; für den Fall $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^m, \mathcal{B}^m)$, $m \in \mathbf{N}$, entsprechend auch Lemma 2.3.3 b) sowie Satz 2.3.6 (vgl. hierzu auch (2.2.90) bis (2.2.96)). In diesem Sinne läßt sich etwa die Grenzwertaussage des Satzes 2.1.5 für Multinomialverteilungen auch als eine schwache Konvergenzaussage formulieren: sind nämlich für festes $k \in \mathbf{N}$ unter den dortigen Voraussetzungen $(N_{1n}, \ldots, N_{kn})$ $\mathfrak{M}(n; p_{1n}, \ldots, p_{kn})$–verteilte Zufallsvektoren ($n \in \mathbf{N}$), so besagt Beziehung (2.1.72) gerade

$$P^{(N_{1n}, \ldots, N_{k-1,n})} \xrightarrow{w} \bigotimes_{i=1}^{k-1} \mathfrak{P}(\lambda_i) \quad (n \to \infty). \tag{2.3.50}$$

Dies ergibt sich z.B. aus (2.2.96), wenn man dort $m = k$, $p_j = p_{jn}$, $1 \le j \le k$ sowie $s_i = e^{t_i}$, $t_i \in \mathbf{R}$, $1 \le i \le k-1$ und $s_k = 1$ wählt.

Zum Abschluß des ersten Kapitels wollen wir uns nun noch mit einem insbesondere im CAD angewandten Aspekt der schwachen Konvergenz befassen, nämlich den sogenannten *Bézier*–Kurven und Flächen. Hierzu benötigen wir allerdings noch einige analytische Hilfsmittel.

[1)] d.h. es existiert eine Metrik ρ auf $\mathcal{X} \times \mathcal{X}$, und $\mathcal{B}$ ist die von den bzgl. ρ offenen Mengen erzeugte (Borel'sche) σ–Algebra

Definition 2.3.3. *(Stetigkeitsmodul)*
Es sei $f : D \longrightarrow \mathbf{R}$ eine auf einem (evtl. unendlichen) Teilintervall $D \subseteq \mathbf{R}$ definierte Funktion. Dann heißt die durch

$$\omega(f;\delta) = \sup_{\substack{x,y\in D \\ |x-y|\le\delta}} |f(x)-f(y)|, \quad \delta > 0 \tag{2.3.51}$$

auf $\mathbf{R}^+$ definierte Abbildung $\omega(f;\cdot)$ der zu f gehörige Stetigkeitsmodul. Der durch

$$\eta(f) = \omega(f;\infty) = \sup_{x,y\in D} |f(x)-f(y)| \tag{2.3.52}$$

gegebene Ausdruck heißt Schwankung von f über D.

Man beachte, daß die auf D *gleichmäßig stetigen* Funktionen f gerade durch die Bedingung $\lim_{\delta\downarrow 0}\omega(f;\delta) = 0$ charakterisiert sind.

Das folgende Lemma stellt eine für die weiteren Untersuchungen fundamentale Abschätzung bereit.

Lemma 2.3.4. *(Stetigkeitsmodul und schwache Konvergenz)*
Es seien D und f wie in Definition 2.3.3 sowie $\{X_n(t)\}_{n\in\mathbf{N},t\in D}$ für jedes feste t unabhängige Zufallsvariablen mit Werten in D derart, daß

$$E(X_n(t)) = t, \qquad 0 < \operatorname{Var}(X_n(t)) = \sigma^2(t) < \infty \qquad (t \in D) \tag{2.3.53}$$

gelte. Ferner bezeiche $S_n(t) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k(t)$, $t \in D$. Dann gilt für jedes $\delta > 0$:

$$\big|E\big[f(S_n(t))\big] - f(t)\big| \le \omega(f;\delta) + \frac{\eta(f)\sigma^2(t)}{n\delta^2}. \tag{2.3.54}$$

Beweis. Wie im Beweis zu Lemma 2.3.2 bezeichne wieder $A_n = \{|S_n(t)-t| \le \delta\}$. Analog zu (2.3.26) ergibt sich

$$\begin{aligned}
\big|E\big[f(S_n(t))\big] - f(t)\big| &\le \int_{A_n} \big|f(S_n(t)) - f(t)\big|\,dP + \int_{A_n^c} \big|f(S_n(t)) - f(t)\big|\,dP \\
&\le \omega(f;\delta)\int_{A_n} dP + \eta(f)\int_{A_n^c} dP \\
&\le \omega(f;\delta) + \eta(f)P\big(|S_n(t)-t| > \delta\big) \\
&\le \omega(f;\delta) + \frac{\eta(f)\sigma^2(t)}{n\delta^2}
\end{aligned} \tag{2.3.55}$$

nach der Tschebyscheff–Ungleichung (2.2.59). ∎

Der Ausdruck auf der linken Seite von (2.3.55) konvergiert demnach für festes $t \in D$ mit wachsendem n insbesondere dann gegen 0, wenn die Funktion f gleichmäßig stetig ist; hierzu hat man lediglich $\delta = \delta_n$ in Abhängigkeit von n so zu

wählen, daß $\delta_n \to 0$, $n \cdot \delta_n^2 \to \infty$ $(n \to \infty)$ gilt (z.B. $\delta_n = n^{-\alpha}$ mit $0 < \alpha < \frac{1}{2}$, $n \in \mathbf{N}$). Wählt man $D = \mathbf{R}$, ergibt Beziehung (2.3.54) also eine Abschätzung der Konvergenzgeschwindigkeit für die *schwache Konvergenz* der arithmetischen Mittel $S_n(t)$ gegen den Erwartungswert t der gegebenen Zufallsvariablen $\{X_n(t)\}$.

Lemma 2.3.4 liefert damit einen bereits auf Bernstein (1912) zurückgehenden Beweis des

Weierstraß'schen Approximationssatzes: *Jede auf einem endlichen, abgeschlossenen Intervall D definierte stetige Funktion f läßt sich gleichmäßig durch Polynome approximieren.*

Beweis. Wir können o.B.d.A. $D = [0,1]$ wählen, da sich jedes endliche, abgeschlossene Intervall $D = [a,b]$ mit $a < b$ durch die lineare Abbildung $G(x) = \frac{x-a}{b-a}$, $x \in \mathbf{R}$, bijektiv auf $[0,1]$ abbilden läßt. Sei nun $\{U_n\}_{n\in\mathbf{N}}$ eine unabhängige Folge $\mathcal{R}([0,1])$–verteilter Zufallsvariablen. Setze

$$X_n(t) = \lfloor t + U_n \rfloor, \quad n \in \mathbf{N},\ t \in [0,1]. \tag{2.3.56}$$

Dann ist jede der Zufallsvariablen $X_n(t)$ $\mathfrak{B}(1,t)$–verteilt wegen

$$\begin{aligned} P(X_n(t) = 1) &= P(t + U_n \geq 1) = P(U_n \geq 1 - t) \\ &= 1 - (1-t) = t, \quad n \in \mathbf{N},\ t \in [0,1]. \end{aligned} \tag{2.3.57}$$

Ferner ist nach (2.1.18) und (2.2.39)

$$B_n(f;t) := E\big[f(S_n(t))\big] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} t^k (1-t)^{n-k} \tag{2.3.58}$$

ein Polynom vom Grad höchstens n (sog. *Bernstein–Polynom*) mit

$$\begin{aligned} \big|B_n(f;t) - f(t)\big| &\leq \omega(f;\delta) + \frac{\eta(f)t(1-t)}{n\delta^2} \\ &\leq \omega(f;\delta) + \frac{\eta(f)}{4n\delta^2}, \qquad n \in \mathbf{N},\ t \in [0,1] \end{aligned} \tag{2.3.59}$$

nach (2.3.54). Da jede auf $[0,1]$ stetige Funktion dort auch gleichmäßig stetig und beschränkt ist (also insbesondere $\eta(f) < \infty$ gilt), folgt nach den obigen Bemerkungen die Behauptung, da die rechte Seite wegen der gleichmäßigen Beschränktheit der Varianzen in t unabhängig von t ist. ∎

Bemerkenswert an der Bernstein'schen Konstruktion ist vor allem die Tatsache, daß die approximierenden Polynome $B_n(f;\cdot)$ zusammen mit einer geeigneten Fehlerabschätzung explizit angegeben werden können.

Wählt man statt reellwertiger, stetiger Funktionen f *vektorwertige*, d.h. insbesondere $\mathbf{R}^2$–wertige Funktionen $\boldsymbol{f}$, so erhält man eine Kurvendarstellung in parametrischer Form mit Parameter $t \in D$. Überträgt man die Konstruktion von Bernsteinpolynomen nun auf diese allgemeinere Situation, so ergeben sich in natürlicher

Weise die durch Bézier 1968 eingeführten gleichnamigen Kurven, die inzwischen ein Standardwerkzeug im CAD darstellen und z.B. auch für die Gestaltung der verschiedenen Schriftarten von TEX verwendet werden.

Modifiziert man die Begriffe "Stetigkeitsmodul" und "Schwankung" für die jetzt betrachtete Situation, d.h. definiert man

$$\omega(\boldsymbol{f};\delta) = \sup_{\substack{x,y\in D\\ |x-y|\le\delta}} \|\boldsymbol{f}(x)-\boldsymbol{f}(y)\|, \quad \delta>0 \tag{2.3.60}$$

sowie

$$\eta(\boldsymbol{f}) = \omega(\boldsymbol{f};\infty) = \sup_{x,y\in D} \|\boldsymbol{f}(x)-\boldsymbol{f}(y)\|, \tag{2.3.61}$$

wobei $\|\boldsymbol{x}\| = \sqrt{x_1^2+x_2^2}$ die Euklidische Norm des Vektors $\boldsymbol{x} = (x_1,x_2) \in \mathbf{R}^2$ bezeichne, so erhält man analog zu Beziehung (2.3.54)

$$\big\|E\big[\boldsymbol{f}(S_n(t))\big] - \boldsymbol{f}(t)\big\| \le \omega(\boldsymbol{f};\delta) + \frac{\eta(\boldsymbol{f})\sigma^2(t)}{n\delta^2} \tag{2.3.62}$$

bzw. spezieller

$$B_n(\boldsymbol{f};t) = E\big[\boldsymbol{f}(S_n(t))\big] = \sum_{k=0}^{n} \boldsymbol{f}\left(\frac{k}{n}\right)\binom{n}{k}t^k(1-t)^{n-k} \tag{2.3.63}$$

mit

$$\big\|B_n(\boldsymbol{f};t) - \boldsymbol{f}(t)\big\| \le \omega(\boldsymbol{f};\delta) + \frac{\eta(\boldsymbol{f})t(1-t)}{n\delta^2}, \tag{2.3.64}$$

wobei noch

$$E(\boldsymbol{X}) = (E(X_1),\ldots,E(X_m)) \tag{2.3.65}$$

für einen m–dimensionalen Zufallsvektor $\boldsymbol{X} = (X_1,\ldots,X_m)$, $m \in \mathbf{N}$, vereinbart sei. Die $\mathbf{R}^2$–wertigen Bernsteinpolynome $B_n(\boldsymbol{f};\cdot)$ approximieren für $n \to \infty$ also ebenfalls die durch $\boldsymbol{f}$ beschriebene Kurve im Sinne des Euklidischen Abstands gleichmäßig, wenn die Abbildung $\boldsymbol{f}$ stetig auf $[0,1]$ ist.

Besondere Bedeutung für graphische Konstruktionen haben hierbei stetige *Polygonzüge*, die durch Vorgabe sogenannter *Knoten* $\boldsymbol{f}_0,\ldots,\boldsymbol{f}_n \in \mathbf{R}^2$, $n \in \mathbf{N}$, bestimmt sind vermöge der Abbildung

$$\boldsymbol{f}(t) = \boldsymbol{f}_{i-1} + (nt-i+1)(\boldsymbol{f}_i - \boldsymbol{f}_{i-1}), \quad \frac{i-1}{n} \le t \le \frac{i}{n},\ 1 \le i \le n. \tag{2.3.66}$$

$\boldsymbol{f}$ ist also stückweise linear mit

$$\boldsymbol{f}\left(\frac{i}{n}\right) = \boldsymbol{f}_i, \quad 0 \le i \le n. \tag{2.3.67}$$

Für solche Polygonzüge gilt speziell

$$\omega(\boldsymbol{f};\delta) \le n\delta L, \quad \eta(\boldsymbol{f}) = \Delta, \tag{2.3.68}$$

wobei $L = \sup_{1 \le i \le n} \|\boldsymbol{f}_i - \boldsymbol{f}_{i-1}\|$ die maximale Kantenlänge des Polygonzugs und Δ den Durchmesser des von den Knoten $\boldsymbol{f}_0, \ldots, \boldsymbol{f}_n$ aufgespannten konvexen Polyeders bezeichnet (vgl. hierzu auch Kapitel 4.3, wo spezieller derartige konvexe Mengen behandelt werden); die Bernsteinpolynome (Bézier–Kurven) haben dann die Form

$$B_n(\boldsymbol{f};t) = \sum_{k=0}^{n} \boldsymbol{f}_k \binom{n}{k} t^k (1-t)^{n-k}, \quad n \in \mathbb{N},\ t \in [0,1] \tag{2.3.69}$$

mit der für alle $n \in \mathbb{N}$ und $t \in [0,1]$ gültigen Fehlerabschätzung

$$\|B_n(\boldsymbol{f};t) - \boldsymbol{f}(t)\| \le n\delta L + \Delta \frac{t(1-t)}{n\delta^2}, \quad \delta > 0. \tag{2.3.70}$$

Eine einfache Rechnung zeigt, daß für $t \in (0,1)$ die rechte Seite minimal bzgl. δ wird, wenn

$$\delta = \sqrt[3]{\frac{2t(1-t)\Delta}{n^2 L}}$$

gewählt wird, woraus die Abschätzung

$$\|B_n(\boldsymbol{f};t) - \boldsymbol{f}(t)\| \le 2\sqrt[3]{t(1-t)n\Delta L^2} \tag{2.3.71}$$

für $n \in \mathbb{N}$, $t \in (0,1)$ resultiert, die aus Stetigkeitsgründen sogar noch für $t \in \{0,1\}$ richtig bleibt. Die Annäherung der Bézier–Kurve an den durch die Knoten erzeugten Polygonzug wird also umso genauer, je kleiner der maximale Abstand L zwischen zwei aufeinanderfolgenden Knoten wird bzw. je "kleiner" das aufgespannte Polyeder ist. Insbesondere fallen — für $t = 0$ bzw. $t = 1$ — Anfangs– und Endpunkt der Bézier–Kurve mit den Anfangs– und Endknoten $\boldsymbol{f}_0$ bzw. $\boldsymbol{f}_n$ zusammen. Aufgrund von Lemma 2.2.1 verläuft ferner die Bézier–Kurve vollständig innerhalb des von den Knoten aufgespannten konvexen Polyeders, da die Zufallsvektoren $\boldsymbol{f}(S_n(t))$, $n \in \mathbb{N}$, $t \in [0,1]$ sämtlich Werte in diesem Polyeder annehmen. Die Knoten $\boldsymbol{f}_0, \ldots, \boldsymbol{f}_n$ steuern also allein durch ihre Lage in $\mathbb{R}^2$ den Verlauf der Beézier–Kurve, wobei noch zusätzlich gilt, daß die Tangentialvektoren der Bézier–Kurve im Anfangs– und Endpunkt mit den Vielfachen $n(\boldsymbol{f}_1 - \boldsymbol{f}_0)$ bzw. $n(\boldsymbol{f}_n - \boldsymbol{f}_{n-1})$ der Anfangs– bzw. Endkantenvektoren zusammenfallen, was insbesondere auf einfache Weise ein "knickfreies" Aneinanderfügen von Bézier–Kurven durch geeignete Wahl der Anschlußknoten ermöglicht. Dies liegt daran, daß — bei beliebigem $\boldsymbol{f}$ — die Bernsteinpolynome $B_n(\boldsymbol{f};\cdot)$ für kleine t die Darstellung

$$B_n(\boldsymbol{f};t) = (1-t)^n \boldsymbol{f}(0) + nt(1-t)^{n-1} \boldsymbol{f}\Big(\frac{1}{n}\Big) + \boldsymbol{R}_n(t)$$

besitzen mit

$$\frac{\|\boldsymbol{R}_n(t)\|}{t} \longrightarrow 0, \quad t \downarrow 0,$$

da $\boldsymbol{R}_n(t)$ nur noch Terme mit Faktor t^k für $k \geq 2$ enthält. Damit erhält man aber

$$\begin{aligned}\frac{1}{t}\big(B_n(\boldsymbol{f};t) - \boldsymbol{f}(0)\big) &= n(1-t)^{n-1}\boldsymbol{f}\Big(\frac{1}{n}\Big) - \frac{1-(1-t)^n}{t}\boldsymbol{f}(0) + \frac{1}{t}\boldsymbol{R}_n(t)\\ &\sim n(1-t)^{n-1}\left(\boldsymbol{f}\Big(\frac{1}{n}\Big) - \boldsymbol{f}(0)\right)\\ &\longrightarrow n\left(\boldsymbol{f}\Big(\frac{1}{n}\Big) - \boldsymbol{f}(0)\right), \quad t \downarrow 0,\end{aligned}$$

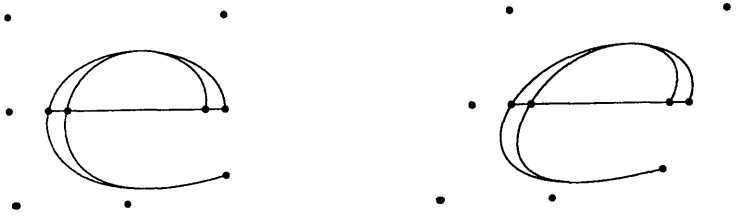
woraus im Falle eines Polygonzugs wegen $\boldsymbol{f}_i = \boldsymbol{f}\big(\frac{i}{n}\big)$, $i = 0, 1$ die Behauptung für die Anfangskante folgt; analog argumentiert man für die Endkante.

Eine weitere für die Anwendungen wichtige Eigenschaft ist die Invarianz der Bézier–Kurven gegen affin–lineare Abbildungen, d.h. ist $\boldsymbol{A}$ eine 2×2–Matrix und $\boldsymbol{b} \in \mathbb{R}^2$ und notiert man alle betrachteten Vektoren als *Spaltenvektoren*, so gilt

$$B_n(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{b};t) = \boldsymbol{A}B_n(\boldsymbol{f};t) + \boldsymbol{b}, \quad n \in \mathbb{N},\ t \in [0,1]. \tag{2.3.72}$$

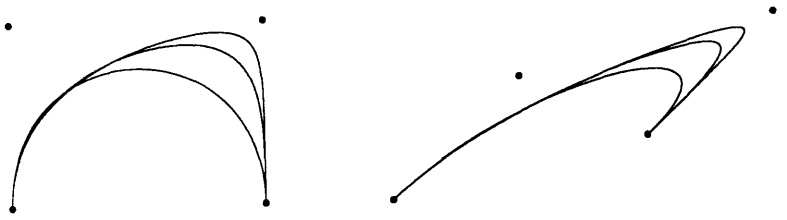
Dies ist eine unmittelbare Konsequenz aus der Linearität des Erwartungswerts (2.2.21) und bedeutet praktisch, daß bei Polygonzügen die Kurve $\boldsymbol{A}B_n(\boldsymbol{f};\cdot) + \boldsymbol{b}$ gerade die Bézier–Kurve zu den affin–linear transformierten Knoten $\boldsymbol{A}\boldsymbol{f}_0 + \boldsymbol{b}, \ldots,$ $\boldsymbol{A}\boldsymbol{f}_n + \boldsymbol{b}$ darstellt, d.h. will man die ursprüngliche Bézier–Kurve affin–linear transformieren, so genügt es, lediglich die Ursprungsknoten zu transformieren.

Die folgende Graphik verdeutlicht diesen Sachverhalt anhand des Buchstabens "e", der hier aus drei verschiedenen Kurvenzügen zusammengesetzt ist und im zweiten Bild mit der Matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 0.3 \\ 1 & 0 \end{pmatrix}$ geschert wurde. Die Knoten sind zur Verdeutlichung mit eingezeichnet.



Eine weitere Einflußmöglichkeit auf den Verlauf der Bézier–Kurve besteht durch sogenannte *multiple* Knoten, d.h. $\boldsymbol{f}_i = \boldsymbol{f}_{i+1} = \ldots = \boldsymbol{f}_{i+k}$ für ein i mit $0 \leq i \leq n-k$, $k \in \mathbb{N}$. Nach den obigen Ausführungen ist zu erwarten, daß die Kurve in der Nähe solcher Knoten dichter am zugehörigen Polygonzug liegt; die folgenden beiden Bilder verdeutlichen dies für $k = 1, 2, 3$, wobei die zweite Darstellung durch lineare Transformation mit $\boldsymbol{A} = \begin{pmatrix} 0.5 & 1 \\ 0.7 & 0.3 \end{pmatrix}$ aus der ersten

hervorgeht.



Besonders eindrucksvoll lassen sich mit dem obigen Zugang auch parametrische Flächendarstellungen erzeugen, deren Form ebenfalls nur durch wenige, räumliche Knoten gesteuert wird. Hierzu betrachten wir Abbildungen $\boldsymbol{f} : [0,1] \times [0,1] \to \mathbf{R}^3$ sowie jeweils stochastisch unabhängige, $\mathcal{R}([0,1])$–verteilte Zufallsvariablen $\{U_n, V_n\}_{n\in\mathbf{N}}$. Definiert man wieder $X_n(t)$, $n \in \mathbf{N}$, $t \in [0,1]$, wie in (2.3.56) sowie analog

$$Y_m(s) = \lfloor s + V_m \rfloor, \quad m \in \mathbf{N},\ s \in [0,1], \tag{2.3.73}$$

und

$$S_n(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t), \quad T_m(s) = \frac{1}{m}\sum_{j=1}^{m} Y_j(s), \quad n,m \in \mathbf{N},\ t,s \in [0,1], \tag{2.3.74}$$

so lassen sich entsprechend Bernsteinpolynome (Bézier–Flächen) $B_{nm}(\boldsymbol{f};t,s)$ definieren durch

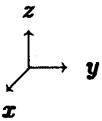
$$\begin{aligned} B_{nm}(\boldsymbol{f};t,s) &= E\big[\boldsymbol{f}(S_n(t),T_m(s))\big] \\ &= \sum_{k=0}^{n}\sum_{j=0}^{m} \boldsymbol{f}\Big(\frac{k}{n},\frac{j}{m}\Big)\binom{n}{k}\binom{m}{j} t^k(1-t)^{n-k}s^j(1-s)^{m-j} \end{aligned} \tag{2.3.75}$$

für $n,m \in \mathbf{N}$, $t,s \in [0,1]$. Bézier–Flächen besitzen dieselben Eigenschaften wie Bézier–Kurven, d.h. insbesondere approximieren sie auf $[0,1] \times [0,1]$ *stetige* Funktionen $\boldsymbol{f}$ gleichmäßig für $n,m \to \infty$. Gibt man $(n+1)\cdot(m+1)$ Knoten $\boldsymbol{f}_{00},\ldots,$ $\boldsymbol{f}_{nm} \in \mathbf{R}^3$ vor, $n,m \in \mathbf{N}$, so kann man analog zu (2.3.69) das zu der von diesen Knoten erzeugten Polyederfläche gehörige Bernsteinpolynom einfacher schreiben als

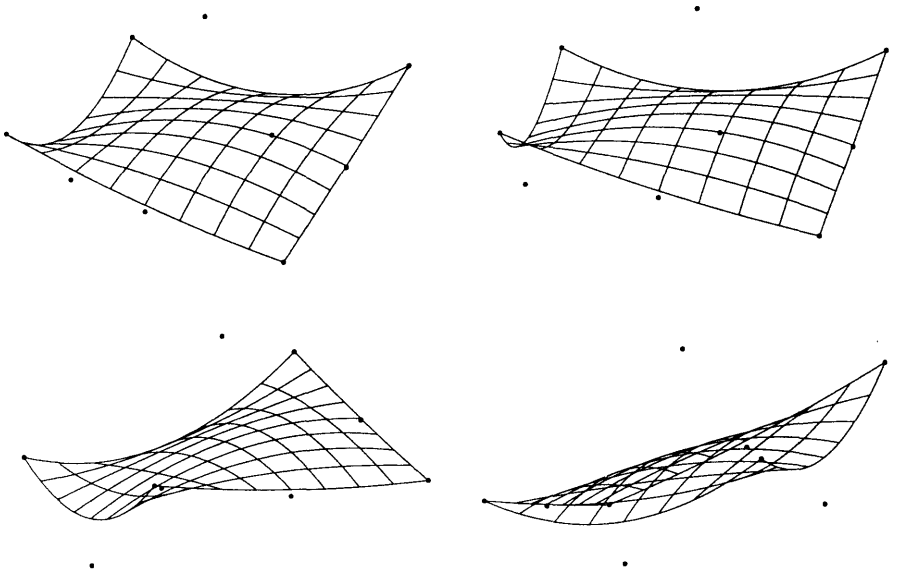
$$B_{nm}(\boldsymbol{f};t,s) = \sum_{k=0}^{n}\sum_{j=0}^{m} \boldsymbol{f}_{kj}\binom{n}{k}\binom{m}{j} t^k(1-t)^{n-k}s^j(1-s)^{m-j} \tag{2.3.76}$$

für $n,m \in \mathbf{N}$, $t,s \in [0,1]$. Auch hier gilt, daß die Bézier–Fläche vollständig in dem von den Knoten aufgespannten konvexen Polyeder verläuft, und daß die Tangentialflächen am Rand mit denen des Polyeders zusammenfallen, was wiederum ein kantenfreies Zusammensetzen mehrerer Bézier–Flächen erlaubt, wenn die Anschlußknoten geeignet gewählt werden. Die Invarianz gegenüber linear–affinen Abbildungen analog zu (2.3.72) bleibt auch hier erhalten; insbesondere lassen sich

damit auf sehr einfache Weise 3D–Darstellungen der Flächen (z.B. durch Netze) bewerkstelligen, da Projektionen von $\mathbf{R}^3$ in $\mathbf{R}^2$ durch geeignete 2×3–Matrizen $\boldsymbol{A}$ beschrieben werden. Man hat dazu lediglich die (dreidimensionalen) Knoten entsprechend zu projizieren, wodurch man zweidimensionale Knoten $\boldsymbol{A}f_{00}, \ldots, \boldsymbol{A}f_{nm}$ erhält; die einzelnen Netzlinien bilden dann Bézier–Kurven der vorher behandelten Art. Wählt man z.B. $\boldsymbol{A} = \begin{pmatrix} -\frac{1}{2}\sqrt{2} & 1 & 0 \\ -\frac{1}{2}\sqrt{2} & 0 & 1 \end{pmatrix}$, so ergibt sich als graphische Darstellung der Einheitsvektoren $\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\boldsymbol{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\boldsymbol{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$:



Entsprechend einfach lassen sich z.B. Rotationen der Bézier–Flächen um den Winkel φ (im Bogenmaß) in der Grundebene graphisch realisieren, indem man die Knoten zunächst mit der Matrix $\boldsymbol{B} = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $0 \leq \varphi \leq 2\pi$ multiplizert und anschließend mit $\boldsymbol{A}$ projiziert. Die folgenden vier Graphiken zeigen eine Bézier–Fläche mit 9 Knoten, die zur Verdeutlichung mit angegeben sind, unter den Drehwinkeln 0°, 20°, 70° und 180°.



Für eine ausführlichere Behandlung von Bézier–Kurven und –Flächen verweisen wir z.B. auf das Buch von Encarnação & Straßer (1986), in dem auch eine einfachere, rekursive graphische Konstruktion von Bézier–Kurven besprochen wird.

2.4. Aufgaben

2.1 Es sei $\{X_n\}_{n\in\mathbf{N}}$ eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Zeigen Sie: $X = \lim_{n\to\infty} X_n$ ist ebenfalls eine Zufallsvariable (d.h. meßbare Abbildung) auf $(\Omega, \mathcal{A}, P)$.

2.2 Es seien X und Y jeweils unabhängige, $\mathfrak{L}(\{1,\dots,n\})$–verteilte Zufallsvariablen, $n \in \mathbf{N}$. Zeigen Sie: die Zufallsvariable $U = X + Y$ ist (diskret) dreiecksverteilt im Sinne von Aufgabe 1.8.
Sind X und Y unabhängige, $\mathcal{R}([0,1])$–verteilte Zufallsvariablen, so ist entsprechend $V = X + Y$ stetig dreiecksverteilt. Wie sehen jeweils die Parameter der Dreiecksverteilungen aus?

2.3 In einem Feld von $N \in \mathbf{N}$ Elementen befinden sich K Schlüsselelemente ($1 \le K \le N$). Aus den Feldelementen wird "zufällig" eine Menge von n Elementen ausgewählt ($1 \le n \le N$). Die Zufallsvariable X bezeichne die Anzahl der Schlüsselelemente, die sich in der ausgewählten Menge befinden. Zeigen Sie:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad 0 \le k \le K.$$

Die Verteilung von X, P^X, heißt auch *hypergeometrische Verteilung*.
Zeigen Sie, daß die Verteilungen aus Beispiel 1.1.3 hierin als Spezialfall enthalten sind.

2.4 (Lineare Maximumsuche) Bei der linearen Suche nach dem größten Element in einer zufälligen Permutation $\eta \in \mathcal{P}_{k,n}$, $1 < k \le n \in \mathbf{N}$ (vgl. Aufgabe 1.10) wird zunächst das erste Element η_1 als Referenzelement abgespeichert; dieses wird sukzessive verglichen mit den nachfolgenden Elementen η_i, $1 < i \le k$, bis erstmalig ein größeres Element $\eta_j > \eta_1$ gefunden wird. Dieses wird als neues Referenzelement gegen das alte η_1 ausgetauscht; die weiteren Vergleiche erfolgen mit η_i, $j < i \le k$ usw. Wird durch Vergleiche kein größeres Element mehr gefunden, ist das aktuelle Referenzelement damit als das gewünschte Maximum identifiziert. Zur Beurteilung der Effizienz dieses Verfahrens ist u.a. die Zufallsvariable S_k von Interesse, die angibt, wie oft eine (Um–)Speicherung des Referenzelementes bis zur Identifikation des Maximums nötig ist. Zeigen Sie unter wesentlicher Verwendung von Aufgabe 1.10, daß S_k $\mathfrak{P}\mathfrak{B}(k; 1, \frac{1}{2}, \dots, \frac{1}{k})$–verteilt ist. Folgern Sie hieraus:

S_k ist approximativ $\mathfrak{P}\left(\sum_{i=1}^{k} \frac{1}{i}\right)$–verteilt mit $\rho\Big(P^{S_k}, \mathfrak{P}\left(\sum_{i=1}^{k} \frac{1}{i}\right)\Big) \le \frac{\pi^2}{6 \ln k}$;

$$P(S_k = 1) = \frac{1}{k}, \quad P(S_k = 2) = \frac{1}{k}\sum_{i=1}^{k-1}\frac{1}{i}, \quad P(S_k = k-1) = \frac{1}{2(k-2)!}, \quad P(S_k = k) = \frac{1}{k!};$$

es gilt die Rekursionsformel $P(S_{k+1} = i) = \frac{1}{k+1}P(S_k = i-1) + \frac{k}{k+1}P(S_k = i), 1 \le i \le k+1$.

(Man vergleiche hierzu auch die Ausführungen in Knuth (1968), 1.2.10 und Kemp (1984), §3.)

2.5 Es seien X, Y unabhängige, je $\mathcal{R}([0,1])$–verteilte Zufallsvariable. Zeigen Sie: der Zufallsvektor $Z = (Z_1, Z_2) = (X+Y-1, X-Y)$ ist $\mathcal{R}(A)$–verteilt über dem Quadrat $A = \{(x,y) \in \mathbf{R}^2 \mid |x| + |y| \le 1\}$ (Skizze!). Sind Z_1 und Z_2 stochastisch unabhängig bzw. unkorreliert? Wie können alternativ $\mathcal{R}(A)$–verteilte Zufallsvektoren durch eine geeignete Verwerfungsmethode erzeugt werden? Wie groß ist dann der Erwartungswert der zugehörigen Stoppzeitenverteilung?

2.6 Es sei X_1 $\mathcal{E}(\lambda)$–verteilt mit $\lambda > 0$, X_2 unabhängig von X_1 $\mathcal{R}((0,1))$–verteilt und $g(x) = e^{-x}$, $x \ge 0$. Zeigen Sie: der durch $\boldsymbol{Y} = \big(g(X_1)\cos(2\pi X_2), g(X_1)\sin(2\pi X_2)\big)$ definierte Zufallsvektor besitzt eine Dichte der Form

$$f_{\boldsymbol{Y}}(y_1, y_2) = \begin{cases} \frac{\lambda}{2\pi}\sqrt{y_1^2 + y_2^2}^{\lambda-2} & \text{für } 0 < y_1^2 + y_2^2 \le 1 \\ 0 & \text{sonst.} \end{cases}$$

Was ergibt sich hier für $\lambda = 2$?
Zeigen sie, daß die Komponenten von Y unkorreliert sind. (Hinweis: benutzen Sie die Symmetrie der Verteilung.)

2.7 (Lineare Maximumsuche) Zeigen Sie, daß für die Anzahl der (Um–)Speicherungen S_k der Referenzelemente in Aufgabe 2.4 gilt:

$$\ln k \le E(S_k) = \sum_{i=1}^{k} \frac{1}{i} \le 1 + \ln k;$$

$$\ln k - \frac{\pi^2}{6} \le \mathrm{Var}\,(S_k) = \sum_{i=1}^{k} \frac{1}{i}\left(1 - \frac{1}{i}\right) \le \ln k.$$

Schließen Sie hieraus mit Hilfe der Tschebyscheff–Ungleichung auf

$$P(S_k > 1 + c + \ln k) \le \frac{\ln k}{c^2} \quad \text{für alle } c > 0.$$

Zeigen Sie damit, daß für $k = 1000$ gilt: $P(S_k > 20) < 0.05$. Macht hier eine Approximation der Verteilung P^{S_k} gemäß Aufgabe 2.4 Sinn?

2.8 (Sortieren durch Maximum–Auswahl) Eine zufällige Permutation $\eta \in \mathcal{P}_{k,n}$, $1 < k \le n \in \mathbf{N}$ (vgl. Aufgabe 1.10 bzw. 2.4)) soll der Größe nach geordnet werden, indem man zunächst durch lineare Maximumsuche das größte Element bestimmt und dieses auf den letzen Platz setzt. Nach Streichung des Maximums aus der Permutation verfährt man mit den übrigen $k-1$ Elementen analog; das nächste Maximum wird auf den vorletzten Platz gesetzt usw. S bezeichne dabei die Anzahl der insgesamt benötigten *Umspeicherungen* der Referenzelemente. Zeigen Sie unter Verwendung von Aufgabe 2.4:

$$E(S) = \sum_{j=1}^{k} E(S_j) - k = (k+1)\sum_{j=1}^{k} \frac{1}{j} - 2k \le (k+1)\ln k - k + 1.$$

(Vergl. hierzu auch Mehlhorn (1988), II.1.1.)

2.9 (**binary search**) Berechnen Sie den Erwartungswert der Anzahl der Schritte X bis zum Abbruch des Verfahrens für die in Aufgabe 1.5 spezifizierte Verteilung der Platznummer des Schlüsselelements. Vergleichen Sie das Ergebnis mit dem entsprechenden Wert für das *lineare* Suchverfahren, bei dem der Reihe nach die Elemente $2^n - 1, 2^n - 2, \ldots, 1$ abgefragt werden. Geben Sie für $n = 2$ eine Verteilung der Platznummer des Schlüsselelements an, bei der das lineare Suchverfahren "im Mittel" besser als binäres Suchen ist.

2.10 Berechnen Sie Erwartungswert und Varianz $\mathcal{E}(\lambda)$– bzw. $\mathcal{G}(p)$–verteilter Zufallsvariablen mit Hilfe von Satz 2.2.1 c) bzw. Lemma 2.2.2 d) ($\lambda > 0$, $0 < p < 1$).

2.11 Bestimmen Sie Erwartungswert und Varianz diskret bzw. stetig dreiecksverteilter Zufallsvariablen (vgl. Aufgabe 1.8).

2.12 (**hybridsort**) Sortiert man in Beispiel 2.1.2 die N_j Elemente in Korb j ($1 \le j \le k$) durch paarweise Vergleiche, so benötigt man insgesamt $S = \sum_{j=1}^{k} \frac{N_j(N_j-1)}{2}$ Vergleiche zum Sortieren des Gesamtfelds. Berechnen Sie Erwartungswert und Varianz von S. Wie verhalten sich diese Größen für große Werte von n?

2.13 Es sei X eine Zufallsvariable, für die die momenterzeugende Funktion Ψ_X an einer Stelle $t > 0$ existieren möge. Zeigen Sie unter Verwendung der Markoff–Ungleichung (2.2.26):

$$P(X > c) \leq \frac{\Psi_X(t)}{e^{tc}}, \quad c \in \mathbf{R}.$$

2.14 Es sei X $\mathcal{N}(0,1)$–verteilt. Zeigen Sie: die Zufallsvariable $Y = e^X$ besitzt eine Dichte f_Y der Form

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{y} \exp\left(-\frac{\ln^2 y}{2}\right) & \text{für } y > 0 \\ 0 & \text{sonst} \end{cases}$$

(sog. *log–Normalverteilung*). Zeigen Sie weiter: Y besitzt Momente $E(Y^k) = \Psi_X(k)$ beliebiger Ordnung ($k \in \mathbf{N}$); die Funktion

$$g(y) = f_Y(y) \cdot (1 + \sin(2\pi \ln y)), \quad y \in \mathbf{R}$$

ist die Dichte einer Verteilung Q; ist Z eine Zufallsvariable mit Verteilung $P^Z = Q$, so besitzt Z dieselben Momente wie Y (d.h. die log–Normalverteilung ist durch ihre Momente *nicht* eindeutig bestimmt).

2.15 Verifizieren Sie die Gültigkeit der in den Tabellen auf Seite 135 angegebenen Formeln für Erwartungswert, Varianz sowie die (moment–)erzeugenden Funktionen der angegebenen Verteilungen.

2.16 Es seien $\{X_\lambda\}$ $\mathfrak{P}(\lambda)$–verteilte Zufallsvariablen ($\lambda > 0$). Zeigen Sie unter Verwendung momenterzeugender Funktionen, daß mit $\lambda \to \infty$ die Verteilung von $\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$ schwach gegen $\mathcal{N}(0,1)$ konvergiert. Wie kann für Werte $\lambda = n\lambda_0$ ($n \in \mathbf{N}$, $\lambda_0 > 0$ fest) dasselbe Ergebnis (für $n \to \infty$) auch aus dem Zentralen Grenzwertsatz 2.3.7 gefolgert werden?

2.17 Es sei $\boldsymbol{f}_0 = (-1,0)$, $\boldsymbol{f}_1 = \ldots = \boldsymbol{f}_{n-1} = (0,1)$ und $\boldsymbol{f}_n = (1,0)$, $n \geq 2$. Geben Sie die zugehörige Bézier–Kurve an und zeigen Sie, daß gilt:

$$\|B_n(\boldsymbol{f}; \tfrac{1}{2}) - \boldsymbol{f}(\tfrac{1}{2})\| = \frac{1}{2^{n-1}}, \quad n \geq 2,$$

wobei $\boldsymbol{f}$ den durch die Knoten erzeugten Polygonzug (Dreieck) bezeichne. Schließen Sie hieraus, daß mit wachsender Knotenzahl n die Bézier–Kurve das durch die Knoten erzeugte Dreieck "gleichmäßig approximiert" (Skizze!). Warum ist dies kein Widerspruch zu

$$\lim_{n\to\infty} \|B_n(\boldsymbol{f}; \tfrac{1}{n}) - \boldsymbol{f}(\tfrac{1}{n})\| = \sqrt{2}\left(1 - \frac{1}{e}\right) > 0?$$

2.18 Es sei $\boldsymbol{f} : [0,1] \longrightarrow \mathbf{R}^2$ eine stetige Abbildung. Zeigen Sie, daß die Folge der zugehörigen Bézier–Kurven gleichmäßig gegen $\boldsymbol{f}$ konvergiert, wenn die Knoten $\boldsymbol{f}_0, \ldots, \boldsymbol{f}_n$, $n \in \mathbf{N}$, so gewählt werden, daß sie auf dem zu $\boldsymbol{f}$ gehörigen Graphen liegen und für den Abstand $L = \sup_{1\leq i\leq n} \|\boldsymbol{f}_i - \boldsymbol{f}_{i-1}\|$ je zweier aufeinanderfolgender Knoten gilt: $\sqrt{n}L \to 0$ ($n \to \infty$).

3. Grundlagen Stochastischer Prozesse

Bei der stochastischen Modellierung komplexer Systeme muß man in der Regel eine Vielzahl gegenseitiger Verflechtungen berücksichtigen. Im Bereich der Informatik trifft dies insbesondere zu in den Bereichen Rechnernetzwerke und Leistungsbewertung von Rechnersystemen, der Bildverarbeitung, der Sprach- und Mustererkennung, der künstlichen Intelligenz (Neuronale Netze, Expertensysteme) usw. Modelle, in denen im wesentlichen nur stochastische *Unabhängigkeit* der beteiligten Zufallsvariablen bzw. Zufallselemente vorausgesetzt wird, sind deshalb kaum geeignet, die tatsächlichen Gegebenheiten solcher Systeme richtig widerzuspiegeln. Im allgemeinen wird man allerdings gar nicht alle möglichen Abhängigkeiten modellmäßig erfassen können; man wird daher versuchen, das Modell aus einigen wenigen Grundtypen von Abhängigkeitsstrukturen, die in der Sprache der Stochastik formalisierbar und — mit nicht allzu großem Aufwand — auch analysierbar sind, zusammenzusetzen. Hierdurch wird zumindest eine in der Praxis hinreichend brauchbare Approximation an die "wahre" Situation möglich.

Eine der einfachsten und zugleich zentralen Abhängigkeitsstrukturen, die vor allem bei zeitorientierten Modellen wie Netzwerken und Bedienungssystemen anzutreffen ist, wird durch die *Markoff'sche* Abhängigkeit von Zufallsvariablen beschrieben (etwa in der Form von Markoff–Ketten). Ihre charakteristische Eigenschaft ist anschaulich die "Gedächtnislosigkeit", d.h. die zukünftige Entwicklung des Systems hängt stochastisch nur von dem momentanen Istzustand ab, unabhängig von der davor liegenden Vergangenheit. Markoff–Ketten sind z.B. fundamental für die Modellierung gewisser probabilistischer Suchalgorithmen wie das in jüngerer Zeit populär gewordene "Simulated Annealing" oder die Simulation Neuronaler Netze. Auch bei der Muster- und Spracherkennung sowie im Zusammenhang mit Expertensystemen spielen Markoff–Modelle eine wichtige Rolle.

3.1. Bedingte Verteilungen und Erwartungswerte

Wesentliche Grundlage für die stochastische Beschreibung von Abhängigkeitsstrukturen bilden die bereits in elementarer Form in den Kapiteln 1 und 2 behandelten bedingten Verteilungen und bedingten Erwartungswerte (vgl. Definition 1.1.9 und 2.2.2). In diesem Abschnitt soll nun schrittweise ein allgemeinerer bedingter Verteilungsbegriff vorgestellt werden, mit dem auch nicht–diskrete Probleme, die typischerweise etwa bei zeitabhängigen Modellen auftreten, behandelt werden können.

Definiton 3.1.1. *(bedingte Verteilung)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{A}$ ein Ereignis. Die in Abhängigkeit von B erklärte Verteilung $P(\cdot \mid B)$ auf $\mathcal{A}$,

$$P(A \mid B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)} & \text{für } P(B) > 0 \\ Q_B & \text{sonst,} \end{cases} \tag{3.1.1}$$

wobei Q_B eine beliebige, feste Verteilung auf $\mathcal{A}$ sei, heißt bedingte Verteilung (unter der Hypothese B).

Im Unterschied zu Definition 1.1.9 ist damit die bedingte Verteilung $P(\cdot \mid B)$ also auch noch für den Fall erklärt, daß $P(B) = 0$ gilt, wobei hier eine gewisse Wahlfreiheit besteht, die hauptsächlich technischer Natur ist, wie das folgende Lemma zeigt. Sie garantiert jedenfalls, daß $P(\cdot \mid B)$ für *jedes* Ereignis $B \in \mathcal{A}$ tatsächlich eine Wahrscheinlichkeitsverteilung ist.

Lemma 3.1.1. *(Satz von der totalen Wahrscheinlichkeit)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\{B_n\}_{n \in \mathbf{N}} \subseteq \mathcal{A}$ eine (höchstens abzählbare) disjunkte Zerlegung von Ω, d.h.

$$\Omega = \bigcup_{n=1}^{\infty} B_n \quad \text{mit} \quad B_i \cap B_j = \emptyset,\ i \neq j,\ i, j \in \mathbf{N}.$$

Dann gilt:

$$P(A) = \sum_{n=1}^{\infty} P(A \mid B_n) P(B_n), \qquad A \in \mathcal{A}. \tag{3.1.2}$$

Ferner ist für jedes Ereignis $A \in \mathcal{A}$ mit $P(A) > 0$

$$P(B_i \mid A) = \frac{P(A \mid B_i) P(B_i)}{\sum_{n=1}^{\infty} P(A \mid B_n) P(B_n)}, \qquad i \in \mathbf{N}. \tag{3.1.3}$$

Beweis. Es ist wegen der Zerlegungseigenschaft der B_n, $n \in \mathbf{N}$,

$$\begin{aligned} P(A) &= P\Big(\bigcup_{n=1}^{\infty} A \cap B_n \Big) = \sum_{n=1}^{\infty} P(A \cap B_n) \\ &= \sum_{n=1}^{\infty} P(A \mid B_n) P(B_n). \end{aligned}$$

Hierbei hat man lediglich zu beachten, daß für Ereignisse B_n mit $P(B_n) = 0$ die Wahl von Q_{B_n} aus (3.1.1) keine Rolle spielt, da die entsprechende bedingte Wahrscheinlichkeit $P(A \mid B_n)$ mit dem Wert $P(B_n) = 0$ multipliziert wird und damit keinen Beitrag zu der Summe in (3.1.2) liefert.
Beziehung (3.1.3) ergibt sich für $P(A) > 0$ unmittelbar aus (3.1.2) wegen

$$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)}, \quad i \in \mathbb{N}.$$

∎

Die in Definition 3.1.1 getroffene Vereinbarung erlaubt es damit auch, durch Wahl der leeren Menge $\emptyset$ für gewisse der Ereignisse B_n, $n \in \mathbb{N}$, entsprechende Aussagen für *endliche* Zerlegungen der Grundmenge Ω zu erhalten.

Beziehung (3.1.3) ist auch als *Satz von Bayes* bekannt; sie ist insbesondere dann von Nutzen, wenn man die Ereignisse B_n, $n \in \mathbb{N}$, als mögliche *Ursachen* für das Ereignis A ansehen kann und man bei Beobachtung von A etwa darauf schließen möchte, welche Ursache B_i, $i \in \mathbb{N}$, das Eintreten von A am wahrscheinlichsten bewirkt hat. Solche Fragestellungen sind z.B. typisch in medizinischen Expertensystemen.

Beispiel 3.1.1. (Expertensystem)
Aus medizinischen Untersuchungen sei bekannt, daß die Symptome A_1 und A_2 bei (genau) drei Krankheiten B_1, B_2 und B_3 auftreten können, und zwar mit (bedingten) Wahrscheinlichkeiten $\pi_{ij} = P(A_j \mid B_i)$, $i = 1,2,3$, $j = 1,2$, die wir zweckmäßigerweise in Matrixform

$$\mathbf{\Pi} = (\pi_{ij}) = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.9 \\ 0.4 & 0.6 \end{pmatrix}$$

notieren. Die Eintrittswahrscheinlichkeiten für die Krankheiten B_1, B_2 und B_3 seien durch den Vektor $\boldsymbol{p} = (0.3 \quad 0.6 \quad 0.1)$ gegeben. Nach dem Satz von der totalen Wahrscheinlichkeit (3.1.2) läßt sich dann der Vektor $\boldsymbol{q} = (P(A_1) \quad P(A_2))$ der Eintrittswahrscheinlichkeiten für die Symptome darstellen als

$$\boldsymbol{q} = \boldsymbol{p} \cdot \mathbf{\Pi} = (0.3 \quad 0.6 \quad 0.1) \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.9 \\ 0.4 & 0.6 \end{pmatrix} = (0.40 \quad 0.72).$$

Für die bedingten Wahrscheinlichkeiten $\pi^*_{ji} = P(B_i \mid A_j)$, $j = 1,2$, $i = 1,2,3$, erhält man daraus nach dem Satz von Bayes (3.1.3)

$$\mathbf{\Pi}^* = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1\overline{6} & 0.75 & 0.08\overline{3} \end{pmatrix}.$$

Beachtenswert hierbei ist, daß zwar — wie erwartet — bei Auftreten des Symptoms A_1 ein Patient am wahrscheinlichsten an der Krankheit B_1 und bei Auftreten von

A_2 am wahrscheinlichsten an B_2 leidet, daß er aber z.B. bei Auftreten des Symptoms A_1 — wenn er *nicht* an B_1 erkrankt ist — häufiger an der Krankheit B_2 als an B_3 leidet, obwohl A_1 häufiger unter der Krankheit B_3 als unter B_2 vorkommt. Dies liegt daran, daß die Krankheit B_3 selbst nur relativ selten auftritt: B_2 kommt etwa dreimal so häufig vor wie B_3.
Welche Aussagen kann man nun beispielsweise treffen, wenn ein Patient zwar Symptom A_1, nicht aber A_2 aufweist? Zunächst lassen sich mit (1.0.3) aus der rechten Spalte der Matrix $\mathbf{\Pi}$ relativ einfach die bedingten Wahrscheinlichkeiten $\begin{pmatrix} P(A_2^c \mid B_1) \\ P(A_2^c \mid B_2) \\ P(A_2^c \mid B_3) \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.1 \\ 0.4 \end{pmatrix}$ bestimmen. Nach dem Satz von Bayes (3.1.3) erhält man dann $\big(P(B_1 \mid A_2^c) \quad P(B_2 \mid A_2^c) \quad P(B_3 \mid A_2^c)\big) = (0.677... \quad 0.193... \quad 0.129...)$. Zusammen mit dem vorigen Ergebnis legt dies nahe, daß der Patient mit hoher Wahrscheinlichkeit an der Krankheit B_1 leidet.
Allerdings lassen sich die beiden Teilergebnisse für $P(B_1 \mid A_1)$ und $P(B_1 \mid A_2^c)$ nicht ohne weiteres kombinieren; hierzu müßten z.B. die bedingten Wahrscheinlichkeiten $P(A_1 \cap A_2 \mid B_i)$, $i = 1,2,3$, für das gleichzeitige Auftreten beider Symptome bekannt sein. Nimmt man an, daß diese für das betrachtete Problem gegeben sind durch den (Spalten–)Vektor $(0.2 \quad 0.1 \quad 0.3)^{tr}$, so kann man wegen $P(A_1 \cap A_2^c \mid B_i) = P(A_1 \mid B_i) - P(A_1 \cap A_2 \mid B_i)$, $i = 1,2,3$ den Satz von Bayes erneut anwenden mit dem Ergebnis

$$\big(P(B_1 \mid A_1 \cap A_2^c) \quad P(B_2 \mid A_1 \cap A_2^c) \quad P(B_3 \mid A_1 \cap A_2^c)\big) = (0.72 \quad 0.24 \quad 0.04).$$

Die Berücksichtigung von mehr Information in Form der gemeinsamen bedingten Verteilung für beide Symptome bringt also hier eine größere Sicherheit in der Diagnose, da der Wert von 0.72 für $P(B_1 \mid A_1 \cap A_2^c)$ die beiden Vergleichswerte $P(B_1 \mid A_1) = 0.6$ und $P(B_1 \mid A_2^c) = 0.677...$ jeweils übertrifft. Dies muß allerdings nicht *zwingend* der Fall sein! Wäre beispielsweise die bedingte gemeinsame Verteilung der Symptome gegeben durch den Vektor $(0.2 \quad 0.05 \quad 0.05)^{tr}$, so ergäbe eine analoge Rechnung $P(B_1 \mid A_1 \cap A_2^c) = 0.59...$, also einen kleineren Wert als das Minimum von $P(B_1 \mid A_1) = 0.6$ und $P(B_1 \mid A_2^c) = 0.677....$ Dies zeigt, daß man bei der Interpretation bedingter Wahrscheinlichkeiten äußerst vorsichtig sein muß und möglichst nur Informationen in Form *gemeinsamer* bedingter Wahrscheinlichkeiten verwenden sollte. ∎

Beispiel 3.1.2. (Ein–Prozessor–System mit einer I/O–Einheit)
Ein System bestehe aus einer CPU und einer Ein– bzw. Ausgabeeinheit (z.B. Diskettenstation). Nach Beendigung eines Programms auf der CPU wird entweder die I/O–Einheit mit Wahrscheinlichkeit $p \in (0,1)$ benutzt oder ein neues Programm (mit Wahrscheinlichkeit $q = 1 - p$) gestartet. Nach Benutzung der I/O–Einheit wird in jedem Fall ein neues Programm gestartet, ebenso beim Einschalten des Systems. A_n, $n \in \mathbb{N}$, bezeichne das Ereignis, daß zu Beginn des n–ten Arbeitszyklus (d.h. entweder CPU– oder I/O–Benutzung) ein neues Programm gestartet wird. Wie groß sind die Wahrscheinlichkeiten $P(A_n)$, $n \in \mathbb{N}$, und wie verhalten sich diese für große n (Langzeitverhalten des Systems)?
Zur Beantwortung dieser Frage können wir wieder Lemma 3.1.1 heranziehen. Nach

der obigen Vorgabe ist

$$\begin{aligned} &P(A_{n+1} \mid A_n) = q, \quad P(A_{n+1}^c \mid A_n) = p, \\ &P(A_{n+1} \mid A_n^c) = 1, \quad P(A_{n+1}^c \mid A_n^c) = 0, \quad n \in \mathbb{N}, \end{aligned} \tag{3.1.4}$$

mit $P(A_1) = 1$. Bezeichnen wir mit $q_n = P(A_n)$, $n \in \mathbb{N}$, die Wahrscheinlichkeit für einen Programmstart zu Beginn des Zyklus n, so erhalten wir mit (3.1.2) die Rekursion

$$\begin{aligned} q_1 = 1, \qquad q_{n+1} &= P(A_{n+1}) \\ &= P(A_{n+1} \mid A_n)P(A_n) + P(A_{n+1} \mid A_n^c)P(A_n^c) \\ &= q \cdot q_n + 1 - q_n = 1 - p \cdot q_n, \quad n \in \mathbb{N}. \end{aligned}$$

Mit vollständiger Induktion läßt sich leicht zeigen, daß die explizite Lösung dieser Rekursion gegeben ist durch

$$P(A_n) = q_n = \sum_{i=0}^{n-1} (-p)^i = \frac{1-(-p)^n}{1+p}, \quad n \in \mathbb{N},$$

mit

$$\lim_{n\to\infty} P(A_n) = \frac{1}{1+p} = \frac{1}{2-q}.$$

Die Ereignisse $\{A_n\}_{n\in\mathbb{N}}$ sind dabei nicht stochastisch unabhängig, denn es ist z.B.

$$P(A_{n+1} \cap A_n) = P(A_{n+1} \mid A_n)P(A_n) = q \cdot q_n \neq q_{n+1} \cdot q_n = P(A_{n+1})P(A_n)$$

für jedes $n \in \mathbb{N}$, $n \geq 2$ (vgl. (1.1.21)); die Gleichung $q = q_{n+1}$ führt nämlich auf $(-p)^{n-1} = 1$, was nur für $n = 1$ richtig ist. ∎

Beispiel 3.1.2 wird im nachfolgenden Abschnitt etwas allgemeiner noch einmal im Zusammenhang mit Markoff-Ketten aufgegriffen (Beispiel 3.2.5)).

Interessiert man sich in dem zuletzt behandelten Beispiel ferner für den Zeitpunkt, zu dem nach Einschalten des Systems erstmalig auf die I/O-Einheit zugegriffen wird, also für die Zufallsvariable (Stoppzeit) $S = \inf\{n \in \mathbb{N} \mid \mathbb{1}_{A_n^c} = 1\}$, so benötigt man die Wahrscheinlichkeiten von *Durchschnitten* der Ereignisse $A_1, \ldots, A_n$, $n \in \mathbb{N}$, denn es gilt analog zu Beziehung (2.1.32):

$$\{S = n\} = \bigcap_{i=1}^{n-1} A_i \cap A_n^c = \bigcap_{i=1}^{n-1} A_i \setminus \bigcap_{i=1}^{n} A_i, \quad n \in \mathbb{N}$$

und somit

$$P(S = n) = P\Big(\bigcap_{i=1}^{n-1} A_i^c \cap A_n\Big) = P\Big(\bigcap_{i=1}^{n-1} A_i\Big) - P\Big(\bigcap_{i=1}^{n} A_i\Big), \quad n \in \mathbb{N}.$$

Die Angabe der bedingten Wahrscheinlichkeiten in (3.1.4) reicht hierfür aber nicht aus; vielmehr benötigt man die bedingten Wahrscheinlichkeiten unter *allen* Durchschnitten vorhergehender Ereignisse, wie das nachfolgende Lemma zeigt.

Lemma 3.1.2. *(Zusammensetzen bedingter Wahrscheinlichkeiten)*
Es sei $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum mit* $A, B, C, A_1, \ldots, A_n \in \mathcal{A}$, $n \in \mathbf{N}$, $P(B \cap C) > 0$. *Dann gilt:*

$$P([A \mid B] \mid C) = P([A \mid C] \mid B) = P(A \mid B \cap C) \tag{3.1.5}$$

$$P(A \mid B \cap C) \cdot P(B \mid C) = P([A \mid B] \mid C) \cdot P(B \mid C) = P(A \cap B \mid C) \tag{3.1.6}$$

$$P\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{k=1}^{n} P\Big(A_k \mid \bigcap_{i=1}^{k-1} A_i\Big)\, P(A_1). \tag{3.1.7}$$

Beweis. Es bezeichne $Q_C(\cdot) = P(\cdot \mid C)$. Dann ist

$$P([A \mid B] \mid C) = Q_C(A \mid B) = \frac{Q_C(A \cap B)}{Q_C(B)} = \frac{P(A \cap B \cap C)}{P(C)} \frac{P(C)}{P(B \cap C)}$$
$$= \frac{P(A \cap B \cap C)}{P(B \cap C)} = P(A \mid B \cap C).$$

Durch Vertauschung von B und C erhält man hieraus (3.1.5). Mit derselben Bezeichnung folgt ferner

$$P(A \mid B \cap C) \cdot P(B \mid C) = P([A \mid B] \mid C) \cdot P(B \mid C) = Q_C(A \mid B) Q_C(B)$$
$$= Q_C(A \cap B) = P(A \cap B \mid C).$$

Dies ist (3.1.6). Beziehung (3.1.7) ergibt sich hieraus induktiv über

$$P\Big(\bigcap_{i=k}^{n} A_i \mid \bigcap_{i=1}^{k-1} A_i\Big) = P\Big(\bigcap_{i=k+1}^{n} A_i \mid \bigcap_{i=1}^{k} A_i\Big) \cdot P\Big(A_k \mid \bigcap_{i=1}^{k-1} A_i\Big), \quad 1 < k < n,$$

falls $P\Big(\bigcap_{i=1}^{n} A_i\Big) > 0$ gilt. Anderenfalls sind beide Seiten von (3.1.7) Null, d.h. Beziehung (3.1.7) ist in jedem Fall gültig. ∎

Das folgende Modell zeigt, daß die Verteilung der Stoppzeit S des letzten Beispiels tatsächlich nicht durch die bedingten Wahrscheinlichkeiten in (3.1.4) allein bestimmt ist:
Wählt man etwa

$$P(A_2 \cap A_3 \cap A_4) = rq^2 \qquad P(A_2 \cap A_3^c \cap A_4) = 0$$
$$P(A_2 \cap A_3 \cap A_4^c) = (1-r)q^2 \qquad P(A_2^c \cap A_3 \cap A_4) = (1-r)q^2 + p^2 q, \quad r \in (0,1),$$

so zeigt eine ausführlichere Rechnung, daß — unabhängig von r — Beziehung (3.1.4) für $n = 1, 2, 3$ erfüllt ist; wegen $P(A_1) = 1$ folgt demnach

$$P(S = 4) = P(A_2 \cap A_3) - P(A_2 \cap A_3 \cap A_4) = q^2 - rq^2 = (1-r)q^2,$$

was offensichtlich von r abhängt.

Beziehung (3.1.5) läßt sich anschaulich wieder mit Lemma 2.1.5 interpretieren: ist nämlich $\{\boldsymbol{X}_n\}_{n\in\mathbf{N}}$ eine unabhängige Folge von Zufallselementen in einem Meßraum $(\mathcal{X},\mathcal{B})$ mit Verteilung $Q = P^{\boldsymbol{X}_n}$, $n \in \mathbf{N}$, und sind B und C Ereignisse mit $Q(B \cap C) > 0$, so liefert wie in (2.1.44) die Folge der Stoppzeiten

$$S_1 = \inf\{k \in \mathbf{N} \mid \boldsymbol{X}_k \in C\}, \quad S_{n+1} = \inf\{k > S_n \mid \boldsymbol{X}_k \in C\}, \quad n \in \mathbf{N},$$

eine unabhängige Folge $\{\boldsymbol{Y}_n\}_{n\in\mathbf{N}} = \{\boldsymbol{X}_{S_n}\}_{n\in\mathbf{N}}$ von Zufallselementen mit Verteilung $P^{\boldsymbol{Y}_n} = Q(\cdot \mid C)$, $n \in \mathbf{N}$; nimmt man nun zu dieser neu gewonnenen Folge die Stoppzeiten

$$T_1 = \inf\{k \in \mathbf{N} \mid \boldsymbol{Y}_k \in B\}, \quad T_{n+1} = \inf\{k > T_n \mid \boldsymbol{Y}_k \in B\}, \quad n \in \mathbf{N},$$

hinzu, so besitzen die Zufallselemente $\boldsymbol{Y}_{T_n}$, $n \in \mathbf{N}$, die Verteilung $Q([\cdot \mid B] \mid C)$. Andererseits werden durch das Stoppen mit $\{T_n\}_{n\in\mathbf{N}}$ gerade diejenigen Beobachtungen ausgewählt, die sowohl in C (durch Stoppen mit $\{S_n\}_{n\in\mathbf{N}}$) als auch in B liegen, d.h. die Folge $\{\boldsymbol{Y}_{T_n}\}_{n\in\mathbf{N}}$ ist *identisch* mit der Folge $\{\boldsymbol{X}_{U_n}\}_{n\in\mathbf{N}}$, wobei

$$U_1 = \inf\{n \in \mathbf{N} \mid \boldsymbol{X}_n \in B \cap C\}, \quad U_{n+1} = \inf\{n > U_n \mid \boldsymbol{X}_n \in B \cap C\}, \quad n \in \mathbf{N}.$$

Will man — ähnlich wie in (3.1.1) — auf die Voraussetzung $P(B \cap C) > 0$ in Lemma 3.1.2 verzichten, ergeben sich bereits in den einfachsten Situationen gewisse Schwierigkeiten bezüglich einer konsistenten Wahl der Ausnahmeverteilungen. Das folgende Beispiel verdeutlicht dies:
Wählt man etwa $\emptyset \subset \Omega \subset \mathbf{R}$ als endliche Menge mit $P = \mathfrak{L}(\Omega)$ sowie eine disjunkte Zerlegung $\Omega = B \cup C$, $B \cap C = \emptyset$, $B, C \neq \emptyset$, so gilt zwar (im Sinne von (1.2.16))

$$P(\cdot \mid B) = \mathfrak{L}(B) =: \mathfrak{L}_B, \quad P(\cdot \mid C) = \mathfrak{L}(C) =: \mathfrak{L}_C,$$

für die bedingten Verteilungen

$$\begin{aligned} P([\cdot \mid B] \mid C) &= \mathfrak{L}_C(\cdot \mid B) \\ P([\cdot \mid C] \mid B) &= \mathfrak{L}_B(\cdot \mid C) \\ P(\cdot \mid B \cap C) &= P(\cdot \mid \emptyset) \end{aligned} \tag{3.1.8}$$

kann gemäß Definition 3.1.1 aber je eine beliebige (feste) Verteilung gewählt werden, so daß für diesen Fall die Gleichheit in (3.1.5) nur dann gilt, wenn alle drei Verteilungen *identisch* gewählt werden, was natürlich hier möglich ist. Die Beziehung (3.1.6) bleibt dagegen auch dann richtig, wenn verschiedene Versionen der bedingten Verteilungen in (3.1.8) gewählt werden, da in jedem Fall $P(B \mid C) = P(A \cap B \mid C) = 0$ gilt.

Die letzten Ausführungen zeigen also, daß es auf die Festlegung der Ausnahmeverteilungen in (3.1.1) dann nicht so sehr ankommt, wenn im wesentlichen Zusammensetzungen bedingter Wahrscheinlichkeiten — wie in (3.1.2), (3.1.5) oder (3.1.6) — von Interesse sind. Eine ähnliche Beobachtung haben wir bereits früher im Zusammenhang mit fast–sicher bestehenden Eigenschaften gemacht: man denke etwa an die Integration von Zufallsvariablen oder Verteilungsdichten, bei denen Abänderungen auf Nullmengen ebenfalls keine Rolle spielen.

Im Hinblick darauf, daß im Gegensatz zu Lemma 3.1.1 bedingte Wahrscheinlichkeiten häufig auch bei *überabzählbar* vielen Bedingungen benötigt werden (etwa in Systemen, die in kontinuierlicher Zeit arbeiten), wollen wir jetzt eine Erweiterung des Begriffs der bedingten Verteilung vornehmen, der eine Beschreibung solcher Situationen und zugleich eine sinnvolle Behandlung der auftretenden Ausnahmeverteilungen erlaubt.

Wir kehren dazu zunächst noch einmal zu der Situation von Lemma 3.1.1 zurück. Die Partition der Grundmenge Ω durch die Mengen $\{B_n\}_{n\in\mathbb{N}}$ erzeugt in kanonischer Weise die σ–Algebra

$$\mathcal{C} = \{ \bigcup_{i\in I} B_i \mid I \subseteq \mathbb{N} \}, \tag{3.1.9}$$

wobei $\bigcup_{i\in\emptyset} B_i = \emptyset$ vereinbart sei. $\mathcal{C}$ ist eine Teil–σ–Algebra von $\mathcal{A}$; wir wollen dies durch die Schreibweise

$$\mathcal{C} \sqsubseteq \mathcal{A} \tag{3.1.10}$$

zum Ausdruck bringen. Definiert man für $A \in \mathcal{A}$ die Zufallsvariable

$$P(A \mid \mathcal{C})(\omega) = \sum_{n=1}^{\infty} P(A \mid B_n) \mathbb{1}_{B_n}(\omega), \quad \omega \in \Omega, \tag{3.1.11}$$

so ist diese $\mathcal{C}$–meßbar, denn $P(A \mid \mathcal{C})$ ist konstant auf allen Mengen B_n, $n \in \mathbb{N}$. Wegen $\mathcal{C} \sqsubseteq \mathcal{A}$ ist sie natürlich auch $\mathcal{A}$–meßbar. Andererseits erhalten wir durch Integration

$$P(A \cap B) = \sum_{\substack{n\in I \\ B=\bigcup_{i\in I} B_i}} P(A \mid B_n) P(B_n) = \int_B P(A \mid \mathcal{C})(\omega)\, dP(\omega), \quad A \in \mathcal{A},\ B \in \mathcal{C}. \tag{3.1.12}$$

Diese Darstellung verdeutlicht noch einmal, daß die Zufallsvariable $P(A \mid \mathcal{C})$ zwar nur fast–sicher eindeutig bestimmt ist, da sie auf den Mengen B_n, $n \in \mathbb{N}$, mit $P(B_n) = 0$ im Sinne von (3.1.1) beliebig festgelegt werden kann, die Integration in (3.1.12) aber unabhängig davon für alle Ereignisse $A \in \mathcal{A}$, $B \in \mathcal{C}$ dieselbe Wahrscheinlichkeit $P(A \cap B)$ ergibt. Die fast–sichere Bestimmtheit von $P(A \mid \mathcal{C})$ bezieht sich dabei auf die auf $\mathcal{C}$ eingeschränkte Verteilung $P_{\mathcal{C}}$.

Es wird sich zeigen, daß der hier gewählte Zugang der Definition bedingter Verteilungen als Zufallsvariable, die bezüglich gewisser Teil–σ–Algebren meßbar sind und die die charakteristische Gleichung (3.1.12) erfüllen, geeignet ist, die oben angesprochenen Problemkreise angemessen zu behandeln.

Definition 3.1.2. *(allgemeine bedingte Wahrscheinlichkeit)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine Teil–σ–Algebra. Jede $\mathcal{C}$–meßbare Zufallsvariable $P(A \mid \mathcal{C})$, $A \in \mathcal{A}$, mit Werten im Intervall $[0,1]$ und $P(\emptyset \mid \mathcal{C}) = 0$, $P(\Omega \mid \mathcal{C}) = 1$, die der Gleichung

$$P(A \cap B) = \int_B P(A \mid \mathcal{C})(\omega)\, dP(\omega) = \int_B P(A \mid \mathcal{C})\, dP \quad \text{für alle } B \in \mathcal{C} \tag{3.1.13}$$

genügt, heißt bedingte Wahrscheinlichkeit von A unter (der σ–Algebra) $\mathcal{C}$.

Die Wahl von $B = \Omega$ in (3.1.13) ergibt dabei eine naheliegende Verallgemeinerung des Satzes von der totalen Wahrscheinlichkeit, Beziehung (3.1.2), für beliebige σ–Algebren $\mathcal{C}$, die man kürzer auch als $P(A) = E\big[P(A \mid \mathcal{C})\big]$, $A \in \mathcal{A}$, ausdrücken kann.

Die aus Beziehung (3.1.13) gewonnenen Lösungen sind dabei sogar $P_{\mathcal{C}}$–fast sicher eindeutig bestimmt, wie das folgende Resultat zeigt.

Lemma 3.1.3. *(fast sichere Eindeutigkeit der bedingten Verteilung)*
In der Situation von Definition 3.1.2 gilt: Sind $P_1(\cdot \mid \mathcal{C})$ und $P_2(\cdot \mid \mathcal{C})$ Versionen der bedingten Verteilung von $P(\cdot \mid \mathcal{C})$, so gilt:

$$P_1(\cdot \mid \mathcal{C}) = P_2(\cdot \mid \mathcal{C}) \quad P_{\mathcal{C}}\text{– fast sicher.}$$

Beweis. Es bezeichne $U = P_1(A \mid \mathcal{C})$, $V = P_2(A \mid \mathcal{C})$, $A \in \mathcal{A}$. Die Menge $B = \{U > V\} = \bigcup_{q \in \mathbb{Q}} \{U > q > V\} = \bigcup_{q \in \mathbb{Q}} \{U > q\} \cap \{V < q\}$ gehört dann zu $\mathcal{C}$; die Zufallsvariable $Z = (U - V) \cdot \mathbb{1}_B$ ist damit nicht–negativ und $\mathcal{C}$–meßbar. Nach Satz 2.2.2 j) und (3.1.13) erhält man somit

$$\int Z \, dP = \int_B U \, dP - \int_B V \, dP = P(A \cap B) - P(A \cap B) = 0,$$

also $Z = 0$ P–fast sicher. Nach Definition der Menge B ist aber $U > V$ auf B; es muß also $P(B) = 0$ gelten und somit $U \leq V$ $P_{\mathcal{C}}$–fast sicher. Analog erhält man $U \geq V$ $P_{\mathcal{C}}$–fast sicher und damit die Behauptung. ∎

Bedingte Wahrscheinlichkeiten sind also i.a. nur $P_{\mathcal{C}}$–fast sicher eindeutig bestimmt und i.a. auch nicht $\mathcal{A}$–meßbar, wenn $\mathcal{C} \neq \mathcal{A}$ gilt. Bei durch abzählbare Partitionen erzeugten Teil–σ–Algebren $\mathcal{C}$ wie in Lemma 3.1.1 ist die Struktur von $P(A \mid \mathcal{C})$ allerdings weitgehend festgelegt; darüberhinaus zeigt Beziehung (3.1.1), daß die gemäß (3.1.11) definierte bedingte Verteilung $P(\cdot \mid \mathcal{C})(\omega)$ für *jedes* $\omega \in \Omega$ tatsächlich ein Wahrscheinlichkeitsmaß ist. Diese Eigenschaft ist nicht ohne weiteres in allen Wahrscheinlichkeitsräumen gültig: zwar erhält man mit den Bezeichnungen von Definition 3.1.2 noch die Gleichung

$$P\Big(\bigcup_{n=1}^{\infty} A_n \,\Big|\, \mathcal{C}\Big) = \sum_{n=1}^{\infty} P(A_n \mid \mathcal{C}) \quad P_{\mathcal{C}}\text{–fast sicher} \tag{3.1.14}$$

für paarweise disjunkte Folgen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$, die Ausnahmemengen hängen aber jeweils von der ganzen Folge $\{A_n\}_{n \in \mathbb{N}}$ ab — von diesen können jedoch insgesamt überabzählbar viele verschiedene existieren. Es ist deshalb u.U. nicht möglich, eine $\mathcal{C}$–meßbare Version von $P(\cdot \mid \mathcal{C})$ zu finden, die $P_{\mathcal{C}}$–fast sicher — mit einer für alle Ereignisse gleichzeitig wählbaren Ausnahmemenge — ein Wahrscheinlichkeitsmaß ist (vgl. auch Aufgabe 33.13 in Billingsley (1986), S. 464, oder Aufgabe 2, §54 in Bauer (1978)). Hierzu muß die Grundmenge Ω im allgemeinen besondere topologische Eigenschaften besitzen, etwa vollständige[1] Metrisierbarkeit und abzählbare

[1] d.h. Konvergenz von Cauchy–Folgen bzgl. der Metrik

Erzeugtheit der Topologie, d.h. jede offene Teilmenge von Ω ist als abzählbare Vereinigung von Basismengen darstellbar (vgl. Bauer (1978), Satz 56.5). Diese Bedingungen sind beispielsweise für $\Omega = \mathbf{R}^m$, $m \in \mathbf{N}$, erfüllt, so daß wir im folgenden stets annehmen können und werden, daß entsprechende bedingte Verteilungen $P(\cdot \mid \mathcal{C})(\omega)$ o.B.d.A. sogar für alle $\omega \in \Omega$ Wahrscheinlichkeitsmaße bilden. Man sagt in diesem Fall auch, die bedingten Verteilungen seien *regulär*.

Besonders wichtig sind in diesem Zusammenhang σ–Algebren, die von Zufallsvariablen bzw. allgemeiner Zufallselementen erzeugt werden; besitzen diese z.B. (Zähl–)Dichten, so lassen sich reguläre bedingte Verteilungen relativ leicht bestimmen.

Definition 3.1.3. *(durch Zufallselemente erzeugte σ–Algebren)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\boldsymbol{Y}$ ein Zufallselement in einem Meßraum $(\mathcal{X}, \mathcal{B})$. Dann heißt

$$\sigma(\boldsymbol{Y}) = \{\boldsymbol{Y}^{-1}(B) \mid B \in \mathcal{B}\} \sqsubseteq \mathcal{A} \tag{3.1.14}$$

die durch $\boldsymbol{Y}$ erzeugte σ–Algebra.

Die Eigenschaft $\sigma(\boldsymbol{Y}) \sqsubseteq \mathcal{A}$ folgt dabei unmittelbar aus der Meßbarkeit von $\boldsymbol{Y}$.

Der folgende wichtige Satz zeigt, daß $\sigma(\boldsymbol{Y})$–meßbare, reellwertige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$— etwa bedingte Verteilungen unter $\mathcal{C} = \sigma(\boldsymbol{Y})$ — eine besonders einfache Struktur besitzen, da sie sich als (meßbare) Transformationen von $\boldsymbol{Y}$ erweisen.

Satz 3.1.1. *(Faktorisierungssatz)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\boldsymbol{Y}$ ein Zufallselement in einem Meßraum $(\mathcal{X}, \mathcal{B})$. $\mathcal{C}$ bezeichne die von $\boldsymbol{Y}$ erzeugte Teil–σ–Algebra von $\mathcal{A}$. Eine reellwertige Zufallsvariable X ist genau dann $\mathcal{C}$–meßbar, wenn eine $\mathcal{B}$–meßbare Abbildung $h : (\mathcal{X}, \mathcal{B}) \longrightarrow (\mathbf{R}, \mathcal{B}^1)$ existiert mit $X = h \circ \boldsymbol{Y}$.

Beweis. Sei zunächst $X = h \circ \boldsymbol{Y}$. Dann ist X $\mathcal{C}$–meßbar wegen

$$X^{-1}(B) = \boldsymbol{Y}^{-1}\big(\underbrace{h^{-1}(B)}_{\in \mathcal{B}}\big) \in \mathcal{C}, \quad B \in \mathcal{B}^1.$$

Zum Beweis der Umkehrung benutzen wir das Prinzip der algebraischen Induktion (vgl. S. 108):
Nimmt X nur endlich viele Werte $\alpha_1, \ldots, \alpha_n$, $n \in \mathbf{N}$, an, so besitzt nach Voraussetzung X eine Darstellung $X = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ mit $A_i = X^{-1}(\{\alpha_i\}) \in \mathcal{C}$, $1 \le i \le n$. Zu jedem i existiert daher eine Menge $B_i \in \mathcal{B}$ mit $A_i = \boldsymbol{Y}^{-1}(B_i)$, $1 \le i \le n$, so daß die Abbildung $h = \sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}$ das Gewünschte leistet.
Ist $X \ge 0$, so existiert eine Folge $\{X_n\}_{n \in \mathbf{N}}$ von nicht–negativen Zufallsvariablen mit je endlich vielen Werten und $X_n \uparrow X$, $n \to \infty$. Nach obigem existiert zu jedem $n \in \mathbf{N}$ eine $\mathcal{B}$–meßbare Abbildung h_n mit $X_n = h_n \circ \boldsymbol{Y}$. Hier kann $h = \sup_{n \in \mathbf{N}} h_n$ gewählt werden.
Im allgemeinen Fall zerlege man X in Positiv– und Negativteil: $X = X^+ - X^-$. Sind nun h^+, h^- $\mathcal{B}$–meßbare Abbildungen mit $X^+ = h^+ \circ \boldsymbol{Y}$, $X^- = h^- \circ \boldsymbol{Y}$, so leistet hier $h = h^+ - h^-$ das Gewünschte; man beachte dabei, daß h^+ und h^- so gewählt werden können, daß nicht zugleich der Wert $h^+(\boldsymbol{x}) = h^-(\boldsymbol{x}) = \infty$ für ein $\boldsymbol{x} \in \mathcal{X}$ angenommen wird. ∎

Für bedingte Verteilungen $P(\cdot \mid \sigma(\boldsymbol{Y}))$ mit Zufallselementen $\boldsymbol{Y}$ schreiben wir im folgenden auch kürzer $P(\cdot \mid \boldsymbol{Y})$. Die nach dem Faktorisierungssatz (3.1.1) für jedes Ereignis $A \in \mathcal{A}$ existierende Abbildung h_A mit

$$P(A \mid \boldsymbol{Y}) = h_A \circ \boldsymbol{Y} \tag{3.1.15}$$

wollen wir einprägsamer auch als

$$h_A(\boldsymbol{y}) = P(A \mid \boldsymbol{Y} = \boldsymbol{y}), \quad \boldsymbol{y} \in \mathcal{X}, \tag{3.1.16}$$

schreiben.

Aufgrund des Transformationssatzes 2.2.3 ist die Definitionsgleichung (3.1.13) in diesem Fall äquivalent zu

$$\begin{aligned} P(A \cap \{\boldsymbol{Y} \in B\}) &= \int_{\{\boldsymbol{Y} \in B\}} P(A \mid \boldsymbol{Y})\, dP \\ &= \int_B P(A \mid \boldsymbol{Y} = \boldsymbol{y})\, dP^{\boldsymbol{Y}}(\boldsymbol{y}), \quad A \in \mathcal{A},\ B \in \mathcal{B}. \end{aligned} \tag{3.1.17}$$

Lassen sich die Abbildungen h_A so wählen, daß $P(\cdot \mid \boldsymbol{Y} = \boldsymbol{y})$ für jedes $\boldsymbol{y} \in \mathcal{X}$ wieder eine Wahrscheinlichkeitsverteilung ist, so spricht man bei $P(\cdot \mid \boldsymbol{Y} = \boldsymbol{y})$ ebenfalls von der (regulären) bedingten Verteilung unter (der Hypothese) $\boldsymbol{Y} = \boldsymbol{y}$.

Der folgende Satz gestattet eine für praktische Zwecke besonders nützliche Konstruktion regulärer bedingter Verteilungen für den Fall, daß die beteiligten Zufallselemente endlich–dimensionale Zufallsvektoren mit (Zähl–)Dichten bilden.

Satz 3.1.2. *(bedingte Verteilungen und Dichten)*
Es sei $\boldsymbol{X}$ ein n–dimensionaler und $\boldsymbol{Y}$ ein m–dimensionaler Zufallsvektor ($n, m \in \mathbb{N}$) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ derart, daß $P^{(\boldsymbol{X},\boldsymbol{Y})}$ eine $n+m$–dimensionale (Zähl–)Dichte $f_{(\boldsymbol{X},\boldsymbol{Y})}$ besitze. In diesem Fall existiert eine reguläre Version der bedingten Verteilung $P^{\boldsymbol{X}}(\cdot \mid \boldsymbol{Y} = \boldsymbol{y})$; eine solche ist etwa gegeben durch die zugehörige bedingte (Zähl–)Dichte

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Y} = \boldsymbol{y}) = \begin{cases} \dfrac{f_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{x}, \boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})}, & \text{falls } f_{\boldsymbol{Y}}(\boldsymbol{y}) > 0 \\ g_{\boldsymbol{Y}}(\boldsymbol{x}), & \text{sonst,} \end{cases} \qquad \boldsymbol{y} \in \mathbb{R}^m. \tag{3.1.18}$$

Hierbei bezeichne $g_{\boldsymbol{Y}}$ eine beliebige (Zähl–)Dichte und $f_{\boldsymbol{Y}}$ die Randdichte von $\boldsymbol{Y}$.

Beweis. Wir zeigen die Aussage zunächst im Fall von Dichten. Für $A \in \mathcal{B}^n$, $B \in \mathcal{B}^m$ gilt definitionsgemäß nach (3.1.18)

$$\begin{aligned} P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B) &= \int_A \int_B f_{(\boldsymbol{X},\boldsymbol{Y})}(x_1, \ldots, x_n, y_1, \ldots, y_m)\, dx_1 \ldots dx_n\, dy_1 \ldots dy_m = \\ \int_B \left[\int_A f_{\boldsymbol{X}}(x_1, \ldots, x_n \mid \boldsymbol{Y} = (y_1, \ldots, y_m))\, dx_1 \ldots dx_n \right] & f_{\boldsymbol{Y}}(y_1, \ldots, y_m)\, dy_1 \ldots dy_m \\ &= \int_B \left[\int_A f_{\boldsymbol{X}}(x_1, \ldots, x_n \mid \boldsymbol{Y} = \boldsymbol{y})\, dx_1 \ldots dx_n \right] dP^{\boldsymbol{Y}}(\boldsymbol{y}); \end{aligned}$$

andererseits gilt nach (3.1.17) auch

$$P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B) = \int_B P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y})\, dP^{\boldsymbol{Y}}(\boldsymbol{y}),$$

woraus durch Vergleich die Behauptung folgt.
Im Fall von Zähldichten argumentiert man analog mit Summation statt Integration. ∎

Beziehung (3.1.18) zeigt, daß im diskreten Fall die reguläre bedingte Zähldichte $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Y} = \boldsymbol{y})$ für $P(\boldsymbol{Y} = \boldsymbol{y}) > 0$ gerade mit der elementaren bedingten Wahrscheinlichkeit $P(\boldsymbol{X} = \boldsymbol{x} \mid \{\boldsymbol{Y} = \boldsymbol{y}\})$, $\boldsymbol{x} \in \mathbf{R}^n$, $\boldsymbol{y} \in \mathbf{R}^m$, übereinstimmt, was die Schreibweise (3.1.16) noch einmal aus anderer Sicht erklärt. Beziehung (3.1.18) gilt dementsprechend auch dann, wenn $\boldsymbol{X}$ und $\boldsymbol{Y}$ diskrete Zufallselemente mit Werten in beliebigen Meßräumen sind.

Eine anschauliche Erklärung von Satz 3.1.2 im Fall stetiger Verteilungen liefert das folgende Resultat.

Satz 3.1.3. *(stetige bedingte Verteilungen)*
In der Situation des Satzes 3.1.2 seien zusätzlich $f_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{x}, \cdot)$ für alle $\boldsymbol{x} \in \mathbf{R}^n$ und $f_{\boldsymbol{Y}}$ stetig in einem Punkt $\boldsymbol{y} \in \mathbf{R}^m$ mit $f_{\boldsymbol{Y}}(\boldsymbol{y}) > 0$. $U_h(\boldsymbol{y}) = \bigtimes_{i=1}^m [y_i - h, y_i + h]$ bezeichne für $h > 0$ die Würfel-Umgebung von $\boldsymbol{y}$ mit Kantenlänge $2h$. Dann gilt für jede Menge $A \in \mathcal{B}^n$:

$$P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}) = \lim_{h \downarrow 0} P\big(\boldsymbol{X} \in A \mid \{\boldsymbol{Y} \in U_h(\boldsymbol{y})\}\big). \tag{3.1.19}$$

Beweis. Nach Voraussetzung ist

$$P\big(\boldsymbol{Y} \in U_h(\boldsymbol{y})\big) = \int_{U_h(\boldsymbol{y})} \cdots \int f_{\boldsymbol{Y}}(z_1, \ldots, z_m)\, dz_1 \ldots dz_m > 0$$

für alle $h > 0$. Damit existiert $P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y})$ als elementare bedingte Wahrscheinlichkeit, d.h. es ist

$$\begin{aligned} P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}) &= \frac{P\big(\{\boldsymbol{X} \in A\} \cap \{\boldsymbol{Y} \in U_h(\boldsymbol{y})\}\big)}{P\big(\boldsymbol{Y} \in U_h(\boldsymbol{y})\big)} \\ &= \frac{\dfrac{1}{(2h)^m} \displaystyle\int_{U_h(\boldsymbol{y})} \cdots \int P\big(\boldsymbol{X} \in A \mid \boldsymbol{Y} = (z_1, \ldots, z_m)\big) f_{\boldsymbol{Y}}(z_1, \ldots, z_m)\, dz_1 \ldots dz_m}{\dfrac{1}{(2h)^m} \displaystyle\int_{U_h(\boldsymbol{y})} \cdots \int f_{\boldsymbol{Y}}(z_1, \ldots, z_m)\, dz_1 \ldots dz_m} \\ &\longrightarrow \frac{P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})} = P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}) \quad (h \downarrow 0); \end{aligned}$$

dies folgt wegen der Stetigkeit der Integranden z.B. aus dem Mittelwertsatz für mehrfache Riemann-Integrale (Heuser (1988), 201.5). ∎

Beziehung (3.1.19) zeigt damit, daß unter den genannten Voraussetzungen die bedingte Wahrscheinlichkeit $P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y})$ ähnlich wie im diskreten Fall interpretiert werden kann: die Bedingung $\{\boldsymbol{Y} \in U_h(\boldsymbol{y})\}$ besagt dabei, daß $\boldsymbol{Y}$ Werte "in der Nähe" von $\boldsymbol{y}$ annimmt. Man beachte jedoch, daß voraussetzungsgemäß $P(\boldsymbol{Y} = \boldsymbol{y}) = 0$ gilt, sogar für alle $\boldsymbol{y} \in \mathbf{R}^m$ wegen der Stetigkeit der Verteilung $P^{\boldsymbol{Y}}$. Eine sinnvolle Definition der bedingten Wahrscheinlichkeiten $P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y})$ gemäß (3.1.1) ist in solchen Fällen also zunächst gar nicht möglich, da sämtliche bedingenden Ereignisse die Wahrscheinlichkeit Null besitzen. Dies rechtfertigt nachträglich noch einmal den hier gewählten Zugang zu bedingten Verteilungen über σ–Algebren wie in Definition 3.1.2.

Die folgenden beiden Rechenbeispiele zeigen, wie man Beziehung (3.1.18) nutzbringend anwenden kann.

a) $X_1, \ldots, X_k$, $k \in \mathbf{N}$, seien unabhängige, $\mathfrak{P}(\lambda_i)$–verteilte Zufallsvariablen mit $\lambda_i > 0$, $1 \leq i \leq k$. Ferner sei $S = \sum_{i=1}^k X_i$. Dann ist S $\mathfrak{P}(\mu)$–verteilt mit $\mu = \sum_{i=1}^k \lambda_i$, und die bedingte Verteilung des Zufallsvektors $(X_1, \ldots, X_k)$ unter $\sigma(S)$ ist eine Multinomialverteilung, gegeben durch

$$P^{(X_1,\ldots,X_k)}(\cdot \mid S = n) = \mathfrak{M}(n; p_1, \ldots, p_k), \quad n \in \mathbf{N}, \tag{3.1.20}$$

mit $p_i = \dfrac{\lambda_i}{\mu}$, $1 \leq i \leq k$.

b) $X_1, \ldots, X_n$, $n \in \mathbf{N}$, seien unabhängige, je $\mathcal{E}(\lambda)$–verteilte Zufallsvariablen mit $\lambda > 0$. Ferner sei für $1 \leq k \leq n$ $S_k = \sum_{i=1}^k X_i$ und $\mathcal{K}_y = \{\boldsymbol{x} \in \mathbf{R}^{n-1} \mid 0 \leq x_1 \leq \ldots \leq x_{n-1} \leq y\}$, $y > 0$. Dann ist der Zufallsvektor $(S_1, \ldots, S_{n-1})$ unter $S_n = y$, $y > 0$, $\mathcal{R}(\mathcal{K}_y)$–verteilt, d.h. es gilt für $n \geq 2$

$$f_{(S_1,\ldots,S_{n-1})}(x_1, \ldots, x_{n-1} \mid S_n = y) = \begin{cases} \dfrac{(n-1)!}{y^{n-1}} & \text{für } (x_1, \ldots, x_{n-1}) \in \mathcal{K}_y \\ 0 & \text{sonst.} \end{cases} \tag{3.1.21}$$

Zum Nachweis von (3.1.20) seien $i_1, \ldots, i_k \in \{0, 1, \ldots, n\}$ mit $\sum_{j=1}^k i_j = n$, $n \in \mathbf{N}$. Für die bedingte Zähldichte von $(X_1, \ldots, X_k)$ unter $\sigma(S)$ ergibt sich dann mit (3.1.18)

$$\begin{aligned} f_{(X_1,\ldots,X_k)}(i_1, \ldots, i_k \mid S = n) &= \frac{P(X_1 = i_1, \ldots, X_k = i_k, S = n)}{P(S = n)} \\ &= \frac{P(X_1 = i_1, \ldots, X_k = i_k)}{P(S = n)} = \frac{\prod_{j=1}^k e^{-\lambda_j} \frac{\lambda_j^{i_j}}{i_j!}}{e^{-\mu} \frac{\mu^n}{n!}} \\ &= \binom{n}{i_1, \ldots, i_k} \prod_{j=1}^k \left(\frac{\lambda}{\mu}\right)^{i_j} = \mathfrak{M}(n; p_1, \ldots, p_k)(i_1, \ldots, i_k). \end{aligned}$$

Zum Nachweis von (3.1.21) benötigen wir zunächst die gemeinsame Dichte des Zufallsvektors $(S_1, \ldots, S_n)$. Mit der Abbildung $G(x_1, \ldots, x_n) = (x_1, x_1 + x_2, \ldots, \sum_{i=1}^n x_i)$, $(x_1, \ldots, x_n) \in \mathbf{R}^n$, erhält man über den Transformationssatz (2.1.8)

$$f_{(S_1,\ldots,S_n)}(y_1, \ldots, y_{n-1}, y) = \lambda^n e^{-\lambda y} \mathbb{1}_{\mathcal{K}_y}(y_1, \ldots, y_{n-1}) \tag{3.1.22}$$

für $(y_1, \ldots, y_{n-1}) \in \mathbf{R}^{n-1}$, $y > 0$, und somit unter Heranziehung von (2.1.82)

$$\begin{aligned} f_{(S_1,\ldots,S_{n-1})}(y_1, \ldots, y_{n-1} \mid S_n = y) &= \frac{f_{(S_1,\ldots,S_n)}(y_1, \ldots, y_{n-1}, y)}{f_{S_n}(y)} \\ &= \frac{\lambda^n e^{-\lambda y}}{\frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y}} = \frac{(n-1)!}{y^{n-1}}, \quad (y_1, \ldots, y_{n-1}) \in \mathcal{K}_y. \end{aligned}$$

Der folgende Satz zeigt u.a., wie Beziehung (3.1.6) auf den Fall allgemeinerer bedingter Verteilungen übertragen werden kann.

Satz 3.1.4. *(Eigenschaften bedingter Verteilungen)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum; $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ seien Zufallselemente mit Werten in Meßräumen $(\mathcal{X}, \mathcal{B})$, $(\mathcal{Y}, \mathcal{D})$, $(\mathcal{Z}, \mathcal{F})$. Dann gilt:

a) *Ist U eine reellwertige Zufallsvariable auf $(\mathcal{Y} \times \mathcal{Z}, \mathcal{D} \otimes \mathcal{F})$ und $P(\cdot \mid \boldsymbol{Z} = \boldsymbol{z})$, $\boldsymbol{z} \in \mathcal{Z}$, regulär, so gilt im Falle der Integrierbarkeit:*

$$\int U(\boldsymbol{y}, \boldsymbol{z})\, dP^{(\boldsymbol{Y},\boldsymbol{Z})}(\boldsymbol{y}, \boldsymbol{z}) = \int \int U(\boldsymbol{y}, \boldsymbol{z})\, dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Z}}(\boldsymbol{z}) \tag{3.1.23}$$

b) *Ist $P(\cdot \mid \boldsymbol{Z} = \boldsymbol{z})$, $\boldsymbol{z} \in \mathcal{Z}$, regulär, so gilt*

$$P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B \mid \boldsymbol{Z} = \boldsymbol{z}) = \int_B P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z}) \tag{3.1.24}$$

$P^{\boldsymbol{Z}}$–fast sicher für alle $A \in \mathcal{B}$, $B \in \mathcal{D}$, $\boldsymbol{z} \in \mathcal{Z}$.

c) *$\boldsymbol{X}$ und $\boldsymbol{Y}$ sind stochastisch unabhängig genau dann, wenn es eine reguläre Version von $P^{\boldsymbol{X}}(\cdot \mid \boldsymbol{Y} = \boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{Y}$, bzw. von $P^{\boldsymbol{Y}}(\cdot \mid \boldsymbol{X} = \boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}$, gibt, die nicht von $\boldsymbol{y}$ bzw. $\boldsymbol{x}$ abhängt.*

Beweis. a) Sei zunächst $U = \mathbb{1}_{D \times F}$ mit $D \in \mathcal{D}, F \in \mathcal{F}$. In diesem Fall ist

$$\begin{aligned} \int U(\boldsymbol{y},\boldsymbol{z})\, dP^{(\boldsymbol{Y},\boldsymbol{Z})}(\boldsymbol{y}, \boldsymbol{z}) &= P\big((\boldsymbol{Y}, \boldsymbol{Z}) \in D \times F\big) = P(\boldsymbol{Y} \in D, \boldsymbol{Z} \in F) \\ &= \int_F P(\boldsymbol{Y} \in D \mid \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Z}}(\boldsymbol{z}) = \int_D \int_F dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Z}}(\boldsymbol{z}) \\ &= \int\int_{D \times F} dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Z}}(\boldsymbol{z}) = \int \int U(\boldsymbol{y}, \boldsymbol{z})\, dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = \boldsymbol{z})\, dP^{\boldsymbol{Z}}(\boldsymbol{z}), \end{aligned}$$

also Beziehung (3.1.23) erfüllt. Da die σ–Algebra $\mathcal{D} \otimes \mathcal{F}$ aber von allen Mengen $\{D \times F \mid D \in \mathcal{D}, F \in \mathcal{F}\}$ erzeugt wird (vgl. Definition 1.4.1), läßt sich zeigen, daß

dies auch dann noch der Fall ist, wenn $U = \mathbb{1}_E$ eine beliebige Indikatorfunktion mit $E \in \mathcal{D} \otimes \mathcal{F}$ ist (das System aller Mengen $E \in \mathcal{D} \otimes \mathcal{F}$, für die (3.1.23) mit $U = \mathbb{1}_E$ gilt, bildet nämlich ein Dynkin–System). Mit algebraischer Induktion ergibt sich daraus schließlich die Behauptung.
b) Mit dem gerade gezeigten Teil a) ergibt sich einerseits

$$\begin{aligned} P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B, \boldsymbol{Z} \in F) &= \int_{B \times F} \int P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = z)\, dP^{(\boldsymbol{Y},\boldsymbol{Z})}(\boldsymbol{y}, z) \\ &= \int_F \int_B P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = z)\, dP^{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{Z} = z)\, dP^{\boldsymbol{Z}}(z), \end{aligned}$$

andererseits nach Definition der bedingten Verteilung

$$P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B, \boldsymbol{Z} \in F) = \int_F P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B \mid \boldsymbol{Z} = z)\, dP^{\boldsymbol{Z}}(z)$$

für alle $A \in \mathcal{A}$, $B \in \mathcal{D}$, $F \in \mathcal{F}$. Durch Vergleich der Integranden folgt damit die Aussage wie im Beweis von Lemma 3.1.3.
c) Seien zunächst $\boldsymbol{X}$ und $\boldsymbol{Y}$ stochastisch unabhängig. Dann ist $P^{\boldsymbol{X}}_{\cdot}(\cdot \mid \boldsymbol{Y} = \boldsymbol{y}) = P^{\boldsymbol{X}}$, $\boldsymbol{y} \in \mathcal{Y}$, eine solche Version, denn es gilt für alle $B \in \mathcal{D}$:

$$\begin{aligned} \int_B P(\boldsymbol{X} \in A)\, dP^{\boldsymbol{Y}}(\boldsymbol{y}) &= P(\boldsymbol{X} \in A) \int_B dP^{\boldsymbol{Y}}(\boldsymbol{y}) = P(\boldsymbol{X} \in A) P(\boldsymbol{Y} \in B) \\ &= P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B). \end{aligned}$$

Damit folgt der erste Teil der Behauptung.
Hängt umgekehrt $Q := P^{\boldsymbol{X}}(\cdot \mid \boldsymbol{Y} = \boldsymbol{y})$ nicht von $\boldsymbol{y}$ ab, so erhält man analog

$$\begin{aligned} P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B) &= \int_B P(\boldsymbol{X} \in A \mid \boldsymbol{Y} = \boldsymbol{y})\, dP^{\boldsymbol{Y}}(\boldsymbol{y}) = Q(A) \int_B dP^{\boldsymbol{Y}}(\boldsymbol{y}) \\ &= P(\boldsymbol{X} \in A) P(\boldsymbol{Y} \in B), \end{aligned}$$

also die stochastische Unabhängigkeit von $\boldsymbol{X}$ und $\boldsymbol{Y}$.
Die restliche Behauptung folgt durch Vertauschung von $\boldsymbol{X}$ und $\boldsymbol{Y}$. ∎

Insbesondere Teil a) des letzten Satzes ist auch von allgemeiner praktischer Bedeutung, da Beziehung (3.1.23) die Berechnung von Erwartungswerten durch iterierte Integration nach den bedingten Verteilungen erlaubt.

Ähnlich wichtig ist das folgende Lemma, welches eine vereinfachte Darstellung bedingter Wahrscheinlichkeiten erlaubt, bei denen Teile der bedingenden Zufallselemente in den zu bewertenden Ereignissen auftreten.

Lemma 3.1.4. *(Ersetzungslemma)*
$\boldsymbol{X}$ und $\boldsymbol{Y}$ seien Zufallselemente auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in Meßräumen $(\mathcal{X}, \mathcal{B})$ bzw. $(\mathcal{Y}, \mathcal{D})$. Die bedingte Verteilung $P(\cdot \mid \boldsymbol{Y} =$

$\boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{Y}$, *sei regulär,* $G : (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{D}) \longrightarrow (\mathcal{Z}, \mathcal{F})$ *sei eine meßbare Abbildung. Dann gilt für alle* $A \in \mathcal{F}$*:*

$$P\big(G(\boldsymbol{X},\boldsymbol{Y}) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big) = P\big((G(\boldsymbol{X},\boldsymbol{y}) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big) \\ := P\big(G(\boldsymbol{X},z) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big)\Big|_{z=\boldsymbol{y}} \quad P^{\boldsymbol{Y}}\text{-f.s.} \tag{3.1.25}$$

Sind insbesondere $\boldsymbol{X}$ *und* $\boldsymbol{Y}$ *stochastisch unabhängig, so gilt auch*

$$P\big(G(\boldsymbol{X},\boldsymbol{Y}) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big) = P\big(G(\boldsymbol{X},\boldsymbol{y}) \in A\big) \quad P^{\boldsymbol{Y}}\text{-f.s.} \tag{3.1.26}$$

Beweis. Wir zeigen die Aussage zunächst für den Fall $(\mathcal{Z},\mathcal{F}) = (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{D})$, $G(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x},\boldsymbol{y})$, $(\boldsymbol{x},\boldsymbol{y}) \in (\mathcal{Z},\mathcal{F})$ mit $A = B \times D$, $B \in \mathcal{B}, D \in \mathcal{D}$. Für beliebige $C \in \mathcal{D}$ ist dann

$$\int_C P\big((\boldsymbol{X},\boldsymbol{Y}) \in B \times D \mid \boldsymbol{Y} = \boldsymbol{y}\big)\, dP^{\boldsymbol{Y}}(\boldsymbol{y}) = P\big((\boldsymbol{X},\boldsymbol{Y}) \in B \times D,\ \boldsymbol{Y} \in C\big) \\ = P\big(\boldsymbol{X} \in B,\ \boldsymbol{Y} \in C \cap D\big) = \int_{C\cap D} P(\boldsymbol{X} \in B \mid \boldsymbol{Y} = \boldsymbol{y})\, dP^{\boldsymbol{Y}}(\boldsymbol{y}) \\ = \int_C \mathbb{1}_D(\boldsymbol{y}) P(\boldsymbol{X} \in B \mid \boldsymbol{Y} = \boldsymbol{y})\, dP^{\boldsymbol{Y}}(\boldsymbol{y}).$$

Durch Vergleich der Integranden ergibt sich also

$$P\big(G(\boldsymbol{X},\boldsymbol{Y}) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big) = P\big((\boldsymbol{X},\boldsymbol{Y}) \in B \times D \mid \boldsymbol{Y} = \boldsymbol{y}\big) \\ = \mathbb{1}_D(\boldsymbol{y}) P(\boldsymbol{X} \in B \mid \boldsymbol{Y} = \boldsymbol{y}) = \begin{cases} P(\boldsymbol{X} \in B \mid \boldsymbol{Y} = \boldsymbol{y}) & \text{für } \boldsymbol{y} \in D \\ 0 & \text{sonst} \end{cases} \\ = P\big((\boldsymbol{X},z) \in B \times D \mid \boldsymbol{Y} = \boldsymbol{y}\big)\Big|_{z=\boldsymbol{y}} = P\big(G(\boldsymbol{X},z) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big)\Big|_{z=\boldsymbol{y}} \quad P^{\boldsymbol{Y}}\text{-f.s.}$$

Mit einer ähnlichen Begründung wie im Beweis zu (3.1.23) folgt dann auch die Gültigkeit der Aussage für beliebige $A \in \mathcal{B} \otimes \mathcal{D}$. Der allgemeine Fall ergibt sich hieraus unmittelbar wegen $\{G(\boldsymbol{X},\boldsymbol{Y}) \in A\} = \{(\boldsymbol{X},\boldsymbol{Y}) \in G^{-1}(A)\}$, $A \in \mathcal{F}$.
Im Falle der stochastischen Unabhängigkeit von $\boldsymbol{X}$ und $\boldsymbol{Y}$ hat man nur die Gültigkeit von

$$P\big((G(\boldsymbol{X},z) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big) = P\big((G(\boldsymbol{X},z) \in A\big)$$

$P^{\boldsymbol{Y}}$-f.s. zu beachten, die aus Satz 3.1.4 c) folgt. ∎

Man beachte, daß der Ausdruck $P\big((G(\boldsymbol{X},\boldsymbol{y}) \in A \mid \boldsymbol{Y} = \boldsymbol{y}\big)$ in (3.1.25) nur im Fall diskreter Verteilungen unmittelbar Sinn macht, während i.a. die angegebene Interpretation unverzichtbar ist, da der Ausdruck im Sinne von (3.1.15) und (3.1.16) zunächst nicht erklärt ist.

Sind beispielsweise X und Y reellwertige Zufallsvariablen und betrachtet man $U = \max(X,Y)$, so gilt mit (3.1.25)

$$P(U \le u \mid Y = y) = P(X \le u,\ Y \le u \mid Y = y) \\ = \mathbb{1}_{(-\infty,u]}(y) P(X \le u \mid Y = y),\ u, y \in \mathbf{R}.$$

Sind X und Y unabhängig, so vereinfacht sich dies mit (3.1.26) noch zu

$$P(U \le u \mid Y = y) = \mathbb{1}_{(-\infty,u]}(y)P(X \le u),\ u, y \in \mathbf{R}.$$

Insbesondere im Zusammenhang mit der rekursiven Erzeugung von Markoff–Ketten wird sich das Ersetzungslemma 3.1.4 als nützlich erweisen.

Ist $\mathcal{C}$ eine Teil–σ–Algebra von $\mathcal{A}$ und die bedingte Verteilung $P(\cdot \mid \mathcal{C})$ regulär, so kann man natürlich alle Begriffsbildungen wie Verteilungsfunktion, Dichte, Erwartungswert usw., die für gewöhnliche Verteilungen existieren, entsprechend auf die bedingte Verteilung übertragen. Im letzteren Fall gelangt man dann etwa zum Begriff des bedingten Erwartungswerts, auf den wir wegen seiner zentralen Bedeutung abschließend noch einmal gesondert eingehen wollen.

Definition 3.1.4. *(bedingter Erwartungswert)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine Teil-σ–Algebra von $\mathcal{A}$; $P(\cdot \mid \mathcal{C})$ sei regulär. Ist X eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, so heißt im Falle der Existenz

$$\begin{aligned} E(X \mid \mathcal{C})(\omega) &= \int X(\eta)\, dP(\eta \mid \mathcal{C})(\omega) \\ &= \int X\, dP(\cdot \mid \mathcal{C})(\omega), \quad \omega \in \Omega, \end{aligned} \tag{3.1.27}$$

oder kürzer auch

$$E(X \mid \mathcal{C}) = \int X\, dP(\cdot \mid \mathcal{C})$$

bedingter Erwartungswert von X unter $\mathcal{C}$. Ist $\mathcal{C} = \sigma(\boldsymbol{Y})$ für ein Zufallselement $\boldsymbol{Y}$ mit Werten in einem Meßraum $(\mathcal{X}, \mathcal{B})$, so heißt entsprechend auch

$$E(X \mid \sigma(\boldsymbol{Y})) = E(X \mid \boldsymbol{Y}) \quad \text{bzw.} \quad E(X \mid \boldsymbol{Y} = \boldsymbol{y}),\ \boldsymbol{y} \in \mathcal{X},$$

bedingter Erwartungswert unter $\boldsymbol{Y}$ bzw. unter $\boldsymbol{Y} = \boldsymbol{y}$.

Für bedingte Erwartungswerte gelten insbesondere die in Abschnitt 2.2 angegebenen Eigenschaften gewöhnlicher Erwartungswerte, wenn man die dort betrachteten Verteilungen, Verteilungsfunktionen, Dichten usw. durch ihre entsprechenden (regulären) bedingten Versionen ersetzt.

Ist speziell $\mathcal{C}$ durch eine Partition der Grundmenge wie in Lemma 3.1.1 erzeugt, so ergibt sich mit den dortigen Bezeichnungen (vgl. auch (3.1.11))

$$E(X \mid \mathcal{C})(\omega) = \sum_{n=1}^{\infty} E(X \mid B_n)\mathbb{1}_{B_n}(\omega), \quad \omega \in \Omega, \tag{3.1.28}$$

wobei $E(X \mid B_n)$, $n \in \mathbf{N}$, wieder den elementaren bedingten Erwartungswert von X unter (der Hypothese) B_n bezeichnet.

In Analogie zu Lemma 3.1.1 von der totalen Wahrscheinlichkeit läßt sich auch der — unbedingte — Erwartungswert von X durch Integration über die bedingten Erwartungswerte zurückerhalten:

$$E(X) = \int E(X \mid \mathcal{C})(\omega)\, dP(\omega) = \int E(X \mid \mathcal{C})\, dP; \tag{3.1.29}$$

das ergibt sich unmittelbar aus (3.1.13) mit $B = \Omega$. Im Falle einer abzählbaren Partition von Ω wie oben bedeutet dies gerade

$$E(X) = \sum_{n=1}^{\infty} E(X \mid B_n)P(B_n), \tag{3.1.30}$$

wie man direkt aus (3.1.28) schließen kann. Das Ersetzungslemma überträgt sich ebenfalls entsprechend: mit den dortigen Bezeichnungen erhält man etwa im Falle einer reellwertigen Abbildung G

$$\begin{aligned} E\big(G(\boldsymbol{X},\boldsymbol{Y}) \mid \boldsymbol{Y} = \boldsymbol{y}\big) &= E\big((G(\boldsymbol{X},\boldsymbol{y}) \mid \boldsymbol{Y} = \boldsymbol{y}\big) \\ &:= E\big(G(\boldsymbol{X},z) \mid \boldsymbol{Y} = \boldsymbol{y}\big)\Big|_{z=\boldsymbol{y}} \quad P^{\boldsymbol{Y}}\text{-f.s.} \end{aligned} \tag{3.1.31}$$

Im Falle der bedingten quadratischen Integrierbarkeit kann man auch die *bedingte Varianz* einer Zufallsvariablen definieren als Varianz bezüglich der entsprechenden bedingten Verteilung, i.Z.: $\operatorname{Var}(X \mid \mathcal{C})$. Will man hieraus die — unbedingte — Varianz von X zurückerhalten, so gelangt man zu der Darstellung

$$\operatorname{Var}(X) = E\big(\operatorname{Var}(X \mid \mathcal{C})\big) + \operatorname{Var}\big(E(X \mid \mathcal{C})\big). \tag{3.1.32}$$

Dies sieht man wie folgt: es ist

$$\begin{aligned} E\big((\operatorname{Var}(X \mid \mathcal{C})\big) &= E\left[E(X^2 \mid \mathcal{C}) - \big[E(X \mid \mathcal{C})\big]^2\right] \\ \operatorname{Var}\big(E(X \mid \mathcal{C})\big) &= E\left[\big[E(X \mid \mathcal{C})\big]^2\right] - \big[E\big(E(X \mid \mathcal{C})\big)\big]^2 \\ &= E\left[\big[E(X \mid \mathcal{C})\big]^2\right] - \big[E(X)\big]^2; \end{aligned}$$

Summation ergibt

$$\begin{aligned} &E\big(\operatorname{Var}(X \mid \mathcal{C})\big) + \operatorname{Var}\big(E(X \mid \mathcal{C})\big) \\ &\quad = E\big(E(X^2 \mid \mathcal{C})\big) - E\left[\big[E(X \mid \mathcal{C})\big]^2\right] + E\left[\big[E(X \mid \mathcal{C})\big]^2\right] - \big[E(X)\big]^2 \\ &\quad = E(X^2) - \big[E(X)\big]^2 = \operatorname{Var}(X). \end{aligned}$$

Das folgende Rechenbeispiel zeigt, daß bedingte Erwartungswerte bzw. Varianzen existieren können, ohne daß dies für die entsprechenden unbedingten Größen gilt:

Es seien X, Y unabhängige, je $\mathcal{R}\big((0,1]\big)$–verteilte Zufallsvariablen. Setzt man $Z = -\dfrac{1}{Y}\ln X$, so ist gemäß (2.1.9) bzw. der anschließenden Bemerkung Z unter $Y = y$, $y > 0$, bedingt $\mathcal{E}(y)$–verteilt. Man erhält also

$$E(Z \mid Y = y) = \frac{1}{y}, \quad \operatorname{Var}(Z \mid Y = y) = \frac{1}{y^2}, \quad y > 0;$$

allerdings ist nach (3.1.29)

$$E(Z) = \int_0^\infty \frac{1}{y} f_Y(y)\, dy = \int_0^1 \frac{1}{y} = \infty,$$

d.h. $E(Z)$ existiert nicht als endlicher Wert, somit auch nicht $\mathrm{Var}(Z)$. Für die Verteilungsfunktion von Z ergibt sich dabei durch Integration

$$\begin{aligned} F_Z(z) &= \int_0^1 P(Z \le z \mid Y = y)\, dP^Y(y) = \int_0^1 (1 - e^{-yz})\, dy \\ &= 1 - \frac{1 - e^{-z}}{z} = \frac{z - 1 + e^{-z}}{z}, \quad z > 0. \end{aligned}$$

Abschließend wollen wir noch darauf hinweisen, daß bedingte Erwartungswerte auch für den Fall nicht–regulärer bedingter Verteilungen definiert werden können, indem man ähnlich wie in Definition 3.1.2 über Teil–σ–Algebren vorgeht.

Definition 3.1.5. *(allgemeiner bedingter Erwartungswert)*
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, X eine nicht–negative bzw. integrierbare Zufallsvariable und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine Teil–σ–Algebra. Jede $\mathcal{C}$–meßbare Zufallsvariable $E(X \mid \mathcal{C})$, die der Gleichung

$$\int_B E(X \mid \mathcal{C})\, dP = \int_B X\, dP \quad \text{für alle } B \in \mathcal{C} \tag{3.1.33}$$

genügt, heißt bedingter Erwartungswert von X unter $\mathcal{C}$.

Auch hier sind die möglichen Versionen des bedingten Erwartungswerts (nur) $P_{\mathcal{C}}$–f.s. eindeutig bestimmt.

Man sieht leicht mit algebraischer Induktion, daß bei Regularität die frühere Definition des bedingten Erwartungswerts hierin enthalten ist:
Für $X = \mathbb{1}_A$, $A \in \mathcal{A}$, ergibt (3.1.33) nämlich

$$\begin{aligned} \int_B E(X \mid \mathcal{C})\, dP &= \int_B \int X\, dP(\cdot \mid \mathcal{C})\, dP \\ &= \int_B \int_A dP(\cdot \mid \mathcal{C})\, dP = \int_B P(A \mid \mathcal{C})\, dP \\ &= P(A \cap B) = \int_B \mathbb{1}_A\, dP = \int_B X\, dP, \quad B \in \mathcal{C}. \end{aligned}$$

Allerdings läßt sich der so definierte bedingte Erwartungswert i.a. nicht mehr durch Integration nach der bedingten Verteilung darstellen. Da dieser allgemeine Zugang für die hier betrachteten Probleme weniger interessant ist, werden wir darauf nicht weiter eingehen; der interessierte Leser sei aber etwa auf Bauer (1978), Kapitel X verwiesen, wo zugleich eine Reihe spezifischer Eigenschaften des bedingten Erwartungswerts angegeben sind.

3.2. Markoff–Ketten

In der Informatik sind Systeme von großem Interesse, die mit einer höchstens abzählbaren Menge von Zuständen $\mathcal{S} = \{s_1, s_2, \dots\}$ beschrieben werden können und bei denen zu aufeinanderfolgenden Zeitpunkten beobachtet wird, in welchem Zustand sich das System befindet. Erinnert sei etwa an sequentielle Automaten, die mit einer endlichen Zustandsmenge arbeiten, an Suchalgorithmen, bei denen die Zustände die Elemente eines geordneten Feldes sein können, oder an Warteschlangenmodelle mit möglichen Anzahlen von zu bedienenden Kunden als Zustandsmenge.

Das Verhalten solcher Systeme läßt sich in vielen Fällen mit stochastischen Hilfsmitteln beschreiben, wobei es oft genügt, Modelle zu betrachten, bei denen die Wahrscheinlichkeit für einen Übergang des Systems von einem Zustand auf einen anderen nur von dem aktuellen und nicht von allen, in der Vergangenheit besuchten Zuständen abhängt. Wenn in einer Warteschlange in jeder Zeiteinheit mit Wahrscheinlichkeit p ein Kunde hinzukommt und mit Wahrscheinlichkeit q ein Kunde bedient wird, so ist die Wahrscheinlichkeit für eine bestimmte Länge der Warteschlange zum Zeitpunkt n, gegeben ihre Länge zum Zeitpunkt $n-1$, unabhängig von der Länge zu den Vorgängerzeitpunkten $n-2, \dots, 0$.

Solche Abhängigkeitsstrukturen werden durch Markoff–Ketten beschrieben.

Definition 3.2.1. *(Markoff–Ketten)*
Eine Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}_0}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in einer höchstens abzählbaren Menge $\mathcal{S} \neq \emptyset$ heißt (diskrete) Markoff–Kette, wenn gilt:

$$\begin{aligned} P\big(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\big) \\ = P\big(X_n = x_n \mid X_{n-1} = x_{n-1}\big) \end{aligned} \tag{3.2.1}$$

für alle $x_0, \dots, x_n \in \mathcal{S}$ mit $P\big(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\big) > 0$. $\mathcal{S}$ heißt Zustandsraum der Markoff–Kette. Für $x, y \in \mathcal{S}$ mit $P(X_{n-1} = x) > 0$ heißt $P(X_n = y \mid X_{n-1} = x)$ Übergangswahrscheinlichkeit im n–ten Schritt von x nach y. Gilt $P(X_{n-1} = x) = 0$, wählt man als Übergangswahrscheinlichkeiten beliebige Zahlen $p_{xy}(n) \geq 0$ mit $\sum_{y\in\mathcal{S}} p_{xy}(n) = 1$.

Die Markoff–Kette $\{X_n\}_{n\in\mathbb{N}_0}$ heißt homogen, wenn ihre Übergangswahrscheinlichkeiten unabhängig von n sind, d.h. $P\big(X_n = x \mid X_{n-1} = y\big) = P\big(X_m = x \mid X_{m-1} = y\big)$ für alle $x, y \in \mathcal{S}$, $n, m \in \mathbb{N}$ mit $P(X_{n-1} = y) > 0$ und $P(X_{m-1} = y) > 0$.

(3.2.1) formalisiert die Vorstellung, daß die Wahrscheinlichkeit für einen Übergang auf x_n nur vom unmittelbaren Vorgänger x_{n-1}, nicht aber von $x_{n-2}, \dots, x_0$ abhängt.

Der Zustandsraum $\mathcal{S}$ wird als höchstens abzählbar angenommen. Ist $\mathcal{S}$ endlich, so werden wir im folgenden die Zustände ohne Einschränkung der Allgemeinheit mit den ersten r natürlichen Zahlen $\mathcal{S} = \{1, 2, \dots, r\}$, $r \in \mathbb{N}$, identifizieren, im abzählbar unendlichen Fall wird $\mathcal{S} = \mathbb{N}$ angenommen.

Die Übergangswahrscheinlichkeiten einer Markoff–Kette lassen sich mit der Abkürzung $p_{ij}(n) = P(X_n = j \mid X_{n-1} = i)$ zu eventuell unendlich–dimensionalen

Matrizen

$$\mathbf{\Pi}_n = \begin{pmatrix} p_{11}(n) & p_{12}(n) & \dots \\ p_{21}(n) & p_{22}(n) & \dots \\ \vdots & \vdots & \end{pmatrix}, \quad n \in \mathbf{N}, \tag{3.2.2}$$

zusammenfassen. Gilt hierbei $P(X_{n-1} = i) = 0$ für einen Zustand i, so wähle man wie oben beliebige Zahlen $(p_{i1}(n), p_{i2}(n), \dots)$ derart, daß $p_{ij}(n) \geq 0$ und $\sum_{j \in \mathcal{S}} p_{ij}(n) = 1$. $\mathbf{\Pi}_n$ heißt dann eine n-te Übergangsmatrix der Markoff-Kette $\{X_n\}_{n \in \mathbf{N}_0}$.

Ist die Markoff-Kette $\{X_n\}_{n \in \mathbf{N}_0}$ homogen im Sinn von Definition 3.2.1, so existiert unabhängig von $n \in \mathbf{N}$ eine stochastische Matrix $\mathbf{\Pi} = (p_{ij})_{i,j \in \mathcal{S}}$ derart, daß $p_{ij} = P(X_n = j \mid X_{n-1} = i)$ für alle $i, j \in \mathcal{S}$, $n \in \mathbf{N}$ mit $P(X_{n-1} = i) > 0$. Betrachte hierzu die Menge $\mathcal{S}_0$ der Zustände, die irgendwann einmal mit positiver Wahrscheinlichkeit angenommen werden,

$$\mathcal{S}_0 = \Big\{ \ell \in \mathcal{S} \mid P\Big(\bigcup_{n=0}^{\infty} \{X_n = \ell\} \Big) > 0 \Big\}.$$

Für alle $i \in \mathcal{S}_0$ existiert ein $n \in \mathbf{N}$ mit $P(X_{n-1} = i) > 0$. Setze nun $p_{ij} = P(X_n = j \mid X_{n-1} = i)$, was wegen der Homogenität nicht von n abhängt. Für $i \notin \mathcal{S}_0$ wählt man wieder beliebige, allerdings von n unabhängige Zahlen $p_{ij} \geq 0$, $\sum_{j \in \mathcal{S}} p_{ij} = 1$.

Matrizen der Form (3.2.2) heißen stochastisch, da ihre Zeilen jeweils die Wahrscheinlichkeitsverteilungen $P^{X_n}(\cdot \mid X_{n-1} = i)$ auf $\big(\mathcal{S}, \mathfrak{P}(\mathcal{S})\big)$ repräsentieren. Für alle Zeilensummen $i \in \mathcal{S}$ und alle $n \in \mathbf{N}$ gilt damit $\sum_{j \in \mathcal{S}} p_{ij}(n) = 1$.

Analog lassen sich alle Randverteilungen P^{X_n} auf $\big(\mathcal{S}, \mathfrak{P}(\mathcal{S})\big)$ mit der Notation $p_j(n) = P(X_n = j)$, $j \in \mathcal{S}$, $n \in \mathbf{N}$, durch stochastische Vektoren

$$\boldsymbol{p}(n) = (p_1(n), p_2(n), \dots)$$

beschreiben, für die $\sum_{j \in \mathcal{S}} p_j(n) = 1$ gilt.

Ist die Mächtigkeit von $\mathcal{S}$ sogar endlich, so wird

$$\mathbf{\Pi}_n = \begin{pmatrix} p_{11}(n) & \dots & p_{1r2}(n) \\ \vdots & & \vdots \\ p_{r1}(n) & \dots & p_{rr}(n) \end{pmatrix}, \quad n \in \mathbf{N}_0,$$

zu einer stochastischen $r \times r$-Matrix und $\boldsymbol{p}(n) = \big(p_1(n), \dots, p_r(n)\big)$ zu einem stochastischen Zeilenvektor der Länge r.

Die aus der Matrizenrechnung für endliche Dimension bekannten Operationen $\boldsymbol{p} \cdot \mathbf{\Pi}$ und $\mathbf{\Phi} \cdot \mathbf{\Pi}$ lassen sich in naheliegender Weise auf den Fall unendlicher Zustandsmengen verallgemeinern. Ist $\boldsymbol{p} = (p_1, p_2, \dots)$ ein stochastischer Vektor, $\mathbf{\Pi} = \big(p_{ij}\big)_{i,j \in \mathbf{N}}$ und $\mathbf{\Phi} = \big(q_{ij}\big)_{i,j \in \mathbf{N}}$ stochastische Matrizen, so definieren wir

$$\begin{aligned} \big(\boldsymbol{p} \cdot \mathbf{\Pi}\big)_i &= \sum_{\ell=1}^{\infty} p_\ell p_{\ell i}, \quad i \in \mathbf{N} \\ \big(\mathbf{\Phi} \cdot \mathbf{\Pi}\big)_{ij} &= \sum_{\ell=1}^{\infty} q_{i\ell} p_{\ell j}, \quad i, j \in \mathbf{N}, \end{aligned} \tag{3.2.3}$$

wobei $(\boldsymbol{p}\cdot\boldsymbol{\Pi})_i$ das i–te Element des Vektors $\boldsymbol{p}\cdot\boldsymbol{\Pi}$ und $(\boldsymbol{\Phi}\cdot\boldsymbol{\Pi})_{ij}$ das (i,j)–te Element der Matrix $\boldsymbol{\Phi}\cdot\boldsymbol{\Pi}$ bezeichnet.

Die Reihen auf der rechten Seite obiger Formeln sind konvergent, da alle Summanden nicht negativ sind und $p_{\ell i} \leq 1$ für alle $\ell, i \in \mathbb{N}$ gilt. $\boldsymbol{p}\cdot\boldsymbol{\Pi}$ bzw. $\boldsymbol{\Phi}\cdot\boldsymbol{\Pi}$ sind wieder ein stochastischer Vektor bzw. eine stochastische Matrix, wie man durch Bilden der Zeilensummen leicht nachweist.

Durch iterierte Anwendung der Regeln (3.2.3) lassen sich höhere Produkte stochastischer Matrizen bilden, für die dann das Assoziativgesetz gilt, also

$$(\boldsymbol{\Pi}\cdot\boldsymbol{\Phi})\cdot\boldsymbol{\Psi} = \boldsymbol{\Pi}\cdot(\boldsymbol{\Phi}\cdot\boldsymbol{\Psi}) = \boldsymbol{\Pi}\cdot\boldsymbol{\Phi}\cdot\boldsymbol{\Psi} \text{ und } (\boldsymbol{p}\cdot\boldsymbol{\Pi})\cdot\boldsymbol{\Phi} = \boldsymbol{p}\cdot(\boldsymbol{\Pi}\cdot\boldsymbol{\Phi}) = \boldsymbol{p}\cdot\boldsymbol{\Pi}\cdot\boldsymbol{\Phi}.$$

Explizit erhält man für das (i,j)–te Element des n–fachen Produkts $\boldsymbol{\Pi}_1\cdot\boldsymbol{\Pi}_2\cdots\boldsymbol{\Pi}_n$ stochastischer Matrizen $\boldsymbol{\Pi}_k = \big(p_{ij}(k)\big)_{i,j\in\mathcal{S}}$, $1 \leq k \leq n$, den folgenden Ausdruck

$$\begin{aligned}\big(\boldsymbol{\Pi}_1\cdot\boldsymbol{\Pi}_2\cdots\boldsymbol{\Pi}_n\big)_{ij} &= \sum_{\ell_1,\ldots,\ell_{n-1}\in\mathcal{S}} p_{i\ell_1}(1)p_{\ell_1\ell_2}(2)\cdots p_{\ell_{n-1}j}(n)\\ &= \sum_{\ell_1,\ldots,\ell_{n-1}\in\mathcal{S}}\ \prod_{j=1}^{n} p_{\ell_{j-1}\ell_j}(j),\end{aligned} \tag{3.2.4}$$

wobei in der letzten Formel $l_0 = i$ und $l_n = j$ zu setzen ist.

Durch (3.2.1) wird lediglich das Übergangsverhalten der Markoff–Kette charakterisiert. Die gemeinsame Verteilung der gesamten Folge $P^{\{X_n\}}$ ist dann nach Festlegung einer *Anfangsverteilung*

$$\boldsymbol{p}(0) = \big(p_1(0), p_2(0), \ldots\big) = \big(P(X_0 = s_1), P(X_0 = s_2), \ldots\big)$$

vollständig gegeben, wie folgendes Lemma zeigt.

Lemma 3.2.1. *(Verteilung einer Markoff–Kette)*
$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine Markoff–Kette mit Übergangsmatrizen $\boldsymbol{\Pi}_n = \big(p_{ij}(n)\big)_{i,j\in\mathcal{S}}$, $n \in \mathbb{N}$, und Anfangsverteilung $\boldsymbol{p}(0) = \big(p_1(0), p_2(0), \ldots\big)$. Dann gilt

a) *für die Verteilung von X_n für alle $n \in \mathbb{N}$*

$$\boldsymbol{p}(n) = \boldsymbol{p}(n-1)\cdot\boldsymbol{\Pi}_n = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi}_1\cdot\boldsymbol{\Pi}_2\cdots\boldsymbol{\Pi}_n, \tag{3.2.5}$$

b) *für die gemeinsame Verteilung von $(X_0, X_1, \ldots, X_n)$ für alle $n \in \mathbb{N}$ und alle $i_0, \ldots, i_n \in \mathcal{S}$*

$$P(X_0 = i_0, \ldots, X_n = i_n) = p_{i_0}(0)\prod_{j=1}^{n} p_{i_{j-1}i_j}(j), \tag{3.2.6}$$

weiterhin für alle $k \in \mathbb{N}$, $k < n$

$$P(X_k = i_k, \ldots, X_n = i_n) = p_{i_k}(k)\prod_{j=k+1}^{n} p_{i_{j-1}i_j}(j), \tag{3.2.7}$$

c) *für die Übergangswahrscheinlichkeiten höherer Stufe für alle $k \in \mathbb{N}_0$, $n \in \mathbb{N}$, $k < n$ und alle $i, j \in \mathcal{S}$ mit $P(X_k = i) > 0$*

$$P(X_n = j \mid X_k = i) = \big(\boldsymbol{\Pi}_{k+1}\cdot\boldsymbol{\Pi}_{k+2}\cdots\boldsymbol{\Pi}_n\big)_{i,j}. \tag{3.2.8}$$

Beweis. a) Für alle $j \in \mathcal{S}$, $n \in \mathbb{N}$ gilt

$$\begin{aligned} p_j(n) = P(X_n = j) &= \sum_{i \in \mathcal{S},\ P(X_{n-1}=i)>0} P(X_n = j, X_{n-1} = i) \\ &= \sum_{i \in \mathcal{S},\ P(X_{n-1}=i)>0} P(X_n = j \mid X_{n-1} = i) P(X_{n-1} = i) \\ &= \sum_{i \in \mathcal{S}} p_{ij}(n) p_i(n-1) = \big(\boldsymbol{p}(n-1) \cdot \boldsymbol{\Pi}_n\big)_j. \end{aligned}$$

Die zweite Identität folgt durch iterierte Anwendung der ersten.

b) Den Beweis führen wir mit vollständiger Induktion. Für $n = 1$ gilt $P(X_0 = i_0, X_1 = i_1) = p_{i_1 i_0}(1) p_{i_0}(0)$ für all $i_0, i_1 \in \mathcal{S}$.
Unter Annahme der Induktionsvoraussetzung schließen wir mit Hilfe der Markoff-Eigenschaft, falls $P(X_0 = i_0, \ldots, X_n = i_n) > 0$

$$\begin{aligned} &P(X_0 = i_0, \ldots, X_{n+1} = i_{n+1}) \\ &\quad = P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) P(X_0 = i_0, \ldots, X_n = i_n) \\ &\quad = P(X_{n+1} = i_{n+1} \mid X_n = i_n) p_{i_0}(0) \prod_{j=1}^{n} p_{i_{j-1} i_j}(j) \end{aligned} \tag{3.2.9}$$

Nach Einsetzen von $p_{i_n i_{n+1}}(n+1)$ folgt (3.2.6).
Gilt $P(X_0 = i_0, \ldots, X_n = i_n) = 0$, so hat die linke Seite von (3.2.9) den Wert Null. Ferner existiert ein $k \in \mathbb{N}_0$, $k \le n$, mit $P(X_0 = i_0) > 0$, $P(X_\ell = i_\ell \mid X_{\ell-1} = i_{\ell-1}) > 0$ für alle $\ell = 1, \ldots, k-1$ und $P(X_k = i_k \mid X_{k-1} = i_{k-1}) = 0$. Anderenfalls würde mit Hilfe der Markoff-Eigenschaft $P(X_0 = i_0, \ldots, X_\ell = i_\ell) > 0$ für alle $\ell = 0, \ldots, n$ folgen. Also existiert ein $k \in \mathbb{N}_0$ mit $P(X_{k-1} = i_{k-1}) > 0$ und $p_{i_{k-1} i_k}(k) = P(X_k = i_k \mid X_{k-1} = i_{k-1}) = 0$, so daß die rechte Seite von (3.2.9), egal wie die frei wählbaren Übergangswahrscheinlichkeiten aussehen, ebenfalls Null ist.
Gleichung (3.2.7) folgt aus (3.2.6) durch Summation über die ersten k Komponenten $0, \ldots, k-1$.

$$\begin{aligned} &P(X_k = i_k, \ldots, X_n = i_n) \\ &\quad = \sum_{i_0, \ldots i_{k-1} \in \mathcal{S}} P(X_0 = i_0, \ldots, X_{k-1} = i_{k-1}, X_k = i_k, \ldots, X_n = i_n) \\ &\quad = \sum_{i_0, \ldots i_{k-1} \in \mathcal{S}} p_{i_0}(0) \prod_{j=1}^{n} p_{i_{j-1} i_j}(j) \\ &\quad = \Big(\prod_{j=k+1}^{n} p_{i_{j-1} i_j}(j) \Big) \Big(\sum_{i_0, \ldots i_{k-1} \in \mathcal{S}} p_{i_0}(0) \prod_{j=1}^{k} p_{i_{j-1} i_j}(j) \Big) \end{aligned}$$

Der zweite Faktor stimmt wegen (3.2.4) und a) überein mit

$$\sum_{i_0 \in \mathcal{S}} \Big(p_{i_0}(0) \sum_{i_1, \ldots i_{k-1} \in \mathcal{S}} \prod_{j=1}^{k} p_{i_{j-1} i_j}(j) \Big) = \sum_{i_0 \in \mathcal{S}} \Big(p_{i_0}(0) \big(\boldsymbol{\Pi}_1 \boldsymbol{\Pi}_2 \cdots \boldsymbol{\Pi}_k\big)_{i_0 i_k} \Big) = p_{i_k}(k).$$

Durch Einsetzen folgt die Behauptung.

c) Es gilt mit $i_k = i$ und $i_n = j$

$$
\begin{aligned}
&P(X_n = j \mid X_k = i) \\
&= \sum_{i_{k+1},\ldots,i_{n-1} \in \mathcal{S}} P(X_k = i, X_{k+1} = i_{k+1}, \ldots, X_{n-1} = i_{n-1}, X_n = j)/P(X_k = i) \\
&= \sum_{i_{k+1},\ldots,i_{n-1} \in \mathcal{S}} \prod_{\ell=k+1}^{n} p_{i_{\ell-1} i_\ell}(j) = \left(\mathbf{\Pi}_{k+1} \cdot \mathbf{\Pi}_{k+2} \cdots \mathbf{\Pi}_n\right)_{ij}
\end{aligned}
$$

wegen (3.2.7) und (3.2.4). ∎

Beziehung (3.2.6) zeigt insbesondere, daß die Verteilungen $P^{(X_0,\ldots,X_n)}$, $n \in \mathbf{N}_0$, eine projektive Familie bilden. Nach Satz 1.4.4 ist somit die Verteilung der gesamten Markoff-Kette eindeutig durch die Anfangsverteilung sowie die Übergangswahrscheinlichkeiten bestimmt.

Formel (3.2.8) gibt die Möglichkeit, Übergangswahrscheinlichkeiten höherer Stufe auch dann festzusetzen, wenn $P(X_k = i) = 0$ gilt, indem man als Übergangswahrscheinlichkeit das (i,j)-te Element des Matrixprodukts $\mathbf{\Pi}_{k+1} \cdots \mathbf{\Pi}_n$ wählt.

Das folgende Lemma zeigt, wie man Markoff-Ketten kanonisch aus einer vorgegebenen Folge von stochastisch unabhängigen Zufallsvariablen und einer Folge von Transformationsfunktionen gewinnen kann.

Lemma 3.2.2. *(rekursive Konstruktion von Markoff-Ketten)*
$(\Omega, \mathcal{A}, P)$sei ein Wahrscheinlichkeitsraum und $X_0 : (\Omega, \mathcal{A}) \to (\mathcal{S}, \mathfrak{P}(\mathcal{S}))$ eine Zufallsvariable. $(\Omega', \mathcal{A}')$ sei ein Meßraum und $\{U_n\}_{n\in\mathbf{N}}$ eine Folge von Zufallselementen $U_n : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$, $n \in \mathbf{N}$, derart daß $\{X_0, U_n\}_{n\in\mathbf{N}}$ stochastisch unabhängig sind. Seien $f_n : (\mathcal{S} \times \Omega', \mathfrak{P}(\mathcal{S}) \otimes \mathcal{A}') \to (\mathcal{S}, \mathfrak{P}(\mathcal{S}))$, $n \in \mathbf{N}$, meßbare Abbildungen. Dann bildet die rekursiv definierte Folge $\{X_n\}_{n\in\mathbf{N}_0}$ mit

$$X_n = f_n(X_{n-1}, U_n), \; n \in \mathbf{N}, \tag{3.2.10}$$

eine Markoff-Kette, deren Übergangswahrscheinlichkeiten $\mathbf{\Pi}_n = \left(p_{ij}(n)\right)_{i,j\in\mathcal{S}}$ gegeben sind durch

$$p_{ij}(n) = P\big(f_n(i, U_n) = j\big), \; i,j \in \mathcal{S}, \; n \in \mathbf{N}. \tag{3.2.11}$$

Ist $f_n = f$ und U_n identisch verteilt für alle $n \in \mathbf{N}$, so ist die Markoff-Kette $\{X_n\}_{n\in\mathbf{N}_0}$ homogen.

Beweis. Sei $\boldsymbol{X}_n = (X_0, \ldots, X_{n-1})$, $n \geq 2$. Dann ist $X_n = G_n(\boldsymbol{X}_n, U_n) := f_n(X_{n-1}, U_n)$, also nach dem Ersetzungslemma 3.1.4

$$
\begin{aligned}
P\big((X_n = i_n \mid \boldsymbol{X}_n = (i_0, \ldots, i_{n-1})\big) &= P\big(G_n(\boldsymbol{X}_n, U_n) = i_n \mid \boldsymbol{X}_n = (i_0, \ldots, i_{n-1})\big) \\
&= P\big((G_n(i_0, \ldots, i_{n-1}, U_n) = i_n\big) \\
&= P\big(f_n(i_{n-1}, U_n) = i_n\big), \quad i_0, \ldots, i_n \in \mathcal{S},
\end{aligned}
$$

da X_n (als Funktion der $X_0, \ldots, X_{n-1}$ und damit der $X_0, U_1, \ldots, U_{n-1}$) und U_n stochastisch unabhängig sind. Folglich bildet $\{X_n\}_{n\in\mathbb{N}_0}$ eine Markoff-Kette mit den angegebenen Übergangswahrscheinlichkeiten (3.2.11).
Ist $f_n = f$ und U_n identisch verteilt für alle $n \in \mathbb{N}$, so ist die Verteilung von $f_n(i, U_n)$ unabhängig von n für alle $i \in \mathcal{S}$, woraus die Homogenität folgt. ∎

Mit wahrscheinlickeitstheoretischen Mitteln läßt sich zeigen, daß sogar *alle* Markoff-Ketten $\{X_n\}$ im wesentlichen rekursiv darstellbar sind, d.h. man kann zu $\{X_n\}_{n\in\mathbb{N}_0}$ eine unabhängige Folge $\{U_n\}_{n\in\mathbb{N}}$ so konstruieren, daß die gegebene Markoff-Kette fast sicher mit der durch (3.2.10) gewonnenen übereinstimmt (vgl. etwa Pfeifer (1989a), Satz A1.1). In dieser Konstruktion sind die Funktionen f_n gegeben durch

$$f_n(x, y) = F^{-1}_{X_n}(y \mid X_{n-1} = x),\ x, y \in \mathbb{R},\ n \in \mathbb{N},$$

wobei $F_{X_n}(\cdot \mid X_{n-1} = x)$ die bedingte Verteilungsfunktion von X_n unter $X_{n-1} = x$ bezeichne. Mit Satz 2.1.1 macht man sich leicht klar, daß diese Wahl der f_n tatsächlich die bedingten Verteilungen $P^{X_n}(\cdot \mid X_{n-1} = x)$ realisiert.

Man macht sich leicht klar, daß die in (3.2.10) gegebene kanonische Konstruktion sogar dann noch zu Markoff-Ketten führt, wenn man lediglich die stochastische Unabhängigkeit von U_n und $(X_0, \ldots, X_{n-1})$ für alle $n \in \mathbb{N}$ fordert.

Wir werden uns im weiteren ausschließlich mit homogenen Markoff-Ketten beschäftigen. Die Abhängigkeit der Übergangswahrscheinlichkeiten von n in (3.2.2) entfällt dann, so daß diese konstant durch die stochastische Matrix

$$\mathbf{\Pi} = (p_{ij})_{i,j\in\mathcal{S}} \text{ mit } p_{ij} = \big(P(X_n = j \mid X_{n-1} = i)\big)_{i,j\in\mathcal{S}},$$

falls $P(X_{n-1} = i) > 0$, beschrieben werden können. Die Verteilung von X_n erhält man aus (3.2.5) speziell zu

$$\boldsymbol{p}(n) = \boldsymbol{p}(0) \cdot \underbrace{\mathbf{\Pi} \cdots \mathbf{\Pi}}_{n-\text{mal}} = \boldsymbol{p}(0) \cdot \mathbf{\Pi}^n. \tag{3.2.12}$$

Für die Untersuchung des stochastischen Verhaltens einer homogenen Markoff-Kette nach n Schritten sind damit die Anfangsverteilung $\boldsymbol{p}(0)$ und die n-te Potenz der Übergangsmatrix $\mathbf{\Pi}$ wichtig. Wir werden später Fälle untersuchen, bei denen im Grenzwert $(n \to \infty)$ der Einfluß der Anfangsverteilung $\boldsymbol{p}(0)$ verschwindet.

Für homogene Markoff-Ketten führen wir noch den Begriff der n-Schritt-Übergangswahrscheinlichkeiten ein.

$$p^{(n)}_{ij} = P(X_n = j \mid X_0 = i),\ i, j \in \mathcal{S} \text{ mit } P(X_0 = i) > 0,\ n \in \mathbb{N}, \tag{3.2.13}$$

ist die Wahrscheinlichkeit, in n Schritten vom Zustand $i \in \mathcal{S}$ in den Zustand $j \in \mathcal{S}$ zu gelangen. Ist $P(X_0 = i) = 0$ für ein $i \in \mathcal{S}$, so setzen wir $p^{(n)}_{ij} = \big(\mathbf{\Pi}^n\big)_{i,j}$. Hierbei gelten die folgenden Zusammenhänge zu Potenzen der zugehörigen Übergangsmatrix $\mathbf{\Pi}$.

Lemma 3.2.3. *(Mehrschritt-Übergangswahrscheinlichkeiten)*
$\{X_n\}_{n\in\mathbf{N}_0}$ sei eine homogene Markoff-Kette mit Übergangsmatrix $\mathbf{\Pi} = (p_{ij})_{i,j\in\mathcal{S}}$. Dann gilt für alle $i \in \mathcal{S}$ mit $P(X_0 = i) > 0$

a) $$p_{ij}^{(n)} = \left(\mathbf{\Pi}^n\right)_{ij} \text{ für alle } j \in \mathcal{S},\ n \in \mathbf{N}, \tag{3.2.14}$$

b) $$p_{ij}^{(n)} = P(X_{k+n} = j \mid X_k = i) \tag{3.2.15}$$
$$\text{für alle } i,j \in \mathcal{S},\ k,n \in \mathbf{N} \text{ mit } P(X_k = i) > 0,$$

c) *(Chapman-Kolmogoroff-Gleichungen)*

$$p_{ij}^{(n+m)} = \sum_{\ell\in\mathcal{S}} p_{i\ell}^{(n)} p_{\ell j}^{(m)} \text{ für alle } i,j \in \mathcal{S},\ m,n \in \mathbf{N}. \tag{3.2.16}$$

Beweis. a) folgt sofort aus (3.2.8) mit $k = 0$, da $\mathbf{\Pi}_1 = \cdots = \mathbf{\Pi}_n = \mathbf{\Pi}$.
Ist $P(X_k = i) > 0$, gilt genauso

$$P(X_{k+n} = j \mid X_k = i) = \left(\mathbf{\Pi}_{k+1}\cdots\mathbf{\Pi}_{k+n}\right)_{i,j} = \left(\mathbf{\Pi}^n\right)_{i,j} = p_{ij}^{(n)}.$$

Dies zeigt b).
Zum Beweis von c) bemerken wir, daß für alle $\ell \in \mathcal{S}$ aus $P(X_0 = i) > 0$ und $P(X_n = \ell) = 0$ für die bedingte Wahrscheinlichkeit $P(X_n = \ell \mid X_0 = i) = 0$ folgt. Also gilt mit a)

$$\begin{aligned} p_{ij}^{(n+m)} &= \left(\mathbf{\Pi}^{n+m}\right)_{i,j} = \sum_{\ell\in\mathcal{S}} \left(\mathbf{\Pi}^n\right)_{i,\ell}\left(\mathbf{\Pi}^m\right)_{\ell,j} \\ &= \sum_{\substack{\ell\in\mathcal{S}\\ P(X_n=\ell)>0}} p_{i\ell}^{(n)}\left(\mathbf{\Pi}^m\right)_{\ell,j} = \sum_{\substack{\ell\in\mathcal{S}\\ P(X_n=\ell)>0}} p_{i\ell}^{(n)} p_{\ell j}^{(m)} = \sum_{\ell\in\mathcal{S}} p_{i\ell}^{(n)} p_{\ell j}^{(m)}. \end{aligned}$$

∎

Beispiel 3.2.1. (Irrfahrt auf einem Gitter, random walk)
Zu festen Zeitpunkten $t_1, t_2, \ldots$ befindet sich ein Teilchen auf einem Gitterpunkt eines eindimensionalen Gitters $\mathcal{G}$, das ist eine Teilmenge von aufeinanderfolgenden ganzen Zahlen oder $\mathbf{Z}$ selbst. Zu jedem nachfolgenden Zeitpunkt bewegt sich das Teilchen unabhängig gemäß einer bestimmten Verteilung um k Stellen, $k \in \mathbf{Z}$, nach links bzw. nach rechts. Bei (einseitig) endlichem Gitter — das größte bzw. kleinste Element des Gitters heißt dann Barriere — kann das Teilchen an den Barrieren absorbiert oder auch reflektiert werden.
X_n, $n \in \mathbf{N}_0$, bezeichne die Position des Teilchens zum Zeitpunkt t_n, und U_n, $n \in \mathbf{N}$, seien stochastisch unabhängige, identisch auf $(\mathbf{Z}, \mathfrak{P}(\mathbf{Z}))$ verteilte Zufallsvariable, stochastisch unabhängig von X_0, die die Bewegung des Teilchens zum Zeitpunkt T_n beschreiben.

a) Endliches Gitter $\mathcal{G} = \{0, \ldots, r\}$ mit absorbierenden Barrieren:
In diesem Fall gilt die Rekursion

$$X_n = \begin{cases} \max\{0, \min\{X_{n-1} + U_n, r\}\}, & \text{falls } 1 \le X_{n-1} \le r-1 \\ X_{n-1}, & \text{falls } X_{n-1} \in \{0, r\} \end{cases},$$

wobei $X_0 = k$, $k \in \mathcal{G}$, als einpunktverteilte Zufallsvariable den Startwert festlegt. Wählt man in Lemma 3.2.2

$$f(i,u) = \begin{cases} \max\{0, \min\{i+u, r\}\}, & \text{falls } 1 \leq i \leq r-1 \\ i, & \text{falls } i \in \{0, r\} \end{cases},$$

so folgt, daß $\{X_n\}_{n \in \mathbb{N}_0}$ eine homogene Markoff–Kette bildet. Besitzen die Zufallsvariablen U_n die Verteilung

$$P(U_n = 1) = p, \quad P(U_n = -1) = 1 - p, \quad 0 < p < 1, \tag{3.2.17}$$

so lautet die Übergangsmatrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1-p & 0 & p & 0 & \dots & 0 \\ 0 & 1-p & 0 & p & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \dots & \dots & 1-p & 0 & p \\ 0 & \dots & \dots & 0 & 0 & 1 \end{pmatrix}$$

b) Unendliches Gitter mit 0 als reflektierende Barriere, $\mathcal{G} = \mathbb{N}_0$:
$\{X_n\}_{n \in \mathbb{N}_0}$ kann auch hier rekursiv gewonnen werden vermöge

$$X_n = |X_{n-1} + U_n|,$$

so daß mit einer stochastisch unabhängigen Startvariablen X_0 wieder eine Markoff–Kette entsteht. Liegt die Verteilung (3.2.17) für U_n vor, so erhält man die unendliche Übergangsmatrix

$$\Pi = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 1-p & 0 & p & 0 & \dots \\ 0 & 1-p & 0 & p & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

■

Beispiel 3.2.2. (Warteschlangenmodell)
An einer Bedienstation kommen zu zufälligen Zeitpunkten Kunden an, die nacheinander während einer Zeitspanne zufälliger Länge bedient werden und die Station dann sofort verlassen. Ist die Bedienstation bei der Ankunft eines Kunden frei, so wird dieser sofort bedient, ist sie bereits besetzt, so muß der Kunde in einer Warteschlange warten, die sich dann mit jeder Neuankunft eines Kunden verlängert und mit jeder abgeschlossenen Bedienung verkürzt.
Ein Anwendungsbeispiel für dieses Modell in der Informatik ist ein Einprozessorsystem als Bedienstation, das Jobs von nicht genau vorhersagbarer Rechenzeit abarbeitet und bei dem neuankommende Jobs eventuell in einer Schlange auf ihre Bearbeitung warten müssen.

Wir interessieren uns nun für die Länge X_n der Warteschlange zu den Zeitpunkten t_n, an denen der n–te Kunde das System verläßt, $n \in \mathbb{N}$. Ist die Warteschlange zum Zeitpunkt t_n leer, so beträgt ihre Länge bei Verlassen des $(n+1)$–ten Kunden gerade die Anzahl während der Bedienzeit des $(n+1)$–ten Kunden neuangekommener Kunden U_n. Im anderen Fall verringert sich ihre Länge um einen Kunden, der die Bedienstation betritt, und erhöht sich ebenfalls um U_n. Aus diesen Überlegungen gilt für die Warteschlangenlänge die folgende Rekursion

$$X_{n+1} = \begin{cases} X_n - 1 + U_n, & \text{falls } X_n > 0 \\ U_n, & \text{falls } X_n = 0 \end{cases}, \quad n \in \mathbb{N}. \tag{3.2.18}$$

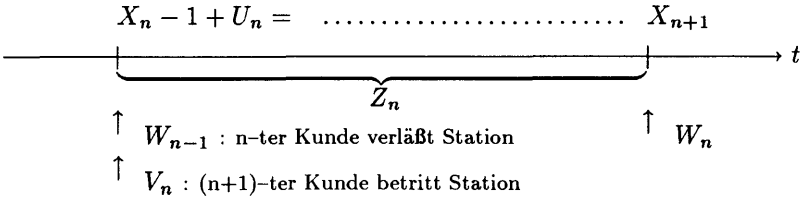
Für die Anzahlen $N_{(t_1,t_2]}$ ankommender Kunden in einem Zeitintervall $(t_1, t_2]$, $0 \le t_1 < t_2$, nehmen wir das folgende Verteilungsmodell an.

a) Für je n disjunkte Intervalle $I_1, I_2, \ldots, I_n$ sind die Ankunftszahlen $N_{I_1}, N_{I_2}, \ldots, N_{I_n}$ stochastisch unabhängige Zufallsvariable.
b) Für alle $0 \le t_1 < t_2$ ist $N_{(t_1,t_2]}$ Poisson–verteilt mit Parameter $\lambda(t_2 - t_1)$, $\lambda > 0$, d.h.

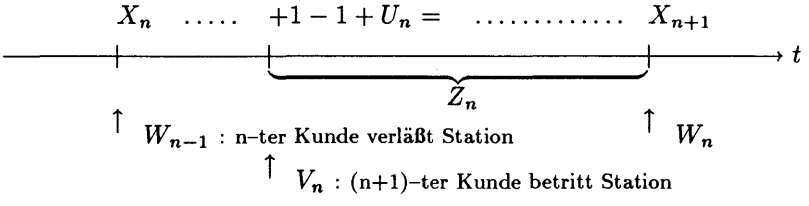
$$P(N_{(t_1,t_2]} = k) = e^{-\lambda(t_2-t_1)} \frac{(\lambda(t_2 - t_1))^k}{k!}, \quad k \in \mathbb{N}_0. \tag{3.2.19}$$

Wir werden später sehen, daß diese Bedingungen gerade den Poisson–Prozeß mit Parameter λ charakterisieren (vgl. auch Beziehungen (2.1.90) bis (2.1.92)). Die Bedienzeiten Z_n, $n \in \mathbb{N}_0$, für den $(n+1)$–ten Kunden seien stochastisch unabhängige, identisch je mit Verteilung P^Z verteilte Zufallsvariable; ferner seien Z_n, $n \in \mathbb{N}_0$, $N_{I_1}, N_{I_2}, \ldots, N_{I_k}$ für alle $k \in \mathbb{N}$ und beliebige disjunkte Intervalle $I_1, I_2, \ldots, I_k$ gemeinsam stochastisch unabhängig[1] .
Folgenden Grafiken veranschaulichen den n–ten Schritt des Bediensystems über der Zeitachse. V_n bezeichnet den zufälligen Zeitpunkt, zu dem der $(n+1)$–te Kunde die Bedienstation betritt und W_n den Zeitpunkt des Verlassens. Es gilt $Z_n = W_n - V_n$.

$X_n > 0$:



$X_n = 0$:



[1] In Abschnitt 3.4 wird gezeigt werden, daß der in (2.1.90) definierte Prozeß dies leistet, wenn man $N_{(s,t]} = N_t - N_s$, $0 \le s \le t$, wählt.

Für die Anzahlen der während der Bedienzeit des $(n+1)$-ten Kunden neuankommenden Kunden U_n, $n \in \mathbf{N}$, gilt $U_n = N_{(V_n, W_n]}$. Wir zeigen:

$$\text{i)} \quad P^{U_n}(\cdot \mid Z_n = z, V_n = v) = P^{U_n}(\cdot \mid Z_n = z) = \mathfrak{P}(\lambda z) \quad \text{f.s.},$$
$$n \in \mathbf{N}, k \in \mathbf{N}_0,\ z, v \geq 0;$$

$$\text{ii)} \quad P^{(U_1,\ldots,U_n)}(\cdot \mid Z_1 = z_1, \ldots, Z_n = z_n, V_1 = v_1, \ldots, V_n = v_n)$$
$$= \bigotimes_{i=1}^{n} P^{U_i}(\cdot \mid Z_i = z_i) = \bigotimes_{i=1}^{n} \mathfrak{P}(\lambda z_i) \quad \text{f.s.},$$
$$n \in \mathbf{N}, k_i \in \mathbf{N}_0,\ z_i, v_i \geq 0,\ 1 \leq i \leq n.$$

Zum Nachweis von i) und ii) verwenden wir das Ersetzungslemma 3.1.4, insbesondere Beziehung (3.1.26). Mit den getroffenen Unabhängigkeitsannahmen erhält man für $k \in \mathbf{N}_0$:

$$P(U_n = k \mid Z_n = z, V_n = v) = P(N_{(V_n, V_n+Z_n]} = k \mid Z_n = z, V_n = v)$$
$$= P(N_{(v,v+z]} = k) = \mathfrak{P}(\lambda z)(\{k\}) \quad \text{f.s.}$$

Dies zeigt i); d.h. die Zufallsvariablen U_n und V_n sind unter Z_n (bedingt) unabhängig und (bedingt) Poisson-verteilt.
Zum Nachweis von ii) seien $k_1, \ldots, k_n \in \mathbf{N}_0$. Es ist dann entsprechend

$$P(U_1 = k_1, \ldots, U_n = k_n \mid Z_1 = z_1, \ldots, Z_n = z_n, V_1 = v_1, \ldots, V_n = v_n)$$
$$= P(N_{(v_1, v_1+z_1]} = k_1, \ldots, N_{(v_n+z_n]} = k_n) = \bigotimes_{i=1}^{n} \mathfrak{P}(\lambda z_i) \quad \text{f.s.},$$

d.h. auch $U_1, \ldots, U_n$ sind (bedingt) unabhängig unter $(Z_1, \ldots, Z_n, V_1, \ldots, V_n)$. Durch Integration erhält man nun aus i) und ii) die — unbedingten — Verteilungen von U_n bzw. $(U_1, \ldots, U_n)$, $n \in \mathbf{N}$:

$$P(U_n = k) = \int P(U_n = k \mid Z_n = z)\, dP^{Z_n}(z)$$
$$= \int e^{-\lambda z} \frac{(\lambda z)^k}{k!}\, dP^{Z}(z), \quad k \in \mathbf{N}_0;$$

$$P(U_1 = k_1, \ldots, U_n = k_n) = \prod_{i=1}^{n} P(U_i = k_i), \quad k_1, \ldots, k_n \in \mathbf{N}_0;$$

d.h. $U_1, \ldots, U_n$ sind stochastisch unabhängig.
Die Voraussetzungen von Lemma 3.2.2 sind damit erfüllt, die dort verwendeten Funktionen f_n lauten hier identisch für alle $n \in \mathbf{N}$

$$f_n(i,j) = f(i,j) = (i-1)^+ + j,\ i, j \in \mathbf{N}_0,$$

wobei $x^+ = \max\{x, 0\}$ wieder den Positivteil von $x \in \mathbf{R}$ bedeutet. Setzt man noch $X_0 = 0$, so folgt insgesamt:

Die durch (3.2.18) definierten Warteschlangenlängen $\{X_n\}_{n\in\mathbb{N}_0}$ zu den Zeitpunkten, an denen der n-te Kunde die Bedienstation verläßt, bilden unter den Verteilungsannahmen (3.2.19) eine homogene Markoff-Kette mit Übergangswahrscheinlichkeiten

$$p_{ij} = P\big((i-1)^+ + U_n = j\big) = P\big(U_n = j - (i-1)^+\big)$$

$$= \begin{cases} \displaystyle\int e^{-\lambda z}\frac{(\lambda z)^j}{j!}dP^Z(z), & \text{falls } i=0,\ j\geq 0, \\ \displaystyle\int e^{-\lambda z}\frac{(\lambda z)^{j-i+1}}{(j-i+1)!}dP^Z(z), & \text{falls } i>0,\ j\geq i-1, \\ 0, & \text{sonst.} \end{cases}$$

■

Wir interessieren uns im weiteren für die Frage, wann die n-ten Randverteilungen P^{X_n} einer Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$ für $n\to\infty$ konvergieren. Wir betrachten hierbei die punktweise Konvergenz der Zähldichten von P^{X_n} gegen die Zähldichte einer Wahrscheinlichkeitsverteilung auf $(\mathcal{S}, \mathfrak{P}(\mathcal{S}))$. Existiert eine solche Limesverteilung als stochastischer Vektor $\boldsymbol{p}_\infty$, wird man nach einer großen Zahl von Übergangsschritten des durch die Markoff-Kette beschriebenen Systems einen Zustand gemäß der asymptotischen Verteilung $\boldsymbol{p}_\infty$ vorfinden.

Beschreibt die Markoff-Kette zum Beispiel einen stochastischen Suchalgorithmus und besitzt sie als Limesverteilung eine Gleichverteilung auf der Menge der optimalen Zustände, so konvergiert die Wahrscheinlichkeit gegen 1, daß der Algorithmus mit fortschreitender Iterationszahl in einen optimalen Zustand läuft.

Wir beschränken uns bei unseren Untersuchungen ausschließlich auf homogene Markoff-Ketten mit konstanter Übergangsmatrix $\boldsymbol{\Pi}$. Nach (3.2.12) gilt $\boldsymbol{p}(n) = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi}^n$ für alle $n\in\mathbb{N}$, so daß die Frage nach der Konvergenz von $\boldsymbol{p}(n)$ eng verbunden ist mit der punktweisen Konvergenz der Matrixpotenzen $\boldsymbol{\Pi}^n$ $(n\to\infty)$.

Allerdings kann auch Konvergenz von $\boldsymbol{p}(n)$ $(n\to\infty)$ vorliegen, ohne daß $\boldsymbol{\Pi}^n$ konvergiert. In diesem Fall ist natürlich der Einfluß der Startverteilung $\boldsymbol{p}(0)$ entscheidend, wie das folgende Beispiel zeigt.

Sei $\mathcal{S}=\{1,2\}$, $\boldsymbol{p}(0)=(0.5,0.5)$, $\boldsymbol{\Pi}=\begin{pmatrix}0&1\\1&0\end{pmatrix}$. Dann gilt

$$\boldsymbol{\Pi}^n = \begin{pmatrix}1&0\\0&1\end{pmatrix},\ \text{falls } n \text{ gerade, und } \boldsymbol{\Pi}^n = \begin{pmatrix}0&1\\1&0\end{pmatrix},\ \text{falls } n \text{ ungerade.}$$

Die Folge $\boldsymbol{\Pi}^n$ konvergiert also nicht, jedoch rechnet man leicht nach, daß $\boldsymbol{p}(0)\cdot\boldsymbol{\Pi} = \boldsymbol{p}(0)$, also

$$\boldsymbol{p}(0)\cdot\boldsymbol{\Pi}^n = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi}\cdot\boldsymbol{\Pi}^{n-1} = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi}^{n-1} = \ldots = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi} = \boldsymbol{p}(0),$$

so daß $\lim_{n\to\infty}\big(\boldsymbol{p}(0)\cdot\boldsymbol{\Pi}^n\big) = \boldsymbol{p}(0)$.

Verteilungen mit der Eigenschaft, bei Multiplikation mit der Übergangsmatrix invariant zu bleiben, behandelt die folgende

Definition 3.2.2. *(stationäre Verteilungen)*
Es sei $\{X_n\}_{n\in\mathbb{N}_0}$ *eine homogene Markoff-Kette mit Zustandsraum* $\mathcal{S}$ *und Übergangsmatrix* $\mathbf{\Pi}$. *Ein stochastischer Vektor* $\boldsymbol{p}$ *heißt stationär (oder invariant), wenn* $\boldsymbol{p}\cdot\mathbf{\Pi}=\boldsymbol{p}$.

Für Markoff-Ketten $\{X_n\}_{n\in\mathbb{N}_0}$ mit stationärer Anfangsverteilung $\boldsymbol{p}(0)$ stimmen damit alle n-ten Randverteilungen $\boldsymbol{p}(n)$ mit $\boldsymbol{p}(0)$ überein, und trivialerweise existiert $\lim_{n\to\infty}\boldsymbol{p}(n)=\boldsymbol{p}(0)$.

Das folgende Lemma stellt einen Zusammenhang zwischen Grenzverteilungen und stationären Verteilungen von Markoff-Ketten mit endlichem Zustandsraum her.

Lemma 3.2.4. *(Konvergenz von Markoff-Ketten)*
$\{X_n\}_{n\in\mathbb{N}_0}$ *sei eine homogene Markoff-Kette mit endlichem Zustandsraum* $\mathcal{S}=\{1,\ldots,r\}$, $r\in\mathbb{N}$. *Konvergiert dann* $\mathbf{\Pi}^n$ *für* $n\to\infty$ *elementweise gegen eine Matrix* $\mathbf{\Pi}^*$, *so ist* $\mathbf{\Pi}^*$ *eine stochastische Matrix und für beliebige Anfangsverteilungen* $\boldsymbol{p}(0)$ *ist die Limesverteilung* $\boldsymbol{p}^*=\lim_{n\to\infty}\boldsymbol{p}(n)=\boldsymbol{p}(0)\cdot\mathbf{\Pi}^*$ *eine stationäre Verteilung.*

Beweis. Nachzuweisen ist, daß alle Zeilensummen von $\mathbf{\Pi}^*=\left(p_{ij}^*\right)_{1\le i,j\le r}$ den Wert 1 haben. Dies folgt aus der Tatsache, daß $\mathbf{\Pi}^n$ für alle $n\in\mathbb{N}$ eine stochastische Matrix ist, so daß mit (3.2.14)

$$\sum_{j=1}^{r}p_{ij}^*=\sum_{j=1}^{r}\lim_{n\to\infty}p_{ij}^{(n)}=\lim_{n\to\infty}\sum_{j=1}^{r}p_{ij}^{(n)}=\lim_{n\to\infty}1=1$$

Ferner gilt

$$\boldsymbol{p}^*=\lim_{n\to\infty}\boldsymbol{p}(n)=\lim_{n\to\infty}\left(\boldsymbol{p}(0)\cdot\mathbf{\Pi}^n\right)=\boldsymbol{p}(0)\cdot\left(\lim_{n\to\infty}\mathbf{\Pi}^n\right)=\boldsymbol{p}(0)\cdot\mathbf{\Pi}^*.$$

Die Stationarität folgt aus

$$\boldsymbol{p}^*\cdot\mathbf{\Pi}=\boldsymbol{p}(0)\cdot\left(\lim_{n\to\infty}\mathbf{\Pi}^n\right)\cdot\mathbf{\Pi}=\boldsymbol{p}(0)\cdot\left(\lim_{n\to\infty}\mathbf{\Pi}^{n+1}\right)=\boldsymbol{p}(0)\cdot\mathbf{\Pi}^*=\boldsymbol{p}^*.\qquad\blacksquare$$

Der Beweis von Lemma 3.2.4 beruht wesentlich auf der Vertauschung von Limiten und Summationen. Dies ist ohne weiteres nur im endlichen Fall durch Anwendung der elementaren Rechenregeln für Grenzwerte möglich. Im Fall eines unendlichen Zustandsraums $\mathcal{S}$ kann selbst bei Konvergenz von $\mathbf{\Pi}^n$ die Grenzmatrix unter Umständen nicht mehr stochastisch sein. Man überlege dies am Beispiel $p_{ij}=1$, falls $j=i+1$, $p_{ij}=0$, sonst, $i,j\in\mathbb{N}$, wo $p_{ij}^{(n)}=1$, falls $j=i+n$, $p_{ij}=0$, sonst, gilt und damit $\lim_{n\to\infty}p_{ij}^{(n)}=0$ für alle $i,j\in\mathbb{N}$.

Eine stationäre Anfangsverteilung einer Markoff-Kette führt zu einem insgesamt stationären Prozeß im folgenden Sinn.

Definition 3.2.3. *(stationäre Folgen von Zufallsvariablen)*
$\{X_n\}_{n\in\mathbb{N}_0}$ *sei eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega,\mathcal{A},P)$. $\{X_n\}_{n\in\mathbb{N}_0}$ *heißt stationär, wenn für alle* $k\in\mathbb{N}$, *für alle Auswahlen* $0\le i_1<i_2<\ldots<i_k\in\mathbb{N}_0$ *von Indizes und für alle* $s\in\mathbb{N}$ *gilt, daß*

$$P^{(X_{i_1+s},X_{i_2+s},\ldots,X_{i_k+s})}=P^{(X_{i_1},X_{i_2},\ldots,X_{i_k})}.\qquad(3.2.20)$$

Bei einer stationären Folge von Zufallsvariablen sind also alle endlich-dimensionalen Randverteilungen invariant gegen eine konstante Indexverschiebung. Egal zu welchem Zeitpunkt und über welche Perioden man den Prozeß betrachtet, sein stochastisches Verhalten bleibt immer das gleiche.

Ist $\{X_n\}_{n\in\mathbb{N}_0}$ nun eine homogene Markoff-Kette mit stationärer Anfangsverteilung $\boldsymbol{p}(0)$, so ist die Folge der Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}_0}$ auch stationär im Sinn von Definition 3.2.3. Denn für die k-ten Randverteilungen $\boldsymbol{p}(k)$ gilt wie oben $\boldsymbol{p}(k) = \boldsymbol{p}(0)\cdot\boldsymbol{\Pi}^k = \boldsymbol{p}(0)$ und mit (3.2.6) und (3.2.7)

$$\begin{aligned} P(X_k = i_k,\ldots,X_n = i_n) &= p_{i_k}(k)\prod_{j=k+1}^{n} p_{i_{j-1}i_j} \\ &= p_{i_k}(0)\prod_{j=k+1}^{n} p_{i_{j-1}i_j} = P(X_0 = i_k,\ldots,X_{n-k} = i_n) \end{aligned}$$

für alle $k,n\in\mathbb{N}$, $i_k,\ldots,i_n\in\mathcal{S}$. Hieraus folgt (3.2.20) durch Integration über die nicht interessierenden Komponenten.

Wir werden uns im weiteren um einfache Bedingungen für die Existenz von Grenzverteilungen bzw. des Grenzwertes $\lim_{n\to\infty}\boldsymbol{\Pi}^n$ einer Markoff-Kette bemühen. Wichtig ist hierzu eine Unterscheidung der Zustände nach der Eigenschaft, mit positiver Wahrscheinlichkeit unendlich oft aufzutreten. Erinnert sei an die Definition (1.1.32) des Limes superior einer Ereignisfolge $\{A_n\}_{n\in\mathbb{N}_0}$, der beschreibt, daß unendlich viele der Ereignisse A_n eintreten: $\limsup_{n\to\infty} A_n = \bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty} A_k$.

Definition 3.2.4. *(rekurrente und transiente Zustände)*
$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine homogene Markoff-Kette. Ein Zustand $i\in\mathcal{S}$ heißt rekurrent, wenn $P\big(\limsup_{n\to\infty}\{X_n = i\}\big) > 0$, anderenfalls heißt $i\in\mathcal{S}$ transient.

Beispiel 3.2.3. (Irrfahrt mit absorbierenden Barrieren)
Wie in Beispiel 3.2.1 betrachten wir die Irrfahrt auf dem endlichen Gitter $\mathcal{S} = \mathcal{G} = \{0,1,\ldots,r\}$ mit absorbierenden Barrieren 0 und r. Die zugehörige Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$ entsteht durch die Rekursion

$$X_n = \begin{cases} X_{n-1} + U_n, & \text{falls } 1 \le X_{n-1} \le r-1 \\ X_{n-1}, & \text{falls } X_{n-1}\in\{0,r\} \end{cases},$$

wobei die Verteilung von U_n durch (3.2.17) gegeben ist.
Die Zustände 0 und r sind für jede Anfangsverteilung P^{X_0} rekurrent. Wir weisen dies für den Zustand 0 nach, für r verläuft der Beweis ganz analog.
Es existiert $i_0\in\mathcal{G}$ mit $P(X_0 = i_0) > 0$. Dann gilt für $0 < p < 1$

$$P(X_{i_0} = 0) = P(X_0 = i_0, U_1 = -1,\ldots,U_{i_0} = -1) = P(X_0 = i_0)\cdot(1-p)^{i_0} > 0.$$

Ferner ist

$$P\big(\limsup_{n\to\infty}\{X_n = 0\}\big) \ge P\big(X_{i_0} = 0\big) > 0,$$

da $\{X_{i_0}=0\}=\bigcap_{n=i_0}^{\infty}\{X_n=0\}\subseteq\bigcap_{n=i_0}^{\infty}\bigcup_{k=n}^{\infty}\{X_k=0\}=\limsup_{n\to\infty}\{X_n=0\}$ gilt unter Beachtung der Monotonie (1.1.36) von P.
Daß die Zustände $1,\dots,r-1$ transient sind, sieht man folgendermaßen ein. Für alle $i\in\{1,\dots r\}$ gilt

$$\begin{aligned}P\big(\limsup_{n\to\infty}\{X_n=i\}\big)&\le P\big(\limsup_{n\to\infty}\{1\le X_n\le r-1\}\big)\\&=P\Big(\bigcap_{n=1}^{\infty}\{1\le X_n\le r-1\}\Big)\le P\Big(\bigcap_{n=1}^{\infty}\Big\{\Big|\sum_{j=1}^{n}U_j\Big|\le r-1\Big\}\Big)\\&\le P\Big(\Big|\sum_{j=1}^{m}U_j\Big|\le r-1\Big)\qquad\text{für alle } m\in\mathbf{N}.\end{aligned}$$

Wegen (3.2.17) besitzen die Zufallsvariablen U_n die Darstellung $U_n=2Y_n-1$ mit stochastisch unabhängigen, je $\mathfrak{B}(1,p)$–verteilten Zufallsvariablen Y_n. Es folgt $\sum_{j=1}^{m}U_j=2\sum_{j=1}^{m}Y_j-m=2S_m-m$, wobei S_m gemäß (2.1.18) $\mathfrak{B}(m,p)$–verteilt ist mit der Zähldichte $P(Y_m=k)=\binom{m}{k}p^k(1-p)^{m-k}$, $k=0,1,\dots,m$. Also

$$\begin{aligned}P\Big(\Big|\sum_{j=1}^{m}U_j\Big|\le r-1\Big)&=P(|2S_m-m|\le r-1)\\&=P\Big(\frac{m-r+1}{2}\le S_m\le\frac{m+r-1}{2}\Big)\longrightarrow 0\qquad(n\to\infty).\end{aligned}$$

Die Konvergenz gilt, da eine Folge $\{a_m\}_{m\in\mathbf{N}}$ existiert mit $\binom{m}{k}p^k(1-p)^{m-k}\le a_m$ für alle $k\in\mathbf{N}_0$, $m\in\mathbf{N}$ und $\lim_{m\to\infty}a_m=0$. Zur Berechnung der obigen Wahrscheinlichkeit wird dann für jedes m die gleiche Anzahl r solcher durch a_m beschränkter Summanden aufaddiert, woraus die Konvergenz gegen 0 folgt. ∎

Satz 3.2.1. *(rekurrente/transiente Zustände)*
Es sei $\{X_n\}_{n\in\mathbf{N}_0}$ eine homogene Markoff–Kette sowie $i\in\mathcal{S}$ ein Zustand mit $P\big(\bigcup_{n=0}^{\infty}\{X_n=i\}\big)>0$ (d.h. $i\in\mathcal{S}$ wird mit positiver Wahrscheinlichkeit irgendwann angenommen).

a) *Wenn i rekurrent ist, gilt $P\big(\limsup_{n\to\infty}\{X_n=i\}\big)=P\big(\bigcup_{n=0}^{\infty}\{X_n=i\}\big)$ und $\sum_{n=1}^{\infty}p_{ii}^{(n)}=\infty$.*

b) *Ist i transient, so folgt $\sum_{n=1}^{\infty}p_{ii}^{(n)}<\infty$.*

Beweis. Mit Hilfe der De Morgan–Regeln und der σ–Additivität (1.1.30) von P folgt

$$\begin{aligned}P\big((\limsup_{n\to\infty}\{X_n=i\})^c\big)&=P\Big(\Big(\bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty}\{X_k=i\}\Big)^c\Big)=P\Big(\bigcup_{n=0}^{\infty}\Big(\bigcap_{k=n}^{\infty}\{X_k\ne i\}\Big)\Big)\\&=P\Big(\{X_k\ne i,k\in\mathbf{N}_0\}\cup\bigcup_{n=0}^{\infty}\{X_n=i,X_k\ne i,k\ge n+1\}\Big)\\&=P(X_k\ne i,k\in\mathbf{N}_0)\\&\quad+\sum_{n\in\mathbf{N}_0,P(X_n=i)>0}P(X_k\ne i,k\ge n+1\mid X_n=i)\cdot P(X_n=i)\end{aligned}\tag{3.2.21}$$

Für obige Summanden gilt für alle $n \in \mathbb{N}_0$ mit $P(X_n = i) > 0$

$$P(X_k \neq i, k \geq n+1 \mid X_n = i) = P\Big(\bigcap_{k=1}^{\infty} \{X_{n+k} \neq i\} \mid X_n = i\Big)$$

$$= P\Big(\bigcap_{\ell=1}^{\infty} \Big(\bigcap_{k=1}^{\ell} \{X_{n+k} \neq i\}\Big) \mid X_n = i\Big) = \lim_{\ell \to \infty} P\Big(\bigcap_{k=1}^{\ell} \{X_{n+k} \neq i\} \mid X_n = i\Big),$$

da $A_\ell = \bigcap_{k=1}^{\ell} \{X_{n+k} \neq i\}$ eine monoton fallende Mengenfolge und das Wahrscheinlichkeitsmaß $P(\cdot \mid X_n = i)$ stetig von oben (vgl. (1.1.39)) ist.
Wie man durch Anwendung von (3.2.7) auf homogene Markoff-Ketten sieht, hängt die bedingte Verteilung $P^{(X_{n+1},\ldots,X_{n+\ell} \mid X_n = i)}$ für alle $\ell \in \mathbb{N}$ nicht von $n \in \mathbb{N}_0$ ab, so daß insgesamt für alle $n \in \mathbb{N}_0$ mit $P(X_n = i) > 0$

$$P(X_k \neq i, k \geq n+1 \mid X_n = i) = \rho(i)$$

unabhängig von $n \in \mathbb{N}_0$ ist. Damit haben wir in (3.2.21)

$$P\big((\limsup_{n\to\infty} \{X_n = i\})^c\big) = P\Big(\Big(\bigcup_{k=0}^{\infty} \{X_k = i\}\Big)^c\Big) + \sum_{n=0}^{\infty} \rho(i) P(X_n = i) \leq 1. \tag{3.2.22}$$

Also muß $\rho(i) = 0$ oder $\sum_{n=0}^{\infty} P(X_n = i) < \infty$ gelten.
a) Sei $i \in \mathcal{S}$ rekurrent mit $P\big(\limsup_{n\to\infty} \{X_n = i\}\big) > 0$. Satz 1.1.3 a) (Borel-Cantelli-Lemma) liefert dann $\sum_{n=0}^{\infty} P(X_n = i) = \infty$, folglich $\rho(i) = 0$. In (3.2.22) erhalten wir durch Komplementbildung

$$P\big(\limsup_{n\to\infty} \{X_n = i\}\big) = P\Big(\bigcup_{n=0}^{\infty} \{X_n = i\}\Big).$$

Dies ist der erste Teil der Behauptung. Insbesondere folgt hieraus für ein $k \in \mathbb{N}_0$ mit $P(X_k = i) > 0$ (ein solches existiert nach Voraussetzung)

$$P\big(\limsup_{n\to\infty} \{X_n = i\} \mid X_k = i\big) = P\Big(\bigcup_{n=0}^{\infty} \{X_n = i\} \mid X_k = i\Big) = 1.$$

Mit Hilfe des Borel-Cantelli-Lemmas, angewendet auf das Wahrscheinlichkeitsmaß $P(\cdot \mid X_k = i)$, schließen wir $\sum_{n=0}^{\infty} P(X_n = i \mid X_k = i) = \infty$, also $\sum_{n=k+1}^{\infty} P(X_n = i \mid X_k = i) = \sum_{n=k+1}^{\infty} p_{ii}^{(n-k)} = \infty$, woraus a) durch Indexverschiebung folgt.
b) Für transientes $i \in \mathcal{S}$ muß $\rho(i) > 0$ gelten, denn anderenfalls würde in (3.2.22) $P\big(\limsup_{n\to\infty} \{X_n = i\}\big) = P\big(\bigcup_{n=0}^{\infty} \{X_n = i\}\big) > 0$ folgen, im Widerspruch zur Definition eines transienten Zustands.
Also gilt $\sum_{n=0}^{\infty} P(X_n = i) < \infty$. Wähle $k \in \mathbb{N}_0$ mit $P(X_k = i) > 0$. Dann folgt

$$\sum_{n=0}^{\infty} P(X_n = i \mid X_k = i) = \frac{1}{P(X_k = i)} \sum_{n=0}^{\infty} P(X_n = i, X_k = i)$$

$$\leq \frac{1}{P(X_k = i)} \sum_{n=0}^{\infty} P(X_n = i) < \infty$$

und

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=k+1}^{\infty} p_{ii}^{(n-k)} = \sum_{n=k+1}^{\infty} P(X_n = i \mid X_k = i) < \infty,$$

womit b) bewiesen ist. ∎

Satz 3.2.1 zeigt insbesondere, daß für homogene Markoff-Ketten $\{X_n\}_{n\in\mathbb{N}_0}$ mit $P\big(\bigcup_{n=0}^{\infty}\{X_n = i\}\big) > 0$ für alle $i \in \mathcal{S}$ die Begriffe rekurrenter bzw. transienter Zustand nur von der Übergangsmatrix $\mathbf{\Pi} = (p_{ij})_{i,j\in\mathcal{S}}$ abhängen. Teil a) und b) von Satz 3.2.1 implizieren ja die Äquivalenzen

$$\begin{aligned} i \in \mathcal{S} \text{ ist rekurrent} \quad &\Longleftrightarrow \quad \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty, \\ i \in \mathcal{S} \text{ ist transient} \quad &\Longleftrightarrow \quad \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty. \end{aligned} \tag{3.2.23}$$

Beispiel 3.2.4. (Irrfahrt auf $\mathbb{Z}$)
Wir betrachten die Irrfahrt auf dem beidseitig unbeschränkten Gitter $\mathcal{G} = \mathbb{Z}$ mit der Übergangsmatrix

$$\mathbf{\Pi} = (p_{ij})_{i,j\in\mathbb{Z}} \text{ mit } p_{ij} = \begin{cases} p, & \text{falls } j = i+1 \\ 1-p, & \text{falls } j = i-1 \\ 0, & \text{sonst} \end{cases},$$

wobei $0 < p < 1$. Offensichtlich existiert für jedes $i \in \mathbb{Z}$ ein Index $n \in \mathbb{N}_0$ mit $P(X_n = i) > 0$, egal welche Startverteilung vorliegt. Damit sind die Voraussetzungen von Satz 3.2.1 erfüllt.
Zur Bestimmung, ob ein Zustand $i \in \mathbb{Z}$ rekurrent oder transient ist, benötigen wir die n-Schritt-Übergangswahrscheinlichkeiten $p_{ii}^{(n)}$, $i \in \mathbb{Z}$, $n \in \mathbb{N}$. Mit (3.2.4) gilt

$$\begin{aligned} p_{ii}^{(n)} &= \sum_{\ell_1,\ldots,\ell_{n-1}\in\mathbb{Z}} p_{i\ell_1}\Big(\prod_{j=2}^{n-1} p_{\ell_{j-1}\ell_j}\Big) p_{\ell_{n-1}i} \\ &= \sum_{u_1,\ldots,u_n\in\{-1,1\},\sum_{j=1}^{n} u_j = 0} p_{i,i+u_1}\Big(\prod_{j=2}^{n} p_{i+u_1+\ldots+u_{j-1},i+u_1+\ldots+u_j}\Big) \\ &= \begin{cases} 0, & \text{falls } n \text{ ungerade} \\ \binom{n}{n/2} p^{n/2}(1-p)^{n/2}, & \text{falls } n \text{ gerade} \end{cases}. \end{aligned}$$

Diese Formel macht man sich auch leicht kombinatorisch klar. Um in n Schritten, ausgehend von der Position i, wieder in i zu landen, muß sich das Teilchen genau $n/2$-mal um eine Einheit nach rechts und $n/2$-mal nach links bewegen. Damit ist für ungerades n keine Rückkehr zum Ausgangsort möglich. Für gerades n gibt es $\binom{n}{n/2}$ verschiedene solcher Wege, und jeder hat wegen der stochastischen Unabhängigkeit der Bewegungen die Wahrscheinlichkeit $p^{n/2}(1-p)^{n/2}$.

Mit Hilfe der Stirling-Formel (vgl. (2.1.87))

$$\lim_{n\to\infty} \frac{n!}{n^{n+1/2}e^{-n}\sqrt{2\pi}} = 1 \tag{3.2.24}$$

(kurz: $n! \sim n^{n+1/2}e^{-n}\sqrt{2\pi}$) zeigen wir für gerade natürliche Zahlen $2n$, $n \in \mathbb{N}$,

$$\begin{aligned} p_{ii}^{(2n)} &= \binom{2n}{n}\big(p(1-p)\big)^n = \frac{2n!}{(n!)^2}\big(p(1-p)\big)^n \\ &\sim \frac{(2n)^{2n+1/2}e^{-2n}\sqrt{2\pi}}{\big(n^{n+1/2}e^{-n}\sqrt{2\pi}\big)^2}\big(p(1-p)\big)^n = \frac{1}{\sqrt{\pi}}\frac{\big(4p(1-p)\big)^n}{\sqrt{n}} \quad . \end{aligned} \tag{3.2.25}$$

Da $4p(1-p) = 1$, falls $p = 1/2$, und $4p(1-p) < 1$, sonst, ist

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi}}\frac{\big(4p(1-p)\big)^n}{\sqrt{n}} \quad \begin{cases} < \infty, & \text{falls } p \neq 1/2 \\ = \infty, & \text{falls } p = 1/2 \end{cases} .$$

Sind nun $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ Folgen positiver Zahlen mit $a_n \sim b_n$, so sind $\sum_{n=1}^{\infty} a_n$ und $\sum_{n=1}^{\infty} b_n$ beide $< \infty$ oder beide $= \infty$, wie man durch Erweitern $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{a_n}{b_n} b_n$ sieht.
Insgesamt gilt

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} p_{ii}^{(2n)} \quad \begin{cases} < \infty, & \text{falls } p \neq 1/2 \\ = \infty, & \text{falls } p = 1/2 \end{cases}$$

und mit Satz 3.2.1: Alle Zustände sind transient (bzw. rekurrent), falls $p \neq 1/2$ (bzw. $p = 1/2$).
Bemerkenswert ist noch, daß auch für $p = 1/2$ aus (3.2.25) $\lim_{n\to\infty} p_{ii}^{(n)} = 0$ folgt. Man nennt Zustände $i \in \mathcal{S}$, die diese Eigenschaften haben, null-rekurrent. Wir werden später hierauf zurückkommen. ∎

Für eine homogene Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$ sei nun $i \in \mathcal{S}$ ein Zustand mit $P(X_k = i) > 0$ für ein $k \in \mathbb{N}_0$.

$$f_{ii}^{(n)} = P\big(X_{k+n} = i, X_{k+j} \neq i, j = 1, \ldots, n-1 \mid X_k = i\big),\ n \in \mathbb{N}_0, \tag{3.2.26}$$

bezeichne die Wahrscheinlichkeit, daß ausgehend von $i \in \mathcal{S}$, nach genau n Schritten erstmalig wieder i angenommen wird. Die Wahrscheinlichkeit in (3.2.26) ist unabhängig von $k \in \mathbb{N}_0$, da die Verteilung $P^{(X_{\ell+1},\ldots,X_{\ell+n})}(\cdot \mid X_\ell = i)$ identisch ist für alle $\ell \in \mathbb{N}_0$ mit $P(X_\ell = i) > 0$, wie man durch Anwendung von (3.2.7) im Fall homogener Markoff-Ketten sieht. Eine Festlegung von $f_{ii}^{(n)}$ erfolgt dann mit einem beliebigen Index $k \in \mathbb{N}_0$, für den die elementare bedingte Wahrscheinlichkeit (3.2.26) definiert ist.

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)} \tag{3.2.27}$$

läßt sich dann als Erwartungswert der ersten Rückkehrzeit nach $i \in \mathcal{S}$ interpretieren.

Genauer gilt mit der Zufallsvariablen $N_{ik} = \min\{r \in \mathbb{N}_0 \mid X_{k+r} = i\}$, daß $\mu_i = \int N_{ik} dP(\cdot \mid X_k = i)$, wobei $P(\cdot \mid X_k = i)$ die elementare bedingte Wahrscheinlichkeitsverteilung unter $\{X_k = i\}$ auf $(\Omega, \mathcal{A})$ bezeichnet.

Für einen rekurrenten Zustand $i \in \mathcal{S}$ gilt $0 < P(\limsup_{n\to\infty}\{X_n = i\}) \leq P\big(\bigcup_{n=0}^{\infty}\{X_n = i\}\big)$, also ist $f_{ii}^{(n)}$ für alle $n \in \mathbb{N}_0$ wohldefiniert.

Definition 3.2.5. *(positiv-rekurrente/null-rekurrente Zustände)*
$i \in \mathcal{S}$ sei ein rekurrenter Zustand der homogenen Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$. i heißt positiv-rekurrent, wenn $\mu_i < \infty$; i heißt null-rekurrent, wenn $\mu_i = \infty$.

Anschaulich gesprochen bedeutet null-rekurrent, daß ein Zustand i zwar mit positiver Wahrscheinlichkeit unendlich oft angenommen wird, man im Mittel aber unendlich viele Zeiteinheiten warten muß, bis ein erneuter Besuch stattfindet.

Den folgenden Satz zitieren wir ohne Beweis. Er findet sich etwa in Ross (1983), Theorem 4.3.1.

Satz 3.2.2. *(Klassifikation von Zuständen)*
$i \in \mathcal{S}$ sei ein rekurrenter Zustand der homogenen Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$ mit Übergangsmatrix $\Pi = (p_{ij})_{i,j\in\mathcal{S}}$. Der Zustand i ist genau dann null-rekurrent, wenn $\lim_{n\to\infty} p_{ii}^{(n)} = 0$ gilt.

In Satz 3.2.2 werden rekurrente Zustände, für die ja $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ gilt, noch einmal nach der Geschwindigkeit der Divergenz dieser Summe unterschieden. Das Kriterium ist, ob die Summanden $p_{ii}^{(n)}$ selbst noch eine Nullfolge bilden.

Wir steuern im folgenden auf die Beschreibung des Grenzverhaltens von Markoff-Ketten im Sinn der punktweisen Konvergenz der Zähldichten $\boldsymbol{p}(n)$ $(n \to \infty)$ zu. Eine wesentliche Rolle spielen dabei die Begriffe "irreduzibel" und "aperiodisch".

Definition 3.2.6. *(irreduzible Übergangsmatrizen)*
Es bezeichne $\Pi = (p_{ij})_{i,j\in\mathcal{S}}$ die Übergangsmatrix einer homogenen Markoff-Kette $\{X_n\}_{n\in\mathbb{N}_0}$. Π heißt irreduzibel, wenn für alle $i, j \in \mathcal{S}$ ein $n \in \mathbb{N}$ und $i_0, i_1, \ldots, i_n \in \mathcal{S}$ existieren mit $i_0 = i$, $i_n = j$ und $p_{i_{k-1}i_k} > 0$ für alle $k = 1, \ldots, n$.

Irreduzibilität einer stochastischen Matrix besagt, daß jeder Zustand $j \in \mathcal{S}$ von jedem anderen $i \in \mathcal{S}$ in einer endlichen Anzahl von Übergängen erreicht werden kann. Wegen (3.2.4) ist die folgende Bedingung äquivalent zu der in Definition 3.2.6 angegebenen.

$$\text{Für alle } i, j \in \mathcal{S} \text{ existiert } n \in \mathbb{N} \text{ mit } p_{ij}^{(n)} > 0. \tag{3.2.28}$$

Wenn $\{X_n\}_{n\in\mathbb{N}_0}$ eine homogene Markoff-Kette mit irreduzibler Übergangsmatrix $\Pi = (p_{ij})_{ij\in\mathcal{S}}$ ist, dann gilt

$$P\Big(\bigcup_{n=1}^{\infty}\{X_n = i\}\Big) > 0 \quad \text{für alle } i \in \mathcal{S}. \tag{3.2.29}$$

Denn bei der Anfangsverteilung $\boldsymbol{p}(0)$ wähle man eine Komponente $i_0 \in \mathcal{S}$ mit $p_{i_0}(0) > 0$. Für beliebiges $i \in \mathcal{S}$ existiert wegen (3.2.28) $n \in \mathbb{N}$ mit $p_{i_0 i}^{(n)} > 0$, und mit (3.2.12) folgt

$$p_i(n) = P(X_n = i) = (\boldsymbol{p}(0)\boldsymbol{\Pi}^n)_i = \sum_{j \in \mathcal{S}} p_j(0) p_{ji}^{(n)} \geq p_{i_0}(0) p_{i_0 i}^{(n)} > 0.$$

Satz 3.2.3. *$\{X_n\}_{n \in \mathbb{N}_0}$ sei eine homogene Markoff-Kette mit irreduzibler Übergangsmatrix $\boldsymbol{\Pi} = (p_{ij})_{i,j \in \mathcal{S}}$. Dann gilt genau eine der folgenden Aussagen:*

a) Alle Zustände $i \in \mathcal{S}$ sind transient, und es gilt

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty \text{ für alle } i, j \in \mathcal{S}.$$

b) Alle Zustände $i \in \mathcal{S}$ sind null-rekurrent mit $P\big(\limsup_{n \to \infty} \{X_n = i\}\big) = 1$ für alle $i \in \mathcal{S}$, und es gilt

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty \quad \text{und} \quad \lim_{n \to \infty} p_{ij}^{(n)} = 0 \text{ für alle } i, j \in \mathcal{S}.$$

c) Alle Zustände $i \in \mathcal{S}$ sind positiv-rekurrent mit $P\big(\limsup_{n \to \infty} \{X_n = i\}\big) = 1$ für alle $i \in \mathcal{S}$, und es gilt

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty \quad \text{und} \quad \limsup_{n \to \infty} p_{ij}^{(n)} > 0 \text{ für alle } i, j \in \mathcal{S}.$$

Beweis. a), b) und c) schließen sich gegenseitig aus; wir zeigen in einer Fallunterscheidung, daß stets eine der drei Aussagen gilt.
Fall 1: Es existiert ein transienter Zustand $k \in \mathcal{S}$.
Nach Satz 3.2.1 b) gilt $\sum_{n=1}^{\infty} p_{kk}^{(n)} < \infty$. Sind nun $i, j \in \mathcal{S}$ beliebige Zustände, so existieren wegen $\boldsymbol{\Pi}$ irreduzibel $s, t \in \mathbb{N}$ mit $p_{ki}^{(s)} > 0$ und $p_{jk}^{(t)} > 0$. Mit der Chapman-Kolmogoroff-Gleichung (3.2.16) schließen wir

$$p_{ki}^{(s)} \cdot p_{ij}^{(n)} \cdot p_{jk}^{(t)} \leq p_{kk}^{(s+n+t)} \text{ für alle } n \in \mathbb{N}, \tag{3.2.30}$$

also

$$p_{ki}^{(s)} \cdot \Big(\sum_{n=1}^{\infty} p_{ij}^{(n)} \Big) \cdot p_{jk}^{(t)} \leq \sum_{n=1}^{\infty} p_{kk}^{(s+n+t)} = \sum_{n=s+t+1}^{\infty} p_{kk}^{(n)} < \infty.$$

Folglich ist $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ für alle $i, j \in \mathcal{S}$, also gilt a).
Fall 2: Es existiert ein null-rekurrenter Zustand $k \in \mathcal{S}$.
Nach Satz 3.2.1 a) und Satz 3.2.2 gilt $\sum_{n=1}^{\infty} p_{kk}^{(n)} = \infty$ und $\lim_{n \to \infty} p_{kk}^{(n)} = 0$. Seien $i, j \in \mathcal{S}$ beliebige Zustände. Auch hier gilt (3.2.30) mit positiven $p_{ki}^{(s)}$ und $p_{jk}^{(t)}$, so daß $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ für alle $i, j \in \mathcal{S}$.

Weiterhin lassen sich $u, v \in \mathcal{S}$ finden mit $p_{ik}^{(u)} > 0$ und $p_{kj}^{(v)} > 0$. Es gilt $p_{ik}^{(u)} \cdot p_{kk}^{(n)} \cdot p_{kj}^{(v)} \leq p_{ij}^{(u+n+v)}$ für alle $n \in \mathbb{N}$, so daß nach Summation der rechten und linken Seite

$$\infty = \sum_{n=1}^{\infty} p_{kk}^{(n)} \leq \sum_{n=1}^{\infty} p_{ij}^{(u+n+v)}, \text{ also } \sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty \text{ für alle } i, j \in \mathcal{S}.$$

Dies gilt insbesondere für beliebiges $i = j$, woraus mit Satz 3.2.1 a) und Satz 3.2.2 die Null–Rekurrenz aller Zustände folgt.
Es bleibt zu zeigen $P\big(\limsup_{n\to\infty}\{X_n = i\}\big) = 1$ für alle $i \in \mathcal{S}$. Dies folgt allein aus der Rekurrenz und Irreduzibilität von $\mathbf{\Pi}$. Wir benutzen hierzu die Argumentation (3.2.22), nach der bei rekurrentem $i \in \mathcal{S}$, falls $P(X_n = i) > 0$, unabhängig von n gilt $\rho(i) = P(X_k \neq i, k \geq n+1 \mid X_n = i) = 0$.
Sei nun $j \in \mathcal{S}$, $j \neq i$, beliebig. Weil $\mathbf{\Pi}$ irreduzibel ist, existieren $m \in \mathbb{N}$ und $i_0, i_1, \ldots, i_m \in \mathcal{S}$ mit $i_0 = i$, $i_m = j$, $i_1, \ldots, i_{m-1} \notin \{i, j\}$ und $p_{i_{\ell-1} i_\ell} > 0$ für alle $\ell = 1, \ldots, m$, derart daß

$$\begin{aligned}
0 = \rho(i) &\geq P(X_k \neq i, k \geq n+m+1,\ X_{n+\ell} = i_\ell, 1 \leq \ell \leq m \mid X_n = i) \\
&= P(X_{n+k} \neq i, k \geq m+1,\ X_{n+\ell} = i_\ell, 0 \leq \ell \leq m)/P(X_n = i) \\
&= P(X_{n+k} \neq i, k \geq m+1 \mid X_{n+\ell} = i_\ell, 0 \leq \ell \leq m) \\
&\qquad \cdot P(X_{n+\ell} = i_\ell, 0 \leq \ell \leq m)/P(X_n = i) \\
&= P(X_{n+k} \neq i, k \geq m+1 \mid X_{n+m} = j) \cdot P(X_{n+\ell} = i_\ell, 0 \leq \ell \leq m)/P(X_n = i).
\end{aligned}$$

Hierbei ist $P(X_n = i) > 0$ und wegen (3.2.7) auch $P(X_{n+\ell} = i_\ell, 0 \leq \ell \leq m) > 0$. Es folgt $P(X_{n+k} \neq i, k \geq m+1 \mid X_{n+m} = j) = 0$ und wegen der Unabhängigkeit dieses Ausdrucks von n und m

$$P(X_\ell \neq i, \ell \geq n+1 \mid X_n = j) = 0 \text{ für alle } n \in \mathbb{N}_0 \text{ mit } P(X_n = j) > 0.$$

Insbesondere

$$P(X_\ell \neq i, \ell \geq 1 \mid X_0 = j) = 0 \text{ für alle } j \in \mathcal{S} \text{ mit } P(X_0 = j) > 0. \tag{3.2.31}$$

Für $i = j$ ist dies die von oben bekannte Aussage $\rho(i) = 0$.
Der Beweis von b) läßt sich nun folgendermaßen führen.

$$P(X_\ell \neq i, \ell \geq 1) = \sum_{j \in \mathcal{S}, P(X_0 = j) > 0} P(X_\ell \neq i, \ell \geq 1 \mid X_0 = j) P(X_0 = j) = 0,$$

also durch Komplementbildung mit Satz 3.2.1 a)

$$P\Big(\bigcup_{n=0}^{\infty} \{X_n = i\}\Big) = P\big(\limsup_{n\to\infty} \{X_n = i\}\big) = 1.$$

Fall 3: Es existiert ein positiv–rekurrenter Zustand.

Dann muß jeder Zustand positiv–rekurrent sein, da nach dem bereits Gezeigten Transienz oder Null–Rekurrenz eine gemeinsame Eigenschaft aller Elemente des Zustandsraums ist. Rekurrenz und Irreduzibilität haben wie im Fall 2 zur Folge, daß $P\big(\limsup_{n\to\infty}\{X_n = i\}\big) = 1$ für alle $i \in \mathcal{S}$. Genauso wie oben folgt auch $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ für alle $i, j \in \mathcal{S}$.

Angenommen es existieren nun $i, j \in \mathcal{S}$ mit $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. Da $\mathbf{\Pi}$ irreduzibel ist, läßt sich ein $m \in \mathbb{N}$ finden mit $p_{ij}^{(m)} > 0$, woraus mit $p_{ii}^{(n)} \cdot p_{ij}^{(m)} \le p_{ij}^{(n+m)}$ folgt, daß $\lim_{n\to\infty} p_{ii}^{(n)} = 0$. Nach Satz 3.2.2 ist i dann null–rekurrent, ein Widerspruch. Dies zeigt c). ∎

Für unser Ziel, die Konvergenz der n-ten Randverteilungen P^{X_n} gegen eine nicht degenerierte Wahrscheinlichkeitsverteilung im Sinn der punktweisen Konvergenz der Zähldichten zu untersuchen, liefert Satz 3.2.3 die folgenden Aussagen.

Korollar 3.2.1. *$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine homogene Markoff–Kette mit irreduzibler Übergangs-Matrix $\mathbf{\Pi}$ und Zustandsraum $\mathcal{S}$. Dann liegt der Fall c) aus Satz 3.2.3 vor, wenn*

a) ein stochastischer Vektor $\boldsymbol{p}^*$ existiert mit $\lim_{n\to\infty} \boldsymbol{p}(n) = \boldsymbol{p}^*$ oder

b) der Zustandsraum $\mathcal{S}$ endlich ist.

Beweis. a) Im Fall a) und b) aus Satz 3.2.3 gilt $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ für alle $i, j \in \mathcal{S}$, so daß $p_j(n) = \sum_{i\in\mathcal{S}} p_i(0) p_{ij}^{(n)}$ für alle $j \in \mathcal{S}$. Da $p_{ij}^{(n)} \le 1$ für alle $i, j \in \mathcal{S}$, $n \in \mathbb{N}$ und $\sum_{i\in\mathcal{S}} p_i(0) = 1$, gilt für den Limes

$$\lim_{n\to\infty} p_j(n) = \lim_{n\to\infty} \sum_{i\in\mathcal{S}} p_i(0) p_{ij}^{(n)} = \sum_{i\in\mathcal{S}} p_i(0) \lim_{n\to\infty} p_{ij}^{(n)} = 0$$

für alle $j \in \mathcal{S}$, also liegt keine Konvergenz gegen einen stochastischen Vektor vor.

b) Im Fall eines endlichen Zustandsraums $\mathcal{S}$ kann ebenso nicht $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ für alle $i, j \in \mathcal{S}$ gelten, da nach Lemma 3.2.4 bei Konvergenz von $\mathbf{\Pi}^n = (p_{ij}^{(n)})_{i,j\in\mathcal{S}}$ die Grenzmatrix stochastisch ist, also nicht die Nullmatrix sein kann. ∎

Selbst im Fall c) von Satz 3.2.3 muß nicht immer Konvergenz von $\mathbf{\Pi}^n$ bzw. $p(n)$ gegen eine stochastische Matrix bzw. einen stochastischen Vektor vorliegen, wie das Beispiel $\mathbf{\Pi} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ zeigt. Als weitere Bedingung ist "Aperiodizität" nötig.

Definition 3.2.7. *(Periode eines Zustands)*
$\mathbf{\Pi} = (p_{ij})_{i,j\in\mathcal{S}}$ sei eine stochastische Matrix. Für $i \in \mathcal{S}$ heißt [1]

$$d(i) = \mathrm{GGT}\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\}$$

die Periode von i. Wenn $d(i) = 1$ ist, heißt i aperiodisch. Gilt $p_{ii}^{(n)} = 0$ für alle $n \in \mathbb{N}$, so wird $d(i) = \infty$ gesetzt.

---

[1] GGT heißt "größter gemeinsamer Teiler"

Lemma 3.2.5. *(Periode)*
$\Pi = (p_{ij})_{i,j\in\mathcal{S}}$ sei eine irreduzible stochastische Matrix. Dann gilt $d(i) = d(j) = d$ für alle $i, j \in \mathcal{S}$. (d heißt dann Periode von Π; bei $d = 1$ heißt Π aperiodisch.)

Beweis. Seien $i, j \in \mathcal{S}$. Es existieren $s, t \in \mathbb{N}$ mit $p_{ij}^{(s)} > 0$ und $p_{ji}^{(t)} > 0$. Es gilt $d(i)|(s+t)$ [1], da $p_{ii}^{(s+t)} \geq p_{ij}^{(s)} \cdot p_{ji}^{(t)} > 0$. Mit der Ungleichung $p_{ii}^{(s+n+t)} \geq p_{ij}^{(s)} \cdot p_{jj}^{(n)} \cdot p_{ji}^{(t)}$ für alle $n \in \mathbb{N}$ schließen wir: Ist $p_{jj}^{(n)} > 0$, gilt $d(i)|(s+n+t)$. Mit $d(i)|(s+t)$ folgt $d(i)|n$, also $d(j) \leq d(i)$. Durch Vertauschen der Rollen von i und j in obiger Argumentation folgt $d(i) \leq d(j)$, also die Behauptung. ∎

Satz 3.2.4. *(positive Rekurrenz und stationäre Verteilung)*
$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine homogene Markoff-Kette mit irreduzibler und aperiodischer Übergangs-Matrix $\Pi = (p_{ij})_{i,j\in\mathcal{S}}$. Dann gilt: Alle Zustände $i \in \mathcal{S}$ sind positiv rekurrent genau dann, wenn eine stationäre Verteilung $\boldsymbol{p}^*$ existiert (also $\boldsymbol{p}^* \cdot \Pi = \boldsymbol{p}^*$ für einen stochastischen Vektor $\boldsymbol{p}^*$). In diesem Fall ist $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots)$ eindeutig bestimmt, und zwar

$$p_j^* = \frac{1}{\mu_j} = \lim_{n\to\infty} p_{ij}^{(n)} \quad \text{für alle } i, j \in \mathcal{S}$$

mit μ_j aus (3.2.27). $\boldsymbol{p}^*$ ist damit unabhängig von der Anfangsverteilung $p(0)$.

Beweis. Wir nehmen an, daß alle Zustände positiv rekurrrent sind (Fall c) in Satz 3.2.3), und benutzen die Tatsache, daß im Fall der Aperiodizität von Π

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} \quad \text{für alle } i, j \in \mathcal{S} \tag{3.2.32}$$

gilt (vgl. Ross (1983), Theorem 4.3.1). Es ist

$$0 < \sum_{j\in\mathcal{S}} \frac{1}{\mu_j} = \sum_{j\in\mathcal{S}} \lim_{n\to\infty} p_{ij}^{(n)} \leq \lim_{n\to\infty} \sum_{j\in\mathcal{S}} p_{ij}^{(n)} = 1,$$

wobei die linke Ungleichung wegen Satz 3.2.3 c) richtig ist. Die rechte Ungleichung gilt im Fall eines endlichen Zustandsraums sogar mit Gleichheit; sie folgt im unendlichen Fall $\mathcal{S} = \mathbb{N}$ aus der Ungleichung $\sum_{j=1}^{r} \frac{1}{\mu_j} = \sum_{j=1}^{r} \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j=1}^{r} p_{ij}^{(n)} \leq 1$ durch Grenzübergang $r \to \infty$.
Definiere $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots)$ durch $p_i^* = \frac{1}{\mu_i} / \sum_{k\in\mathcal{S}} \frac{1}{\mu_k}$. $\boldsymbol{p}^*$ ist offensichtlich ein stochastischer Vektor, für den

$$(\boldsymbol{p}^* \cdot \Pi)_j = \sum_{i\in\mathcal{S}} p_i^* p_{ij} = \frac{1}{\sum_{k\in\mathcal{S}} 1/\mu_k} \sum_{i\in\mathcal{S}} \frac{1}{\mu_i} p_{ij}$$

gilt. Mit

$$\sum_{i=1}^{r} \frac{1}{\mu_i} p_{ij} = \lim_{n\to\infty} \sum_{i=1}^{r} p_{ji}^{(n)} p_{ij} \leq \lim_{n\to\infty} \sum_{i\in\mathcal{S}} p_{ji}^{(n)} p_{ij} = \lim_{n\to\infty} p_{jj}^{(n+1)} = \frac{1}{\mu_j}$$

[1] $d(i)|(s+1)$ bedeutet "$d(i)$ ist Teiler von $(s+1)$"

folgt nach Grenzübergang $r \to \infty$

$$(\boldsymbol{p}^* \cdot \boldsymbol{\Pi})_j = \frac{1}{\mu_j} \Big/ \sum_{k \in \mathcal{S}} \frac{1}{\mu_k} \leq p_j^*. \tag{3.2.33}$$

Da $\boldsymbol{p}^* \cdot \boldsymbol{\Pi}$ und $\boldsymbol{p}^*$ beides stochastische Vektoren mit Zeilensumme 1 sind, gilt Gleichheit in (3.2.33), also $\boldsymbol{p}^* \cdot \boldsymbol{\Pi} = \boldsymbol{p}^*$, und damit ist $\boldsymbol{p}^*$ stationäre Verteilung.
Zum Beweis der umgekehrten Richtung bemerken wir zunächst, daß für eine stationäre Verteilung $\boldsymbol{p}^*$ für alle $n \in \mathbb{N}$ die Identität

$$\boldsymbol{p}^* \cdot \boldsymbol{\Pi}^n = \boldsymbol{p}^* \cdot \boldsymbol{\Pi} \cdot \boldsymbol{\Pi}^{n-1} = \boldsymbol{p}^* \cdot \boldsymbol{\Pi}^{n-1} = \cdots = \boldsymbol{p}^* \cdot \boldsymbol{\Pi} = \boldsymbol{p}^*$$

gilt.
Angenommen, es existiert ein nicht positiv–rekurrenter Zustand. Dann gilt a) oder b) aus Satz 3.2.3 mit der Konsequenz, daß $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ ist für alle $i, j \in \mathcal{S}$. Es folgt wie im Beweis von Korollar 3.2.1

$$(\boldsymbol{p}^* \cdot \boldsymbol{\Pi}^n)_j = \sum_{i \in \mathcal{S}} p_i^* p_{ij}^{(n)} \to 0 \quad (n \to \infty),$$

im Widerspruch zu $\boldsymbol{p}^* \cdot \boldsymbol{\Pi}^n = \boldsymbol{p}^*$ für alle $n \in \mathbb{N}$.
Zu zeigen bleibt $p_j^* = 1/\mu_j$ für alle $j \in \mathcal{S}$, wenn $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots)$ eine stationäre Verteilung ist. Wie oben gilt $\boldsymbol{p}^* \cdot \boldsymbol{\Pi}^n = \boldsymbol{p}^*$ für alle $n \in \mathbb{N}$ und mit (3.2.32)

$$p_j^* = \sum_{i \in \mathcal{S}} p_i^* p_{ij}^{(n)} \underset{(n\to\infty)}{\longrightarrow} \sum_{i \in \mathcal{S}} p_i^* \frac{1}{\mu_j} = \frac{1}{\mu_j},$$

da $\sum_{i \in \mathcal{S}} p_i^* = 1$. Dies zeigt die Behauptung. ∎

Ist $\{X_n\}_{n \in \mathbb{N}_0}$ eine homogene Markoff–Kette mit irreduzibler und aperiodischer Übergangsmatrix $\boldsymbol{\Pi}$, so konvergiert bei positiv–rekurrenten Zuständen nach Satz 3.2.4 die Matrix $\boldsymbol{\Pi}^n$ elementweise gegen eine stochastische Matrix $\boldsymbol{\Pi}^*$, deren Zeilen identisch sind und die eindeutige, stationäre Verteilung $\boldsymbol{p}^*$ repräsentieren. Für einen endlichen Zustandsraum $\mathcal{S}$ liegt wegen Korollar 3.2.1 b) der Fall c) aus Satz 3.2.3, also positive Rekurrenz aller Zustände bei irreduzibler Übergangsmatrix vor. Es gilt damit das folgende

Korollar 3.2.2. *$\{X_n\}_{n \in \mathbb{N}_0}$ sei eine homogene Markoff–Kette mit endlichem Zustandsraum $\mathcal{S}$ und irreduzibler und aperiodischer Übergangsmatrix $\boldsymbol{\Pi}$. Dann sind alle Zustände positiv rekurrent, es existiert eine eindeutige stationäre Verteilung $\boldsymbol{p}^* = (p_1^*, \ldots, p_r^*)$ mit $p_i^* = 1/\mu_i$, $i = 1, \ldots, r$, und es gilt*

$$\lim_{n\to\infty} \boldsymbol{\Pi}^n = \begin{pmatrix} p_1^* & p_2^* & \cdots & p_r^* \\ \vdots & \vdots & & \vdots \\ p_1^* & p_2^* & \cdots & p_r^* \end{pmatrix}. \tag{3.2.34}$$

Ist die Übergangsmatrix irreduzibel, reicht die folgende, oft einfach nachzuprüfende Bedingung aus.

Korollar 3.2.3. *$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine homogene Markoff-Kette mit endlichem Zustandsraum $\mathcal{S}$ und irreduzibler Übergangsmatrix $\mathbf{\Pi}$. Wenn dann $n \in \mathbb{N}$ existiert derart, daß $\mathbf{\Pi}^n$ eine Spalte mit lauter positiven Elementen besitzt, so gelten die Aussagen von Korollar 3.2.2. Ferner existiert dann $m \in \mathbb{N}$ so, daß $\mathbf{\Pi}^m$ primitiv ist, d.h. lauter positive Elemente besitzt.*

Beweis. Wir zeigen zunächst die Aperiodizität von $\mathbf{\Pi}$, wodurch die Voraussetzungen von Korollar 3.2.2 erfüllt sind. Gelte also $p_{jk}^{(n)} > 0$ für ein $k \in \mathcal{S}$, $n \in \mathbb{N}$ und alle $j \in \mathcal{S}$. Mit vollständiger Induktion folgt

$$p_{kk}^{(s)} > 0 \quad \text{für alle } s \geq n, \tag{3.2.35}$$

und zwar so: Der Induktionsanfang gilt nach Voraussetzung. Sei nun $p_{kk}^{(s)} > 0$. Da $\mathbf{\Pi}$ irreduzibel ist, existiert ein Zustand $j \in \mathcal{S}$, $j \neq k$ mit $p_{kj} > 0$, und daher folgt mit (3.2.16)

$$p_{kk}^{(s+1)} \geq p_{kj} p_{jk}^{(s)} > 0.$$

Dies ist der Induktionsschluß, und (3.2.35) ist bewiesen. Für beliebige Primzahlen $u_1, u_2 \in \mathbb{N}$ mit $n \leq u_1 < u_2$ gilt also $p_{kk}^{(u_1)} > 0$ und $p_{kk}^{(u_2)} > 0$, so daß $\mathrm{GGT}\{\ell \in \mathbb{N} \mid p_{kk} > 0\} \leq \mathrm{GGT}\{u_1, u_2\} = 1$ ist.
Da $\mathbf{\Pi}$ irreduzibel ist, existiert für alle $i, j \in \mathcal{S}$ eine Potenz $n(i,j) \in \mathbb{N}$ mit $p_{ij}^{(n(i,j))} > 0$. Wähle $m = 2\max\{n(i,j) \mid i,j \in \mathcal{S}\} + n$. Dann gilt mit (3.2.16) für alle $i, j \in \mathcal{S}$

$$p_{ij}^{(m)} \geq p_{ik}^{(n(i,k))} \cdot p_{kk}^{(m-n(i,k)-n(k,j))} \cdot p_{kj}^{(n(k,j))} > 0,$$

da $m - n(i,k) - n(k,j) \geq n$ ist,und wegen (3.2.35). Der Beweis ist damit vollständig geführt. ∎

Das folgende Beispiel greift noch einmal Beispiel 3.1.2 in etwas allgemeinerer Form auf.

Beispiel 3.2.5. (Ein–Prozessor–System mit r I/O-Einheiten)
Ein System bestehe aus einer CPU (Zustand 0) und r Ein- bzw. Ausgabeeinheiten (Zustände $1, \ldots, r$). Beim Ablauf von Programmen auf diesem System sei bekannt, daß nach der Bearbeitung eines Programms auf der CPU unabhängig von der Vorvergangenheit mit Wahrscheinlichkeit $p_j > 0$ die I/O-Einheit Nr. j, $1 \leq j \leq r$, benutzt wird oder daß es mit Wahrscheinlichkeit $p_0 > 0$ sofort endet und damit für ein neues Programm im System Platz macht. Hieraus ergibt sich $\sum_{j=0}^{r} p_j = 1$. Nach Beendigung eines Programms beginnt sofort ein neues, das das gleiche Anforderungsverhalten bzgl. der I/O–Geräte zeigt.
Zur Beschreibung der Arbeitsweise des Systems ist eine homogene Markoff–Kette mit Zustandsraum $\mathcal{S} = \{0, 1, \ldots, r\}$ und Übergangsmatrix

$$\mathbf{\Pi} = \begin{pmatrix} p_0 & p_1 & \cdots & p_r \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} \tag{3.2.36}$$

geeignet. Π ist offensichtlich irreduzibel und besitzt eine positive Spalte. Mit Korollar 3.2.3 existiert eine eindeutige stationäre Verteilung $\boldsymbol{p}^* = (p_0^*, \ldots, p_r^*)$, und $\lim_{n\to\infty} \Pi^n$ besitzt $\boldsymbol{p}^*$ als identische Zeilen.
Zur expliziten Bestimmung von $\boldsymbol{p}^*$ lösen wir das Gleichungssystem $\boldsymbol{x} \cdot \Pi = \boldsymbol{x}$ in der Menge der stochastischen Vektoren $\{\boldsymbol{x} = (x_0, \ldots, x_r) \mid \sum_{i=0}^{r} x_i = 1, x_i \geq 0\}$ also

$$x_0 = x_0 p_0 + \sum_{i=1}^{r} x_i \quad \text{und} \quad x_j = x_0 p_j,\ j = 1, \ldots, r.$$

Da $\sum_{i=1}^{r} x_i = 1 - x_0$, liefert die erste Gleichung $x_0 = x_0 p_0 + 1 - x_0$ mit der Lösung $x_0 = 1/(2 - p_0)$. Hieraus folgt $x_j = p_j/(2 - p_0)$, $j = 1, \ldots, r$. Die eindeutige stationäre Verteilung lautet damit

$$\boldsymbol{p}^* = \left(\frac{1}{2 - p_0}, \frac{p_1}{2 - p_0}, \ldots, \frac{p_r}{2 - p_0} \right),$$

sie stimmt zugleich überein mit allen Zeilen der Matrix $\Pi^* = \lim_{n\to\infty} \Pi^n$. ∎

Man vergleiche dies mit dem in Beispiel 3.1.2 hergeleiteten Ergebnis.

In Kapitel 2.1 wurde die erste Eintrittszeit S in eine Menge B bei Folgen stochastisch unabhängiger Zufallsvariablen behandelt (vgl. (2.1.34) und (2.1.44)). Allgemein wollen wir für eine beliebige Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}_0}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in einem Meßraum $(\mathcal{X}, \mathcal{B})$ und $B \in \mathcal{B}$ definieren:

$$S(B) = \min\{n \in \mathbb{N}_0 \mid X_n \in B\} \tag{3.2.37}$$

heißt erste Eintrittszeit von $\{X_n\}_{n\in\mathbb{N}_0}$ in die Menge B. Es wird $S(B) = \infty$ gesetzt, falls $\{n \in \mathbb{N}_0 \mid X_n \in B\} = \emptyset$. (2.1.32) zeigt, daß $S(B)$ meßbar bezüglich $(\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}, \mathfrak{P}(\overline{\mathbb{N}}_0))$, also eine Zufallsvariable ist.

Beschreibt etwa $\{X_n\}_{n\in\mathbb{N}_0}$ einen stochastischen Suchalgorithmus und B die Menge der optimalen Elemente, so gibt die erste Eintrittszeit $S(B)$ an, wann erstmalig ein optimales Element gefunden wird. Der Erwartungswert von $S(B)$ ist die Kenngröße, die bei der Average-Case-Analyse dieses Algorithmus relevant ist.

Natürlich ist (3.2.37) wohldefiniert, wenn $\{X_n\}_{n\in\mathbb{N}_0}$ eine Markoff-Kette bildet. Der Erwartungswert der ersten Eintrittszeit läßt sich für homogene Markoff-Ketten mit endlichem Zustandsraum $\mathcal{S} = \{1, \ldots, r\}$ explizit mit relativ wenig Aufwand berechnen.

Wir benötigen hierzu die folgende Notation, wobei $\Pi = (p_{ij})_{i,j\in\mathcal{S}}$ eine $(r \times r)$-Matrix, $\boldsymbol{p} = (p_1, \ldots, p_r)$ ein Zeilenvektor und $M \subset \mathcal{S}$ eine Teilmenge von $\mathcal{S}$ ist.

$$\Pi_M = (p_{ij})_{i,j\in M} \quad \text{und} \quad \boldsymbol{p}_M = (p_i)_{i\in M} \tag{3.2.38}$$

bezeichnen die Teilmatrix bzw. den Teilvektor, die aus den Komponenten mit Indizes in der Menge M zusammengesetzt sind.

Lemma 3.2.6. *(erwartete erste Eintrittszeit bei Markoff-Ketten)*
$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine homogene Markoff-Kette mit endlichem Zustandsraum $S = \{1,\dots,r\}$, Übergangsmatrix $\Pi = (p_{ij})_{i,j\in S}$ und einer Anfangsverteilung gegeben durch $\boldsymbol{p}(0) = \big(p_1(0),\dots,p_r(0)\big)$. Es sei $B \subseteq S$ und $S(B)$ die in (3.2.37) definierte erste Eintrittszeit in die Menge B. Sind dann alle (eventuell auch komplexen) Eigenwerte der Matrix Π_{B^c} betragsmäßig kleiner als 1, so existiert $E\big(S(B)\big)$ und es gilt

$$E\big(S(B)\big) = \boldsymbol{p}_{B^c}(0)\cdot\big(I_b - \Pi_{B^c}\big)^{-1}\cdot \mathbf{1}_b, \tag{3.2.39}$$

wobei $b = \#(B^c)$, I_b die $b\times b$-Einheitsmatrix und $\mathbf{1}_b$ den Vektor $(1,\dots,1)^{tr}\in\mathbf{R}^b$ bezeichnen.

Beweis. Wegen (2.2.17) gilt

$$E\big(S(B)\big) = \sum_{n=0}^{\infty} P\big(S(B) > n\big) = \sum_{n=0}^{\infty} P(X_0\notin B, X_1\notin B,\dots,X_n\notin B).$$

Durch Anwendung der Formel (3.2.4) auf die Matrizen Π_{B^c} erhalten wir mit (3.2.6)

$$\begin{aligned} P(X_0\notin B,\dots,X_n\notin B) &= \sum_{i_0,\dots,i_n\in B^c} P(X_0 = i_0,\dots,X_n = i_n)\\ &= \sum_{i_0,\dots,i_n\in B^c} p_{i_0}(0)\prod_{j=1}^{n} p_{i_{j-1}i_j} = \sum_{i_0,i_n\in B^c}\Big(p_{i_0}(0)\sum_{i_1,\dots,i_{n-1}\in B^c}\prod_{j=1}^{n} p_{i_{j-1}i_j}\Big)\\ &= \sum_{i_0,i_n\in B^c} p_{i_0}(0)\big(\Pi_{B^c}^{n}\big)_{i_0 i_n} = \boldsymbol{p}_{B^c}(0)\cdot\Pi_{B^c}^{n}\cdot\mathbf{1}_b. \end{aligned}$$

Aus $(I-A)^{-1} = \sum_{n=0}^{\infty} A^n$, falls alle Eigenwerte der Matrix A betragsmäßig kleiner als 1 sind (vgl. Hunter (1983), Theorem 4.5.4), folgt

$$E\big(S(B)\big) = \boldsymbol{p}_{B^c}(0)\cdot\Big(\sum_{n=0}^{\infty}\Pi_{B^c}^{n}\Big)\cdot\mathbf{1}_b = \boldsymbol{p}_{B^c}(0)\cdot\big(I_b-\Pi_{B^c}\big)^{-1}\cdot\mathbf{1}_b.$$

Dies ist die Behauptung. ∎

Die Voraussetzungen von Lemma 3.2.6 sind etwa dann erfüllt, wenn alle Zeilensummen von Π_{B^c} kleiner als 1 sind. Denn nach dem Satz von Perron–Frobenius liegen alle Eigenwerte der Matrix $A = (a_{ij})_{1\le i,j\le r}$ in einer der Kreisscheiben der komplexen Ebene mit Mittelpunkt a_{ii} und Radius $\sum_{j\neq i}|a_{ij}|$. Für jeden Eigenwert λ von Π_{B^c} gilt damit $|\lambda - p_{ii}|\le\sum_{j\in B^c, j\neq i}|p_{ij}|$ für ein $i\in B^c$, also

$$|\lambda|\le|\lambda-p_{ii}|+|p_{ii}|\le\sum_{j\in B^c}|p_{ij}| = \sum_{j\in B^c}p_{ij} < 1.$$

Beispiel 3.2.6. (Ein–Prozessor–System)
Wir betrachten das Ein–Prozessor–System mit r Ein– bzw. Ausgabegeräten aus

Beispiel 3.2.5 und Übergangsmatrix (3.2.36) und fragen nach dem Erwartungswert der ersten Eintrittszeit in Zustand j, $1 \le j \le r$, wenn man in Zustand 0 startet. In der Realität ist das die erwartete Zeit bis zur ersten Nachfrage nach Peripheriegerät Nr. j nach Start des Programms im Prozessor. Wir betrachten lediglich $j = 1$; die Fälle $j = 2, \dots, r$ verlaufen völlig analog.
Sei also $B = \{1\}$. Alle Eigenwerte der Matrix

$$\Pi_{B^c} = \begin{pmatrix} p_0 & p_2 & \cdots & p_r \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

liegen in einer der beiden Kreisscheiben der komplexen Ebene mit Mittelpunkt p_0 und Radius $(1 - p_0 - p_1)$ bzw. mit Mittelpunkt 0 und Radius 1. Also gilt $|\lambda| \le 1$ für alle Eigenwerte λ von Π_{B^c}.
Angenommen es existiert ein (komplexer) Eigenwert λ mit $|\lambda| = 1$. Dann ist die Matrix

$$I_r - \lambda \Pi_{B^c} = \begin{pmatrix} 1 - \lambda p_0 & -\lambda p_2 & \cdots & -\lambda p_r \\ -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\lambda & 0 & \cdots & 1 \end{pmatrix}$$

wegen $\det(I_r - \lambda \Pi_{B^c}) = 0$ singulär. Die letzten $r - 1$ Spalten sind offensichtlich linear unabhängig, so daß

$$1 - \lambda p_0 = \lambda^2 (p_2 + \cdots + p_r)$$

folgt. Ist $\lambda = x + iy$ und $\bar{\lambda} = x - iy$ der konjugiert komplexe Eigenwert, so erhält man nach Multiplikation beider Seiten dieser Gleichung mit $\bar{\lambda}$, da $\lambda\bar{\lambda} = 1$, daß $p_0 = \bar{\lambda} - \lambda(p_2 + \cdots + p_r)$, also $-y - y(p_2 + \cdots + p_r) = -y(1 + p_2 + \cdots + p_r) = 0$, also $y = 0$. Dies bedeutet $p_0 = x(1 - p_2 - \cdots - p_r) = x(p_0 + p_1)$, also $x = p_0/(p_0 + p_1) < 1$ und folglich $x^2 + y^2 < 1$ im Widerspruch zur Annahme $|\lambda| = 1$.
Insgesamt sind alle Eigenwerte von Π_{B^c} betragsmäßig kleiner als 1 und wegen Lemma 3.2.6 folgt

$$E\big(S(\{1\})\big) = (1, 0 \dots, 0) \cdot \begin{pmatrix} 1 - p_0 & -p_2 & \cdots & p_r \\ -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Wie man durch Ausmultiplizieren nachrechnet, ist die Blockmatrix

$$\begin{pmatrix} \frac{1}{p_1} & \frac{1}{p_1}(p_2, \dots, p_r) \\ \frac{1}{p_1} e_{r-1} & I_{r-1} + \frac{1}{p_1} e_{r-1} \cdot (p_2, \dots, p_r) \end{pmatrix}$$

die oben verlangte Inverse, so daß $E\big(S(\{1\})\big) = (1 + p_2 + \cdots + p_r)/p_1$ bzw. allgemein

$$E\big(S(\{j\})\big) = \frac{1}{p_j}\Big(1 + \sum_{\substack{\ell=1 \\ \ell \ne j}}^{r} p_\ell\Big), \quad j = 1, \dots, r,$$

die erwartete Zahl der Zeiteinheiten bis zur Nachfrage des Geräts Nr. j, ist. ∎

Weitere Beispiele zur Anwendung von Markoff–Modellen in der Informatik findet man z.B. in Gläßer (1987) (automatische Spracherkennung), Ritter, Martinez & Schulten (1990) (Neuronale Netze) oder Lauritzen & Spiegelhalter (1988) (Expertensysteme).

3.3. Simulated Annealing

Ein Optimierungsverfahren, das der Evolution unserer biologischen Welt unterliegt, ist "Mutation und Selektion". Durch Mutation wird bei der Entstehung von Lebewesen zufällig das vorhandene Muster verändert, anschließend entscheidet die Umwelt, ob die Veränderung zu einer verbesserten Funktionsfähigkeit und damit zu einer erhöhten Fortpflanzungsrate unter Beibehaltung der veränderten Struktur führt. Im anderen Fall wird die neue Lösung wieder verworfen, d.h. die Mutation stirbt aus. Beachtenswert ist, daß auch schon einmal "schlechtere" Lösungen vorübergehend akzeptiert werden können, aus denen dann in späteren Schritten von anderer Basis aus wieder bessere hervorgehen können. Diese Verknüpfung von zufälligem "lokalem" Suchen nach besseren Möglichkeiten und einer zielgerichteten Auswahl der neuen Varianten hat in der Natur offensichtlich sehr erfolgreich gewirkt und gut angepaßte Lebensformen hervorgebracht.

Das gleiche Grundprinzip — quantitativ schon viel besser beschreibbar— findet man bei der Abkühlung von Metallen vom flüsigen zu einem möglichst energiearmen, festen Zustand. Beim langsamen Abkühlen von flüssigem Metall stellen sich nach und nach Zustände geringerer Energie ein. Trotz der absinkenden Temperatur können schon einmal wieder höhere Energiezustände, also schlechtere Lösungen, auftreten.

Die Modellierung dieses physikalischen Phänomens führte zur Entwicklung des Annealing–Algorithmus durch Metropolis u.a. (1953) und Kirkpatrick (1983); hieraus bezieht der Algorithmus auch seinen Namen (annealing = härten, ausglühen). Andere Bezeichnungen sind Monte–Carlo–Annealing oder Statistical Cooling.

Eine wesentliche Idee ist, daß mit gewisser abnehmender Wahrscheinlichkeit im Verlauf des Algorithmus auch wieder schlechtere Zustände akzeptiert werden. Dies verhindert, daß der Algorithmus in lokalen Optima steckenbleibt. Ist nämlich einmal ein lokal optimaler Zustand erreicht, reicht unter Umständen die Palette der möglichen Mutationen nicht aus, dieses lokale Optimum wieder zu verlassen, wenn man nicht auch in Zwischenschritten wieder schlechtere Zustände akzeptiert. Die zufällige Mutation oder Zustandsänderung und das zufällige Akzeptieren von neuen Zuständen bilden den stochastischen Anteil des Algorithmus; seine Implementierung auf Computern hängt damit wesentlich von einem effizienten Zufallsgenerator ab.

Wir werden im folgenden ein allgemeines Markoff–Modell für den Simulated Annealing–Algorithmus herleiten und hiermit sein stochastisches Verhalten analysieren. Dieses Modell wird sehr allgemein sein, so daß Simulated Annealing in vielen Anwendungsfällen als heuristisches Optimierungsverfahren eingesetzt werden kann, ohne zunächst die Struktur des speziellen Problems zu durchleuchten. Dies ist als Hauptvorteil des Algorithmus zu sehen: er stellt bei kombinatorischen Optimierungsproblemen ein flexibles und weithin anwendbares Hilfsmittel dar. Sein Nachteil ist, daß er manchmal selbst nach einer Vielzahl von Iterationen das globale Optimum nicht erreicht, obwohl er schon nach relativ wenigen Iterationen sehr nahe

am Optimum liegt. Das Verfahren bietet sich insbesondere an, wenn der Anwender auch schon mit suboptimalen Lösungen zufrieden ist. Natürlich wird Simulated Annealing häufig in Rechenaufwand und Laufzeit von Algorithmen unterboten, die auf die spezielle Struktur des zu bearbeitenden Problems genau zugeschnitten sind.

Das zu behandelnde kombinatorische Optimierungsproblem kann allgemein beschrieben werden mit einer endlichen Menge von möglichen Konfigurationen oder Zuständen S, die wir im folgenden ohne Einschränkung der Allgemeinheit mit $S = \{1, \ldots, r\} \subset \mathbf{N}$ identifizieren wollen. Jeder Konfiguration $i \in S$ wird vermöge einer Abbildung $f : S \to \mathbf{R}$ eine relle Zahl (Kosten) zugeordnet. Das Ziel ist nun, aus der in der Regel riesigen Konfigurationenmenge S einen Zustand i_{opt} zu finden, der die Kosten minimiert, d.h.

$$\underset{i \in S}{\text{minimiere}}\, f(i) \tag{3.3.1}$$

$S_{\text{opt}} = \{i \in S \mid f(i) \leq f(j) \text{ für alle } j = 1, \ldots, r\}$ bezeichne die Menge der optimalen Zustände.

Simulated Annealing ändert jetzt in jedem Schritt die aktuelle Konfiguration zufällig in eine mögliche Nachbarkonfiguration. Diese wird als neue Ausgangskonfiguration akzeptiert, wenn sie geringere Kosten hat. Mit einer gewissen vorgegebenen Wahrscheinlichkeit wird sie aber auch dann akzeptiert, wenn sie höhere Kosten trägt. Diese Wahrscheinlichkeit ist zu Beginn des Algorithmus groß, wird im Verlauf aber langsam herabgesetzt und konvergiert schließlich gegen Null. Wie oben beschrieben, wird hierdurch frühzeitiges Steckenbleiben des Algorithmus in lokalen Optima verhindert. Dieses Verhalten werden wir mit Hilfe von Markoff–Ketten beschreiben. Weiterführende Darstellungen finden sich etwa in Arbeiten von Aarts & van Laarhoven (1985), (1987).

$$\begin{aligned}&\{X_n^{(c)}\}_{n \in \mathbf{N}},\ c > 0, \text{ sei eine Familie von homogenen Markoff–Ketten}\\ &\text{auf einem Wahrscheinlichkeitsraum } (\Omega, \mathcal{A}, P) \text{ mit Zustandsraum}\\ &S = \{1, \ldots, r\},\ r \in \mathbf{N}, \text{ und Übergangsmatrizen}\\ &\mathbf{\Pi}(c) = \big(p_{ij}(c)\big)_{1 \leq i,j \leq r} = \big(P(X_n^{(c)} = j \mid X_{n-1}^{(c)} = i)\big)_{1 \leq i,j \leq r}.\end{aligned} \tag{3.3.2}$$

$c \in \mathbf{R}, c > 0$, hat die Rolle eines Kontrollparameters, er bestimmt, in welchem Maß die Wahrscheinlichkeit für das Akzeptieren von Zuständen mit schlechterem Zielfunktionswert im Verlauf des Algorithmus fällt. Der gesamte Annealing–Prozeß wird durch "Aneinanderhängen" von mit abnehmenden Werten von c simulierten, jeweils homogenen Markoff–Ketten realisiert.

$p_{ij}(c)$ ist die Wahrscheinlichkeit, daß der Algorithmus bei gegebenem Wert von c, ausgehend von $i \in S$, im nächsten Schritt auf die Konfiguration $j \in S$ übergeht. Im weiteren nehmen wir an, daß diese Übergangswahrscheinlichkeiten folgendermaßen zusammengesetzt sind.

$$p_{ij}(c) = \begin{cases} g_{ij}(c) a_{ij}(c), & \text{falls } j \neq i, \\ 1 - \sum_{\substack{\ell=1 \\ \ell \neq i}}^{r} g_{i\ell}(c) a_{i\ell}(c), & \text{falls } j = i, \end{cases}, \quad 1 \leq i, j \leq r, \tag{3.3.3}$$

wobei $0 < a_{ij}(c) \leq 1$, $0 \leq g_{ij}(c)$, $\sum_{\ell=1}^{r} g_{i\ell}(c) = 1$ für alle $i,j = 1\ldots,r$, $c > 0$ Wahrscheinlichkeiten mit folgender Interpretation sind. Die stochastische Matrix $G(c) = \big(g_{ij}(c)\big)_{1\leq i,j\leq r}$ enthält die Wahrscheinlichkeiten, ausgehend vom Zustand $i \in S$ im nächsten Schritt die Konfiguration $j \in S$ zu generieren. Unabhängig wird dann ein Zufallsexperiment durchgeführt, bei dem mit Wahrscheinlichkeit $a_{ij}(c)$ entschieden wird, den Zustand j auch tatsächlich als aktuellen zu akzeptieren. Die Elemente der Mengen $S_i = \{j \in S \mid g_{ij}(c) > 0\}$ heißen Nachbarzustände von $i \in S$. Wie man leicht nachrechnet, ist die durch (3.3.3) gegebene Matrix $\mathbf{\Pi}(c) = \big(p_{ij}(c)\big)_{1\leq i,j\leq r}$ eine stochastische Matrix.

Markoff–Ketten $\{X_n\}_{n\in\mathbf{N}}$ mit den Übergangswahrscheinlichkeiten (3.3.3) erhält man durch folgende Konstruktion, die den Prozeß der zufälligen Generierung und unabhängigen, zufälligen Annahme des generierten Zustands modelliert. Zur Vereinfachung der Notation wird der Parameter c bei den auftretenden Zufallsvariablen nicht explizit vermerkt. Y_i, $1 \leq i \leq r$, seien Zufallsvariable mit Werten in S und Verteilung $P(Y_i = j) = g_{i,j}(c)$. $\{Y_i = j\}$ beschreibt das Ereignis, daß vom Zustand i aus aus Konfiguration j erzeugt wird. Ferner seien Z_{ij}, $1 \leq i,j \leq r$, je auf $\{0,1\}$ binomialverteilte Zufallsvariable mit Parameter $a_{ij}(c)$, d.h. $P(Z_{ij} = 1) = a_{ij}(c) = 1 - P(Z_{ij} = 0)$. $\{Z_{ij} = 1\}$ bedeutet, daß vom Zustand i aus der Zustand j akzeptiert wird. Alle Zufallsvariablen Y_{ij}, Z_i, $1 \leq i,j \leq r$, seien stochastisch unabhängig, wodurch ausgedrückt wird, daß man in jedem Iterationsschritt unabhängig einen neuen Zustand generiert und in einem unabhängigen Zufallsexperiment mit Wahrscheinlichkeit $a_{ij}(c)$ über dessen Annahme entscheidet. Definiert man jetzt rekursiv

$$\begin{aligned} &X_n = Z_{X_{n-1},Y_{X_{n-1}}} Y_{X_{n-1}} + \big(1 - Z_{X_{n-1},Y_{X_{n-1}}}\big) X_{n-1}, \\ &X_0 = 0,\ Z_{0,i} = 1,\ 1 \leq i \leq r,\ Y_0 \text{ eine Zufallsvariable} \\ &\text{mit Werten in } S \text{ und gewünschter Anfangsverteilung,} \end{aligned} \tag{3.3.4}$$

so bildet $\{X_n\}_{n\in\mathbf{N}}$ eine Markoff–Kette mit Übergangswahrscheinlichkeiten (3.3.3), denn für $i \neq j$ gilt

$$\begin{aligned} &P(X_n = j \mid X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1) = P(X_n = j \mid X_{n-1} = i) \\ &= P\big(Z_{i,Y_i} Y_i + (1 - Z_{i,Y_i}) i = j\big) = P(Z_{i,Y_i} = 1, Y_i = j) \\ &= P(Z_{i,j} = 1, Y_i = j) = P(Z_{ij} = 1) P(Y_i = j) = a_{ij}(c) g_{ij}(c). \end{aligned}$$

Der zweite Teil der Formel aus (3.3.3) für $i = j$ ergibt sich hieraus als Komplementärwahrscheinlichkeit.

Man beachte, daß die Zufallsvariablen in (3.3.4) "zufällige Indizes" erhalten haben. Das in (3.3.4) auftretende Ereignis $\{Y_{X_{n-1}} = k\}$ etwa läßt sich schreiben als $\bigcap_{j=1}^{r} \{Y_j = k, X_{n-1} = j\}$, so daß wegen (1.1.27) und (1.1.28) die derart gebildeten Abbildungen meßbar, also Zufallsvariable sind.

Ausgangspunkt für das folgende ist jetzt das Markoff–Modell (3.3.2) mit den Übergangswahrscheinlichkeiten (3.3.3). Es stellt sich die Frage nach Bedingungen an die Wahrscheinlichkeiten $g_{ij}(c)$ und $a_{ij}(c)$ derart, daß der Algorithmus für eine

beliebige Startverteilung $\boldsymbol{q}_0$ mit $n \to \infty$ und $c \downarrow 0$ in einen optimalen Punkt läuft, genauer, daß die folgenden Limesverteilungen existieren

$$\lim_{n\to\infty} \boldsymbol{q}_0 \boldsymbol{\Pi}^n(c) = \boldsymbol{u}(c) \quad \text{und} \quad \lim_{c\downarrow 0} \boldsymbol{u}(c) = \boldsymbol{u}, \tag{3.3.5}$$

wobei $\boldsymbol{u} = (u_1, \ldots, u_r)$, $u_i \geq 0$, und $\sum_{i\in S_{\text{opt}}} u_i = 1$.

Die folgenden Bedingungen stellen sicher, daß der erste Grenzwert in (3.3.5) existiert.

$$\begin{aligned} &\text{Für alle } i,j \in \{1,\ldots,r\},\ c > 0 \text{ existieren } k = k(i,j) \in \mathbb{N}, \\ &\{l_1,\ldots,l_k\} \subseteq \{1,\ldots,r\} \text{ mit } l_1 = i, l_k = j \text{ so,} \\ &\text{daß } g_{l_t,l_{t+1}}(c) > 0 \text{ für alle } t = 1,\ldots,k-1 \end{aligned} \tag{3.3.6}$$

und

$$g_{ii}(c) > 0 \text{ für alle } i = 1,\ldots,r. \tag{3.3.7}$$

Lemma 3.3.1. *(Primitivität der Übergangsmatrix)*
Unter den in (3.3.6) und (3.3.7) getroffenen Annahmen ist die Übergangsmatrix $\boldsymbol{\Pi}(c) = \big(p_{ij}(c)\big)_{1\leq i,j\leq r}$ *primitiv, d.h. es existiert* $m \in \mathbb{N}$ *derart, daß* $\boldsymbol{\Pi}^m(c)$ *lauter positive Elemente besitzt.*

Beweis. Für $k \in \mathbb{N}$ bezeichne $\boldsymbol{\Pi}^k(c) = \big(p_{ij}^{(k)}(c)\big)_{1\leq i,j\leq r}$ die Matrix der k–Schritt Übergangswahrscheinlichkeiten. Nach Voraussetzung gilt $a_{i,j}(c) > 0$, so daß für alle $i \neq j$ Bedingung (3.3.6) auch für $p_{ij}(c)$ anstelle von $g_{ij}(c)$, also für die Elemente der Übergangsmatrix $\boldsymbol{\Pi}(c)$ gilt. Wählt man nun $m = \max\{k(i,j) \mid i,j = 1,\ldots,r\}$, $k(i,j)$ aus (3.3.6), so folgt für alle $i,j \in \{1,\ldots,r\}$, daß

$$p_{ij}^{(m)}(c) \geq p_{ij}^{(k(i,j))}(c) p_{jj}^{(m-k(i,j))}(c) \geq \Big(\prod_{t=1}^{k(i,j)-1} p_{l_t l_{t+1}(c)} \Big) p_{jj}^{m-k(i,j)}(c) > 0, \tag{3.3.8}$$

wobei $l_1,\ldots,l_{k(i,j)}$ die Indizes aus Bedingung (3.3.6) sind. Hierbei gilt $p_{jj}(c) > 0$, da wegen (3.3.7) $\sum_{\ell=1,\ell\neq j}^r g_{j\ell}(c) a_{j\ell}(c) \leq \sum_{\ell=1,\ell\neq j}^r g_{j\ell}(c) = 1 - g_{jj}(c) < 1$. Die zugehörige Markoff–Kette $\{X_n^{(c)}\}$ ist damit insgesamt irreduzibel.
Anschaulich besagt Gleichung (3.3.8), daß man mit positiver Wahrscheinlichkeit von i nach j in $k(i,j)$ Schritten kommt und mit positiver Wahrscheinlichkeit für die restlichen $m - k(i,j)$ Schritte in j verbleibt. Die Ungleichungen sind eine einfache Konsequenz der Chapman–Kolmogoroff–Gleichungen (3.1.14). ∎

Unter den Annahmen (3.3.6) und (3.3.7) folgt mit Lemma 3.3.1 für alle $c > 0$ aus Korollar 3.2.3 die Existenz einer Limesmatrix

$$\lim_{n\to\infty} \boldsymbol{\Pi}^n(c) = \boldsymbol{\Pi}_\infty(c) = \begin{pmatrix} \boldsymbol{u}(c) \\ \vdots \\ \boldsymbol{u}(c) \end{pmatrix}, \tag{3.3.9}$$

die lauter identische Zeilen $\boldsymbol{u}(c)$ besitzt. $\boldsymbol{u}(c) = \big(u_1(c), \ldots, u_r(c)\big)$ ist gleichzeitig die eindeutige, stationäre Verteilung der zugehörigen Markoff–Kette $\{X_n\}_{n\in\mathbb{N}}$, und es gilt $\lim_{n\to\infty} \boldsymbol{q}_0 \boldsymbol{\Pi}^n(c) = \boldsymbol{u}(c)$ für jede beliebige Anfangsverteilung $\boldsymbol{q}_0$. Der erste Grenzwert in (3.3.5) existiert also in der geeigneten Form.

Die folgenden $a_{ij}(c)$ bilden ein Beispiel für eine konkrete Wahl von Annahmewahrscheinlichkeiten.

$$a_{ij}(c) = \min\Big\{1, \exp\Big(-\frac{f(j) - f(i)}{c}\Big)\Big\}, \quad i, j \in S, \ c > 0, \tag{3.3.10}$$

wobei f die Kostenfunktion aus (3.3.1) ist.

Mit diesen $a_{ij}(c)$ wird vom Zustand i aus mit Wahrscheinlichkeit 1 der Zustand j akzeptiert, wenn $f(j) \le f(i)$ gilt, also

$$f(j) \le f(i) \quad \Longrightarrow \quad a_{ij}(c) = 1 \text{ für alle } i, j \in S, \ c > 0. \tag{3.3.11}$$

Ist aber $f(j) > f(i)$, so wird $j \in S$ noch mit positiver Wahrscheinlichkeit akzeptiert, die um so kleiner ist je größer die Differenz zwischen $f(j)$ und $f(i)$ lautet und je kleiner der Wert von c ist.

Man überprüft leicht, daß $a_{ij}(c)$ aus (3.3.10) die folgenden weiteren Eigenschaften besitzt.

$$\begin{aligned} f(j) > f(i) \quad &\Longrightarrow \quad 0 < a_{ij}(c) < 1 \text{ für alle } i, j \in S \\ &\text{sowie } \lim_{c\to\infty} a_{ij}(c) = 1 \text{ und } \lim_{c\downarrow 0} a_{ij}(c) = 0. \end{aligned} \tag{3.3.12}$$

Ferner gilt für alle $i, j, k \in S, \ c > 0$

$$f(i) \le f(j) \le f(k) \quad \Longrightarrow \quad a_{ik}(c) = a_{ij}(c) a_{jk}(c), \tag{3.3.13}$$

wie man anhand von (3.3.10) durch Fallunterscheidung überprüft.

Die folgende Bedingung fordert noch die Symmetrie des Generierungsprozesses:

$$g_{ij}(c) = g_{ji}(c) \text{ für alle } i, j \in S, \ c > 0. \tag{3.3.14}$$

Der folgende Satz lehrt, daß beliebige Annahmewahrscheinlichkeiten $a_{ij}(c)$, sofern sie nur (3.3.11), (3.3.12) und (3.3.13) erfüllen, die Konvergenz der stationären Verteilungen $\boldsymbol{u}(c) = \big(u_1(c), \ldots, u_r(c)\big)$ $(c \to \infty)$ gegen eine Grenzverteilung mit positiven Wahrscheinlichkeiten nur auf optimalen Punkten $i \in S_{\text{opt}}$ sicherstellen.

Satz 3.3.1. *(stationäre Verteilung)*
Für die Generierungswahrscheinlichkeiten $g_{ij}(c)$ gelte (3.3.6), (3.3.7) und (3.3.14), für die Annahmewahrscheinlichkeiten $a_{ij}(c)$ (3.3.11) und (3.3.13), $1 \le i, j \le r, c > 0$. Dann besitzt für alle $c > 0$ die Markoff–Kette $\{X_n^{(c)}\}_{n\in\mathbb{N}}$ mit Übergangsmatrix $\boldsymbol{\Pi}(c)$ aus (3.3.3) eine eindeutige stationäre Verteilung $\boldsymbol{u}(c) = \big(u_1(c), \ldots, u_r(c)\big)$ mit

$$u_i(c) = \frac{a_{i_0 i}(c)}{\sum_{j=1}^{r} a_{i_0 j}(c)}, \quad i = 1, \ldots, r, \tag{3.3.15}$$

wobei $i_0 \in S_{\text{opt}}$ eine Konfiguration mit minimalen Kosten ist. Gilt zusätzlich (3.3.12), so folgt

$$\lim_{c\downarrow 0} \boldsymbol{u}(c) = \boldsymbol{u} = (u_1, \ldots, u_r), \ \textit{mit}$$

$$u_i = \begin{cases} \big(\#(S_{\text{opt}})\big)^{-1}, & \textit{falls } i \in S_{\text{opt}} \\ 0, & \textit{sonst} \end{cases},$$

also eine Gleichverteilung auf der Menge der optimalen Zustände.

Beweis. Wie oben gezeigt, stellen (3.3.6) und (3.3.7) über Lemma 3.3.1 die Existenz und Eindeutigkeit einer stationären Verteilung sicher. Jede stationäre Verteilung $\boldsymbol{u}(c)$ ist linker Eigenvektor von $\boldsymbol{\Pi}(c)$ zum Eigenwert 1, also $\boldsymbol{u}(c) = \boldsymbol{u}(c)\boldsymbol{\Pi}(c)$. Wir weisen im folgenden nach, daß der stochastische Vektor $\boldsymbol{u}(c)$ aus (3.3.15) diesem Gleichungssystem genügt, also die eindeutige stationäre Verteilung repräsentiert.
Bezeichne $\alpha(c) = \left(\sum_{j=1}^{r} a_{i_0 j}(c)\right)^{-1}$ den Normierungsfaktor. Für alle $j \neq k \in \{1, \ldots, r\}$ gilt dann wegen (3.3.14)

$$u_j(c)p_{jk}(c) = \alpha(c)a_{i_0 j}(c)g_{jk}(c)a_{jk}(c) = \alpha(c)g_{kj}(c)a_{i_0 j}(c)a_{jk}(c).$$

Weiterhin folgt mit (3.3.11) und (3.3.13) in beiden Fällen $f(i_0) \leq f(j) \leq f(k)$ und $f(i_0) \leq f(k) \leq f(j)$ die Gleichheit $a_{i_0 j}(c)a_{jk}(c) = a_{i_0 k}(c)a_{kj}(c)$, also

$$u_j(c)p_{jk}(c) = \alpha(c)g_{kj}(c)a_{i_0 k}(c)a_{kj}(c) = u_k(c)p_{kj}(c). \tag{3.3.16}$$

Gleichung (3.3.16) gilt trivialerweise auch für $k = j$, so daß

$$\begin{aligned}\boldsymbol{u}(c)\boldsymbol{\Pi}(c) &= \Big(\sum_{j=1}^{r} u_j(c)p_{j1}(c), \ldots, \sum_{j=1}^{r} u_j(c)p_{jr}(c)\Big) \\ &= \Big(\sum_{j=1}^{r} u_1(c)p_{1j}(c), \ldots, \sum_{j=1}^{r} u_r(c)p_{rj}(c)\Big) = \big(u_1(c), \ldots, u_r(c)\big) = \boldsymbol{u}(c).\end{aligned}$$

Dies zeigt den ersten Teil der Behauptung.
Die Konvergenz von $u_i(c)$ gegen den angegebenen Limes folgt sofort aus (3.3.11): " $a_{i_0 i}(c) = 1$, falls $i \in S_{\text{opt}}$ für alle $c > 0$ " und (3.3.12): " $\lim_{c \downarrow 0} a_{i_0 j}(c) = 0$, falls $j \notin S_{\text{opt}}$ ". ∎

Bei Verwendung der konkreten $a_{ij}(c)$ aus (3.3.10) lassen sich die $u_i(c)$ aus (3.3.15) weiter spezifizieren zu

$$u_i(c) = \frac{\exp\{(f(i_0) - f(i))/c\}}{\sum_{j=1}^{r} \exp\{(f(i_o) - f(j))/c\}}. \tag{3.3.17}$$

Der Algorithmus wird nun durch Simulation der Markoff–Ketten $\{X_n^{(c)}\}_{n \in \mathbb{N}}$ für fallenden Parameter c unter Ausnutzung der Konstruktion (3.3.4) implementiert.

Wir verwenden im folgenden ausschließlich die Annahmewahrscheinlichkeiten $a_{ij}(c)$ aus (3.3.10), die für festes $c > 0$ zu den stationären Verteilungen $u_i(c) = \exp\{(f(i_0) - f(i))/c\} / \sum_{j=1}^{r} \exp\{(f(i_o) - f(j))/c\}$ aus (3.3.17) führen. $u_i(c)$ ist dann für alle $i = 1, \ldots, r$ als Funktion von c stetig, so daß kleine Änderungen von c auch nur kleine Änderungen in der Verteilung $(u_1(c), \ldots, u_r(c))$ bewirken. Simuliert man jetzt für festes $c_1 > 0$ eine große Anzahl von Realisationen der Markoff–Kette $\{X_n^{(c_1)}\}_{n \in \mathbb{N}}$, so liegt die Verteilung des zuletzt erhaltenen Zustands

$\tilde{\boldsymbol{u}}(c_1)$ nahe an der stationären Verteilung $\boldsymbol{u}(c_1)$. Diese wiederum liegt für ein geringfügig kleineres $c_2 < c_1$ in der Nähe der stationären Verteilung $\boldsymbol{u}(c_2)$ der Markoff–Kette $\{X_n^{(c_2)}\}_{n\in\mathbf{N}}$ und läßt hoffen, daß die stationäre Verteilung $\boldsymbol{u}(c_2)$ nach einer genügenden Anzahl von Realisationen der Markoff–Kette $\{X_n^{(c_2)}\}_{n\in\mathbf{N}}$ mit Startverteilung $\boldsymbol{u}(c_1)$ gut approximiert wird.

Für die praktische Implementierung bedeutet dies: Wähle einen Startparameter c_0 und eine Startzustand i_0 (dies entspricht einer Einpunktverteilung in i_0 als Startverteilung). Generiere entsprechend den Wahrscheinlichkeiten $g_{ij}(c_0)$ sukzessive Nachbarzustände und entscheide mit Hilfe der Annahmewahrscheinlichkeiten $a_{ij}(c_0)$ über deren Annahme. Führe dieses Verfahren ℓ_0–mal durch; der zuletzt erhaltene Zustand sei i_{c_0}. Verkleinere den Wert von c_0 zu c_1, benutze i_{c_0} als Startzustand und wiederhole obiges Verfahren mit dem neuen Wert des Kontrollparameters c_1. Der Endzustand nach ℓ_1 Iterationen sei i_{c_1}. Mit diesem wird für $c_2 < c_1$ ein neuer Zyklus mit i_{c_1} als Startzustand begonnen, und anschließend werden weitere solcher Zyklen aneinandergehängt.

Um das theoretische Konvergenzresultat bei der praktischen Simulation umzusetzen, muß darauf geachtet werden, daß die Generierungswahrscheinlichkeiten $g_{ij}(c)$ für alle $c > 0$ symmetrisch sind (vgl. (3.3.14)) und den Bedingungen (3.3.6) und (3.3.7) genügen. $g_{ij}(c)$ wird in der Regel unabhängig von c festgesetzt.

Entscheidend für das Funktionieren des Simulated Annealing–Algorithmus ist die a–priori Festlegung der folgenden Größen:

1) des Startwertes c_0,
2) der Verminderungsvorschrift $c_{k+1} = v(c_k)$, $k = 0, 1, \ldots$,
3) der Anzahl ℓ_k von Simulationen der Markoff–Kette $\{X_n^{(c_k)}\}_{n\in\mathbf{N}}$,
4) eines Stopkriteriums.

Diese Größen lassen sich nach Implementation des Algorithmus häufig interaktiv und experimentell zum Erzielen guter Ergebnisse festsetzen. Dieses empirische Vorgehen ist bei vielen Beispielen empfehlenswert. Dennoch finden sich in der Literatur einige globale Empfehlungen, die bei polynomialen Laufzeiten des Algorithmus zu fast optimalen Ergebnissen führen (vgl. Aarts & van Laarhoven (1985)). Auf die technischen Details, die teilweise tiefliegende Ergebnisse aus der Theorie der Markoff–Kette benutzen, wollen wir hier verzichten.

1) Mit $\beta \approx 10$ ist $c_0 = \beta \cdot \max\{f(j) - f(i) \mid i, j = 1, \ldots, r\}$ ein heuristischer Startwert. Steht obiges Maximum nicht zur Verfügung, so genügt auch eine grobe Abschätzung.

2) Wähle $c_{k+1} = c_k\Big(1 + \frac{\ln(1+\varepsilon)c_k}{3\sigma(c_k)}\Big)^{-1}$, $k \in \mathbf{N}_0$, wobei $\sigma(c) = \sum_{i=1}^r u_i(c)\big(f(i)\big)^2 - \big(\sum_{i=1}^r u_i(c) f(i)\big)^2$ die Varianz der Kostenfunktion unter der stationären Verteilung $\boldsymbol{u}(c)$ bezeichnet. Die Konstante $\varepsilon > 0$ stammt aus der Bedingung, daß die stationären Verteilungen $\boldsymbol{u}(c_k)$ und $\boldsymbol{u}(c_{k+1})$ nahe beieinanderliegen, genauer $\frac{1}{1+\varepsilon} < \frac{u_i(c_k)}{u_i(c_{k+1})} < 1 + \varepsilon$ für alle $i = 1, \ldots, r$. Sie wird vom Benutzer als positive Zahl nahe bei Null festgelegt.
Bei typischen Anwendungen ist r jedoch so groß, daß $\sigma(c)$ nicht explizit berechnet werden kann. Für "gutartige" Kostenfunktionen f ist aber $\hat{\sigma}(c) = \frac{1}{t}\sum_{j=1}^t f^2(i_j) - \big(\frac{1}{t}\sum_{j=1}^t f(i_j)\big)^2$ ein Schätzer für $\sigma(c)$, wobei $i_1, \ldots, i_t$ die Zustände sind, die bei der Simulation der Markoff–Kette $\{X_n(c)\}_{n\in\mathbf{N}}$ erzeugt

wurden und t die Anzahl der Simulationen ist. $\sigma(c_k)$ wird dann in obiger Formel durch die Approximation $\hat{\sigma}(c_k)$ ersetzt.
Eine sehr einfache Möglichkeit ist, den Kontrollparameter geometrisch zu verkleinern, d.h. $c_{k+1} = \alpha c_k$, $k \in \mathbb{N}_0$, wobei α–Werte zwischen 0.8 und 0.99 in verschiedenen Beispielen gute Erfolge gebracht haben.

3) Wähle $\ell_k = \max\{\#(S_i) \mid i \in S\}$, unabhängig von k, wobei S_i die Menge der Nachbarzustände von $i \in S$ ist.

4) Der Algorithmus wird abgebrochen, wenn die Differenz der Zielfunktionswerte nach Ende des $(k+1)$–ten bzw. des k–ten Simulationszyklus $\Delta_k = f(i_{c_k}) - f(i_{c_{k+1}})$ keine lohnenswerte Verkleinerung mehr bringt, etwa wenn $\Delta_k / f(i_0) < \varepsilon$ für ein vorgegebenes $\varepsilon > 0$. Um zufällige kleine Schwankungen in frühem Stadium des Algorithmus auszugleichen und einen ungewollten frühen Abbruch zu vermeiden, kann man in Δ_k statt der Kosten $f(c_{i_k})$ gleitende Mittel $\bar{f}_k = \frac{1}{2t+1} \sum_{j=k-t}^{k+t} f(c_{i_j})$, $t \in \mathbb{N}$, verwenden. Eine weitere Möglichkeit ist, den Verlauf der Kosten online graphisch darzustellen und den Abbruch per Hand vorzunehmen, wenn die Reduzierung der Kosten über mehrere Zyklen nicht mehr lohnt.

Der Algorithmus in PASCAL–Notation hat folgende Gestalt.

```
PROCEDURE simulated annealing;
  VAR c,delta:  REAL;
    k,l:  INTEGER;
    a,b:  state;
BEGIN
  init(l,c,a);
  REPEAT BEGIN
    FOR k := 1 TO l DO BEGIN
      generate(a,b);
      delta := f(b) - f(a);
      IF delta <= 0 THEN a := b
        ELSE IF exp(-delta/c) > random THEN a := b
      END;
    diminish(c)
  END;
  UNTIL stop(delta)
END;                                                    (3.3.18)
```

Er arbeitet mit dem Datentyp `state`, der vom konkreten Problem abhängt und die möglichen Zustände der Menge S charakterisiert. Die Prozeduren `init`, `diminish` und die Bool'sche Funktion `stop` sind noch zu spezifizieren. `init` initialisiert den Wert c_0, den Startzustand a und die Länge der homogenen Markoff–Ketten ℓ. `generate(a,b)` erzeugt ausgehend von a gemäß der Wahrscheinlichkeiten $g_{ab}(c) = g_{ab}$ einen neuen Zustand b. `diminish` enthält die Vorschrift zur Verminderung des Kontrollwerts c, und der Wert `stop = TRUE` bewirkt den Abbruch des Algorithmus. Der Zufallsgenerator `random` liefert eine auf $[0,1]$ gleichverteilte Zufallszahl.

Simulated Annealing ist bei der Lösung einer Vielzahl von kombinatorischen Optimierungsproblemen getestet worden. Besonders hervorzuheben sind hier die

Anwendungen beim VLSI–design (VLSI = very large scale integration), bei der Bildverarbeitung und bei der digitalen Signalkodierung. Ein ausführlicher Literaturüberblick und detailierte Studien finden sich in den Arbeiten von Aarts & van Laarhoven (1985, 1987). Wir wollen hier ein einfach formulierbares, diskretes Optimierungsproblem aus dem Bereich der statistischen Versuchsplanung behandeln, mit dem der Leser nach kurzer Programmierarbeit Erfahrung mit Simulated Annealing gewinnen kann.

Beispiel 3.3.1. (Simulated Annealing bei Versuchsplanung)
Bei einem statistischen Experiment werden p Getreidesorten auf einer Auswahl von Parzellen aus q Feldern verschiedener Bodengüte ausgesät und die Erträge nach Ablauf einer Wachstumsperiode gemessen. Aus diesen mit Zufallsfehlern behafteten Messungen soll die Ertragsfähigkeit der einzelnen Getreidesorten ermittelt werden. Man kann die Effizienz einer Schätzung des erwarteten Ertrags der einzelnen Sorten verbessern, wenn der Einflußfaktor Bodengüte möglichst ausgemittelt wird. Werden nun genau $m \leq pq$ Versuche gemacht, so müssen m Aussaaten auf die q Bodenqualitäten verteilt werden. Die Menge aller möglichen solcher Konstellationen wird durch die Menge der zugehörigen Inzidenzmatrizen

$$S = \mathfrak{V}(p,q,m) = \Big\{ N = (n_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}} \mid n_{ij} \in \{0,1\}, \sum_{1 \leq i \leq p,\ 1 \leq j \leq q} n_{ij} = m \Big\}$$

beschrieben, wobei "$n_{ij} = 1$" bedeutet, Getreidesorte i auf einer Parzelle der Bodengüte j auszusäen. Zur Maximierung der Effizienz von Schätzungen für die Getreideeffekte muß das Problem

$$\max_{N \in \mathfrak{V}(p,q,m)} \left\{ \det\big(K(P - NQ^{-1}N^{tr})K^{tr}\big) \right\} \tag{3.3.19}$$

gelöst werden, wobei $P = \mathrm{diag}(n_{1\cdot}, \ldots, n_{p\cdot})$ die Diagonalmatrix der Zeilensummen von N und $Q = \mathrm{diag}(n_{\cdot 1}, \ldots, n_{\cdot q})$ die Diagonalmatrix der Spaltensummen von N ist. K ist eine $(p-1)\times p$–Matrix, deren Zeilen ein beliebiges Orthonormalsystem auf $(1,\ldots,1)^{tr} \in \mathbf{R}^p$ bilden. Falls Q nicht invertierbar ist, wird der Zielfunktionswert in (3.3.19) zu Null gesetzt. (3.3.19) läßt sich leicht in ein Minimierungsproblem vom Typ (3.3.1) umschreiben, indem man zum Negativen der Zielfunktion übergeht.
Die Mächtigkeit des Konfigurationsraums S beträgt $r = \binom{pq}{m}$. Ein realistisches Beispiel für die Parameterwerte ist etwa $p = 5$, $q = 10$ und $m = 20$, was zu $r \approx 4.713 \cdot 10^{13}$ führt.
Jeder Versuchsplan läßt sich mit Hilfe des Datentyps

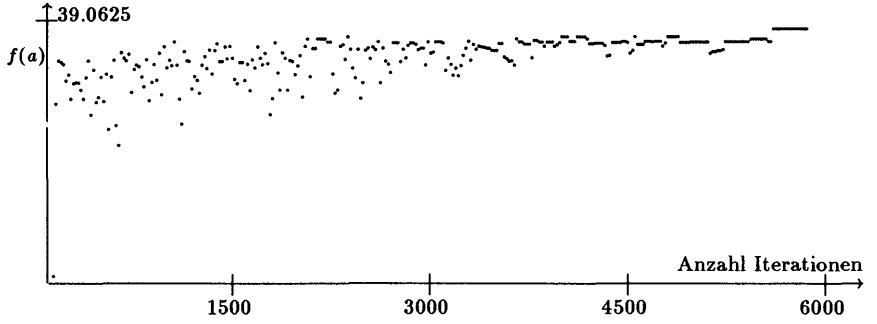
```
state = ARRAY[1..p*q] OF INTEGER
```

beschreiben, indem die Zeilen der Inzidenzmatrix N zu einem eindimensionalen Feld der Länge $p \cdot q$ aneinandergefügt werden.

`generate(a,b)` in (3.3.18) kann etwa folgendermaßen realisiert werden.

```
PROCEDURE generate(a:  state; VAR b:  state);
  VAR rn,j,s:  INTEGER;
BEGIN
  b := a;
  rn:=random(m); j:=0; s:=0;
  WHILE s < rn DO BEGIN
    j := j + 1; s := s + b[j]
    END;
  rn := random(p*q);
  IF b[rn] = 0 THEN BEGIN
    b[rn] := 1; b[j] := 0
    END;
END;
```

Hierbei wird zunächst eines der mit "1" besetzten Elemente zufällig ausgesucht, anschließend zufällig ein beliebiges Element des gesamten Feldes, und falls dieses mit "0" besetzt ist, wird der Wert der Elemente vertauscht. Mit diesem Verfahren ist die Wahrscheinlichkeit, im gleichen Zustand zu verbleiben, positiv (vgl. (3.3.7)), und auch die Symmetriebedingung (3.3.14) ist sichergestellt.
Bei der Programmierung der prozeduren `init`, `diminish` und `stop` kann man den obigen Empfehlungen 1) bis 4) folgen; zur Berechnung der Zielfunktionswerte `f(a)` in (3.3.18) empfiehlt sich die Verwendung von Bibliotheksroutinen.
Mit Hilfe einer graphischen Darstellung des Verlaufs der Zielfunktionswerte kann nun das Verhalten des Algorithmus für verschiedene Parameterkonstellationen von c und ℓ studiert werden. Ein typischer Verlauf ist in der folgenden Graphik für das Beispiel $p = 5,\ q = 10,\ m = 20$ dargestellt.



Die Länge der homogenen Teilketten beträgt in diesem Beispiel $\ell = 20$. Die Punkte geben die Größe des Zielfunktionswerts bei jeder 20-ten Iteration, also bei jeder Veränderung des Kontrollparameters c, an. Dieser wird geometrisch verkleinert mit $c_{k+1} = 0.99 \cdot c_k$, $k \in \mathbb{N}_0$, wobei der Startwert $c_0 = 5$ beträgt. 39.0625 ist der optimale Zielfunktionswert, wie sich in diesem Beispiel aus der Theorie der Versuchsplanung ergibt. Er wird in dem hier dargestellten Versuchslauf durch Simulated Annealing nach 5360 Iterationen erreicht. Der zugehörige Zustand bleibt dann bis zur 20000-ten Iteration aktuell. ∎

Wie oben beschrieben —und in Programm (3.3.18) realisiert—, werden bei der Durchführung des Simulated Annealing–Algorithmus "Stücke" $\{X_n^{(c_k)}\}_{n=1}^{\ell_k}$ der Länge ℓ_k von jeweils homogenen Markoff–Ketten aneinandergehängt. Bezeichne

$$\{X_n^*\}_{n\in\mathbb{N}} = \{X_1^{(c_0)},\ldots,X_{\ell_0}^{(c_0)},X_1^{(c_1)},\ldots,X_{\ell_1}^{(c_1)},\ \ldots\ ,X_1^{(c_k)},\ldots,X_{\ell_k}^{(c_k)},\ \ldots\}$$

die solchermaßen konstruierte Markoff–Kette insgesamt. $\{X_n^*\}_{n\in\mathbb{N}}$ ist nicht mehr homogen, da sich die Übergangswahrscheinlichkeiten mit den Werten von c_k ändern.

Satz 3.3.1 macht nun eine Aussage über stationäre Verteilungen der "Stücke", wenn diese unendlich lang werden, und eine Konvergenzaussage über die stationären Verteilungen, wenn der Kontrollparameter c beliebig klein wird. Hieraus läßt sich jedoch keine Aussage über die stationäre Verteilung der gesamten Folge $\{X_n^*\}_{n\in\mathbb{N}}$ gewinnen, die ja nur aus homogenen Ketten endlicher Länge zusammengesetzt ist. Man darf lediglich hoffen, daß für große Werte von ℓ_k und geschickte Wahl der c_k eine stationäre Verteilung nahe bei der Gleichverteilung auf den optimalen Zuständen erreicht wird, was in der praktischen Anwendung des Algorithmus ja auch bestätigt wird.

Wünschenswert bleiben also Aussagen über die Existenz einer Limesverteilung $\boldsymbol{u}^* = (u_1^*,\ldots,u_r^*)$ der inhomogenen Markoff–Kette $\{X_n^*\}_{n\in\mathbb{N}}$ mit

$$\lim_{n\to\infty} P(X_n^* = j) = u_j^*,\ j = 1,\ldots,r \quad \text{und} \quad \sum_{j\in S_{\text{opt}}} u_j^* = 1, \tag{3.3.20}$$

die für beliebige Startverteilungen erreicht wird.

Ein naheliegendes Verfahren ist, nach jeder Simulation den Kontrollparameter c zu verkleinern, d.h. $\ell_k = 1$ für alle $k \in \mathbb{N}$ zu wählen. In diesem Fall läßt sich unter bestimmten Voraussetzungen, die in geeigneter Weise die Kostenfunktion, die Kontrollwerte c_k und die Übergangswahrscheinlichkeiten $p_{ij}(c_k)$ verknüpfen, ein Konvergenzresultat vom Typ (3.3.20) erzielen. Die hierfür benötigten Hilfsmittel sind tiefliegende Ergebnisse aus der Theorie inhomogener Markoff–Ketten, deren Bereitstellung den Rahmen des Buchs sprengen würden. Der interessiert Leser sei auf die Übersichtsarbeit von Aarts & van Laarhoven (1987) verwiesen.

3.4. Markoff- und Punktprozesse

In den beiden vorigen Abschnitten wurden Markoff-Modelle betrachtet, die von einem diskreten Zeitparameter abhingen, d.h. also *Folgen* von Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}_0}$, die der Markoff-Eigenschaft (3.2.1) genügen. Will man entsprechende Abhängigkeitsstrukturen in kontiniuierlicher Zeit modellieren, gelangt man zum Begriff des Markoff-Prozesses, den wir hier allerdings nur für den Fall eines abzählbaren Zustandsraums $\mathcal{S}$ präzisieren wollen.

Definition 3.4.1. *(Markoff-Prozeß)*
Es sei $\{X_t\}_{t\geq 0}$ eine Familie von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in einer abzählbaren Menge $\mathcal{S}$. $\{X_t\}_{t\geq 0}$ heißt Markoff-Prozeß, wenn für jede aufsteigende Folge $\{t_n\}_{n\in\mathbf{N}_0}\subset \mathbf{R}^+$ die zugehörige Folge $\{X_{t_n}\}_{n\in\mathbf{N}_0}$ eine Markoff-Kette bildet. Der Markoff-Prozeß heißt (zeitlich) homogen, wenn für alle solche Folgen $\{t_n\}_{n\in\mathbf{N}_0}$ mit $t_{n+1} - t_n = const$, $n \in \mathbf{N}_0$, die Markoff-Kette $\{X_{t_n}\}_{n\in\mathbf{N}_0}$ homogen im Sinne von Definition 3.2.1 ist.

Solche Markoff-Prozesse sind z.B. Grundlage für die Bedienungs- und Warteschlangentheorie, wo etwa $\mathcal{S} = \mathbf{N}_0$ ist; die Zufallsvariable X_t, $t \geq 0$, gibt dann z.B. an, wieviele Kunden (Programme) sich zur Zeit t im System befinden oder wieviele Kunden (Programme) zur Zeit t auf ihre Bedienung (Bearbeitung) warten.

Einer der einfachsten und zugleich wichtigsten Markoff-Prozesse dieser Art ist der sogenannte *Poisson*-Prozeß, auf den wir bereits früher — in (2.1.90) und Beispiel 3.2.2 — kurz eingegangen sind. Seine charakteristischen Eigenschaften sollen im folgenden ausführlicher behandelt werden.

Definition 3.4.2. *(homogener Poisson-Prozeß)*
$\{X_n\}_{n\in\mathbf{N}}$ sei eine Folge unabhängiger, je $\mathcal{E}(\lambda)$-verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit $\lambda > 0$ und $T_0 \equiv 0$, $T_n = \sum_{i=1}^{n} X_i$, $n \in \mathbf{N}$. Der durch

$$N_t = \#\{n \in \mathbf{N} \mid T_n \leq t\}, \quad t \geq 0, \tag{3.4.1}$$

definierte Prozeß heißt Poisson-Prozeß mit Parameter λ. Die Folge $\{T_n\}_{n\in\mathbf{N}_0}$ heißt auch Folge der Ankunftszeiten des Prozesses.

Man beachte, daß aufgrund von (3.4.1) $N_0 = 0$ nur P-fast sicher gilt; zur Umgehung einiger technischer Schwierigkeiten wollen wir daher annehmen, daß im folgenden die Zufallsvariablen $\{X_n\}_{n\in\mathbf{N}}$ so auf $(\Omega, \mathcal{A}, P)$ definiert sind, daß stets $X_n > 0$, $n \in \mathbf{N}$, gilt. Dies bedeutet natürlich keine Einschränkung der Allgemeinheit, da (meßbare) Abänderungen auf P-Nullmengen für die Verteilung der Zufallsvariablen bedeutungslos sind.

Zum Nachweis der Markoff-Eigenschaft des Poisson-Prozesses benötigen wir eine charakteristische Eigenschaft, die als *Unabhängigkeit der Zuwächse* des Prozesses bezeichnet wird, d.h. die Tatsache, daß für alle aufsteigenden Folgen von Zeitpunkten $\{t_n\}_{n\in\mathbf{N}_0}\subset \mathbf{R}^+$ die Zuwächse $\{N_{t_{n+1}} - N_{t_n}\}_{n\in\mathbf{N}_0}$ stochastisch unabhängig sind. Zur Vereinfachung der Schreibweise wollen wir dabei wie bereits in Beispiel 3.2.2 die Notation

$$N_{(s,t]} = N_t - N_s, \quad 0 \leq s \leq t, \tag{3.4.2}$$

verwenden; insbesondere ist damit gemäß der obigen Vereinbarung $N_\emptyset = 0$.

Satz 3.4.1. *(unabhängige Zuwächse von Poisson–Prozessen)*
Es sei $\{N_t\}_{t\geq 0}$ ein Poisson–Prozeß im Sinne von Definition 3.4.2. Dann besitzt $\{N_t\}_{t\geq 0}$ unabhängige, Poisson–verteilte Zuwächse, d.h. für jede aufsteigende Folge von Zeitpunkten $\{t_n\}_{n\in\mathbf{N}_0}\subset \mathbf{R}^+$ gilt: die Zuwächse $\{N_{(t_n,t_{n+1}]}\}_{n\in\mathbf{N}_0}$ sind stochastisch unabhängig mit

$$P^{N_{(s,t]}} = \mathfrak{P}\big((t-s)\lambda\big), \quad 0\leq s<t. \tag{3.4.3}$$

Beweis. Zum Beweis benötigen wir die Beziehung

$$\int_{a<x_1<\ldots<x_n<b}\cdots\int dx_1\ldots dx_n = \frac{(b-a)^n}{n!}, \quad a<b,\ n\in\mathbf{N}. \tag{3.4.4}$$

Diese ergibt sich leicht z.B. mit vollständiger Induktion: für $n=1$ ist dies trivial; gilt (3.4.4) für $n\in\mathbf{N}$, so erhält man

$$\int_{a<x_1<\ldots<x_n<x_{n+1}<b}\cdots\int dx_1\ldots dx_n\,dx_{n+1} = \int_a^b \frac{(x_{n+1}-a)^n}{n!}\,dx_{n+1} = \frac{(b-a)^{n+1}}{(n+1)!}$$

für $a<b$, d.h. (3.4.4) gilt dann auch für $n+1$. Dabei können natürlich die Relationszeichen "$<$" wahlweise auch durch "$\leq$" ersetzt werden.
Sei nun für $m\in\mathbf{N}$ $0=t_0<t_1<\ldots<t_m$ und $k_1,\ldots,k_m\in\mathbf{N}$. Zur Abkürzung wählen wir $n_i=\sum_{j=1}^{i}k_j$, $i\in\mathbf{N}$ sowie die Mengen

$$B_1 = \big\{(x_1,\ldots,x_{n_1})\in\mathbf{R}^{k_1}\mid 0<x_1<\ldots<x_{n_1}\leq t_1\big\}$$
$$B_i = \big\{(x_{n_{i-1}+1},\ldots,x_{n_i})\in\mathbf{R}^{k_i}\mid t_{i-1}<x_{n_{i-1}+1}<\ldots\leq x_{n_i}<t_i\big\},\ 2\leq i\leq m$$
$$B_{m+1} = (t_m,\infty).$$

Nach Definition von $\{N_t\mid t\geq 0\}$ (vgl. auch (2.1.91)) erhält man

$$\bigcap_{i=1}^{m}\{N_{(t_{i-1},t_i]}=k_i\}$$
$$= \big\{(T_1,\ldots,T_{n_1})\in B_1\big\}\cap\bigcap_{i=2}^{m}\big\{(T_{n_{i-1}+1},\ldots,T_{n_i}\in B_i)\big\}\cap\{T_{m+1}\in B_{m+1}\}$$

und damit unter Verwendung von (3.1.22)

$$P\Big(\bigcap_{i=1}^{m}\{N_{(t_{i-1},t_i]}=k_i\}\Big) = \int_{B_{m+1}}\cdots\int_{B_1}\lambda^{n_m+1}e^{-\lambda y}\,dx_1\ldots dx_{n_m}\,dy$$
$$= \int_{t_m}^{\infty}\prod_{i=1}^{m}\frac{(t_i-t_{i-1})^{k_i}}{k_i!}\lambda^{n_m+1}e^{-\lambda y}\,dy = \prod_{i=1}^{m}\frac{(t_i-t_{i-1})^{k_i}}{k_i!}\lambda^{n_m}e^{-\lambda t_m}$$
$$= \prod_{i=1}^{m}\frac{\big(\lambda(t_i-t_{i-1})\big)^{k_i}}{k_i!}e^{-\lambda(t_i-t_{i-1})} = \bigotimes_{i=1}^{m}\mathfrak{P}\big(\lambda(t_i-t_{i-1})\big)\big(\{(k_1,\ldots,k_m)\}\big).$$

Entsprechend argumentiert man für den Fall, daß gewisse der k_i Null sind; es entfallen dann die Integrationen über die (in diesem Fall nicht definierten) Mengen B_i.
Damit ergibt sich aber unmittelbar die Behauptung. ∎

Nunmehr können wir die Markoff-Eigenschaft von Poisson-Prozessen leicht anhand des Rekursions-Lemmas 3.2.2 nachweisen.

Satz 3.4.2. *(Markoff-Eigenschaft von Poisson-Prozessen)*
Es sei $\{N_t\}_{t\geq 0}$ ein Poisson-Prozeß im Sinne von Definition 3.4.2. Dann ist $\{N_t\}_{t\geq 0}$ ein homogener Markoff-Prozeß mit

$$P(N_t = n \mid N_s = k) = P(N_{(s,t]} = n-k) = \frac{(\lambda(t-s))^{n-k}}{(n-k)!} e^{-\lambda(t-s)}, \qquad (3.4.5)$$
$$0 \leq s < t,\ k,n \in \mathbb{N}_0,\ k \leq n.$$

Beweis. Ist $\{t_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}^+$ eine aufsteigende Folge von Zeitpunkten mit o.B.d.A. $t_0 = 0$, so gilt

$$N_{t_n} = N_{t_{n-1}} + D_n, \quad D_n := N_{(t_{n-1},t_n]},\ n \in \mathbb{N},$$

wobei $N_{t_0} = 0$ und $\{D_n\}_{n\in\mathbb{N}}$ nach Satz 3.4.1 eine unabhängige Folge von Zufallsvariablen ist. Gemäß dem Rekursions-Lemma 3.2.2 ist damit aber $\{N_{t_n}\}_{n\in\mathbb{N}_0}$ eine Markoff-Kette; insbesondere sind die D_n, $n \in \mathbb{N}$, identisch Poisson-verteilt, die Markoff-Kette ist also sogar homogen.
Beziehung (3.4.5) folgt unmittelbar aus dem Ersetzungslemma 3.1.4, Beziehung (3.1.16), wegen

$$P(N_t = n \mid N_s = k) = P(N_s + N_{(s,t]} = n \mid N_s = k) = P(N_{(s,t]} = n-k)$$
$$= \frac{(\lambda(t-s))^{n-k}}{(n-k)!} e^{-\lambda(t-s)}, \quad 0 \leq s < t,\ k,n \in \mathbb{N}_0,\ k \leq n,$$

nach (3.4.3). ∎

Das folgende Lemma ermöglicht eine andere Darstellung von Poisson-Prozessen, die sich insbesondere im Hinblick auf den Punktprozeßzugang als nützlich erweisen wird.

Lemma 3.4.1. *(Ordnungsstatistiken)*
Es seien $X_1, \ldots, X_n$, $n \in \mathbb{N}$, stochastisch unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit einer Dichte f. Bezeichnet $\boldsymbol{Y} = (X_{(1)}, \ldots, X_{(n)})$ den Vektor der nach aufsteigender Größe geordneten Komponenten von $(X_1, \ldots, X_n)$ (sog. Ordnungsstatistik zu $(X_1, \ldots, X_n)$), so besitzt $\boldsymbol{Y}$ eine Dichte $f_{\boldsymbol{Y}}$ der Form

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = \begin{cases} n! \prod_{i=1}^n f(y_i) & \text{falls } y_1 \leq \ldots \leq y_n \\ 0 & \text{sonst.} \end{cases} \qquad (3.4.6)$$

Beweis. Es bezeichne $\Sigma_n = Perm_n^n(\{1,\dots,n\}; o.W.)$ die Menge der Permutationen der Menge $\{1,\dots,n\}$, $n \in \mathbb{N}$. Definiert man die Zufallsvariable S vermöge

$$S = \begin{cases} \sigma \in \Sigma_n & \text{für } X_{\sigma(1)} < \dots < X_{\sigma(n)} \\ (1\ 2\ \dots\ n) & \text{sonst,} \end{cases}$$

so gilt $\boldsymbol{Y} = (X_{S(1)},\dots,X_{S(n)})$ P–fast sicher, da wegen der Stetigkeit der Verteilung der Zufallsvariablen $X_1,\dots,X_n$ Bindungen nur mit Wahrscheinlichkeit 0 auftreten, d.h. es ist — unter Verwendung von (3.1.23) und (3.1.26) —

$$P(X_i = X_j) = \int P(X_i = x)\, dP^{X_j}(x) = 0, \quad 1 \le i,j \le n,\ i \ne j.$$

Mit der Abkürzung $\mathcal{K}_n = \{(x_1,\dots,x_n) \mid x_1 \le \dots \le x_n\}$ ergibt sich nun für $A \in \mathcal{B}^n$:

$$\begin{aligned} P\big((X_{(1)},\dots,X_{(n)}) \in A\big) &= P\Big(\bigcup_{\sigma\in\Sigma_n} \{S=\sigma\} \cap \{(X_{\sigma(1)},\dots,X_{\sigma(n)}) \in A\}\Big) \\ &= \sum_{\sigma\in\Sigma_n} P\big((X_{\sigma(1)},\dots,X_{\sigma(n)}) \in \mathcal{K}_n \cap A\big) \\ &= n!\, P\big((X_1,\dots,X_n) \in \mathcal{K}_n \cap A\big) \\ &= n! \int_A \cdots \int \mathbb{1}_{\mathcal{K}_n}(y_1,\dots,y_n) \prod_{i=1}^n f(y_i)\, dy_1 \dots dy_n, \end{aligned}$$

also die Behauptung. ∎

Insbesondere besitzen die Ankunftszeiten $\{T_n\}_{n\in\mathbb{N}}$ eines Poisson–Prozesses mit Parameter $\lambda > 0$ die durch (3.1.21) gegebene bedingte Dichte

$$f_{(T_1,\dots,T_n)}(y_1,\dots,y_n \mid T_{n+1} = y) = \begin{cases} \dfrac{n!}{y^n} & \text{für } 0 < y_1 \le \dots \le y_n \le y \\ 0 & \text{sonst,} \end{cases} \qquad (n \in \mathbb{N})$$

d.h. $(T_1,\dots,T_n)$ verhält sich unter der Bedingung $T_{n+1} = y$, $y > 0$, wie die Ordnungsstatistik von n unabhängigen, je $\mathcal{R}\big((0,y]\big)$–verteilten Zufallsvariablen. Der folgende Satz zeigt, daß diese Eigenschaft erhalten bleibt, wenn die Bedingung $T_{n+1} = y$, $y > 0$, durch die Bedingung $N_y = n$ ersetzt wird (d.h. im Zeitintervall $(0,y]$ werden genau n Ankünfte beobachtet).

Satz 3.4.3. *(bedingte Unabhängigkeit von Ankunftszeiten)*
Es sei $\{N_t\}_{t\ge 0}$ ein Poisson–Prozeß mit Parameter λ, $\{T_n\}_{n\in\mathbb{N}_0}$ bezeichne die Folge der Ankunftszeiten. Dann gilt für jede aufsteigende Folge $\{t_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}^+$ mit $t_0 = 0$ und jede Folge $\{k_n\}_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$ mit $k_0 = 0$ und $s_j = \sum_{i=0}^j k_i$, $j \in \mathbb{N}_0$:

$$\begin{aligned} &f_{(T_1,\dots,T_{s_n})}(y_1,\dots,y_{s_n} \mid N_{(t_{i-1},t_i]} = k_i,\ 1 \le i \le n) \\ &\quad = \prod_{i=1}^n f_{(T_{s_{i-1}+1},\dots,T_{s_i})}(y_{s_{i-1}+1},\dots,y_{s_i} \mid N_{(t_{i-1},t_i]} = k_i) \\ &\quad = \prod_{i=1}^n \frac{k_i!}{(t_i - t_{i-1})^{k_i}}, \quad t_{i-1} < y_{k_{i-1}+1} < \dots < y_{k_i} \le t_i,\ 1 \le i \le n, \end{aligned} \tag{3.4.7}$$

d.h. die Ankunftszeiten $T_1, \dots, T_{s_n}$ verhalten sich unter der Bedingung, daß in den Zeitintervallen $(t_{i-1}, t_i]$, $1 \le i \le n$, jeweils genau k_i Ankünfte eintreten, wie die geordneten Werte unabhängiger, jeweils über $(t_{i-1}, t_i]$ stetig gleichverteilter Zufallsvariablen.

Beweis. Zur Abkürzung setzen wir

$$\mathcal{K}_n = \{(y_1, \dots, y_{s_n}) \mid t_{i-1} < y_{k_{i-1}+1} < \dots < y_{k_i} \le t_i,\ 1 \le i \le n\}.$$

Für beliebige Mengen $A_1, \dots, A_n \in \mathcal{B}^1$, $1 \le i \le n$, gilt dann unter Heranziehung von Satz 3.1.4

$$P\Big(\bigcap_{i=1}^{n} \{(T_{s_{i-1}+1}, \dots, T_{s_i}) \in A_i \mid N_{(t_{i-1},t_i]} = k_i,\ 1 \le i \le n\Big)$$

$$= P\Big(\bigcap_{i=1}^{n} \{(T_{s_{i-1}+1}, \dots, T_{s_i}) \in A_i \mid \bigcap_{i=1}^{n} \{T_{s_i} \le t_i < T_{s_i+1}\}\Big)$$

$$= \frac{P\Big((T_1, \dots, T_{s_n}) \in \bigtimes_{i=1}^{n} (A_i \cap (t_{i-1}, t_i]),\ T_{s_n+1} > t_n\Big)}{P\big(N_{(t_{i-1},t_i]} = k_i,\ 1 \le i \le n\big)}$$

$$= \frac{\displaystyle\int_{t_n}^{\infty} \int \cdots \int f_{(T_1,\dots,T_{s_n})}(y_1, \dots, y_{s_n} \mid T_{s_n+1} = y)\, dy_1 \dots dy_{s_n}\, dP^{T_{s_n+1}}(y)}{\big(\bigtimes_{i=1}^{n} A_i\big) \cap \mathcal{K}_n}$$

$$= \frac{\displaystyle\int \cdots \int \int_{t_n}^{\infty} \frac{s_n!}{y^{s_n}} \frac{\lambda^{s_n+1}}{s_n!} y^{s_n+1} e^{-\lambda y}\, dy_1 \dots dy_{s_n}\, dy}{\big(\bigtimes_{i=1}^{n} A_i\big) \cap \mathcal{K}_n}$$

$$= \frac{\big(\bigtimes_{i=1}^{n} A_i\big) \cap \mathcal{K}_n}{\displaystyle\prod_{i=1}^{n} e^{-\lambda(t_i - t_{i-1})} \frac{\big(\lambda(t_i - t_{i-1})\big)^{k_i}}{k_i!}}$$

$$= \int_{A_n} \cdots \int_{A_1} \mathbb{1}_{\mathcal{K}_n}(y_1, \dots, y_{s_n}) \prod_{i=1}^{n} \frac{k_i!}{(t_i - t_{i-1})^{k_i}}\, dy_1 \dots dy_{s_n},$$

woraus die Behauptung folgt. ∎

Der letzte Satz erlaubt damit z.B. die folgende alternative Konstruktion von Poisson-Prozessen mit Parameter $\lambda > 0$:

a) Erzeuge eine unabhängige Folge $\mathfrak{P}(\lambda)$-verteilter Zufallsvariablen $\{N_n\}_{n \in \mathbb{N}}$; setze $S_0 = 0$, $S_n = \sum_{i=1}^{n} N_i$, $n \in \mathbb{N}$.

b) Erzeuge eine Folge (auch von $\{N_n\}_{n\in\mathbf{N}}$) unabhängiger $\mathcal{R}\big((0,1]\big)$–verteilter Zufallsvariablen $\{X_k\}_{k\in\mathbf{N}}$; setze

$$Y_k = X_k + n - 1, \quad S_{n-1} < k \leq S_n, \quad \text{sofern} \quad N_n > 0. \tag{3.4.8}$$

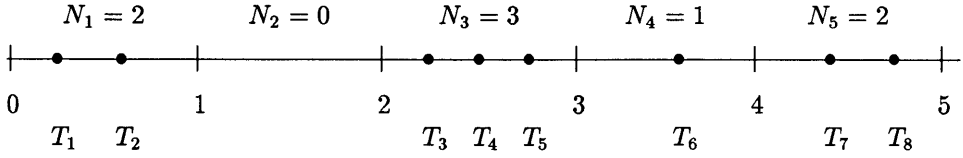
c) Sortiere die Folge $\{Y_k\}_{k\in\mathbf{N}}$ der Größe nach, d.h. bilde

$$T_0 = 0, \quad T_n = \min\{Y_k > T_{n-1} \mid k \in \mathbf{N}\}, \quad n \in \mathbf{N}.$$

Dann ist $\{T_n\}_{n\in\mathbf{N}_0}$ die Ankunftszeitenfolge eines Poisson–Prozesses mit Parameter λ.

Das gerade beschriebene Verfahren erzeugt also in Schritt a) in allen Intervallen $(n-1,n]$, $n \in \mathbf{N}$, zunächst die *Anzahl* der Ankünfte N_n, in Schritt b) deren (ungeordnete) *Position* gemäß einer stetigen Gleichverteilung — sofern $N_n > 0$ gilt, also überhaupt Ankünfte eintreten —, und schließlich in Schritt c) die zeitlich richtig geordnete Ankunftszeitenfolge durch Sortierung.

Die folgende Graphik zeigt eine mögliche Realisation des gerade beschriebenen Verfahrens.



Natürlich kann man statt der Intervalle $(n-1,n]$, $n \in \mathbf{N}$, auch jede andere Intervallpartition von $\mathbf{R}^+$ betrachten; die Schritte a) und b) sind dann gemäß Satz 3.4.3 entsprechend zu modifizieren.

Dieses Modell ist vor allem dann von Interesse, wenn man Poisson–Prozesse nur für eine begrenzte Zeit simulieren will, etwa im Intervall $(a,b]$, $0 \leq a < b$. In diesem Fall lassen sich die Schritte a) und b) noch dahingehend vereinfachen, daß man lediglich *eine* $\mathfrak{P}\big(\lambda(b-a)\big)$–verteilte Zufallsvariable N erzeugt und anschließend — sofern $N = n > 0$ gilt — n davon unabhängige $\mathcal{R}\big((a,b]\big)$–verteilte Zufallsvariablen $X_1,\ldots,X_n$. Die Ordnungsstatistik $(X_{(1)},\ldots,X_{(n)})$ ist dann ein Repräsentant der Ankunftszeiten in diesem Intervall.

Die gerade angestellten Überlegungen verdeutlichen noch einmal aus anderer Sicht, warum Poisson–Prozesse besonders dann betrachtet werden, wenn es um die Modellierung "zeitlich rein zufälliger" Phänomene geht, auch wenn durch die Markoff–Eigenschaft insgesamt gewisse Abhängigkeiten miterfaßt werden.

Satz 3.4.3 ist auch ein wesentlicher Grund dafür, daß man voneinander unabhängige Poisson–Prozesse zu einem neuen Poisson–Prozeß überlagern bzw. einen Poisson–Prozeß durch geeignete Auswahl in zwei voneinander unabhängige Poisson–Prozesse aufteilen kann; eine Eigenschaft, die insbesondere bei der Modellierung von Rechnernetzen äußerst nützlich ist. Zur Präzisierung dieses Sachverhalts werden die folgenden beiden Lemmata benötigt.

Lemma 3.4.2. *(Zerlegung von Poisson–Verteilungen)*
Es sei N eine $\mathfrak{P}(\lambda)$–verteilte Zufallsvariable mit $\lambda > 0$ sowie $\{I_n\}_{n\in\mathbf{N}}$ eine (auch

von N) unabhängige Folge $\mathfrak{B}(1,p)$-verteilter Zufallsvariablen mit $p = 1-q \in (0,1)$. Dann gilt:
Die durch

$$S = \sum_{i=1}^{N} I_i, \quad T = \sum_{i=1}^{N} (1 - I_i) = N - S \tag{3.4.9}$$

definierten Zufallsvariablen[1] *sind stochastisch unabhängig und je Poisson-verteilt mit*

$$P^S = \mathfrak{P}(\lambda p), \quad P^T = \mathfrak{P}(\lambda q). \tag{3.4.10}$$

Beweis. Für $k, m \in \mathbb{N}_0, k+m \leq 1$ gilt mit dem Ersetzungslemma 3.1.4

$$\begin{aligned}
P(S=k, T=m) &= P(S=k, N-S=m) = P(S=k, N=k+m) \\
&= P(S=k \mid N=k+m) P(N=k+m) \\
&= P\Big(\sum_{i=1}^{k+m} I_i = k\Big) P(N=k+m) \\
&= \mathfrak{B}(k+m,p)(\{k\}) \cdot \mathfrak{P}(\lambda)(\{k+m\}) \\
&= \binom{k+m}{k} p^k q^m e^{-\lambda} \frac{\lambda^{k+m}}{(k+m)!} \\
&= \frac{(\lambda p)^k}{k!} \frac{(\lambda q)^m}{m!} e^{-(p+q)\lambda} = \mathfrak{P}(\lambda p)(\{k\}) \cdot \mathfrak{P}(\lambda q)(\{m\});
\end{aligned}$$

analog für $k+m=0$, womit die Aussage bewiesen ist. ∎

Lemma 3.4.2 ist damit gewissermaßen das Gegenstück zu Lemma 2.1.9, denn wegen $p+q=1$ ergibt (3.4.10) gerade wieder

$$\mathfrak{P}(\lambda) = P^N = P^S * P^T = \mathfrak{P}(\lambda p) * \mathfrak{P}(\lambda q).$$

Lemma 3.4.3. *(zufällige Auswahl von Ordnungsstatistiken)*
Es seien für $n \in \mathbb{N}$ $X_1, \ldots, X_n$ unabhängige, je $\mathcal{R}((a,b])$-verteilte Zufallsvariablen mit $a < b$, $a, b \in \mathbb{R}$. Die Kombination η sei unabhängig von $X_1, \ldots, X_n$ und für $1 \leq k \leq n$ über $Komb_k^n(\{1, \ldots, n\}; o.W.)$ gleichverteilt, d.h. es gelte

$$P(\eta = (i_1, \ldots, i_k)) = \frac{1}{\binom{n}{k}}, \quad (i_1, \ldots, i_k) \in Komb_k^n(\{1, \ldots, n\}; o.W.).$$

Dann besitzt die zufällige k-Auswahl $\boldsymbol{X} = (X_{(\eta_1)}, \ldots, X_{(\eta_k)})$ aus der Ordnungsstatistik $(X_{(1)}, \ldots, X_{(n)})$ dieselbe Verteilung wie die (evtl. kleinere) Ordnungsstatistik $(X_{(1)}, \ldots, X_{(k)})$, d.h. $\boldsymbol{X}$ besitzt eine Dichte der Form

$$f_{\boldsymbol{X}}(y_1, \ldots, y_k) = \begin{cases} \dfrac{k!}{(b-a)^k} & \text{für } a < y_1 < \ldots < y_k < b \\ 0 & \text{sonst.} \end{cases}$$

[1] hierbei ist eine Summation über einen leeren Indexbereich als Null zu werten

Beweis. Sei zunächst $(i_1,\dots,i_k) \in Komb_k^n(\{1,\dots,n\}; o.W.)$ fest. Durch Integration von (3.4.6) nach den restlichen Komponenten von $(X_{(1)},\dots,X_{(n)})$ ergibt sich dann mit der Wahl $y_0 = a$, $y_{k+1} = b$, $i_0 = 0$, $i_{k+1} = n+1$:

$$\begin{aligned} f_{(X_{(i_1)},\dots,X_{(i_k)})}(y_1,\dots,y_k) &= \frac{n!}{(b-a)^n} \prod_{j=1}^{k+1} \frac{(y_j - y_{j-1})^{i_j - i_{j-1}-1}}{(i_j - i_{j-1}-1)!} \\ &= \frac{n!}{(n-k)!(b-a)^k} \binom{n-k}{m_1,\dots,m_{k+1}} \prod_{j=1}^{k+1} \left(\frac{y_j - y_{j-1}}{b-a}\right)^{i_j - i_{j-1}-1} \\ &= \frac{n!}{(n-k)!(b-a)^k}\, \mathfrak{M}\left(n-k; p_1,\dots,p_{k+1}\right)\left(\{(m_1,\dots,m_k)\}\right) \end{aligned}$$

mit $m_j = i_j - i_{j-1} - 1$ und $p_j = \frac{y_j - y_{j-1}}{b-a}$, $1 \le j \le k+1$, für $a < y_1 < \dots < y_k < b$. Durch Summation über alle Werte von η erhält man also

$$\begin{aligned} f_{\boldsymbol{X}}(y_1,\dots,y_k) &= \sum_{1 \le i_1 < \dots < i_k \le n} f_{\boldsymbol{X}}(y_1,\dots,y_k \mid \eta = (i_1,\dots,i_k)) P(\eta = (i_1,\dots,i_k)) \\ &= \sum_{1 \le i_1 < \dots < i_k \le n} \frac{k!}{(b-a)^k}\, \mathfrak{M}\left(n-k; p_1,\dots,p_{k+1}\right)\left(\{(m_1,\dots,m_k)\}\right) \\ &= \frac{k!}{(b-a)^k}, \quad a < y_1 < \dots < y_k < b. \end{aligned}$$

Hieraus folgt die Behauptung. ∎

Satz 3.4.4. *(Überlagerung und Aufteilung von Poisson-Prozessen)*
Es seien $\{N_t\}_{t\ge 0}$ und $\{M_t\}_{t\ge 0}$ voneinander unabhängige Poisson-Prozesse mit Parametern $\lambda, \mu > 0$ und Ankunftszeitenfolgen $\{T_n\}_{n\in\mathbb{N}_0}$ bzw. $\{S_n\}_{n\in\mathbb{N}_0}$. Die durch

$$U_0 = 0, \quad U_n = \min\{T_k, S_k > U_{n-1} \mid k \in \mathbb{N}\}, \quad n \in \mathbb{N} \tag{3.4.11}$$

definierte Überlagerung von $\{T_n\}_{n\in\mathbb{N}_0}$ und $\{S_n\}_{n\in\mathbb{N}_0}$ bildet dann die Ankunftszeitenfolge eines Poisson-Prozesses mit Parameter $\lambda + \mu$.
Ist ferner $\{I_k\}_{k\in\mathbb{N}}$ eine von $\{N_t\}_{t\ge 0}$ unabhängige Folge $\mathfrak{B}(1,p)$-verteilter Zufallsvariablen mit $p = 1 - q \in (0,1)$, und teilt man die Ankunftszeitenfolge $\{T_n\}_{n\in\mathbb{N}_0}$ zufällig auf gemäß

$$U_0 = V_0 = 0, \quad \begin{aligned} U_n &= \min\{T_k > U_{n-1} \mid I_k = 1,\ k \in \mathbb{N}\} \\ V_n &= \min\{T_k > V_{n-1} \mid I_k = 0,\ k \in \mathbb{N}\}, \end{aligned} \quad (n \in \mathbb{N}) \tag{3.4.12}$$

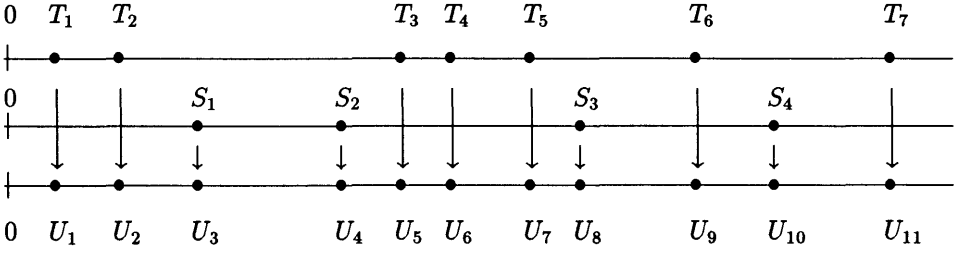
d.h. wählt man aus der Folge $\{T_n\}_{n\in\mathbb{N}_0}$ Ankunftszeiten unabhängig voneinander mit Wahrscheinlichkeit p (bzw. q) aus, so bilden $\{U_n\}_{n\in\mathbb{N}_0}$ und $\{V_n\}_{n\in\mathbb{N}_0}$ die Ankunftszeiten zweier stochastisch unabhängiger Poisson-Prozesse mit Parametern λp bzw. λq.

Beweis. Wir betrachten die in (3.4.8) durchgeführte Alternativ-Konstruktion von Poisson-Prozessen.

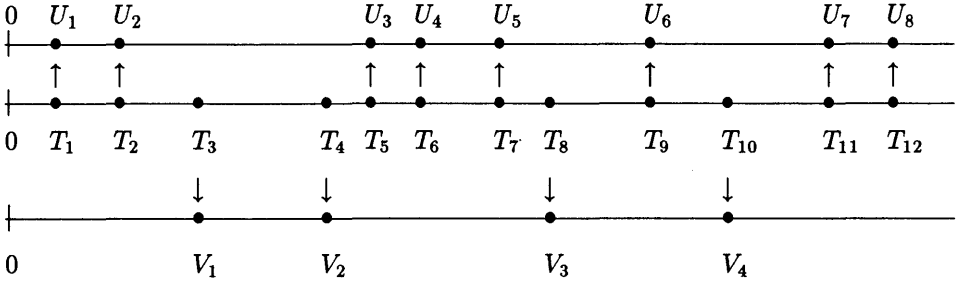
Der erste Teil des Satzes ist dann eine Konsequenz der Faltungsstabilität der Poisson–Verteilung (vgl. Lemma 2.1.9), da durch die Überlagerung der Ankunftszeiten in jedem der Intervalle $(n-1,n]$, $n \in \mathbb{N}$, je zwei voneinander unabhängige Poisson–verteilte Anzahlen von Ankünften addiert werden; die Gesamtanzahl der Ankünfte in $(n-1,n]$ ist damit $\mathfrak{P}(\lambda+\mu)$–verteilt. Da für beide Ausgangsprozesse die Position der Ankünfte — bei bekannter Anzahl — jeweils stetig gleichverteilt ist, gilt dies ebenso für die Überlagerung. Damit ergibt sich der erste Teil der Behauptung.
Für den zweiten Teil argumentiert man analog unter Verwendung von Lemma 3.4.2 und Lemma 3.4.3: liegen in $(n-1,n]$ nämlich $N \in \mathbb{N}$ Ankünfte, so werden diese aufgrund des angegebenen Auswahlmechanismus gerade in $\sum_{i=1}^{N} I_i$ Ankünfte des einen sowie $\sum_{i=1}^{N}(1-I_i)$ Ankünfte des anderen Prozesses zerlegt. Man hat jetzt nur noch zu beachten, daß für alle $m \in \mathbb{N}$ die bedingte Verteilung der Kombination η mit

$$\eta_1 = \min\{1 \le i \le m \mid I_i = 1\}, \quad \eta_j = \min\{i > \eta_{j-1} \mid I_i = 1\},\ 2 \le j \le m,$$

$\mathfrak{L}\left(Komb_k^m(\{1,\dots,m\}; o.W.)\right)$–verteilt ist unter der Bedingung $S_m = \sum_{i=1}^{m} I_i = k$, $1 \le k \le m$, d.h. die verbleibenden Ankunftszeiten jeweils wieder Ordnungsstatistiken stetig gleichverteilter Zufallsvariablen bilden. ∎



Überlagerung von Poisson–Prozessen

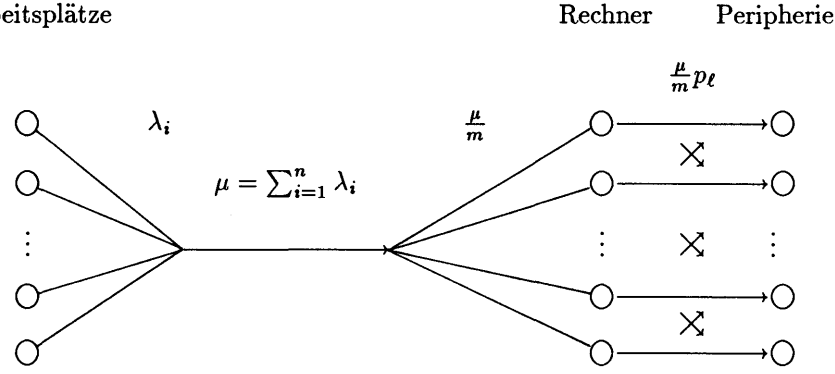


Aufteilung von Poisson–Prozessen

Beispiel 3.4.1. (Netzwerke)
Wir betrachten ein System mit $n \in \mathbb{N}$ Arbeitsplätzen und $m \in \mathbb{N}$ parallel arbeitenden Rechnereinheiten. Von den Arbeitsplätzen aus werden Programme im Zeittakt von Poisson–Prozessen mit Parametern $\lambda_1, \ldots, \lambda_n$ an die Rechnereinheiten übergeben. Ankommende Programme werden zufällig mit gleicher Wahrscheinlichkeit auf die Rechnereinheiten verteilt. Weiterhin stehen $k \in \mathbb{N}$ Peripherie–Geräte (Drucker, Diskettenstationen etc.) zur Verfügung. $p_1, \ldots, p_k$ seien die Wahrscheinlichkeiten dafür, daß ein Programm nach Beendigung auf eines dieser Geräte zugreift; $1 - \sum_{\ell=1}^{k} p_\ell$ bezeichne die Wahrscheinlichkeit dafür, daß ein Programm aufgrund von Fehlern abgebrochen wird. Vernachlässigt man zunächst die Bearbeitungszeit der Programme, so ergeben sich insgesamt folgende Poisson–Prozesse zwischen den einzelnen Einheiten:

Ausgang	Eingang	Parameter
Arbeitsplatz i		λ_i
Arbeitsplätze gesamt		$\mu := \sum_{i=1}^{n} \lambda_i$
	Rechnereinheit j	$\frac{\mu}{m}$
Rechnereinheit j	Peripheriegerät ℓ	$\frac{\mu}{m} p_\ell$
	Peripheriegerät ℓ	μp_ℓ
	Peripherie gesamt	$\mu \sum_{\ell=1}^{k} p_\ell$

Die folgende Graphik verdeutlicht dies noch einmal:



■

Wir behandeln im folgenden allgemein homogene Markoff–Prozesse. Hierbei stellt sich wie bei Markoff–Ketten die Frage, ob das stochastische Verhalten eines

solchen Prozesses nicht schon allein durch die Angabe von Übergangswahrscheinlichkeiten, etwa wie in (3.4.5), charakterisiert ist. Bei zeitlich homogenen Markoff-Prozessen benötigt man dazu eine geeignete Familie $\{\mathbf{\Pi}^t\}_{t\geq 0}$ stochastischer Matrizen mit Elementen $p_{ij}^{(t)}$, $i,j \in \mathcal{S}$, $t \geq 0$, die das Übergangsverhalten des Prozesses $\{X_t\}_{t\geq 0}$ vermöge

$$P(X_t = j \mid X_s = i) = p_{ij}^{(t-s)} \quad i,j \in \mathcal{S},\ 0 \leq s \leq t, \tag{3.4.13}$$

beschreiben, wobei zweckmäßigerweise $\mathbf{\Pi}^0 = I$ (Einheitsmatrix) zu setzen ist. Im Falle eines Poisson-Prozesses mit Parameter $\lambda > 0$ ist etwa nach (3.4.5)

$$p_{ij}^{(t)} = \begin{cases} \dfrac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t} & \text{für } 0 \leq i \leq j \\ 0 & \text{sonst,} \end{cases} \qquad t > 0.$$

Die Matrix $\mathbf{\Pi} = \mathbf{\Pi}^1$ entspricht dabei den Einschritt-Übergangswahrscheinlichkeiten der aus dem Prozeß gewonnenen (homogenen) Markoff-Ketten $\{X_{t_n}\}_{n\in\mathbb{N}_0}$ mit $t_n - t_{n-1} = 1$, $n \in \mathbb{N}$. Insbesondere übertragen sich die *Chapman-Kolmogoroff-Gleichungen* (Lemma 3.2.3 c)) unmittelbar auf den allgemeineren Fall, d.h. es gilt in Matrixnotation:

$$\mathbf{\Pi}^{s+t} = \mathbf{\Pi}^s \mathbf{\Pi}^t, \quad 0 \leq s \leq t. \tag{3.4.14}$$

Dies ist evident aufgrund der Definition 3.4.1 homogener Markoff-Prozesse; man spricht daher bei der Abbildung $t \mapsto \mathbf{\Pi}^t$, $t \geq 0$, auch von einer *Matrixhalbgruppe*. Allerdings läßt sich nicht zu jeder beliebigen stochastischen Matrix $\mathbf{\Pi}$ ein Markoff-Prozeß $\{X_t\}_{t\geq 0}$ finden, für den $\mathbf{\Pi}^1 = \mathbf{\Pi}$ gilt. Die Matrix $\mathbf{\Pi} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (d.h. $\#\mathcal{S} = 2$) liefert hierfür ein Beispiel: für $t = 1/2$ müßte nach (3.4.14) nämlich gelten $\mathbf{\Psi}^2 = \mathbf{\Pi}$ mit $\mathbf{\Psi} = \mathbf{\Pi}^{1/2}$; es gibt aber keine Lösung dieser quadratischen Matrixgleichung, bei der $\mathbf{\Psi}$ reelle Komponenten enthält. Allgemeiner gilt dies sogar für alle Matrizen $\mathbf{\Pi} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$ mit $0 \leq p < 1/2$; die Eigenwerte von $\mathbf{\Pi}$ sind hier nämlich gegeben durch $p \pm |1-p|$, also 1 und $2p-1 < 0$. $\mathbf{\Pi}$ ist also indefinit; eine reelle Lösung der obigen quadratischen Matrixgleichung kann somit nicht existieren. Anschaulich bedeutet der zuletzt genannte Sachverhalt ein zu starkes Oszillieren der Markoff-Ketten mit einer solchen Übergangsmatrix $\mathbf{\Pi}$, da wegen $p < 1/2$ Änderungen des momentanen Zustands häufiger auftreten als ein Verharren in diesem. Für $1/2 \leq p \leq 1$ existiert allerdings eine Lösung des Problems; sie ist gegeben durch

$$\mathbf{\Pi}^t = \frac{1}{2} \begin{pmatrix} 1+(2p-1)^t & 1-(2p-1)^t \\ 1-(2p-1)^t & 1+(2p-1)^t \end{pmatrix}, \qquad t \geq 0. \tag{3.4.15}$$

Allgemeiner läßt sich immer dann eine Lösung des Problems finden, wenn eine reelle Matrix $\boldsymbol{Q} = (q_{ij})_{i,j\in\mathcal{S}}$ existiert mit

$$e^{\boldsymbol{Q}} := \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{Q}^k = I + \boldsymbol{Q} + \frac{1}{2}\boldsymbol{Q}^2 + \ldots = \mathbf{\Pi}; \tag{3.4.16}$$

in diesem Fall erhält man die Lösung

$$\mathbf{\Pi}^t = e^{t\boldsymbol{Q}} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\boldsymbol{Q}^k = I + t\boldsymbol{Q} + \frac{t^2}{2}\boldsymbol{Q}^2 + \ldots, \quad t \geq 0. \tag{3.4.17}$$

Definition 3.4.3. *(Intensitätsmatrix)*
Es sei $\{X_t\}_{t\geq 0}$ ein homogener Markoff-Prozeß mit Übergangsmatrizen $\{\mathbf{\Pi}^t\}_{t\geq 0}$ der Form (3.4.13). Eine reelle Matrix $\boldsymbol{Q}$ mit

$$e^{t\boldsymbol{Q}} = \mathbf{\Pi}^t, \quad t \geq 0,$$

heißt Intensitätsmatrix des Prozesses.

Die Intensitätsmatrix ist im Falle der Existenz eindeutig bestimmt und kann aufgrund von (3.4.17) aus der Matrixhalbgruppe durch Differenzieren erhalten werden vermöge

$$\boldsymbol{Q} = \lim_{t \downarrow 0} \frac{1}{t}(\mathbf{\Pi}^t - I) \tag{3.4.18}$$

bzw. elementweise

$$q_{ij} = \begin{cases} \lim\limits_{t\downarrow 0} \frac{1}{t} p_{ij}^{(t)} = \lim\limits_{t\downarrow 0} \frac{1}{t} P(X_t = j \mid X_0 = i) & \text{für } i \neq j \\ \lim\limits_{t\downarrow 0} \frac{1}{t}(p_{ii}^{(t)} - 1) = -\lim\limits_{t\downarrow 0} \frac{1}{t} P(X_t \neq i \mid X_0 = i) & \text{für } i = j, \end{cases} \qquad i, j \in \mathcal{S}. \tag{3.4.19}$$

Insbesondere gilt stets

$$q_{ii} \leq 0 \qquad \text{und} \qquad \sum_{j\in\mathcal{S}} q_{ij} = 0, \quad i \in \mathcal{S}; \tag{3.4.20}$$

letzteres ergibt sich aus der Tatsache, daß alle Zeilensummen von $\mathbf{\Pi}^t - I$, $t > 0$, Null ergeben. Man beachte dabei, daß wegen der Homogenität $P(X_t = j \mid X_0 = i) = P(X_{t+s} = j \mid X_s = i)$, $s, t \geq 0$, $i, j \in \mathcal{S}$, gilt. Für einen Poisson-Prozeß mit Parameter $\lambda > 0$ erhält man so etwa

$$q_{ij} = \begin{cases} \lambda & \text{für } j = i+1 \\ -\lambda & \text{für } j = i \\ 0 & \text{sonst,} \end{cases} \qquad i \in \mathbb{N}_0. \tag{3.4.21}$$

Man spricht daher auch von einem Poisson-Prozeß mit *Intensität* $\lambda > 0$. Für das in (3.4.15) angesprochene Beispiel ergibt sich im Fall $1/2 < p \leq 1$ analog

$$\boldsymbol{Q} = -\ln(2p-1)\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3.4.22}$$

Der Sonderfall $p = 1/2$ läßt sich darunter nicht subsumieren, da hier $\mathbf{\Pi}^t = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ gilt für alle $t > 0$, also die verlangte Differenzierbarkeit im Nullpunkt nicht gegeben ist; er entspricht der Situation einer *unabhängigen* Familie

$\{X_t\}_{t>0}$ von je $\mathfrak{B}(1,\frac{1}{2})$–verteilten Zufallsvariablen, die natürlich trivialerweise eine homogene Markoff–Kette bilden.

Wegen $\mathbf{\Pi}^t = e^{t\boldsymbol{Q}} \approx I + t\boldsymbol{Q}$ für kleine t nach (3.4.17) erhält man noch die Näherungsformeln

$$P(X_{t+s}=j \mid X_s=i) \approx \begin{cases} tq_{ij} & \text{für } i \neq j \\ 1-t(-q_{ii}) & \text{für } i=j, \end{cases} \quad s,t \geq 0,\ i,j \in \mathcal{S},\ t \text{ klein.}$$

Hat man eine Intensitätsmatrix $\boldsymbol{Q}$ zu Verfügung, d.h. eine Matrix $\boldsymbol{Q}$ mit der Darstellung (3.4.16), stellt sich allerdings noch die Frage nach der Existenz eines Markoff–Prozesses mit den Übergangsmatrizen $\mathbf{\Pi}^t = e^{t\boldsymbol{Q}}$, $t \geq 0$. Nicht jede Intensitätsmatrix $\boldsymbol{Q}$ gestattet nämlich eine eindeutige Konstruktion eines entsprechenden Markoff–Prozesses über die gesamte Zeitachse $[0,\infty)$; beispielsweise gibt es Probleme, wenn $\boldsymbol{Q}$ die Form

$$q_{ij} = \begin{cases} \lambda_j > 0 & \text{für } j = i+1 \\ -\lambda_j & \text{für } j = i \\ 0 & \text{sonst,} \end{cases} \qquad i \in \mathbb{N}_0, \tag{3.4.23}$$

besitzt mit einer konvergenten Reihe $\sum_{j=1}^{\infty} \frac{1}{\lambda_j} < \infty$. In diesem Fall divergiert nämlich der Prozeß f.s. bereits nach endlicher Zeit; man spricht daher auch von einer *Explosion* des Markoff–Prozesses.

"Gutartige" Intensitätsmatrizen erlauben allerdings Darstellungen von zugehörigen Markoff–Prozessen, wie der folgende Satz zeigt.

Satz 3.4.5. *(Struktur homogener Markoff–Prozesse)*
Es sei $\boldsymbol{Q} = (q_{ij})_{i,j\in\mathcal{S}}$ eine Intensitätsmatrix, die die Bedingungen

$$0 < \nu = \inf_{i\in\mathcal{S}}\{\lambda_i\}, \quad \mu = \sup_{i\in\mathcal{S}}\{\lambda_i\} < \infty, \quad \lambda_i = -q_{ii},\ i \in \mathcal{S}, \tag{3.4.24}$$

erfülle. Dann gibt es einen homogenen Markoff–Prozeß $\{X_t\}_{t\geq 0}$ auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$, dessen Übergangsmatrizen durch $\mathbf{\Pi}^t = e^{t\boldsymbol{Q}}$ gegeben sind. $\{X_t\}_{t\geq 0}$ kann kanonisch wie folgt dargestellt werden:

Es sei $\{Y_n\}_{n\in\mathbb{N}_0}$ eine homogene Markoff–Kette auf $(\Omega,\mathcal{A},P)$ mit Zustandraum $\mathcal{S}$ und Übergangswahrscheinlichkeiten

$$P(Y_{n+1}=j \mid Y_n=i) = \begin{cases} \frac{q_{ij}}{\lambda_i} & \text{für } j \neq i \\ 0 & \text{für } j=i, \end{cases} \qquad i,j \in \mathcal{S} \tag{3.4.25}$$

und Anfangsverteilung $P(Y_0 = s) = 1$. Ferner sei $\{U_k\}_{k\in\mathbb{N}_0}$ eine (auch von $\{Y_n\}_{n\in\mathbb{N}_0}$) unabhängige Folge $\mathcal{R}\big((0,1]\big)$–verteilter Zufallsvariablen. Die Zwischenankunftszeiten $\Delta_0 = 0$, $\Delta_k = T_k - T_{k-1}$, $k \in \mathbb{N}$, des Prozesses $\{X_t\}_{t\geq 0}$ können dann dargestellt werden vermöge

$$\Delta_{k+1} = -\frac{1}{\lambda_{Y_k}} \ln(U_k), \quad k \in \mathbb{N}; \tag{3.4.26}$$

der Prozeß selbst besitzt damit die Darstellung

$$X_t = Y_k \quad \text{für} \quad T_k = \sum_{j=0}^{k} \Delta_j \leq t < \sum_{j=0}^{k+1} \Delta_j = T_{k+1}, \tag{3.4.27}$$

d.h. unter der Bedingung $Y_k = j_k \in \mathcal{S}$, $k \in \mathbb{N}$ fest, verweilt der Prozeß jeweils eine $\mathcal{E}(\lambda_{j_k})$-verteilte Zeit Δ_{k+1} im Zustand j_k (vgl. (2.1.9)) und geht dann unabhängig von der Verweilzeit Δ_{k+1} gemäß den Übergangswahrscheinlichkeiten (3.4.25) in den Zustand Y_{k+1} über. Die Folge $\{T_n\}_{n\in\mathbb{N}_0}$ mit $T_0 = 0$, $N_0 = s \in \mathcal{S}$ fest heißt wieder Ankunftszeitenfolge des Prozesses; sie besitzt die Eigenschaft $T_n \to \infty$ P-fast sicher für $n \to \infty$.

Den technisch aufwendigen Beweis, der im wesentlichen auf der bereits in Abschnitt 2.1 betrachteten Gedächtnislosigkeit der Exponentialverteilung beruht, können wir hier nicht führen; eine tiefergehende Behandlung der Problematik mit analytischen Methoden findet sich z.B. in Breiman (1968), Abschnitt 15.5 und Çinlar (1975), Abschnitt 8.3. Eine elementare Beweisskizze gibt Pflug (1986), S. 35/36. Allerdings läßt sich die Wahl der Übergangswahrscheinlichkeiten in (3.4.25) leicht durch die folgende Plausibilitätsbetrachtung motivieren:

Da das Übergangsverhalten der Markoff-Kette $\{Y_n\}_{n\in\mathbb{N}_0}$ mit den Ankunftszeiten des Prozesses gekoppelt ist, genügt es, hierfür den Ausdruck

$$q_{ij}^* = \lim_{h\downarrow 0} P(X_{s+h} = j \mid X_s = i, X_{s+h} \neq i), \quad i,j \in \mathcal{S},\ i \neq j, \quad s \geq 0,$$

zu betrachten, sofern dieser existiert. Aufgrund von (3.1.5) ergibt eine einfache Rechnung nun

$$\begin{aligned} q_{ij}^* &= \lim_{h\downarrow 0} \frac{P(X_{s+h} = j \mid X_s = i)}{P(X_{s+h} \neq i \mid X_s = i)} = \lim_{h\downarrow 0} \frac{\frac{1}{h}P(X_{s+h} = j \mid X_s = i)}{\frac{1}{h}P(X_{s+h} \neq i \mid X_s = i)} \\ &= \lim_{h\downarrow 0} \frac{\frac{1}{h}P(X_h = j \mid X_0 = i)}{\frac{1}{h}P(X_h \neq i \mid X_0 = i)} = \frac{q_{ij}}{-q_{ii}} \end{aligned}$$

nach (3.4.19), also gerade den in (3.4.25) angegebenen Ausdruck.

Man beachte, daß aufgrund von (3.4.20) die in (3.4.25) angegebenen Größen tatsächlich *Wahrscheinlichkeiten* sind.

Satz 3.4.5 läßt sich sogar noch dahingehend erweitern, daß praktisch jeder homogene Markoff-Prozeß $\{X_t\}_{t\geq 0}$ mit Zustandsraum $\mathcal{S}$ und $X_0 = s \in \mathcal{S}$, der zwischen den Ankunftszeiten $\{T_n\}_{n\in\mathbb{N}_0}$ konstant ist mit $T_n \to \infty$, $n \to \infty$, eine wie oben beschriebene Struktur besitzt, d.h. die Zwischenankunftszeiten bedingt exponentiell verteilt sind und die Folge der Sprünge $\{X_{T_n}\}_{n\in\mathbb{N}_0}$ eine homogene Markoff-Kette mit Übergangswahrscheinlichkeiten (3.4.25) bildet. Die Markoff-Kette $\{X_{T_n}\}_{n\in\mathbb{N}_0}$ bzw. $\{Y_n\}_{n\in\mathbb{N}_0}$ aus Satz 3.4.5 heißt deshalb auch *eingebettete Markoff-Kette* zum Markoff-Prozeß $\{X_t\}_{t\geq 0}$.

Satz 3.4.5 erklärt damit in gewisser Weise auch, warum unter den Bedingungen (3.4.23) der Prozeß nach endlicher Zeit explodiert, da die Zwischenankunftszeiten

$\{\Delta_k\}_{k\in\mathbb{N}_0}$ hier ebenfalls exponentialverteilt sind mit $E(\Delta_k) = \frac{1}{\lambda_k}$, $k \in \mathbb{N}_0$. Die Aussage ergibt sich dann aus dem Satz von der monotonen Konvergenz (2.2.23).

Andererseits kann man auch noch Markoff–Prozesse behandeln, deren eingebettete Markoff–Kette absorbierende Zustände besitzt; dies entspricht dem Fall $\lambda_i = 0$ für ein $i \in \mathcal{S}$. In dieser Situation kann die Ankunftszeitenfolge degenerieren, also mit positiver Wahrscheinlichkeit nur noch endlich viele verschiedene Werte liefern (vgl. Aufgabe 3.9).

Natürlich kann man auch Markoff–Prozesse mit beliebigen Anfangsverteilungen P^{X_0} betrachten; Satz 3.4.5 gilt dann entsprechend, indem man für die eingebettete Markoff–Kette die Anfangsverteilung $P^{Y_0} = P^{X_0}$ wählt.

Satz 3.4.5 zeigt damit, daß homogene Markoff–Prozesse mit abzählbarem Zustandsraum $\mathcal{S}$ im wesentlichen aus homogenen Markoff–Ketten durch geeignete zeitstetige Interpolation über bedingte Exponentialverteilungen für die Zwischenankunftszeiten entstehen. Das stochastische Verhalten solcher Markoff–Prozesse kann aber durchaus von dem der eingebetteten Markoff–Kette abweichen; wir wollen dies hier nur exemplarisch an dem in (3.4.16) beschriebenen Beispiel zeigen:
Einerseits folgt direkt aus Beziehung (3.4.16)

$$\lim_{t\to\infty} \mathbf{\Pi}^t = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix},$$

d.h. die Verteilung des zugehörigen Markoff–Prozesses konvergiert gegen die Gleichverteilung $\mathfrak{B}(1,1/2)$ über $\{0,1\}$; andererseits besitzt die eingebettete Markoff–Kette die Übergangsmatrix $\mathbf{\Phi} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, deren stationäre Verteilung zwar ebenfalls durch $\mathfrak{B}(1,1/2)$ gegeben ist, wie man leicht nachrechnet; für die Matrixpotenzen erhält man aber $\mathbf{\Phi}^n = \begin{cases} \mathbf{\Phi} & \text{für } n \text{ ungerade} \\ I & \text{für } n \text{ gerade} \end{cases}$, so daß die eingebettete Markoff–Kette i.a. *keine* Grenzverteilung besitzt (man beachte, daß hier $\mathbf{\Phi}$ periodisch, also Korollar 3.2.2 nicht anwendbar ist).

Für den Fall, daß die eingebettete Markoff–Kette eine aperiodische und irreduzible Übergangsmatrix besitzt, stimmt das Grenzverhalten des Prozesses allerdings mit demjenigen der eingebetteten Markoff–Kette überein.

Satz 3.4.6. *(Grenzverhalten eines homogenen Markoff–Prozesses)*
Es sei $\{X_t\}_{t\geq 0}$ ein homogener Markoff–Prozeß auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Zustandsraum $\mathcal{S}$, Intensitätsmatrix $\boldsymbol{Q}$, Übergangsmatrizen $\mathbf{\Pi}^t = e^{t\boldsymbol{Q}}$, $t \geq 0$, und eingebetteter Markoff–Kette $\{Y_n\}_{n\in\mathbb{N}_0}$ mit Übergangsmatrix $\mathbf{\Phi} = \{\phi_{ij}\}_{i,j\in\mathcal{S}}$, gegeben durch (3.4.25), also

$$\phi_{ij} = \begin{cases} \frac{q_{ij}}{\lambda_i} & \text{für } i \neq j \\ 0 & \text{für } i = j, \end{cases} \qquad i,j \in \mathcal{S}.$$

Ist $\mathcal{S}$ endlich bzw. sind alle Zustände bezüglich der Markoff-Kette $\{Y_n\}_{n\in\mathbb{N}_0}$ positiv-rekurrent und ist $\boldsymbol{\Phi}$ aperiodisch und irreduzibel, so existiert

$$\lim_{t\to\infty} \boldsymbol{\Pi}^t = \lim_{n\to\infty} \boldsymbol{\Phi}^n = \begin{pmatrix} \boldsymbol{p} & \cdots & \boldsymbol{p} & \cdots \end{pmatrix}^{tr}, \tag{3.4.28}$$

wobei $\boldsymbol{p}$ die nach Korollar 3.2.2 bzw. Satz 3.2.4 eindeutig bestimmte stationäre Verteilung der eingebetteten Markoff-Kette bezeichnet. Ist $\boldsymbol{\Phi}$ lediglich irreduzibel, so existiert ebenfalls

$$\lim_{t\to\infty} \boldsymbol{\Pi}^t = \begin{pmatrix} \boldsymbol{p} & \cdots & \boldsymbol{p} & \cdots \end{pmatrix}^{tr};$$

$\boldsymbol{p}$ ist dann stationäre Verteilung des Prozesses, d.h. es gilt $\boldsymbol{p}\cdot\boldsymbol{\Pi}^t = \boldsymbol{p}$ für alle $t > 0$. $\boldsymbol{p}$ ist in beiden Fällen zugleich (normierte) Lösung des Gleichungssystems

$$\boldsymbol{pQ} = \boldsymbol{0}, \tag{3.4.29}$$

wobei $\boldsymbol{0}$ den Nullvektor bezeichne.

Beweis. Den ersten Teil der Aussage können wir hier nicht beweisen; er folgt z.B. aus Grenzwertsätzen für Semi-Markoff-Prozesse, vgl. Ross (1984), Abschnitte 4.8 und 5.5. Der zweite Teil folgt aus Beziehung (3.4.17), da $\boldsymbol{p}$ unter den angegebenen Bedingungen auch stationäre Verteilung der aus dem Prozeß gewonnenen Markoff-Ketten $\{X_{t_n}\}_{n\in\mathbb{N}_0}$ mit $t_{n+1} - t_n = const$, $n \in \mathbb{N}_0$, sein muß, also die Beziehung $\boldsymbol{p\Pi} = \boldsymbol{p}e^{t\boldsymbol{Q}} = \boldsymbol{p}$ für alle $t > 0$ erfüllt, welche aber äquivalent zu $\boldsymbol{pQ} = \boldsymbol{0}$ ist. Damit ist der Satz bewiesen. ∎

Wir wollen nun den letzten Satz benutzen, um das zeitliche Verhalten eines einfachen Bedienungsystems zu analysieren.

Beispiel 3.4.2. (Warteschlangenmodell)
Wir betrachten ein Rechnersystem, in das im Zeittakt eines Poisson-Prozesses mit Intensität $\lambda > 0$ Programme durch die Anwender eingegeben werden. Das System arbeitet unabhängig vom Eingabestrom die Jobs im Zeittakt eines Poisson-Prozesses mit Intensität $\mu > 0$ ab, d.h. die Bearbeitungszeiten für die einzelnen Programme seien unabhängig voneinander und jeweils $\mathcal{E}(\mu)$-exponentialverteilt. Noch nicht bearbeitete Programme werden in eine Warteschlange gestellt, die nach oben durch die Zahl $M \in \mathbb{N}$ begrenzt sei; sollten mehr Programme das System erreichen, gehen die überzähligen Programme — unter Ausgabe einer Fehlermeldung — verloren. Von Interesse ist das Langzeitverhalten der Warteschlange.
Aus den gemachten Annahmen folgt, daß die Intensitätsmatrix $\boldsymbol{Q}$ des zugehörigen Markoff-Prozesses, der die Anzahl der zur Zeit $t \geq 0$ auf Bearbeitung wartenden Programme beschreibt, gegeben ist durch die $(M+1)\times(M+1)$-Matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -\mu & \mu \end{pmatrix};$$

da der Zustandsraum $\mathcal{S} = \{0, 1, \ldots, M\}$ endlich und die Matrix Φ irreduzibel ist, ergibt sich die Grenzverteilung der Warteschlangenlänge $\boldsymbol{p} = (p_0, \ldots, p_M)$ aus dem Gleichungssystem $\boldsymbol{pQ} = \boldsymbol{0}$, also

$$\begin{aligned} -\lambda p_0 + \mu p_1 &= 0 \\ \lambda p_{i-1} - (\lambda + \mu) p_i + \mu p_{i+1} &= 0, \quad 1 \leq i \leq M-1 \\ -\mu p_{M-1} + \mu p_M &= 0. \end{aligned} \tag{3.4.30}$$

Bezeichnet man zur Abkürzung $p_0 = \alpha$, so erhält man rekursiv aus (3.4.30) die Menge aller (nicht-normierten) Lösungen $\boldsymbol{p}^*$ zu

$$p_i^* = \left(\frac{\lambda}{\mu}\right)^i \cdot \alpha, \quad 0 \leq i \leq M.$$

Durch Summation über i erhält man also die Grenzverteilung $\boldsymbol{p}$ zu

$$p_i = \begin{cases} \left(\frac{\lambda}{\mu}\right)^i \dfrac{1 - \frac{\lambda}{\mu}}{\left(1 - \frac{\lambda}{\mu}\right)^{M+1}} & \text{für } \lambda < \mu \\ \dfrac{1}{M+1} & \text{für } \lambda = \mu \\ \left(\frac{\lambda}{\mu}\right)^i \dfrac{\frac{\lambda}{\mu} - 1}{\left(\frac{\lambda}{\mu} - 1\right)^{M+1}} & \text{für } \lambda > \mu. \end{cases} \tag{3.4.31}$$

Für $\lambda = \mu$ ist z.B. die Warteschlangenlänge asymtotisch $\mathfrak{L}(\{0, 1, \ldots, M\})$-verteilt.
■

Man beachte, daß die Matrix Φ hier *periodisch* ist, also i.a. die eingebettete Markoff-Kette wiederum keine Grenzverteilung besitzt!

Verzichtet man in dem obigen Beispiel auf eine Beschränkung der Warteschlange, so läßt sich noch zeigen, daß die Zustände $i \in \mathcal{S} = \mathbb{N}_0$ der eingebetteten Markoff-Kette sämtlich positiv-rekurrent sind, solange $\lambda < \mu$ gilt, also Programme schneller bearbeitet werden als neue in das System gelangen (sog. $M/M/1$-System). In diesem Fall gilt

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix};$$

die Grenzverteilung der Warteschlangenlänge ergibt sich dann analog zu

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right), \quad i \in \mathbb{N}_0,$$

d.h. $\boldsymbol{p}$ ist eine geometrische Verteilung $\mathfrak{G}^+(p)$ mit Parameter $p = \frac{\lambda}{\mu}$. Dieses Ergebnis folgt im übrigen auch aus (3.4.31) bei Grenzübergang $M \to \infty$. Für $\lambda = \mu$ ergibt sich dagegen ein oszillierendes Verhalten der Warteschlangenlänge, d.h. eine Grenzverteilung existiert nicht; bei $\lambda > \mu$ strebt die Warteschlangenlänge sogar f.s. gegen ∞ (siehe dazu auch Ross (1984), Abschnitt 5.5).

Warteschlangenmodelle spielen auch eine besondere Rolle bei der Leistungsbewertung von Rechnersystemen, siehe etwa Bolch (1989).

In den bisherigen Ausführungen haben wir nur die Modellierung zeitlich homogener Phänomene behandelt. Poisson–Prozesse bieten durch geeignete Zeittransformationen aber auch die Möglichkeit, nicht–homogene Phänomene angemessen zu modellieren. Wir betrachten dazu eine Abbildung $A : [0, \infty) \longrightarrow [0, \infty)$ mit den Eigenschaften:

$$\begin{aligned} &a) \quad A(0) = 0 \\ &b) \quad A(s) \leq A(t), \quad 0 \leq s \leq t \\ &c) \quad \lim_{s \downarrow t} A(s) = A(t), \quad t \geq 0 \\ &d) \quad \lim_{t \to \infty} A(t) = \infty. \end{aligned} \tag{3.4.32}$$

A erfüllt also bis auf d) gerade die Eigenschaften einer Verteilungsfunktion über $\mathbb{R}^+$.

Definition 3.4.4. *(inhomogener Poisson–Prozeß)*
Es sei $\{X_t\}_{t \geq 0}$ ein homogener Poisson–Prozeß mit Intensität $\lambda = 1$ und A eine Abbildung mit den in (3.4.32) angegebenen Eigenschaften. Der aus $\{X_t\}_{t \geq 0}$ vermöge

$$Y_t = X_{A(t)}, \quad t \geq 0 \tag{3.4.33}$$

gewonnene Markoff–Prozeß heißt inhomogener Poisson–Prozeß mit Zeittransformation A.

Die Markoff–Eigenschaft ergibt sich dabei unmittelbar aus Definition 3.4.1, da A schwach monoton ist und $\{X_t\}_{t \geq 0}$ selbst die Markoff–Eigenschaft besitzt. Für $A(t) = t$, $t \geq 0$, erhält man natürlich wieder einen homogenen Poisson–Prozeß mit Intensität 1. Man beachte, daß mit A auch jedes Vielfache λA mit $\lambda > 0$ eine Zeittransformation im Sinne von (3.4.2) ist.

Das folgende Lemma beschreibt die charakteristischen Eigenschaften inhomogener Poisson–Prozesse.

Lemma 3.4.4. *(Eigenschaften inhomogener Poisson–Prozesse)*
Es sei $\{Y_t\}_{t \geq 0}$ ein inhomogener Poisson–Prozeß mit Zeittransformation A. Dann gilt:

a) $P(Y_t = j \mid Y_s = i) = e^{-\left(A(t) - A(s)\right)} \dfrac{\left(A(t) - A(s)\right)^{j-i}}{(j-i)!}, \qquad 0 \leq s < t,\ i, j \in \mathbb{N}_0,\ i \leq j;$

b) Der Prozeß $\{Y_t\}_{t \geq 0}$ besitzt unabhängige und Poisson–verteilte Zuwächse, d.h. es gilt: für jede aufsteigende Folge von Zeitpunkten $\{t_n\}_{n \in \mathbb{N}_0} \subset \mathbb{R}^+$ sind die Zuwächse $\left\{Y_{(t_n, t_{n+1}]}\right\}_{n \in \mathbb{N}_0}$ stochastisch unabhängig mit

$$P^{Y_{(s,t]}} = \mathfrak{P}\left(A(t) - A(s)\right), \quad 0 \leq s < t. \tag{3.4.34}$$

Beweis. Dies folgt unmittelbar aus (3.4.5) und Satz 3.4.1, wenn dort jeweils λ durch 1 und s, t durch $A(s), A(t)$ ersetzt werden. ∎

Die Zeittransformation A gibt also aufgrund von (3.4.34) zugleich die erwartete Anzahl der Ankünfte des inhomogenen Poisson–Prozesses im Intervall $(s,t]$ an, und zwar vermöge

$$E\big(Y_{(s,t]}\big) = A(t) - A(s),\ 0 \le s < t. \tag{3.4.35}$$

Die in Beziehung (3.4.8) angedeutete Alternativ–Konstruktion homogener Poisson–Prozesse läßt sich ebenfalls in natürlicher Weise auf den inhomogenen Fall übertragen. Ein Analogon zu Satz 3.4.3 für diesen Fall ist

Satz 3.4.7. *(Ankunftszeiten bei inhomogenen Poisson–Prozessen)*
Es sei $\{Y_t\}_{t\ge 0}$ ein Poisson–Prozeß mit Zeittransformation A, $\{T_n\}_{n\in\mathbb{N}_0}$ bezeichne die Folge der Ankunftszeiten. Ferner sei $\{t_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}^+$ eine aufsteigende Folge von Zeitpunkten. Dann verhalten sich die Ankunftszeiten $T_1, \dots, T_{s_n}$ unter der Bedingung, daß in den Zeitintervallen $(t_{i-1}, t_i]$, $1 \le i \le n$, jeweils genau k_i Ankünfte eintreten, wie die geordneten Werte unabhängiger, jeweils über $(t_{i-1}, t_i]$ identisch verteilter Zufallsvariablen mit Verteilungsfunktion

$$F_i(x) = \begin{cases} 0 & \text{für } x < t_{i-1} \\ \dfrac{A(x) - A(t_{i-1})}{A(t_i) - A(t_{i-1})} & \text{für } t_{i-1} \le x \le t_i \\ 1 & \text{für } x > t_i. \end{cases} \tag{3.4.36}$$

Beweis. Dies ergibt sich aus Satz 3.4.3 und Satz 2.1.1, wenn man beachtet, daß nach Definition von $\{Y_t\}_{t\ge 0}$ die Folge $\{A^{-1}(T_n)\}_{n\in\mathbb{N}_0}$ fast sicher die Ankunftszeitenfolge eines homogenen Poisson–Prozesses mit Intensität 1 bildet, wobei A^{-1} wieder die Pseudo–Inverse von A bezeichnet. ∎

Einen inhomogenen Poisson–Prozeß mit Zeittransformation A kann man also auch folgendermaßen konstruieren:

a) Erzeuge eine unabhängige Folge $\mathfrak{P}(\lambda_n)$–verteilter Zufallsvariablen N_n mit $\lambda_n = A(n+1) - A(n)$, $n \in \mathbb{N}$; setze $S_0 = 0$, $S_n = \sum_{i=1}^{n} N_i$, $n \in \mathbb{N}$.

b) Erzeuge eine Folge (auch von $\{N_n\}_{n\in\mathbb{N}}$) unabhängiger $\mathcal{R}\big((0,1]\big)$–verteilter Zufallsvariablen $\{X_k\}_{k\in\mathbb{N}}$; setze

$$Y_k = F_n^{-1}(X_k) + n - 1, \quad S_{n-1} < k \le S_n, \quad \text{sofern} \quad N_n > 0, \tag{3.4.37}$$

wobei F_n die Verteilungsfunktion aus (3.4.36) mit $t_i = i$, $i \in \mathbb{N}_0$, bezeichne.

c) Sortiere die Folge $\{Y_k\}_{k\in\mathbb{N}}$ der Größe nach, d.h. bilde

$$T_0 = 0, \quad T_n = \min\{Y_k > T_{n-1} \mid k \in \mathbb{N}\}, \quad n \in \mathbb{N}.$$

Dann ist $\{T_n\}_{n\in\mathbb{N}_0}$ die Ankunftszeitenfolge eines inhomogenen Poisson–Prozesses mit Zeittransformation A.

Das gerade beschriebene Verfahren erzeugt also wieder in Schritt a) in allen Intervallen $(n-1, n]$, $n \in \mathbb{N}$, zunächst die *Anzahl* der Ankünfte N_n, in Schritt b) deren (ungeordnete) *Position* — sofern $N_n > 0$ gilt, also überhaupt Ankünfte

eintreten —, und schließlich in Schritt c) die zeitlich richtig geordnete Ankunftszeitenfolge durch Sortierung.

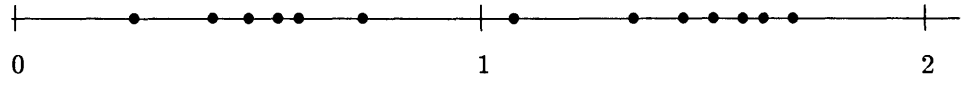
Will man beispielsweise die tageszeitlich unterschiedliche Auslastung von Rechnersystemen durch inhomogene Poisson–Prozesse beschreiben, so kann man etwa

$$A(t) = \int_0^t f(u)\,du, \quad t \geq 0, \tag{3.4.38}$$

wählen, wobei $f \geq 0$ z.B. eine stetige, periodische Funktion ist. Für $f(u) = 2\pi\lambda \cdot (1 - \cos(2\pi u))$, $u \geq 0$, erhält man beispielsweise

$$A(t) = \lambda \cdot \big(t - \sin(2\pi t)\big), \quad t \geq 0;$$

eine typische Realisation des zugehörigen inhomogenen Poisson–Prozesses ist in der folgenden Graphik wiedergegeben:



Für inhomogene Poisson–Prozesse läßt sich auch Satz 3.4.4 entsprechend formulieren.

Satz 3.4.8. *(Überlagerung und Aufteilung von inhomogenen Poisson–Prozessen) Es seien* $\{Y_t\}_{t\geq 0}$ *und* $\{Z_t\}_{t\geq 0}$ *voneinander unabhängige Poisson–Prozesse mit Zeittransformationen* λA *und* μA, $\lambda, \mu > 0$, *und Ankunftszeitenfolgen* $\{T_n\}_{n\in\mathbf{N}_0}$ *bzw.* $\{S_n\}_{n\in\mathbf{N}_0}$. *Die durch*

$$U_0 = 0, \quad U_n = \min\{T_k, S_k > U_{n-1} \mid k \in \mathbf{N}\}, \quad n \in \mathbf{N} \tag{3.4.39}$$

definierte Überlagerung von $\{T_n\}_{n\in\mathbf{N}_0}$ *und* $\{S_n\}_{n\in\mathbf{N}_0}$ *bildet dann die Ankunftszeitenfolge eines Poisson–Prozesses mit Zeittransformation* $(\lambda + \mu)A$.
Ist ferner $\{I_k\}_{k\in\mathbf{N}}$ *eine von* $\{N_t\}_{t\geq 0}$ *unabhängige Folge* $\mathfrak{B}(1,p)$*–verteilter Zufallsvariablen mit* $p = 1 - q \in (0,1)$, *und teilt man die Ankunftszeitenfolge* $\{T_n\}_{n\in\mathbf{N}_0}$ *zufällig auf gemäß*

$$U_0 = V_0 = 0, \quad \begin{matrix} U_n = \min\{T_k > U_{n-1} \mid I_k = 1,\ k \in \mathbf{N}\} \\ V_n = \min\{T_k > V_{n-1} \mid I_k = 0,\ k \in \mathbf{N}\}, \end{matrix} \quad (n \in \mathbf{N}) \tag{3.4.40}$$

d.h. wählt man aus der Folge $\{T_n\}_{n\in\mathbf{N}_0}$ *Ankunftszeiten unabhängig voneinander mit Wahrscheinlichkeit* p *(bzw.* q*) aus, so bilden* $\{U_n\}_{n\in\mathbf{N}_0}$ *und* $\{V_n\}_{n\in\mathbf{N}_0}$ *die Ankunftszeiten zweier stochastisch unabhängiger Poisson–Prozesse mit Zeittransformationen* pA *bzw.* qA.

Der Beweis dieses Satzes verläuft völlig analog zu dem des Satzes 3.4.4.

Betrachtet man die in (3.4.8) bzw. (3.4.37) durchgeführte Alternativ–Konstruktion Poisson'scher Prozesse weniger unter dem zeitlichen Aspekt als vielmehr im Hinblick auf die möglichen Punktmuster über der reellen Halbachse $\mathbf{R}^+$, so

gelangt man zwangsläufig zu dem Konzept der *Punktprozesse*. Es stellt sich die Frage, ob das vorgestellte Konstruktionsverfahren nicht ähnlich auch in höherdimensionalen Räumen, insbesondere dem $\mathbf{R}^m$, $m \in \mathbf{N}$, funktioniert, indem man die gesamte Grundmenge in geeignete disjunkte Stücke — etwa Intervalle der in (1.4.7) bis (1.4.10) betrachteten Art — zerlegt und dann in jedem Stück eine Poisson-verteilte Anzahl von Punkten nach einer von der Partitionsmenge abhängenden Verteilung erzeugt. Solche Modelle sind z.B. bei Bildverarbeitungsproblemen, etwa in der Computertomographie, von Interesse. Eine Schwierigkeit bei der Behandlung höherdimensionaler Räume liegt allerdings darin begründet, daß für $m > 1$ keine "natürliche" Ordnungsstruktur mehr gegeben ist, die für $m = 1$ ja gerade die Modellierung zeitlicher Phänomene so leicht ermöglichte. Andererseits induziert die Unabhängigkeit der Zuwächse von Poisson–Prozessen eine Additivitätseigenschaft, die der in Beziehung (1.1.30) formulierten σ–Additivität von Wahrscheinlichkeitsverteilungen nahekommt. Es hat sich herausgestellt, daß ein solcher Zugang zur Beschreibung von Punktprozessen sehr geeignet ist. Wir benötigen dazu allerdings den etwas allgemeineren Begriff eines Maßes.

Definition 3.4.5. *(allgemeines Maß)*
Es sei $(\Omega, \mathcal{A})$ ein Meßraum. Eine σ–additive Abbildung $\mu : \mathcal{A} \longrightarrow [0, \infty]$, d.h. mit der Eigenschaft

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \tag{3.4.41}$$

für jede Familie $\{A_n\}_{n\in\mathbf{N}}$ paarweise disjunkter Mengen heißt Maß auf $\mathcal{A}$. μ heißt σ–endlich, wenn es eine disjunkte Zerlegung von Ω, $\Omega = \bigcup_{n=1}^{\infty} B_n$, $\{B_n\}_{n\in\mathbf{N}} \subseteq \mathcal{A}$, gibt mit $\mu(B_n) < \infty$ für alle $n \in \mathbf{N}$.

Man beachte, daß sich aus der σ–Additivität von μ wie bei Wahrscheinlichkeitsmaßen entsprechend

$$\mu(\emptyset) = 0$$

ergibt. Die Menge aller Maße über einer σ–Algebra $\mathcal{A}$ bildet darüberhinaus einen *konvexen Kegel*, d.h. für Maße μ, ν und nichtnegative Zahlen α, β ist auch $\alpha\mu + \beta\nu$ ein Maß.

Eine Wahrscheinlichkeitsverteilung P auf $\mathcal{A}$ ist also in diesem Sinne gerade ein *normiertes* Maß. Man beachte, daß bei der Definition eines Maßes der Wert ∞ ausdrücklich zugelassen ist. Ein einfaches derartiges Maß ist bereits das *abzählende* Maß μ, welches definiert ist durch

$$\mu(A) = \#A, \quad A \in \mathcal{A};$$

μ ist offenbar genau dann nicht–endlich, wenn die Grundmenge Ω nicht–endlich ist. Es ist allerdings σ–endlich, wenn Ω höchstens abzählbar ist. Ein insbesondere auch in der Analysis wichtiges (ebenfalls σ–endliches) Maß ist das *Lebesgue*–Maß, welches durch die (1.0.8) einschließende Forderung

$$\mu\big((a,b]\big) = b - a \quad \text{für alle } a < b,\ a, b \in \mathbf{R} \tag{3.4.42}$$

auf $\mathcal{B}^1$ festgelegt ist und über die Beziehung

$$\mathcal{R}(A) = \frac{\mu(\cdot \cap A)}{\mu(A)}, \quad A \in \mathcal{B}^1,\ \mu(A) > 0,$$

mit der stetigen Gleichverteilung über einer Borel–Menge A in Zusammenhang steht.

Ähnlich wie bei Wahrscheinlichkeitsmaßen lassen sich auch für allgemeine Maße Produktmaße definieren; für $\mathcal{A} = \mathcal{B}^m$, $m \in \mathbb{N}$, ergibt sich damit das Lebesgue–Maß $\mu^{(m)}$ vermöge

$$\begin{aligned}\mu^{(m)}\Big(\bigtimes_{i=1}^{m}(a_i,b_i]\Big) &= \prod_{i=1}^{m}\mu\big((a_i,b_i]\big)\\ &= \prod_{i=1}^{m}(b_i-a_i), \quad a_i<b_i,\ a_i,b_i\in\mathbb{R},\ 1\le i\le m.\end{aligned} \tag{3.4.43}$$

Das Lebesgue–Maß ist damit die natürliche Erweiterung der Begriffe "Fläche" und "Volumen", die zunächst nur sinnvoll für Rechtecke bzw. Quader definiert sind, auf alle 2– bzw. 3–dimensionalen Borel'schen Mengen.

Analog zu dem Begriff der Dichte einer Wahrscheinlichkeitsverteilung läßt sich auch der Begriff der *Dichte eines Maßes* prägen.

Definition 3.4.6. *(Dichte eines Maßes)*
Es sei μ ein Maß auf $\mathcal{B}^m$, $m \in \mathbb{N}$, im Sinne von Definition 3.4.5. Man sagt, μ besitzte eine Dichte $f : \mathbb{R}^m \longrightarrow [0,\infty]$, wenn gilt:

$$\mu\Big(\bigtimes_{i=1}^{m}(a_i,b_i]\Big) = \int_{a_m}^{b_m}\cdots\int_{a_1}^{b_1} f(x_1,\ldots,x_m)\,dx_1\ldots dx_m \tag{3.4.44}$$

für alle $a_i < b_i$, $a_i, b_i \in \mathbb{R}$, $1 \le i \le m$.

Das Lebesgue–Maß besitzt also insbesondere die konstante Dichte $f \equiv 1$; allgemeiner ist ein Maß mit einer Dichte f σ–endlich, wenn f *endlich* ist.

Die grundlegende Idee zur Beschreibung von Punktprozessen besteht nun darin, diese als Zufallselemente ξ in der Menge der abzählend–diskreten Maße auf einer geeigneten σ–Algebra $\mathcal{B}$ eines Meßraums $(\mathcal{X},\mathcal{B})$ — typischerweise $\mathcal{X} = \mathbb{R}^m$, $\mathcal{B} = \mathcal{B}^m$, $m \in \mathbb{N}$ — aufzufassen, wobei man hierunter alle Maße μ zu verstehen hat, für die

$$\mu(B) \in \overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\} \quad \text{für alle } B \in \mathcal{B}$$

gilt. Für jede Menge $B \in \mathcal{B}$ ist dann $\xi(B)$ eine (gewöhnliche) Zufallsvariable; sie gibt anschaulich an, wieviele "Punkte" der Menge $\mathcal{X}$ in der Menge B liegen. Die σ–Additivität von ξ bedeutet hier, daß für jede Folge $\{B_n\}_{n\in\mathbb{N}}$ paarweise disjunkter Mengen

$$\xi\Big(\bigcup_{n=1}^{\infty}B_n\Big) = \sum_{n=1}^{\infty}\xi(B_n)$$

gilt, also die — zufällige — Gesamtzahl der in der disjunkten Vereinigung $\bigcup_{n=1}^{\infty} B_n$ befindlichen Punkte gerade die Summe der in den jeweiligen Mengen B_n, $n \in \mathbb{N}$, liegenden Punktanzahlen $\xi(B_n)$ ist. Die Beschreibung von Punktprozessen als abzählend–diskrete Maße hat also den Vorteil, die zufälligen Punktkonfigurationen in verschiedenen Teilmengen des Bildraums $\mathcal{X}$ simultan zu erfassen.

Als Wertebereich der Abbildungen ξ auf $(\Omega, \mathcal{A})$ wählt man dementsprechend die Menge $\mathcal{M}$ aller abzählend-diskreten Maße auf der σ-Algebra $\mathcal{B}$ des Bildraumes $\mathcal{X}$. Allerdings müssen wir noch eine geeignete σ-Algebra $\mathfrak{M}$ über $\mathcal{M}$ spezifizieren, damit ξ tatsächlich ein Zufallselement, also eine meßbare Abbildung von $(\Omega, \mathcal{A})$ nach $(\mathcal{M}, \mathfrak{M})$ ist. Diese σ-Algebra muß dabei so gewählt sein, daß für jede Menge $B \in \mathcal{B}$ die Abbildung $\xi(B)$ eine Zufallsvariable im üblichen Sinne ist. Wir betrachten dazu die *Evolutionsabbildungen* τ_B, $B \in \mathcal{B}$, auf $\mathcal{M}$, die definiert sind durch

$$\tau_B(\mu) = \mu(B), \quad B \in \mathcal{B},\ \mu \in \mathcal{M}. \tag{3.4.45}$$

Es bietet sich dann an, für $\mathfrak{M}$ die kleinste σ-Algebra über $\mathcal{M}$ zu wählen, bezüglich der alle Evolutionsabbildungen meßbar sind, d.h. die von dem Mengensystem

$$\mathfrak{E} = \{\tau_B^{-1}(E) \mid B \in \mathcal{B},\ E \in \mathcal{B}^1\} \subseteq \mathfrak{P}(\mathcal{M}) \tag{3.4.46}$$

über $\mathcal{M}$ erzeugte σ-Algebra $\sigma(\mathfrak{E})$.

Definition 3.4.7. *(Punktprozeß)*
Es seien $(\Omega, \mathcal{A})$, $(\mathcal{X}, \mathcal{B})$ Meßräume und $(\mathcal{M}, \mathfrak{M})$ der Meßraum der abzählend-diskreten Maße auf $\mathcal{B}$ im Sinne von (3.4.46). Ein Zufallselement $\xi : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{M}, \mathfrak{M})$ heißt dann Punktprozeß über $\mathcal{X}$.

Zur Vereinfachung der Darstellung wollen wir dabei anstatt $\xi(\omega)(B)$ oder $\xi(B)(\omega)$ stets $\xi(B, \omega)$ für $B \in \mathcal{B}$ bzw. $\omega \in \Omega$ schreiben.

Lemma 3.4.5. *(Eigenschaften von Punktprozessen)*
Es sei ξ ein Punktprozeß über $\mathcal{X}$ im Sinne von Definition 3.4.7. Dann gilt:
a) $\xi(B, \cdot)$ ist für jedes $B \in \mathcal{B}$ eine Zufallsvariable auf $(\Omega, \mathcal{A})$;
b) $\xi(\cdot, \omega)$ ist für jedes $\omega \in \Omega$ ein abzählend-diskretes Maß auf $\mathcal{B}$.

Beweis. a) Es sei $B \in \mathcal{B}$ fest. Wir zeigen:

$$A = \{\omega \in \Omega \mid \xi(B, \omega) \in E\} \in \mathcal{A} \quad \text{für alle } E \in \mathcal{B}^1.$$

Dies folgt aber aus der Darstellung $\xi(B, \omega) = \tau_B\big(\xi(\cdot, \omega)\big)$, $\omega \in \Omega$, also

$$A = \{\omega \in \Omega \mid \xi(\cdot, \omega) \in \tau_B^{-1}(E)\} = \xi^{-1}\big(\underbrace{\tau_B^{-1}(E)}_{\in \mathfrak{M}}\big) \in \mathcal{A}$$

nach Konstruktion von $\mathfrak{M}$.
b) Dies ergibt sich unmittelbar aus der Definition eines Punktprozesses. ∎

Für die Beschreibung von Punktprozessen ξ ist u.a. die erwartete Anzahl $E\big(\xi(B)\big)$ von Punkten in den Mengen $B \in \mathcal{B}$ von Interesse, da diese Größen einen Eindruck von der durchschnittlichen Konzentration der Punkte über dem Raum $\mathcal{X}$ geben.

Lemma 3.4.6. *(Intensitätsmaß)*
Es sei ξ ein Punktprozeß über $\mathcal{X}$ im Sinne von Definition 3.4.7. Dann ist die Abbildung

$$E\xi : \mathcal{B} \longrightarrow [0, \infty] : B \mapsto E\big(\xi(B)\big) \tag{3.4.47}$$

ein Maß auf $\mathcal{B}$. $E\xi$ heißt Intensitätsmaß von ξ.

Beweis. Zu zeigen ist lediglich die σ–Additivität von $E\xi$. Sei hierzu $\{B_n\}_{n\in\mathbb{N}}$ eine paarweise disjunkte Mengenfolge aus $\mathcal{B}$. Mit $C_n = \bigcup_{i=1}^{n} B_i$, $n \in \mathbb{N}$, gilt $C_n \uparrow C_\infty = \bigcup_{i=1}^{\infty} B_i$ und daher wegen der σ–Additivität von ξ auch $0 \le \xi(C_n) \uparrow \xi(C_\infty)$, $n \to \infty$. Nach dem Satz von der monotonen Konvergenz (2.2.23) folgt demnach

$$\begin{aligned} E\xi\Big(\bigcup_{i=1}^{\infty} B_i\Big) &= E\big[\lim_{n\to\infty} \xi(C_n)\big] \\ &= \lim_{n\to\infty} E\big[\xi(C_n)\big] = \lim_{n\to\infty} E\Big[\sum_{i=1}^{n} \xi(B_i)\Big] \\ &= \sum_{i=1}^{\infty} E\big[\xi(B_i)\big] = \sum_{i=1}^{\infty} E\xi(B_i), \end{aligned}$$

was zu zeigen war. ∎

Allerdings ist i.a. $E\xi$ kein abzählend–diskretes Maß mehr wie $\xi(\cdot,\omega)$, $\omega \in \Omega$.

Will man nun die Konstruktion Poisson'scher Prozesse gemäß (3.4.8) oder (3.4.37) auf beliebige Meßräume $(\mathcal{X},\mathcal{B})$ übertragen, so stellt man leicht fest, daß hierzu lediglich die Angabe eines geeigneten Intensitätsmaßes erforderlich ist, wenn man Poisson–Prozesse allgemein wie folgt definiert.

Definition 3.4.8. *(Poisson'scher Punktprozeß)*
Es sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum und $(\mathcal{X},\mathcal{B})$ ein Meßraum; μ sei ein Maß auf $\mathcal{B}$. $(\mathcal{M},\mathfrak{M})$ bezeichne wieder den Meßraum der abzählend–diskreten Maße auf $\mathcal{B}$. Ein Punktprozeß $\xi : (\Omega,\mathcal{A}) \longrightarrow (\mathcal{M},\mathfrak{M})$ heißt Poisson–Prozeß mit Intensitätsmaß μ, wenn gilt:

a) Für jede Menge $B \in \mathcal{B}$ ist die Zufallsvariable $\xi(B)$ Poisson–verteilt[1] *mit Parameter $\mu(B)$;*

b) für jede Familie paarweise disjunkter Mengen $\{B_n\}_{n\in\mathbb{N}} \subseteq \mathcal{B}$ ist die Folge der Zufallsvariablen $\{\xi(B_n)\}_{n\in\mathbb{N}}$ stochastisch unabhängig (Unabhängigkeit der Zuwächse).

Mit maßtheoretischen Methoden läßt sich zeigen, daß durch diese beiden Forderungen die Verteilung P^ξ des Punktprozesses eindeutig bestimmt ist (vgl. Kallenberg (1986), Daley & Vere–Jones (1988) oder Karr (1986)). Das folgende Ergebnis zeigt, wie man bei σ–endlichem Intensitätsmaß Poisson–Prozesse leicht konstruieren kann.

Lemma 3.4.7. *(Konstruktion von Poisson'schen Punktprozessen)*
Es sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum und $(\mathcal{X},\mathcal{B})$ ein Meßraum; $\mu \neq 0$ sei ein σ–endliches Maß auf $\mathcal{B}$. $\{B_n\}_{n\in I} \subseteq \mathcal{B}$ sei eine höchstens abzählbare disjunkte Zerlegung von $\mathcal{X}$ mit[2] *$0 < \mu(B_n) < \infty$ für alle $n \in I$, wobei $I \subseteq \mathbb{N}$ eine geeignete Indexmenge sei. Ferner seien N_n, $n \in I$, unabhängige, jeweils $\mathfrak{P}\big(\mu(B_n)\big)$–verteilte*

---

[1] Für $\mu(B)=0$ bedeute dies $P(\xi(B)=0)=1$, für $\mu(B)=\infty$ $P(\xi(B)=\infty)=1$.

[2] Dies kann durch Vereinigung von Mengen B_n mit $\mu(B_n)=0$ und Mengen B_m mit $\mu(B_m)>0$ stets erreicht werden.

Zufallsvariablen und $\{X_{nk}\}_{n\in I,k\in\mathbf{N}}$ *(auch von* $\{N_n\}_{n\in I}$*) unabhängige Zufallselemente mit Werten in* $(\mathcal{X},\mathcal{B})$ *und Verteilung*

$$P^{X_{nk}} = \frac{\mu(\cdot\cap B_n)}{\mu(B_n)}, \quad n\in I, k\in\mathbf{N}. \tag{3.4.48}$$

Dann ist der durch

$$\xi(B) = \sum_{n\in I}\sum_{k=1}^{N_n} \mathbb{1}_{\{X_{nk}\in B\}}, \quad B\in\mathcal{B}, \tag{3.4.49}$$

definierte Punktprozeß ein Poisson–Prozeß mit Intensitätsmaß $E\xi = \mu$.

Beweis. Wir zeigen die Eigenschaften a) und b) aus Definition 3.4.8.
a) Nach Lemma 3.4.2 und (3.4.48) ist für jedes $n\in I$ die Zufallsvariable $Y_n = \sum_{k=1}^{N_n}\mathbb{1}_{\{X_{nk}\in B\}}$ Poisson–verteilt mit Parameter

$$E(Y_n) = E(N_n)\cdot P(X_{nk}\in B) = \mu(B_n)\frac{\mu(B\cap B_n)}{\mu(B_n)} = \mu(B\cap B_n).$$

Nach Konstruktion sind die Y_n, $n\in I$, stochastisch unabhängig; wegen der Faltungsstabilität der Poisson–Verteilung (Lemma 2.1.9) ist damit jede der Zufallsvariablen $S_m = \sum_{i\in\mathcal{I}_m} Y_i$, $I_m = I\cap\{1,2,\dots,m\}$, $m\in\mathbf{N}$, Poisson–verteilt mit Parameter

$$\lambda_m = \sum_{i\in\mathcal{I}_m}\mu(B_i\cap B),\ m\in\mathbf{N}.$$

Für endliches I folgt damit die Behauptung. Anderenfalls können wir o.B.d.A. $I=\mathbf{N}$ annehmen. Nach Voraussetzung konvergiert λ_m mit $m\to\infty$ gegen

$$\lambda_\infty = \sum_{i=1}^{\infty}\mu(B_i\cap B) = \mu\Big(\bigcup_{i=1}^{\infty} B_i\cap B\Big) = \mu(\Omega\cap B) = \mu(B)\in(0,\infty);$$

demnach konvergiert die Folge $\{S_m\}_{m\in\mathbf{N}}$ mit $m\to\infty$ schwach gegen $\xi(B)$ mit Verteilung $\mathfrak{P}(\lambda_\infty) = \mathfrak{P}(\mu(B))$. Hieraus folgt die Behauptung.
b) Nach Lemma 3.4.2 sind für $C\in\mathcal{B}$ und alle $n\in\mathbf{N}$ die Zufallsvariablen $\sum_{k=1}^{N_n}\mathbb{1}_{\{X_{nk}\in C\}}$ und $\sum_{k=1}^{N_n}\mathbb{1}_{\{X_{nk}\in C^c\}}$ stochastisch unabhängig, also nach Summation auch $\xi(C)$ und $\xi(C^c)$. Für abzählbar viele paarweise disjunkte Ereignisse $\{C_n\}_{n\in\mathbf{N}}\subseteq\mathcal{B}$ argumentiert man analog unter mehrfacher Heranziehung von Lemma 3.4.2 und Lemma 2.1.7. ∎

Lemma 3.4.7 umfaßt die durch Beziehung (3.4.8) gegebene Konstruktion eines homogenen Poisson–Prozesses ξ für $(\mathcal{X},\mathcal{B}) = (\mathbf{R}^+,\mathbf{R}^+\cap\mathcal{B}^1)$ mit dem Intensitätsmaß $E\xi = \mu_{|\mathcal{B}}$, wobei $\mu_{|\mathcal{B}}$ das auf die σ–Algebra $\mathcal{B}$ eingeschränkte Lebesgue–Maß bezeichne. Entsprechendes gilt für die Konstruktion in (3.4.37): hier ist das Intensitätsmaß $E\xi$ gegeben durch die Zeittransformation

$$E\xi\big((a,b]\big) = A(b) - A(a), \quad 0\le a<b,\ a,b\in\mathbf{R}^+$$

(vgl. auch (3.4.36)).

Besonders einfach wird die Konstruktion eines Poisson'schen Punktprozesses ξ bei *endlichem* Intensitätsmaß $\mu = E\xi$ mit $\lambda = \mu(\Omega) > 0$, da man hier $I = \{1\}$ wählen kann. Es genügt in diesem Fall die Angabe einer $\mathfrak{P}(\lambda)$-verteilten Zufallsvariablen N und einer davon unabhängigen Folge $\{X_n\}_{n\in\mathbb{N}}$ mit Werten in $(\mathcal{X}, \mathcal{B})$ und Verteilung $P^{X_n} = \frac{1}{\lambda}\mu$; der gewünschte Prozeß besitzt dann die Darstellung

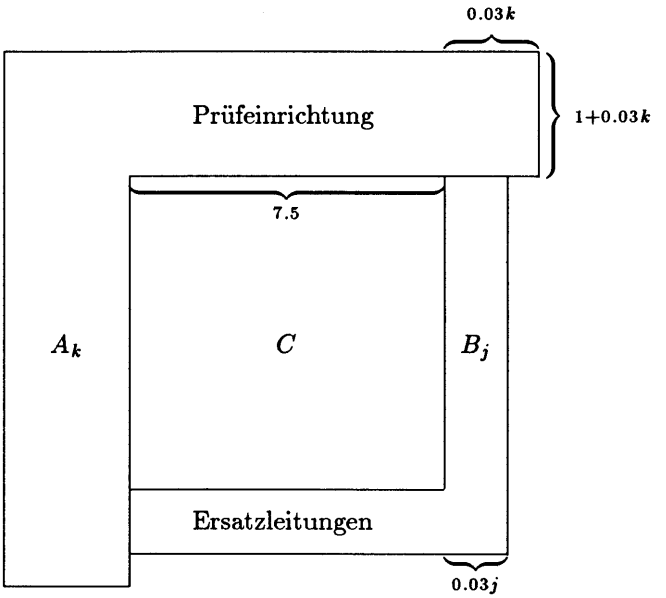
$$\xi(B) = \sum_{n=1}^{N} \mathbb{1}_{\{X_n \in B\}}, \quad B \in \mathcal{B}.$$

Die Konstruktion in Lemma 3.4.7 ist demnach aufgrund der Eigenschaft b) der Definition 3.4.8 nichts anderes als eine Zusammensetzung voneinander stochastisch unabhängiger Poisson–Prozesse $\xi_n = \xi(\cdot \cap B_n)$, $n \in \mathbb{N}$, mit den in (3.4.48) verwendeten *endlichen* Intensitätsmaßen $E\xi_n = \mu(\cdot \cap B_n)$, $n \in \mathbb{N}$, auf den disjunkten Teilstücken B_n von Ω.

Das folgende Beispiel ist im wesentlichen Weiß (1987), Beispiel 4.41, entnommen.

Beispiel 3.4.3. (Fehlertolerante Speicherbausteine)
Bei der Produktion von Speicherbausteinen für Computer werden auf einem Siliziumplättchen (Chip) quadratische Gitter von sich kreuzenden Schreib–Leseleitungen mit je einem speichernden Element in den Kreuzungspunkten angebracht. Der gesamte Baustein arbeitet bereits dann fehlerhaft, wenn in nur einem dieser Speicherelemente Fremdkörper im Silizium sitzen. Zur Reduktion des Prüfaufwands werden i.a. die Bausteine mit selbständig arbeitenden Prüfeinrichtungen versehen. Aus praktischen Erfahrungen weiß man, daß bei den heutigen Fertigungsmethoden mit durchschnittlich vier Fremdkörpereinschlüssen je cm^2 Fläche zu rechnen ist, also ein großer Teil der Fertigung praktisch Ausschuß darstellt. Man geht daher zunehmend zu "selbstreparierenden" Speicherbausteinen über. Anstelle der ursprünglich $2s$ Schreib–Leseleitungen verwendet man $2(s+k)$ Schreib–Leseleitungen ($s, k \in \mathbb{N}$) und gestaltet die Prüfeinrichtung so, daß diese stets jene zwei Leitungen, welche sich in einem als fehlerhaft erkannten Speicherelement kreuzen, automatisch durch zwei noch nicht verwendete Ersatzleitungen ersetzt. Falls die im Bereich der Ersatzleitungen vorhandenen Speicherelemente fehlerfrei arbeiten, kann man so bis zu k Fehlern automatisch beheben; defekte Ersatzleitungen können allerdings nicht mehr ersetzt werden.
Da mit zunehmendem k sowohl der Platz für die Prüfeinrichtung (die stets fehlerfrei arbeiten soll) als auch der Platz für die Ersatzleitungen zunimmt, darf man k nicht beliebig groß machen. Wie groß sollte k sein, um bei einem quadratischen Chip mit den angegebenen Abmessungen (in mm) und einem Platzbedarf von 0.03 mm je zusätzlicher Leitung möglichst wenig Ausschuß zu produzieren?
Zur Beantwortung der Frage nehmen wir an, daß die zufälligen Punkte, in denen Fremdkörpereinschlüsse vorliegen, einen homogenen Poisson–Prozeß ξ mit Intensitätsmaß $\nu = E\xi = 0.04\mu$ bilden, wobei μ das zweidimesionale Lebesgue–Maß (vgl. (3.4.43)) bezeichne. Für jede Borel–Menge $B \in \mathcal{B}^1$ mit $\mu(B) = 100$ (mm^2) ist dann $\nu(B) = E\xi(B) = 4$, d.h. die erwartete Zahl von Punkten je 100 mm^2 Fläche beträgt gerade 4. Bezeichnet A_k die durch k Prüfleitungen und B_j die durch $j \le k$ Ersatzleitungen beanspruchte Fläche sowie C die durch die ursprünglich gegebenen

$2s$ Schreib–Leseleitungen beanspruchte Fläche, so ergibt sich mit den angegebenen Abmessungen (in mm)



$$\begin{aligned}\mu(A_k) &= 16 + 0.57k + 0.0027k^2\\ \mu(B_j) &= 0.45j + 0.0009j^2 \qquad (\text{mm}^2).\\ \mu(C) &= 56.25\end{aligned}$$

Das Ereignis E_k, daß der fehlertolerante Baustein mit $2k$ Ersatzleitungen fehlerfrei arbeitet, läßt sich dann wie folgt darstellen:

$$E_k = \underbrace{\{\xi(A_k) = 0\}}_{\substack{\text{Prüfeinrichtung}\\ \text{fehlerfrei}}} \cap \bigcup_{j=0}^{k} \Big[\underbrace{\{\xi(C) = j\}}_{\substack{j \text{ Einschlüsse}\\ \text{in } C}} \cap \underbrace{\{\xi(B_j) = 0\}}_{\substack{j \text{ Ersatzleitungen}\\ \text{fehlerfrei}}} \Big],$$

d.h. es gilt aufgrund der Unabhängigkeit der Zuwächse des Poisson–Prozesses

$$\begin{aligned}P(E_k) &= P\Big(\{\xi(A_k) = 0\} \cap \bigcup_{j=0}^{k} \big[\{\xi(C) = j\} \cap \{\xi(B_j) = 0\}\big]\Big)\\ &= \exp\big(-\nu(A_k)\big)\Big[\sum_{j=0}^{k} \exp\big(-\nu(C)\big)\frac{(\nu(C))^j}{j!}\exp\big(-\nu(B_j)\big)\Big].\end{aligned}$$

Die folgende Tabelle enthält die entsprechenden Werte für $k = 0, 1, \ldots, 8$:

k	0	1	2	3	4	5	6	7	8
$P(E_k)$	0.056	0.174	0.300	0.386	0.428	**0.439**	0.437	0.429	0.419

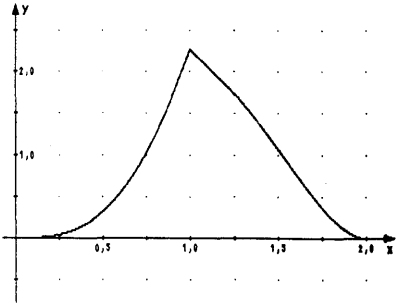
Der günstigste Fall wird also bei $k = 5$ erreicht, d.h. bei 10 zusätzlichen Ersatzleitungen. Der Prozentsatz an Ausschuß verringert sich damit von ursprünglich $100 \cdot P(E_0^c) = 94.4\%$ auf $100 \cdot P(E_5^c) = 56.1\%$! ∎

Beispiel 3.4.4. (Computertomographie)
Bei der Positronen-Emissions-Tomographie (PET) wird einem Patienten kurzlebiges radioakives (neutronenarmes) Material injiziert, welches sich in einem bestimmten Organ anreichert und näherungsweise gemäß einem dreidimensionalen Poisson'schen Punktprozeß zerfällt. Die beim Zerfall des Materials freiwerdenden Positronen vereinigen sich nach kurzer Abbremsung im Gewebe je mit einem Elektron; beide Teilchen wandeln dabei ihre Masse in Strahlungsenergie um, d.h. sie zerstrahlen unter Entstehung zweier Gammaquanten, die einander entgegengesetzt auseinanderfliegen und beide die gleiche Energie von 511 keV haben. Werden diese beiden Gammaquanten mit zwei Detektoren (die üblicherweise außerhalb des Körpers ringförmig um den Patienten angeordnet sind) in zeitlicher Koinzidenz nachgewiesen, so weiß man, daß das Zerfallsereignis auf der Verbindungslinie der beiden Detektoren stattgefunden hat[1)]. Die so erhaltenen Daten lassen sich dann schichtweise zu Schnittbildern des Organs zusammensetzen. Die innerhalb dünner disjunkter Schnitte freigesetzten Positronen bilden dabei näherungsweise voneinander unabhängige, zweidimensionale Poisson-Prozeße ξ mit einem Intensitätsmaß $E\xi = \mu$, welches eine Dichte f besitzt, die die Massenverteilung des Organs in dem Schnitt repräsentiert; d.h. färbt man eine Fläche mit Grauwerten proportional zu der Dichte f ein, so erhält man ein Bild des entsprechenden Schnitts.
Die folgenden Simulationen zeigen Realisationen eines Poisson'schen Punktprozesses für ein fiktives Organ mit einem Intensitätsmaß mit der rotationssymmetrischen Dichte $f(x,y) = c\frac{g(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}$ (vgl. (2.1.98)), wobei $c > 0$ eine Konstante und g die Dichte der Zufallsvariablen $X = 1 - \sqrt{U} + \sqrt[4]{V}$ mit unabhängigen, je $\mathcal{R}([0,1])$-verteilten Zufallsvariablen bezeichne, die aufgrund des Faltungslemmas 2.1.11 gegeben ist durch

$$g(x) = \begin{cases} \frac{8}{3}x^3 - \frac{2}{5}x^5 & \text{für } 0 \le x \le 1 \\ \frac{64}{15} - 2x - \frac{8}{3}(x-1)^3 + 2x(x-1)^4 - \frac{8}{5}(x-1)^5 & \text{für } 1 \le x \le 2 \\ 0 & \text{sonst.} \end{cases}$$

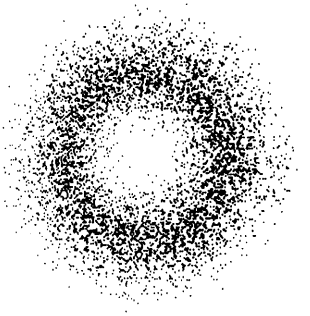
[1)] nach Heiss, Herholz, Pawlik, Szelies, Wagner & Wienhard (1988). PET wird z.B. zur Messung des Glukosestoffwechsels im Gehirn eingesetzt; siehe ebenda.

Da für $c = 1$ f gerade eine Wahrscheinlichkeitsdichte darstellt, erhält man also $E\xi(\mathbf{R}^2) = c$, d.h. c ist die erwartete Gesamtanzahl von Punkten (Zerfällen) des Poisson–Prozesses.

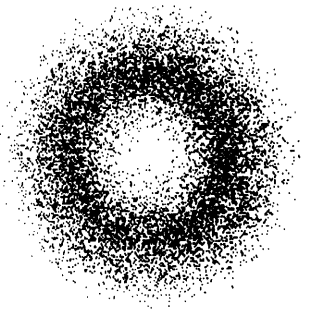


Skizze der Dichte g

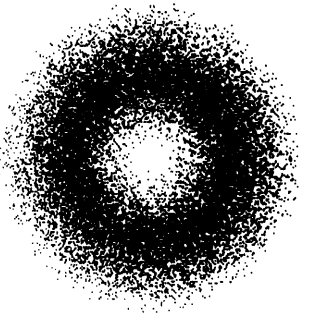
$c = 3000$



$c = 6000$



$c = 12000$



$c = 24000$

Mit statistischen Methoden läßt sich zeigen, daß durch die Beobachtung solcher Punktkonfigurationen die Dichte f des zugrundeliegenden Intensitätsmaßes — und damit das Schnittbild des Organs — konsistent geschätzt werden kann. Allerdings ergeben sich gewisse Schwierigkeiten dadurch, daß man aufgrund der Datenerfassung nicht den Punktprozeß selbst beobachten kann, sondern lediglich Randsum-

menhäufigkeiten für die einzelnen Schnitte. Zur Schätzung des Bildes werden daher aufwendigere Verfahren benötigt; vgl. hierzu etwa die Arbeiten von Geman & McClure (1987) oder Vardi, Shepp & Kaufman (1985). ∎

Ähnlich wie bei den eindimensionalen Poisson–Prozessen lassen sich auch allgemeine Poisson'sche Punktprozesse überlagern bzw. zufällig aufteilen.

Satz 3.4.9. *(Überlagerung und Aufteilung von Punktprozessen)*

a) *Es seien mit den Bezeichnungen aus Definition 3.4.7 ξ_1 und ξ_2 voneinander unabhängige Poisson'sche Punktprozesse über $\mathcal{X}$ mit Intensitätsmaßen $E\xi_1$ und $E\xi_2$. Dann ist auch die Überlagerung $\xi = \xi_1 + \xi_2$ ein Poisson'scher Punktprozeß über $\mathcal{X}$ mit Intensitätsmaß $E\xi = E\xi_1 + E\xi_2$.*

b) *Ist ξ ein Poisson'scher Punktprozeß mit der Darstellung (3.4.49) und sind mit den dortigen Bezeichnungen die Zufallsvariablen $\{J_{nk}\}_{n\in I, k\in\mathbb{N}}$ (auch von ξ) stochastisch unabhängig und identisch $\mathfrak{B}(1,p)$–verteilt mit $p \in (0,1)$, so sind die Punktprozesse ξ_1 und ξ_2, definiert durch*

$$\xi_1(B) = \sum_{n\in I}\sum_{k=1}^{N_n} J_{nk}\cdot \mathbb{1}_{\{X_{nk}\in B\}} \qquad B \in \mathcal{B},$$

$$\xi_2(B) = \sum_{n\in I}\sum_{k=1}^{N_n} (1-J_{nk})\cdot \mathbb{1}_{\{X_{nk}\in B\}},$$

voneinander unabhängige Poisson–Prozesse mit Intensitätsmaßen

$$E\xi_1 = p\cdot E\xi, \quad E\xi_2 = (1-p)\cdot E\xi.$$

Beweis. a) Für jede Menge $B \in \mathcal{B}$ mit $0 < E\xi_1(B) < \infty$, $0 < E\xi_2(B) < \infty$ sind $\xi_1(B)$ und $\xi_2(B)$ stochastisch unabhängige, $\mathfrak{P}\left(E\xi_1(B)\right)$– bzw. $\mathfrak{P}\left(E\xi_2(B)\right)$–verteilte Zufallsvariablen, d.h. $\xi(B)$ ist ebenfalls Poisson–verteilt mit Parameter $E\xi_1(B) + E\xi_2(B)$. Damit ist die Eigenschaft a) der Definition 3.4.8 erfüllt. Die Eigenschaft b) folgt unmittelbar aus der Unabhängigkeit der Zuwächse von ξ_1 und ξ_2 sowie Lemma 2.1.7.

b) Dies folgt aus Lemma 3.4.2, da die Anzahlen der Punkte des Prozesses ξ in den Zerlegungsmengen B_n, $n \in \mathbb{N}$, jeweils Poisson–verteilt sind mit Parameter $E\xi(B_n) = \mu(B_n)$. ∎

Der letzte Satz ist z.B. in der Computertomographie von Bedeutung, da in dem Ringempfänger durch Streuung und Absorption von Elementarteilchen lediglich "zufällige" Auswahlen der ursprünglich vorhandenen Positronen registriert werden können. Aufgrund des Teils b) von Satz 3.4.9 bleibt aber der tatsächlich gemessene Punktprozeß ein Poisson–Prozeß, allderdings mit einem um einen bestimmten Faktor geringeren Intensitätsmaß. Dadurch wird jedoch die Bildanalyse nicht wesentlich beeinflußt, da das Verhältnis der Grauwerte untereinander konstant bleibt (vgl. hierzu auch die in der obigen Simulation erhaltenen Bilder mit dem variablen Parameter c, der gerade diesen Effekt verdeutlicht).

Neuere Methoden der Analyse speziell digitaler Bilder werden auch in Ripley (1988), Kapitel 5.4 (Simulated Annealing), und 6 behandelt.

3.5. Aufgaben

3.1 Es seien X und Y stochastisch unabhängige, $\mathfrak{B}(n,p)$- bzw. $\mathfrak{B}(m,p)$-verteilte Zufallsvariablen mit $0 < p < 1$ und $n, m \in \mathbf{N}$. Zeigen Sie:
Die bedingten Verteilungen $P^X(\cdot \mid X+Y=r)$ bzw. $P^Y(\cdot \mid X+Y=r)$, $0 \le r \le n+m$, sind jeweils hypergeometrische Verteilungen (vgl. Aufgabe 2.3), d.h. es gilt

$$P(X=k \mid X+Y=r) = \frac{\binom{n}{k}\binom{m}{r-k}}{\binom{n+m}{r}}, \quad P(Y=j \mid X+Y=r) = \frac{\binom{m}{j}\binom{n}{r-j}}{\binom{n+m}{r}}$$

für $0 \le k \le n$, $0 \le j \le m$, unabhängig von p. Interpretieren sie dieses Ergebnis kombinatorisch.

3.2 Es seien X und Y stochastisch unabhängige, $\overline{\mathfrak{B}}^+(n,p)$- bzw. $\overline{\mathfrak{B}}^+(m,p)$-verteilte Zufallsvariablen mit $0 < p < 1$ und $n, m \in \mathbf{N}$. Zeigen Sie:
Die bedingten Verteilungen $P^X(\cdot \mid X+Y=r)$ bzw. $P^Y(\cdot \mid X+Y=r)$, $r \in \mathbf{N}_0$, sind gegeben durch

$$P(X=k \mid X+Y=r) = P(Y=k \mid X+Y=r) = \frac{n+m-1}{r+n+m-1} \frac{\binom{r}{k}\binom{n+m-2}{n-1}}{\binom{r+n+m-2}{k+n-1}}$$

für $0 \le k \le r$. Was ergibt sich hier für $n = m = 1$, was für $n = 1$, $m = 2$? Geben sie eine kombinatorische Deutung für das Ergebnis.

3.3 X, Y seien Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit einer gemeinsamen Dichte $f_{(X,Y)}$. Zeigen Sie die Gültigkeit der folgenden stetigen Version von Lemma 3.1.1:

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x \mid Y=y)\, dP^Y(y)$$

$$f_Y(y \mid X=x) = \frac{f_X(x \mid Y=y) f_Y(y)}{\int_{-\infty}^{\infty} f_X(x \mid Y=z)\, dP^Y(z)} \quad P^X\text{-f.s.}$$

Zeigen Sie hiermit für den Fall $P^X(\cdot \mid Y=y) = \mathcal{E}(y)$, $y > 0$, und $P^Y = \mathcal{E}(\lambda)$, daß $P^Y(\cdot \mid X=x) = \mathcal{E}(2, x+\lambda)$, $x > 0$, gilt.

3.4 Es seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Zeigen Sie mit Hilfe des Ersetzungslemmas 3.1.4:

$$E(X \cdot Y \mid Y=y) = y \cdot E(X \mid Y=y) \quad P^Y\text{-f.s.} \quad \text{bzw.} \quad E(X \cdot Y \mid Y) = Y \cdot E(X \mid Y) \quad P\text{-f.s.}$$

Zeigen Sie hiermit für das in Aufgabe 3.3 angegebene Verteilungsbeispiel:

$$E(X \cdot Y \mid Y) = 1, \quad E(X \cdot Y \mid X) = \frac{2X}{X+\lambda}, \quad E(X \cdot Y) = 1 \quad P\text{-f.s.}$$

Existiert hier $\mathrm{Kov}(X,Y)$ bzw. $\mathrm{Korr}(X,Y)$?

3.5 $\{X_n\}_{n \in \mathbf{N}_0}$ sei eine homogene Markoff-Kette mit Zustandsraum $\mathcal{S} = \{1,2,3,4\}$ und Übergangsmatrix

$$\mathbf{\Pi} = \begin{pmatrix} 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0.2 & 0.8 \\ 0.7 & 0.3 & 0 & 0 \end{pmatrix}.$$

Bestimmen Sie alle stationären Anfangsverteilungen und $\lim_{n \to \infty} \mathbf{\Pi}^n$.

3.6 In Verallgemeinerung von Beispiel 3.2.4 betrachte man die Irrfahrt auf dem d-dimensionalen Gitter $\mathcal{S} = \mathbf{Z}^d = \{(i_1, \ldots, i_d) \mid i_1, \ldots, i_d \in \mathbf{Z}\}$, $d \geq 2$, mit Übergangswahrscheinlichkeiten

$$p_{(i_1,\ldots,i_d),(j_1,\ldots,j_d)} = \begin{cases} \dfrac{1}{2d} & \text{für } \sum_{\ell=1}^{d} |i_\ell - j_\ell| = 1 \\ 0 & \text{sonst} \end{cases}.$$

Zeigen Sie: Für $d = 2$ sind alle Zustände null-rekurrent, für $d \geq 3$ sind alle Zustände transient.

3.7 Zur Erhöhung der Ausfallsicherheit sind in einer elektronischen Schaltung zwei gleichartige Bauteile eingebaut, von denen jedes die Funktion des anderen übernehmen kann. Wir betrachten das System zu äquidistanten Zeittakten. Fällt ein Bauteil aus, wird zu Beginn des nächsten Zeittakts auf das andere umgeschaltet und das erste ersetzt. Die Ausfallwahrscheinlichkeit für ein Bauteil in Betrieb während eines Zeittakts betrage p, $0 < p < 1$. Die Ersetzung eines defekten Bauteils dauert genau zwei Zeittakte. Die zu Beginn eines Zeittakts möglichen Zustände lassen sich durch Paare (i, j) beschreiben, wobei $i \in \mathbf{N}_0$ die Anzahl intakter Bauteile und $j \in \mathbf{N}_0$ die Anzahl der zur Instandsetzung eines defekten Bauteils verwendeten Zeiteinheiten bedeuten, also $\mathcal{S} = \{(2,0),(1,0),(1,1),(0,1)\}$.
Bestimmen Sie die Übergangswahrscheinlichkeiten für die einzelnen Zustände, wenn

a) defekte Bauteile nur nacheinander ausgetauscht werden können.

b) defekte Bauteile auch gleichzeitig ersetzt werden können.

Die Schaltung sei sehr lange in Betrieb. Mit welcher Wahrscheinlichkeit (in Anhängigkeit von p) ist sie unter Instandsetzungsregel a) bzw. b) funktionsfähig (mindestens ein Bauteil intakt) ?

3.8 Eine Nachrichtenquelle erzeuge eine unabhängige Folge $\{X_n\}_{n\in\mathbf{N}}$ je $\mathfrak{B}(1,p)$-verteilter Zufallsvariablen mit $p \in [0,1]$. $S = \inf\{n \geq 2 \mid X_n = X_{n-1}\}$ sei der Zeitpunkt der erstmaligen Wiederholung eines Signals.
Geben Sie ein homogenes Markoff-Modell an, in dem S die erste Eintrittszeit in einen geeigneten absorbierenden Zustand ist, und zeigen Sie damit: $E(S) = \dfrac{2 + p(1-p)}{1 - p(1-p)}$. Zeigen Sie ferner, daß $E(S)$ maximal ist für $p = \frac{1}{2}$ mit Wert $E(S) = 3$.

3.9 Ein Zentralrechner kann von $n \in \mathbf{N}$ Arbeitsplätzen aus mit Programmen beschickt werden. Es sei angenommen, daß jeder Arbeitsplatz unabhängig von den übrigen *genau ein* Programm nach einer $\mathcal{E}(\lambda)$-exponentialverteilten Bearbeitungszeit an die Zentraleinheit überstellt. Zeigen Sie:

a) Die Ankunftszeiten $T_1, \ldots, T_n$ bilden die Ordnungsstatistiken unabhängiger, $\mathcal{E}(\lambda)$-verteilter Zufallsvariablen.

b) Die Zwischenankunftszeiten $\Delta_1 = T_1, \Delta_2 = T_2 - T_1, \ldots, \Delta_n = T_n - T_{n-1}$ sind unabhängige, exponentialverteilte Zufallsvariablen mit $P^{\Delta_k} = \mathcal{E}\big((n-k+1)\lambda\big)$ (Hinweis: Transformationssatz 2.1.10).

c) Der durch

$$N_t = \#\{k \in \mathbf{N} \mid T_k \leq t\}, \quad t \geq 0,$$

definierte stochastische Prozeß ist ein homogener Markoff-Prozeß mit Intensitätsmatrix $\boldsymbol{Q} = \{q_{ij}\}$, gegeben durch

$$q_{ij} = \begin{cases} (n-i)\lambda & \text{für } j = i+1 \leq n \\ -(n-i)\lambda & \text{für } j = i < n \\ 0 & \text{sonst,} \end{cases} \qquad i \in \mathbf{N}_0$$

(Hinweis: Ausführungen im Anschluß an den Struktursatz 3.4.5).

d) Für $0 \leq a < b$ ist $N_{(a,b]} = N_b - N_a$ $\mathfrak{B}\big(n, e^{-\lambda a} - e^{-\lambda b}\big)$-binomialverteilt (d.h. $\{N_t\}_{t\geq 0}$ ist *kein* Poisson-Prozeß); das Intensitätsmaß μ des zugehörigen Punktprozesses ist gegeben durch

$$\mu\big((a,b]\big) = \begin{cases} n\big(e^{-\lambda a} - e^{-\lambda b}\big) & \text{für } 0 \leq a < b \\ 0 & \text{sonst.} \end{cases}$$

e) Für die Ankunftszeit T_n des letzten Programms im Zentralrechner gilt:

$$E(T_n) = \frac{1}{\lambda}\sum_{k=1}^{n}\frac{1}{k} \approx \frac{1}{\lambda}\ln n \qquad \mathrm{Var}(T_n) = \frac{1}{\lambda^2}\sum_{k=1}^{n}\frac{1}{k^2} \approx \frac{\pi^2}{6\lambda^2}.$$

f) Es sei $\lambda = \frac{\nu}{n}$, $\nu > 0$. Zeigen Sie, daß sich für große n der Prozeß $\{N_t\}_{t\geq 0}$ näherungsweise wie ein inhomogener Poisson-Prozeß $\{N_t^*\}_{t\geq 0}$ mit dem in d) angegebenen Intensitätsmaß μ verhält, d.h. es gilt für alle $0 \leq a < b$:

$$\rho\Big(P^{N_{(a,b]}}, P^{N^*_{(a,b]}}\Big) \leq \exp\Big(-\frac{\nu a}{n}\Big)\Big(1-\exp\Big(-\frac{\nu(b-a)}{n}\Big)\Big) \leq 1-\exp\Big(-\frac{\nu b}{n}\Big) \leq \frac{\nu b}{n}.$$

3.10 (radioaktiver Zerfall) Zur Untersuchung des Hirnstoffwechsels mittels PET werden Patienten radioaktiv markierte Glucoseverbindungen injiziert. Als Tracer werden dabei u.a. die Isotopen Fluor (F), Kohlenstoff (C), Stickstoff (N) oder Sauerstoff (O) eingesetzt:

Isotop	^{18}F	^{11}C	^{13}N	^{15}O
Halbwertszeit (in min)	120	20.4	10	2

Mit physikalischen Überlegungen läßt sich zeigen, daß die Lebensdauerverteilung des Isotops (in min) gegeben ist durch eine $\mathcal{E}(\lambda)$-Exponentialverteilung mit

$$\lambda = \ln 2\; h^{-1} \approx 0.693\; h^{-1}$$

(vgl. Pfeifer (1989a), Beispiel 0.1 und S. 145f.). Zeigen Sie, daß der in Aufgabe 3.9 definierte Markoff-Prozeß $\{N_t\}_{t\geq 0}$ den zugehörigen radioaktiven Zerfallsprozeß beschreibt, wobei n die Anzahl der injizierten Isotope ist.
Nach dem Gesetz von Avogadro enthält ein Mol[1] eines Stoffes stets dieselbe Anzahl von Atomen, nämlich ca. $6 \cdot 10^{23}$. Zeigen Sie, daß bei Injektion von $n \triangleq 10^{-14}$ Mol des Isotops für den Zerfallszeitpunkt T_n des *letzten* Isotops gilt:

$$E(T_n) \approx 32.5\,h \;\text{ (min)}, \qquad \mathrm{Var}(T_n) \approx 3.42\,h^2,$$

und berechnen Sie die entsprechenden Werte für die in der Tabelle angegebenen Isotope.

3.11 Es sei ξ ein Poisson'scher Punktprozeß über $\mathcal{X} = \mathbf{R}^2$ mit σ-endlichem Intensitätsmaß $E\xi = \mu$ und $\mu(\mathbf{R}^2) = \infty$). X sei der Abstand des am nächsten zum Nullpunkt gelegenen Punktes von diesem. Zeigen Sie:

$$F_X(r) = 1 - P(X > r) = 1 - P\big(\xi(K_r) = 0\big) = 1 - e^{-\mu(K_r)}, \quad r > 0,$$

wobei $K_r = K^a(0,0;r)$ den abgeschlossenen Kreis um den Nullpunkt mit Radius $r > 0$ bezeichne.
Zeigen sie speziell für den Fall, daß μ das c-fache des Lebesgue-Maßes ist ($c > 0$ fest):

$$F_X(r) = 1 - e^{-c\pi r^2}, \quad r > 0$$
$$E(X) = \int_0^\infty e^{-c\pi r^2}\,dr = \frac{1}{2\sqrt{c}}.$$

Was ergibt sich entsprechend, wenn für den Meßraum $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^3, \mathcal{B}^3)$ gewählt wird (vgl. Weiß (1987), Beispiel 4.42)?

---

[1] d.h. Molekulargewicht in g

3.12 Es sei ξ ein Poisson'scher Punktprozeß über $\mathcal{X} = \mathbf{R}^2$ mit *endlichem* Intensitätsmaß $\mu = E\xi$ und $\mu(\mathbf{R}^2) > 0$. X sei wieder der Abstand des am nächsten zum Nullpunkt gelegenen Punktes von diesem, Y der Abstand des am weitesten davon entfernten Punktes von diesem (Skizze!). Zeigen Sie:

$$F_X(r) = 1 - P(X > r) = 1 - P\big(\xi(K_r) = 0 \mid \xi(\mathbf{R}^2) \geq 1\big) = \frac{1 - e^{-\mu(K_r)}}{1 - e^{-\mu(\mathbf{R}^2)}}$$

$$F_Y(r) = P(Y \leq r) = P\big(\xi(K_r^c) = 0 \mid \xi(\mathbf{R}^2) \geq 1\big) = \frac{e^{\mu(K_r)} - 1}{e^{\mu(\mathbf{R}^2)} - 1},$$

wobei $K_r = K^a(0,0;r)$ wieder den abgeschlossenen Kreis um den Nullpunkt mit Radius $r > 0$ bezeichne.
Leiten Sie hieraus für das Intensitätsmaß μ mit der Dichte

$$f(x,y) = c \cdot e^{-(x^2+y^2)}, \quad x, y \in \mathbf{R}$$

($c > 0$ fest) speziell ab:

$$F_X(r) = \frac{e^{c\pi} - e^{c\pi e^{-r^2}}}{e^{c\pi} - 1}, \qquad F_Y(r) = \frac{e^{c\pi\left(1-e^{-r^2}\right)} - 1}{e^{c\pi} - 1}, \quad r > 0.$$

Skizzieren Sie beide Verteilungsfunktionen.

3.13 Es sei ξ ein Poisson'scher Punktprozeß über dem Meßraum $(\mathcal{X}, \mathcal{B})$ mit Intensitätsmaß $E\xi = \mu$ und T eine meßbare Abbildung von $(\mathcal{X}, \mathcal{B})$ in den Meßraum $(\mathcal{Y}, \mathcal{C})$. Zeigen Sie:

a) Durch
$$\mu^T(C) = \mu\big(T^{-1}(C)\big), \quad C \in \mathcal{C},$$
wird ein Maß auf $\mathcal{C}$ definiert (sog. *Bildmaß* von μ).

b) Der durch
$$\xi^T(C) = \xi\big(T^{-1}(C)\big), \quad C \in \mathcal{C},$$
über $\mathcal{Y}$ definierte Punktprozeß ξ^T ist ein Poisson–Prozeß mit Intensitätsmaß $E(\xi^T) = \mu^T$.
(Hinweis: Benutzen sie Definition 3.4.8.)

3.14 Es sei ξ ein Poisson'scher Punktprozeß über $\mathbf{R}^3$ mit Intensitätsmaß $E(\xi) = \mu$. $T : \mathbf{R}^3 \longrightarrow \mathbf{R}^2$ sei die orthogonale Projektion von $\mathbf{R}^3$ in $\mathbf{R}^2$, d.h. $T(x,y,z) = (x,y)$, $x, y, z \in \mathbf{R}$. Zeigen Sie unter Verwendung von Aufgabe 3.12:
Der Poisson–Prozeß ξ^T besitzt das Intensitätsmaß $E\xi^T = \mu^T$, gegeben durch

$$\mu^T(B) = \mu(B \times \mathbf{R}), \quad B \in \mathcal{B}^2.$$

Was bedeutet dies für Bildauswertungen im Bereich der Computertomographie? Diskutieren sie diese Frage am Beispiel des Intensitätsmaßes μ mit der durch

$$f(x,y,z) = \begin{cases} (x^2 + y^2 + 6z^2) & \text{für } (x,y,z) \in K^a(0,0,0;1) \\ 0 & \text{sonst,} \end{cases} \qquad x, y, z \in \mathbf{R},$$

gegebenen Dichte. Zeigen Sie, daß in diesem Fall μ^T die Dichte

$$f_T(x,y) = \begin{cases} 2\sqrt{1 - x^2 - y^2} & \text{für } (x,y) \in K^a(0,0;1) \\ 0 & \text{sonst,} \end{cases} \qquad x, y \in \mathbf{R},$$

besitzt (Skizze!).

4. Probabilistische Analyse von Algorithmen

Häufig lassen sich Problemstellungen mit verschiedenen Algorithmen bearbeiten. Die Aufgabe, ein vorgegebenes Feld nach einem bestimmten Merkmal zu sortieren, kann man durch Vergleich jedes Elements mit jedem anderen lösen; nach kurzem Nachdenken findet man jedoch Methoden, die schneller zum Ziel führen. Das Maximum einer reellen Funktion zu bestimmen, die auf einem Gitter $\{1, 2, \ldots, n\}$ definiert ist und von der man weiß, daß sie erst monoton wächst und dann monoton fällt (unimodular ist), schafft man durch Vergleich aller Funktionswerte auf dem Gitter, aber auch hier wird jemand, der die Struktur des Problems berücksichtigt, mit einem Bruchteil der Zeit auskommen.

Unser Anliegen ist nun, Algorithmen bezüglich Ihrer Effizienz zu bewerten. Schon eine Präzisierung dieser Fragestellung ist nicht einfach. Von der mathematischen Formulierung eines Algorithmus mit abstrakten Größen bis zum lauffähigen Maschinencode auf einem bestimmten Rechner ist in der Regel ein weiter Weg. Der abstrakten Formulierung entgegenkommende Programmiersprachen erfordern die Übersetzung durch einen Compiler. Durch wieviele Maschinenbefehle genau dabei ein komplexes Statement der höheren Sprache ersetzt wird, entzieht sich meist der Kenntnis dessen, der einen Algorithmus entwickelt hat. Man kann jedoch annehmen, daß sich die Anzahl der elementaren Befehle proportional zu den in der Programmiersprache formulierten verhält, also höchstens um einen (eventuell sehr großen) Faktor höher als die der ursprünglichen Befehle ist. Zählt man nun, wie oft eine Befehlszeile des Quellcodes abgearbeitet wird, ist dies bis auf einen konstanten Faktor die Anzahl elementarer Operationen eines Prozessors und repräsentiert — wieder bis auf einen konstanten Faktor, der von der Schnelligkeit der Maschine abhängt — die Laufzeit des Algorithmus.

Wir werden uns um eine Form der Analyse bemühen, die diese technischen Details in einer globalen Konstante berücksichtigt und eine Charakterisierung des Verhaltens maschinenunabhängig erlaubt. Die geeignete Notation hierfür ist das Landau'sche O-Symbol.

Definition 4.1. *$\{c_n\}_{n\in\mathbb{N}}$ und $\{x_n\}_{n\in\mathbb{N}}$, $x_n > 0$, seien Folgen reeller Zahlen. Dann heißt $c_n = O(x_n)$, wenn $n_0 \in \mathbb{N}$ und eine von n unabhängige Konstante $C \geq 0$ existieren mit $|c_n/x_n| \leq C$ für alle $n \in \mathbb{N}$, $n \geq n_0$.*

Ist in obiger Definition c_n die Laufzeit eines Algorithmus bei Eingabelänge n und $x_n = n$ für alle $n \in \mathbb{N}$, so bedeutet $c_n = O(n)$, daß die Laufzeit höchstens proportional zu n wächst. Die Umsetzung komplexer Statements in Maschinenbefehle und die Schnelligkeit der Ausführrung auf verschiedenen Rechnern wird hierbei durch die Konstante C subsumiert.

Die Laufzeit eines gegebenen Algorithmus hängt natürlich vom speziellen Input ab, der bearbeitet werden soll. Sie ist also eine reellwertige Abbildung, die in der Regel jedoch nicht geschlossen, in einfacher Form dargestellt werden kann. Bei den Algorithmen, die wir betrachten, ist zur Bewertung nicht die gesamte Variabilität des Inputs interessant, sondern nur eine charakteristische Größe, die sich meist in der Anzahl n von zu bearbeitenden, in einer bestimmten Datenstruktur abstrahierten Objekten ausdrückt.

4.1. Sortier– und Suchverfahren

In diesem Abschnitt wollen wir exemplarisch das stochastische Verhalten einiger wichtiger Algorithmen untersuchen, wobei gegebenenfalls auch die Möglichkeit nicht–gleichverteilter Inputs berücksichtigt werden soll. Derartig allgemeine Fragestellungen entziehen sich allerdings in der Regel einer rein kombinatorischen Behandlung, die bisher überwiegend in der Literatur zu finden ist (vgl. etwa Kemp (1984) oder Aigner (1988)). Wir wollen deshalb hier auch allgemeinere wahrscheinlichkeitstheoretische Modelle und Methoden vorstellen, mit denen solche Probleme behandelt werden können.

Beispiel 4.1.1. (`quicksort`)
Ein oft eingesetztes und für viele Anwendungen schnelles Sortierverfahren ist der Algorithmus `quicksort`, der auf C.A.R. Hoare (1962) zurückgeht. Die Effizienz des Verfahrens wurde nicht durch den Einsatz komplizierter Datenstrukturen erreicht, sondern durch eine raffinierte Implementierung einer im Grunde einfachen Idee: der "teile und beherrsche (devide and conquer) –Strategie ". Wir führen das Verfahren an einem Feld

```
a: array[0..n] of integer
```

vor, wobei `a[0]` die Rolle eines Platzhalters zum Abbruch des Algorithmus spielt. `a[0]` wird mit der kleinsten negativen Integerzahl besetzt, die zur Verfügung steht. In der Praxis werden die einzelnen Feldelemente natürlich mit aufwendigeren Strukturen besetzt sein, etwa mit Daten des Typs `record`, die neben einem zu sortierenden Schlüsselelement noch weitere Informationen tragen. Obwohl dann das Umspeichern eines Datenblocks aufwendiger ist als im hier betrachteten Fall, bleibt die Funktionsweise des Algorithmus in ihrem Kern die gleiche.

Zusätzlich wollen wir annehmen, daß alle Schlüsselelemente verschieden sind. Diese Zusatzvoraussetzung wird die spätere Analyse in einigen Punkten vereinfachen.

Die Idee des Algorithmus `quicksort` läßt sich folgendermaßen kurz skizzieren: Teile das Feld in zwei Teilfelder mit einem trennenden Element `a[i]` so auf, daß

(i) `a[i]` an der richtigen Stelle des Feldes steht,
(ii) alle Elemente im linken Teilfeld (links von `a[i]`) kleiner als `a[i]` sind und
(iii) alle Elemente im rechten Teilfeld (rechts von `a[i]`) größer als `a[i]` sind.

Wiederhole dann dieses Verfahren mit allen autretenden Teilfeldern bis zur vollständigen Sortierung des gesamten Felds. Dieses Vorgehen legt eine rekursive Formulierung des Programms nahe, die in einem ersten Entwurf folgendermaßen realisiert wird.

```
procedure quicksort(l,r: integer);
var i,j,v,t: integer;
begin
    if r>l then begin
        {Teile das Feld a[l]..a[r] mit
          trennendem Element a[i], l ≤ i ≤ r,
          mit obigen Eigenschaften (i)–(iii) auf.}
        quicksort(l,i-1);
        quicksort(i+1,r);
        end;
end;
```

(4.1.1)

Der kommentierte Teil in obigem Entwurf wird mit den folgenden Schritten umgesetzt.

a) Wähle im Feld `a[l],..,a[r]` ein beliebiges Element, das aufteilen und an der richtigen Position stehen soll, etwa `a[r]`.
b) Durchsuche das Feld von links beginnend mit zwei Zeigern `i` und `j` bis ein Element `> a[r]` und von rechts beginnend bis ein Element `< a[r]` gefunden ist. Tausche diese Elemente aus und setze das Verfahren fort, bis der linke Zeiger i größer oder gleich als der rechte j ist.
c) Tausche dann den Inhalt der Elemente `a[i]` und `a[r]` aus.

Eine ausgefeilte Implementierung von `quicksort` und eine ausführliche Besprechung des Programms und seiner Feinheiten findet sich in Sedgewick (1988). Hier werden auch einige Verbesserungen des Algorithmus vorgeschlagen, die vor allem bei der Sortierung kleiner Restfelder angebracht werden können.

Wir untersuchen die stochastischen Eigenschaften des Basisalgorithmus in der folgenden Version:

```
 1:      procedure quicksort(l,r:  integer);
 2:      var i,j,v,t:  integer;
 3:      begin
 4:          if r>l then begin
 5:              v:=a[r]; i:=l-1; j:=r;
 6:              repeat
 7:                  repeat i:=i+1 until a[i]>=v;
 8:                  repeat j:=j-1 until a[j]<=v;
 9:                  if i<j then begin
10:                      t:=a[i]; a[i]:=a[j]; a[j]:=t;
11:                      end;
12:                  until j<=i;
13:              t:=a[i]; a[i]:=a[r]; a[r]:=t;
14:              quicksort(l,i-1);
15:              quicksort(i+1,r);
16:              end;
17:      end;
```

(4.1.2)

Die Anzahl der Umspeicherungen im "inneren loop" in den Zeilen 7 und 8 von (4.1.2) ist höchstens so groß wie die Anzahl der Vergleiche. Wir untersuchen deshalb lediglich die Anzahl der Vergleiche, die in diesen beiden Zeilen durchgeführt werden. Der zusätzliche Aufwand wächst linear in n, so daß höchstens ein additiver Term der Größenordnung $O(n)$ hinzukommt.

Für ein (Teil-) Feld der Länge n mit Grenzen ℓ und r werden 7 und 8 bei jedem Aufruf zusammen n- oder $(n+1)$-mal durchlaufen. Hieraus resultiert für die Gesamtanzahl der Vergleiche v_n, $n \in \mathbb{N}$, in den Zeilen 7 und 8 bei der Sortierung eines Felds der Länge n die Rekursion

$$n + \min_{1\le k\le n}\{v_{k-1}+v_{n-k}\} \le v_n \le n+1+\max_{1\le k\le n}\{v_{k-1}+v_{n-k}\},\ n\ge 2, \quad \text{wobei } v_0 = v_1 = 0. \tag{4.1.3}$$

Hieraus lassen sich induktiv untere und obere Schranken für v_n gewinnen.

Lemma 4.1.1. *Bezeichne v_n, $n \in \mathbb{N}$, die Gesamtanzahl der Vergleiche in den Zeilen 7 und 8 von (4.1.2), die der Rekursion (4.1.3) genügt. Dann gilt für alle $n \in \mathbb{N}$*

$$n \cdot \ln n - 0.65 \leq v_n \leq \frac{(n+1)(n+2)}{2} - 3$$

Beweis. Wir zeigen beide Ungleichungen mit vollständiger Induktion und beginnen mit der rechten. Für $n = 1$ gilt $v_1 = 0 \leq 2 \cdot 3/2 - 3 = 0$. Unter der Voraussetzung, daß die Aussage für $1, \ldots, n$ gilt, folgt mit (4.1.3)

$$\begin{aligned} v_{n+1} &\leq n + 2 + \max_{1 \leq k \leq n+1} \{v_{k-1} + v_{n+1-k}\} \\ &= n + 2 + \max\{v_0 + v_n, \max_{1 \leq k \leq n} \{v_k + v_{n-k}\}\} \\ &\leq n + 2 + \max\Big\{\frac{(n+1)(n+2)}{2} - 3, \\ &\qquad \max_{1 \leq k \leq n} \Big\{\frac{(k+1)(k+2)}{2} - 3 + \frac{(n-k+1)(n-k+2)}{2} - 3\Big\}\Big\}. \end{aligned}$$

Die Summe der Zähler $(k+1)(k+2) + (n-k+1)(n-k+2) = 2k^2 - 2nk + (n+1)(n+2) + 2$ ist offensichtlich maximal für $k = n$, so daß

$$\begin{aligned} v_{n+1} &\leq n + 2 + \max\Big\{\frac{(n+1)(n+2)}{2} - 3, \frac{(n+1)(n+2)}{2} - 5\Big\} \\ &= n + 2 + \frac{(n+1)(n+2)}{2} - 3 = \frac{(n+2)(n+3)}{2} - 3. \end{aligned}$$

Zum Nachweis der linken Ungleichung wird (4.1.3) sukzessive für die ersten zehn natürlichen Zahlen ausgewertet.

v_n	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
$\geq$	0	2	3	6	8	11	13	17	20	24
$n \cdot \ln n$	0	1.386	3.296	5.545	8.047	10.751	13.621	16.636	19.775	23.026

In der zweiten Zeile der obigen Tabelle sind die hieraus entstehenden unteren Schranken für v_n angegeben, in der dritten finden sich die Werte von $n \cdot \ln n$ auf vier Stellen gerundet. Damit ist der Induktionsanfang für $n = 1, \ldots, 10$ geführt. Mit der Induktionsvorraussetzung schließen wir

$$\begin{aligned} v_{n+1} &\geq n + 1 + \min_{1 \leq k \leq n+1} \{v_{k-1} + v_{n+1-k}\} \\ &= n + 1 + \min\{v_n, \min_{1 \leq k \leq n} \{v_k + v_{n-k}\}\} \\ &\geq n + 1 + \min\{n \cdot \ln n - 0.65, \\ &\qquad \min_{1 \leq k \leq n} \{k \cdot \ln k - 0.65 + (n-k)\ln(n-k) - 0.65\}\} \end{aligned}$$

Nach Division der nichtkonstanten Terme durch n wird das zweite Minimierungsproblem auf die Entropieungleichung in Satz 5.1.1 a) zurückgeführt. Eine untere Schranke ergibt sich für $k/n = \frac{1}{2}$ also $k = \frac{n}{2}$. Dies eingesetzt liefert

$$\begin{aligned} v_{n+1} &\geq n + 1 + \min\{n \cdot \ln n - 0.65, n \cdot \ln(n/2) - 1.3\} = n - 0.3 + n \cdot \ln(n/2) \\ &\geq (n+1)\ln(n+1) - 0.65, \text{ falls } n \geq 10, \end{aligned}$$

da $n + 0.35 + n \cdot \ln(n/2) - (n+1)\ln(n+1) \geq 0$ für alle $n \geq 10$, wie man durch Diskussion der konvexen Funktion $f(x) = x + 0.35 + x \cdot \ln(x/2) - (x+1)\ln(x+1)$ sieht. ∎

Obere und untere Schranken für die Gesamtanzahl der Umspeicherungen sind also von der Ordnung $O(n^2)$ und $O(n \cdot \ln n)$. Diese werden auch erreicht, wie die folgenden Überlegungen zeigen.

Der ungünstigste Fall tritt bei bereits sortiertem Feld auf. Dann eliminiert `quicksort` in jeder Iteration genau ein Element, nämlich jeweils das letzte. Für die Anzahl der Vergleiche gilt dann

$$v_n = (n+1) + n + (n-1) + \cdots + 3 = \sum_{i=1}^{n+1} i - 3 = \frac{(n+1)(n+2)}{2} - 3,$$

d.h. die obere Schranke aus Lemma 4.1.1 wird genau erreicht.

Zur Untersuchung des besten Falls betrachten wir ein Feld der Länge $n = 2^m$, $m \in \mathbb{N}$. Dieses sei so sortiert, daß in jedem Aufruf von `quicksort` die Variable `v=a[r]` eines der mittleren Elemente von `a[l]..a[r]` enthält (vgl. Aufgabe 4.2). Dann gilt

$$v_n \leq 2v_{n/2} + n, \; n \geq 2, \quad v_1 = 0,$$

also $v_{2^m} \leq 2v_{2^{m-1}} + 2^m$ und $v_{2^m}/2^m \leq v_{2^{m-1}}/2^{m-1} + 1$. Durch iterierte Anwendung der letzten Ungleichung folgt

$$\frac{v_{2^m}}{2^m} \leq v_1 + m = m, \text{ d.h. } v_n \leq n \cdot \log_2 n = O(n \cdot \ln n).$$

Die untere Schranke in Lemma 4.1.1 wird also ebenfalls asymptotisch angenommen.

Günstigster und ungünstigster Fall treten jedoch nur mit sehr kleiner Wahrscheinlichkeit auf (vgl. Beispiel 1.1.4); im folgenden wird daher unter einem geeigneten stochastischen Modell die erwartete Anzahl von Vergleichen in `quicksort` ermittelt.

Wir nehmen zunächst an, daß jede Anordnung von verschiedenen Elementen eines bestimmten Universums im Feld `a[1]..a[n]` gleichwahrscheinlich ist. Diese Situation läßt sich durch den Wahrscheinlichkeitsraum

$$\begin{aligned} &\Omega^{(n)} = Perm_n^n(\{1,\ldots,n\}; o.W.), \mathcal{A} = \mathfrak{P}(\Omega^{(n)}) \quad \text{und} \\ &P(\{\omega\}) = \frac{1}{n!} \text{ für alle } \omega = (\omega_1,\ldots,\omega_n) \in \Omega^{(n)} \end{aligned} \tag{4.1.4}$$

beschreiben. Ein Problem ist, ob sich dieses Grundmodell auf die entstehenden Teilfelder überträgt. Dies kann in folgendem Sinn positiv beantwortet werden.

Wir betrachten hierzu die (meßbare) Abbildung $\boldsymbol{T}^{(n)} = (T_1,\ldots,T_n) : \Omega^{(n)} \to \Omega^{(n)}$, die beschreibt, welche Umsortierung nach einem Aufruf von `quicksort` stattgefunden hat, also

$$\boldsymbol{T}^{(n)} = (T_1,\ldots,T_n) : \Omega^{(n)} \to \Omega^{(n)} \text{ mit } \boldsymbol{T}^{(n)}(\omega) = \omega', \tag{4.1.5}$$

wobei $\omega' \in Perm_n^n(\{1,\ldots,n\}; o.W.)$ die Permutation ist, die bei einem Aufruf von `quicksort` nach Ausführung der Zeile 13 von (4.1.2) als Feldbelegung vorgefunden wird. Weiterhin bezeichne Z_n die Projektion auf die n–te Komponente, also

$$Z_n : \Omega^{(n)} \to \{1,\ldots,n\} : (\omega_1,\ldots,\omega_n) \mapsto \omega_n.$$

Lemma 4.1.2. *Unter dem stochastischen Modell (4.1.4) gilt für die Zufallsvariablen $T_1,\ldots,T_n$ für alle $k \leq n \in \mathbb{N}$, daß $(T_1,\ldots,T_{k-1})$ und $(T_{k+1},\ldots,T_n)$ bedingt stochastisch unabhängig sind, gegeben $\{Z_n = k\}$.*
$P^{(T_1,\ldots,T_{k-1}|Z_n=k)}$ ist eine Laplace–Verteilung auf $Perm_{k-1}^{k-1}(\{1,\ldots,k-1\}; o.W.)$ und $P^{(T_{k+1},\ldots,T_n|Z_n=k)}$ eine Laplace–Verteilung auf $Perm_{n-k}^{n-k}(\{k+1,\ldots,n\}; o.W.)$.

Beweis. Für $1 \leq k \leq n$ bezeichne $B_k = \{\omega \in \Omega^{(n)} \mid Z_n(\omega) = k\}$. Sei $\omega' = (\omega'_1,\ldots,\omega'_{k-1},\omega'_k,\omega'_{k+1}\ldots,\omega'_n)$ ein Element aus $Perm_n^n(\{1,\ldots,n\}; o.W.)$ mit $\omega'_1,\ldots,\omega'_{k-1} < k$, $\omega'_k = k$ und $\omega'_{k+1},\ldots,\omega'_n > k$. Dann gilt

$$\#(\{\omega \in B_k \mid \boldsymbol{T}^{(n)}(\omega) = \omega'\} = \sum_{\ell=0}^{\min\{k-1,n-k\}} \binom{k-1}{\ell}\binom{n-k}{\ell} = \binom{n-1}{k-1}.$$

Obige Menge setzt sich zusammen aus den Permutationen $\omega = (\omega_1,\ldots,\omega_n) \in Perm_n^n(\{1,\ldots,n\}; o.W.)$, für die $\omega_n = k$ und für $\ell = 0,\ldots,\min\{k-1,n-k\}$ genau ℓ der Elemente $\omega_1,\ldots,\omega_{k-1} > k$ und genau ℓ der Elemente $\omega_{k+1},\ldots,\omega_n < k$. Dies liefert als Mächtigkeit die Summe der Binomialkoeffizienten. Die zweite Gleichheit wird in Aufgabe ??? bewiesen. Es folgt

$$\begin{aligned}
&P(T_1 = \omega'_1,\ldots,T_{k-1} = \omega'_{k-1}, T_{k+1} = \omega'_{k+1},\ldots,T_n = \omega'_n \mid Z_n = k)\\
&= \begin{cases} \dfrac{n}{n!}\binom{n-1}{k-1}, & \text{falls } \omega'_1,\ldots,\omega'_{k-1} < k < \omega'_{k+1},\ldots,\omega'_n\\ 0, & \text{sonst}\end{cases}\\
&= \begin{cases} \dfrac{1}{(k-1)!}\cdot\dfrac{1}{(n-k)!}, & \text{falls } (\omega'_1,\ldots,\omega'_{k-1}) \in Perm_{k-1}^{k-1}(\{1,\ldots,k-1\}; o.W.)\\ & \text{und } (\omega'_{k+1},\ldots,\omega'_n) \in Perm_{n-k}^{n-k}(\{k+1,\ldots,n\}; o.W.)\\ 0, & \text{sonst.}\end{cases}
\end{aligned}$$

Hieraus folgen die Behauptungen. ∎

V_ℓ, $\ell \in \mathbb{N}$, bezeichne die Anzahl der Vergleiche, die `quicksort` in den Zeilen 7 und 8 insgesamt bei einem Aufruf für ein Feld der Länge ℓ durchführt. Formal kann V_ℓ als Abbildung

$$V_\ell : \bigcup_{1\leq i_1<\cdots<i_\ell\leq n} Perm_\ell^\ell(\{i_1,\ldots,i_\ell\}; o.W.) \longrightarrow \mathbb{N}$$

geschrieben werden. Unter den Bezeichnungen (4.1.4) und (4.1.5) ergibt sich die folgende Rekursion, die den ungünstigeren Fall von $(n+1)$ Vergleichen in einem

Feld der Länge n betrachtet. Für $1 \leq k \leq n$ und $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ mit $\omega_n = k$ gilt

$$V_n(\omega) = n + 1 + V_{k-1}\big((T_1, \ldots, T_{k-1})(\omega)\big) + V_{n-k}\big((T_{k+1}, \ldots, T_n)(\omega)\big), \; n \geq 2,$$
$$\text{mit } V_0 = 0, \; V_1 = 0.$$

Insgesamt folgt die Darstellung

$$\begin{aligned} V_n(\omega) &= \sum_{k=1}^{n} \Big[n + 1 + V_{k-1}\big((T_1, \ldots, T_{k-1})(\omega)\big) + V_{n-k}\big((T_{k+1}, \ldots, T_n)(\omega)\big)\Big] \\ &\qquad\qquad \cdot \mathbb{1}_{\{Z_n = k\}}(\omega) \\ &= n + 1 + \sum_{k=1}^{n} V_{k-1}\big((T_1, \ldots, T_{k-1})(\omega)\big) \cdot \mathbb{1}_{\{Z_n = k\}}(\omega) \\ &\quad + \sum_{k=1}^{n} V_{n-k}\big((T_{k+1}, \ldots, T_n)(\omega)\big) \cdot \mathbb{1}_{\{Z_n = k\}}(\omega), \quad n \geq 2, \\ &\qquad \text{mit } V_0 = 0, \; V_1 = 0. \end{aligned} \tag{4.1.6}$$

Nach Lemma 4.1.2 und Lemma 2.1.7 sind die Zufallsvariablen $V_{k-1}\big((T_1, \ldots, T_{k-1})\big)$ und $V_{n-k}\big((T_{k+1}, \ldots, T_n)\big)$ bedingt stochastisch unabhängig, gegeben $\{Z_n = k\}$.

Das folgende Verteilungsmodell, das auf Rösler (1988) zurückgeht, ist damit geeignet, das stochastische Verhalten von `quicksort` zu beschreiben. $\{X_n\}_{n \in \mathbb{N}}$, $\{X'_n\}_{n \in \mathbb{N}}$ mit $E(X_n) = E(X'_n)$, $n \in \mathbb{N}$, $X_0 = X_1 = 0$, $X'_0 = X'_1 = 0$ und $\{Z_n\}_{n \in \mathbb{N}}$ seien stochastisch unabhängige Zufallsvariable, Z_n gleichverteilt auf $\{1, \ldots, n\}$, die der Verteilungsrekursion

$$X_n \stackrel{\mathcal{D}}{=} n + 1 + X_{Z_n - 1} + X'_{n - Z_n}, \quad n \geq 2. \tag{4.1.7}$$

genügen. " $\stackrel{\mathcal{D}}{=}$ " bedeutet hierbei *verteilungsgleich*, d.h. die Zufallsvariablen auf der linken und rechten Seite besitzen die gleiche Verteilung. $X_{Z_n - 1}$ und $X'_{n - Z_n}$ sind nach Konstruktion bedingt stochastisch unabhängig, gegeben $\{Z_n = k\}$. Ferner gilt $X_{Z_n - 1} = \sum_{k=1}^{n} X_{k-1} \cdot \mathbb{1}_{\{Z_n = k\}}$ und $X'_{n - Z_n} = \sum_{k=1}^{n} X'_{n-k} \cdot \mathbb{1}_{\{Z_n = k\}}$.

Wir bestimmen im folgenden den Erwartungswert der Zufallsvariablen X_n unter Modell (4.1.7). Dieser hängt lediglich von der Verteilung der Zufallsvariablen X_n ab; er repräsentiert den erwarteten Aufwand von `quicksort` bei der Anzahl der Vergleiche in den Zeilen 7 und 8 der Prozedur (4.1.2).

Satz 4.1.1. *Für die erwartete Anzahl der Vergleiche $E(X_n)$ in den Zeilen 7 und 8 von (4.1.2) zur Sortierung eines Felds der Länge $n \geq 3$ durch* `quicksort` *gilt unter dem stochastischen Modell (4.1.7)*

$$\begin{aligned} E(X_n) &= 2(n+1) \cdot H_{n+1} - 3(n+1) \\ &= 2(n+1) \cdot \ln(n+1) + (2\gamma - 3)(n+1) + 1 + \mathrm{O}(\tfrac{1}{n}), \end{aligned} \tag{4.1.8}$$

wobei $C = 0.577215...$ die Euler'sche Konstante und $H_n = \sum_{k=1}^{n} \frac{1}{k}$ die Harmonische Reihe bezeichnen.

Beweis. Die Zufallsvariablen auf der rechten Seite von (4.1.7) besitzen für alle $n \geq 2$ die Darstellung

$$n+1+\sum_{k=0}^{n-1} X_k \cdot \mathbb{1}_{\{Z_n=k+1\}} + \sum_{k=0}^{n-1} X'_{n-k-1} \cdot \mathbb{1}_{\{Z_n=k+1\}},$$

wobei $X_1, \ldots, X_{n-1}, X'_1, \ldots, X'_{n-1}, Z_n$ stochastisch unabhängig sind. Mit (2.2.21) und (2.2.27) folgt für die Erwartungswerte $\mu_n = E(X_n) = E(X'_n)$, $n \in \mathbb{N}$, daß $\mu_0 = 0$, $\mu_1 = 0$ und $\mu_n = n+1+\sum_{k=0}^{n-1} \mu_k P(Z_n = k+1) + \sum_{k=0}^{n-1} \mu_{n-k-1} P(Z_n = k+1)$, also die Rekursion $\mu_0 = 0$, $\mu_1 = 0$ und

$$\mu_n = n+1+\frac{2}{n}\sum_{k=0}^{n-1} \mu_k, \quad n \geq 2.$$

Dies liefert die folgende einfachere rekursive Darstellung

$$\begin{aligned}\mu_{n+1} &= n+2+\frac{2}{n+1}\sum_{k=0}^{n} \mu_k = n+2+\frac{n}{n+1}\Big(\frac{2}{n}\sum_{k=0}^{n-1} \mu_k\Big) + \frac{2}{n+1}\mu_n \\ &= n+2+\frac{n}{n+1}\mu_n - n + \frac{2}{n+1}\mu_n = 2+\frac{n+2}{n+1}\mu_n, \quad n \geq 2,\end{aligned}$$

wobei $\mu_0 = \mu_1 = 0$, $\mu_2 = 3$. Ein einfacher Induktionsbeweis zeigt, daß

$$\mu_n = 2(n+1)\sum_{i=2}^{n} \frac{1}{i+1} = 2(n+1)\big(H_{n+1} - \tfrac{3}{2}\big), \quad n \geq 3,$$

woraus die erste Gleichheit folgt. Aus der asymptotischen Entwicklung der harmonischen Reihe $H_n = \ln n + \gamma + \frac{1}{2n} + \mathrm{O}\big(\frac{1}{n^2}\big)$, $n \in \mathbb{N}$, erhalten wir

$$\mu_n = 2(n+1)\cdot \ln(n+1) + 2\gamma(n+1) + 1 - 3(n+1) + \mathrm{O}\big(\tfrac{1}{n}\big).$$

Hieraus folgt der zweite Teil der Behauptung. ∎

Die erste Gleichung in (4.1.8) gestattet die folgende einfache Abschätzung des erwarteten Aufwands von `quicksort`. Für alle $n \geq 2$ gilt $\sum_{k=2}^{n} \frac{1}{k} \leq \int_1^n \frac{1}{x} dx = \ln n$, da die linke Seite eine Riemann-Untersumme der Funktion $f(x) = \frac{1}{x}$ im Intervall $[1, n]$ ist. (4.1.8) liefert dann

$$E(X_n) \leq 2(n+1)\cdot \ln(n+1), \quad n \geq 3. \tag{4.1.9}$$

Ein Vergleich dieser Schranke mit der unteren Schranke aus Lemma 4.1.1 zeigt, daß der erwartete Aufwand im Mittel asymptotisch nur doppelt so groß ist wie im günstigsten Fall. Dies ist wohl der Grund, daß `quicksort` in den meisten Fällen schnelle Laufzeiten erzielt.

Allerdings geht die Aussage (4.1.9) von einer Gleichverteilung auf der Menge aller möglichen Anfangsanordnungen aus. Liegen häufig schon teilweise sortierte

Felder vor, ist diese Voraussetzung sicher gestört. Hier hilft eine randomisierte Version von `quicksort`, bei der in jedem Schritt des äußeren Loop nicht das letzte Element des zu bearbeitenden Teilfeldes als trennendes Element gewählt wird, sondern zufällig und unabhängig eines aus `a[l] ... a[r]`. Zur Implementation dieser Idee wird Zeile 5 von (4.1.2) durch den Block

```
rr:=l+random(r-l+1);
t:=a[r]; a[r]:=a[rr]; a[rr]:=t;
v:=a[r]; i:=l-1; j:=r;
```

ersetzt, wobei `random(r-l+1)` einen zufälligen Index gemäß einer Laplace–Verteilung auf $\{0,1,\ldots,r-l\}$ liefert. Der erhöhte Aufwand bei der Bestimmung von `rr` und drei zusätzlichen Umspeicherungen wächst dabei höchstens linear mit der Feldgröße n.

Für jede Ausgangsanordnung folgt die Anzahl der Vergleiche der randomisierten Version dem Verteilungsmodell (4.1.7). Unabhängig von der Eingabefolge ergibt sich mit Satz 4.1.1 stets ein erwarteter Aufwand der maximalen Größenordnung $O(n \cdot \ln n)$. ∎

Beispiel 4.1.2. (`hybridsort`)
Wir betrachten die in Beispiel 2.1.2 gegebene Situation, allerdings sogleich unter der allgemeineren Annahme, daß die zu sortierenden Zahlen $x_1,\ldots,x_n$ Realisationen unabhängiger, aber beliebig über $(0,1]$–verteilter Zufallsvariablen $X_1,\ldots,X_n$ mit Verteilungsfunktion F sind. Der Zufallsvektor $(N_1,\ldots,N_k)$ der nach Ausführung des Schrittes a) von `hybridsort` in den k Körben befindlichen Zahlen ist also gemäß den Ausführungen auf S. 85 $\mathfrak{M}(n;p_1,\ldots,p_k)$–multinomialverteilt mit

$$p_j = F\left(\frac{j}{k}\right) - F\left(\frac{j-1}{k}\right), \quad 1 \le j \le k. \tag{4.1.10}$$

Wir nehmen nun zunächst an, daß die Zahlen in den k Körben durch paarweise Vergleiche (also dem ungünstigsten Verfahren) sortiert werden. Dann ist die Gesamtzahl der benötigten Sortierschritte gegeben durch die Zufallsvariable

$$S = \sum_{j=1}^{k} \binom{N_j}{2} = \sum_{j=1}^{k} \frac{N_j(N_j-1)}{2} \tag{4.1.11}$$

(vgl. hierzu auch Aufgabe 2.12). Aufgrund von (2.2.91) ergibt sich also für den Erwartungswert von S durch zweimaliges partielles Differenzieren der wahrscheinlichkeitserzeugenden Funktion nach derselben Variablen

$$E(S) = \frac{1}{2}\sum_{j=1}^{k} E\big(N_j(N_j-1)\big) = \frac{n(n-1)}{2}\sum_{j=1}^{k} p_j^2. \tag{4.1.12}$$

Im Falle einer stetigen Gleichverteilung für den Input ergibt sich somit wegen $p_j = \frac{1}{k}$, $1 \le j \le k$:

$$E(S) = k \cdot \frac{n(n-1)}{2k^2} = \frac{n(n-1)}{2k} \approx \frac{n}{2\alpha} \tag{4.1.13}$$

für große n, also eine mittlere Komplexität von $O(n)$. Leider zeigen die nachfolgenden Überlegungen jedoch, daß diese günstige Aussage nur für den Fall einer Inputverteilung gilt, für die die p_j konstant, also gleich $\frac{1}{k}$ für alle $1 \le j \le k$ sind. Hierzu betrachten wir die Funktion

$$g(p_1,\ldots,p_k) = \sum_{j=1}^{k} p_j^2 \text{ unter } h(p_1,\ldots,p_k) = \sum_{j=1}^{k} p_j = 1, \quad p_1,\ldots,p_k \ge 0.$$

Durch partielles Differenzieren der Lagrange–Funktion

$$L(p_1,\ldots,p_k) = g(p_1,\ldots,p_k) + \lambda h(p_1,\ldots,p_k), \quad \lambda \in \mathbf{R},$$

nach allen Variablen und Nullsetzen ergibt sich:

$$2p_i + \lambda = 0, \quad 1 \le i \le k,$$

also $p_i = const$ für alle $1 \le i \le k$ und somit wegen der Nebenbedingung $p_i = \frac{1}{k}$, $1 \le i \le k$. Da die Funktion g konvex ist, liegt für diese Wahl der p_i nicht nur ein relatives, sondern sogar ein absolutes Minimum von g unter der spezifizierten Nebenbedingung vor (vgl. etwa Heuser (1988), Abschnitt 174). Dies zeigt, daß die günstige Komplexität von $O(n)$ für große n mit dem in (4.1.13) gegebenen Faktor bei `hybridsort` praktisch *nur* bei einer stetigen Gleichverteilung für den Input erreicht wird, wenn in den einzelnen Körben durch paarweise Vergleiche sortiert wird. Das folgende Resultat zeigt allerdings auch, daß die Komplexität von $O(n)$ — mit einem evtl. sehr großen Faktor — erhalten bleibt, wenn die Inputverteilung eine auf dem Intervall $[0,1]$ stetige, *beschränkte* Dichte f besitzt. In diesem Fall gilt:

$$E(S) \approx \frac{n}{2\alpha} \int_0^1 f^2(x)\,dx \le \frac{M^2 n}{2\alpha} = O(n) \quad (n \to \infty) \tag{4.1.14}$$

mit $M = \sup_{0<x<1} f(x)$. Nach dem Mittelwertsatz für Integrale (vgl. Heuser (1988), 85.5) exisitiert nämlich für alle $1 \le j \le k$ eine Zahl $\xi_j \in [\frac{j-1}{k}, \frac{j}{k}]$ mit

$$F\left(\frac{j}{k}\right) - F\left(\frac{j-1}{k}\right) = \int_{\frac{j-1}{k}}^{\frac{j}{k}} f(x)\,dx = \frac{1}{k} f(\xi_j), \quad 1 \le j \le k,$$

so daß

$$k \sum_{j=1}^{k} p_j^2 = \sum_{j=1}^{k} f(\xi_j) \left[F\left(\frac{j}{k}\right) - F\left(\frac{j-1}{k}\right) \right]$$

gilt; dieser Ausdruck strebt mit $n \to \infty$ (und damit auch $k \to \infty$) aber gegen

$$E(f(X)) = \int_0^1 f^2(x)\,dx,$$

wenn X eine Zufallsvariable mit der Verteilungsfunktion F bezeichnet (vgl. den Beweis zu Satz 2.2.4 b)), womit wegen (4.1.12) Beziehung (4.1.14) gezeigt ist. Bei polynomialen Verteilungen über (0,1), d.h. Verteilungen mit

$$F_\beta(x) = x^\beta \qquad \text{bzw.} \qquad f_\beta(x) = \beta x^{\beta-1}, \quad 0 < x < 1 \quad (\beta \ge 1), \tag{4.1.15}$$

ergibt sich damit etwa

$$E(S) \approx \frac{\beta^2}{2(\beta-1)\alpha} n \qquad (n \to \infty),$$

was für große Werte von β näherungsweise linear in β wächst.
Für über (0,1) unbeschränkte Dichten kann die Komplexität jedoch beliebig nahe an $O(n^2)$ herankommen; betrachtet man z.B. in (4.1.15) den Fall $0 < \beta < 1/2$, so ergibt eine analoge Rechnung

$$E(S) \approx \frac{\beta^2}{2\alpha^{1+2\beta}(1-2\beta)} n^{2(1-\beta)} = O\big(n^{2(1-\beta)}\big) \qquad (n \to \infty).$$

Weitere Beispiele hierzu werden in den Aufgaben 4.4 bis 4.6 behandelt.
Sortiert man die einzelnen Körbe nicht rein vergleichsorientiert, sondern z.B. mit der gerade behandelten Prozedur **quicksort**, fällt die entsprechende Average-Case-Analyse etwas günstiger aus. Wegen der vorausgesetzten stochastischen Unabhängigkeit von $X_1, \ldots, X_n$ verhalten sich nämlich die in den Körben befindlichen Zahlen ebenfalls wie Realisationen unabhängiger Zufallsvariablen mit den entsprechenden bedingten Verteilungen. Genauer gilt: Bezeichen $\mathcal{K}_j = \big(\frac{j-1}{k}, \frac{j}{k}\big]$, $1 \le j \le k$, die einzelnen Körbe und definiert man für alle Indizes $j \in \{1, \ldots, n\}$ mit $N_j > 0$ die Stoppzeiten

$$S_1(j) = \min\{\ell \mid X_\ell \in \mathcal{K}_j\}, \quad S_i(j) = \min\{\ell > S_{i-1}(j) \mid X_\ell \in \mathcal{K}_j\}, \quad 1 < i < N_j,$$

so bilden also für die gerade betrachteten Indizes j die Zufallselemente $(X_{S_1(j)}, \ldots, X_{S_{N_j}(j)})$ die in den (nicht-leeren) Korb $\mathcal{K}_j$ abgelegten Zahlen. Unter Verwendung offenkundiger Modifikationen von Satz 2.1.2 und Lemma 2.1.5 folgt dann, daß die Zufallsvariablen $X_{S_1(j)}, \ldots, X_{S_{N_j}(j)}$ unter der Bedingung $N_j = m_j$, $1 \le m_j \le n$, stochastisch unabhängig sind mit der (bedingten) Verteilung

$$P^{X_{S_i(j)}}(\cdot \mid N_j = m_j) = P^X(\cdot \mid X \in \mathcal{K}_j), \quad 1 \le i \le m_j.$$

Die Anzahlen $A_1, \ldots, A_k$ der nötigen Sortierschritte in den einzelnen Körben sind daher — bei bekannten Belegungszahlen $N_1, \ldots, N_k$ — (bedingt) stochastisch unabhängig mit einer Verteilung, die lediglich von den Anzahlen $N_1, \ldots, N_k$ abhängt. Nach den Ausführungen in Beispiel 4.1.1 erhält man somit für den erwarteten Sortieraufwand in den einzelnen Körben die Abschätzung

$$E(A_j \mid N_j = m_j) \le 2(m_j + 1)\ln(m_j + 1),$$

wobei wir jetzt sogar den Fall $m_j = 0$ zulassen können (d.h. kein Sortieraufwand bei leerem Korb). Für den gesamten Sortieraufwand $S = \sum_{j=1}^{k} A_j$ ergibt sich damit die Abschätzung

$$E(S) = \sum_{j=1}^{k} E(A_j) = \sum_{j=1}^{k} E\big[E(A_j \mid N_j)\big] \le 2 \sum_{j=1}^{k} E\big((N_j + 1)\ln(N_j + 1)\big). \quad (4.1.16)$$

Zur weiteren Abschätzung von (4.1.16) benötigen wir genauere Kenntnisse der Größe

$$E\big((N+1)\ln(N+1)\big)$$

für eine $\mathfrak{B}(n,p)$–verteilte Zufallsvariable N mit $p \in [0,1]$. Hier gilt aber:

$$(1+np)\ln(1+np) \le E\big((N+1)\ln(N+1)\big) \le (1+np)\ln(n+1). \qquad (4.1.17)$$

Die linke Seite folgt dabei aus der Jensen'schen Ungleichung, Lemma 2.2.1, da die Funktion $g(x) = x\ln x$ für $x \ge 1$ konvex ist. Die rechte Seite folgt aus der Abschätzung $(N+1)\ln(N+1) \le (N+1)\ln(n+1)$. Man erhält damit insgesamt

$$\begin{aligned} E(S) &\le 2\sum_{j=1}^{k}(1+np_j)\ln(n+1) \le 2(n+k)\ln(n+1) \\ &\approx 2(1+\alpha)n\ln n = \mathrm{O}(n\ln n) \quad (n\to\infty), \end{aligned}$$

unabhängig von der Inputverteilung.
Wir wollen hier noch die naheliegende Frage untersuchen, ob im Fall eines stetig gleichverteilten Inputs die Komplexität $\mathrm{O}(n)$ aus (4.1.13) nicht vielleicht verbessert werden kann, wenn `quicksort` zur Sortierung der Körbe verwendet wird. Dies ist allerdings nicht möglich, wie die unteren Schranken in Lemma 4.1.1 und (4.1.17) bzw. die Jensen'sche Ungleichung zeigen: hier gilt nämlich

$$E(S) \ge n\ln\left(\frac{n}{k}\right) - k \approx (-\ln\alpha - \alpha)n = \mathrm{O}(n) \quad (n\to\infty)$$

für $0 < \alpha < 0.567....$
Die bisherigen Ausführungen zeigen also, daß das durchschnittliche Verhalten von `hybridsort` sehr stark davon abhängt, welche Inputverteilung gegeben bzw. welches Sortierverfahren in den einzelnen Körben angewendet wird. Ist die Inputverteilung bekannt, läßt sich allerdings sogar mit paarweise vergleichender Sortierung eine Komplexität von $\mathrm{O}(n)$, $n\to\infty$, erreichen, wenn man die Körbe $\mathcal{K}_1,\dots,\mathcal{K}_k$ so wählt, daß die Wahrscheinlichkeiten $p_1,\dots,p_k$ alle gleich groß sind, d.h.

$$\mathcal{K}_j = \left(F^{-1}\Big(\frac{j-1}{k}\Big), F^{-1}\Big(\frac{j}{k}\Big)\right], \quad 1 \le j \le k.$$

Dies ergibt sich aus Lemma 2.1.2 d), da

$$F^{-1}\Big(\frac{j-1}{k}\Big) < X \le F^{-1}\Big(\frac{j}{k}\Big) \quad \Longleftrightarrow \quad \frac{j-1}{k} < F(X) \le \frac{j}{k}$$

für alle $1 \le j \le k$ gilt und nach Satz 2.1.1 $F(X)$ gerade $\mathcal{R}\big((0,1]\big)$–verteilt ist. Für polynomiale Verteilungen mit $\beta > 0$ bedeutet dies z.B.

$$\mathcal{K}_j = \left(\Big(\frac{j-1}{k}\Big)^{\frac{1}{\beta}}, \Big(\frac{j}{k}\Big)^{\frac{1}{\beta}}\right], \quad 1 \le j \le k.$$

■

Beispiel 4.1.3. (`max-search` und `straightselection`)
Eine fundamentale Prozedur für viele Algorithmen ist das Aufsuchen des maximalen Elements in einem angeordneten Feld durch sukzessive Vergleiche — ähnlich dem Schritt b) in der Prozedur `quicksort` nach Beziehung (4.1.1). Eine mehr oder weniger ausführliche Analyse dieses Verfahrens, das wir hier kurz `max-search` nennen wollen, ist daher in fast allen einschlägigen Lehrbüchern über Sortierprobleme zu finden, etwa in Knuth (1973), Abschnitt 1.2.10 ("Algorithm M"), Kemp (1984), Kapitel 3 ("Algorithm MAX") oder Mehlhorn (1988), Abschnitt II.1.1 ("Sortieren durch Auswahl"), um exemplarisch einige davon zu nennen; vgl. hierzu auch die Aufgaben 1.10, 2.4, 2.7 und 2.8.
Für eine (zufällige) Permutation η eines Feldes $\Omega = \{\omega_1, \ldots, \omega_n\}$, $n \in \mathbb{N}$, mit $\omega_1 < \ldots < \omega_n$ benötigt man jedenfalls $n-1$ *Vergleichsschritte*, um die Position des maximalen Elements ω_n innerhalb der Permutation η zu lokalisieren, wenn man die Werte $\omega_1, \ldots, \omega_n$ nicht exakt kennt. (Aigner (1988) gibt hierfür eine anschauliche Erklärung, indem er die Prozedur mit einem Tennismatch vergleicht; siehe dort Proposition 4.2 und 4.3.) Sortiert man die Permutation η also durch Auswahl des größten, zweitgrößten usw. Elements (Prozedur `straightselection`), so benötigt man exakt

$$(n-1) + (n-2) + \ldots + 1 = \frac{n(n-1)}{2} = \binom{n}{2}$$

Vergleichsschritte; `straightselection` ist daher — etwa im Vergleich zu `qicksort` — nicht besonders effektiv. Interessanter ist hier vielmehr die Anzahl der benötigten *(Um-)speicherungen* der jeweiligen Referenzelemente. Da die Prozedur `straightselection` durch rekursive Anwendung von `max-search` entsteht, indem man sukzessiv das größte, zweitgrößte usw. Element $\omega_n, \omega_{n-1}, \ldots$ aus dem Feld bzw. der Permutation η streicht, läßt sich eine Analyse des stochastischen Verhaltens von `straightselection` unmittelbar auf eine solche von `max-search` zurückführen.
Hierzu betrachten wir für jede Permutation η die Rekursion

$$X_1 = 1, \quad X_{k+1} = \begin{cases} 1 & \text{falls } \eta_{k+1} > x_k \\ 0 & \text{sonst,} \end{cases} \qquad 1 \le k \le n-1; \tag{4.1.18}$$

$S = \sum_{k=1}^{n} X_k$ ist dann die Anzahl der verschiedenen Referenzelemente und damit der nötigen (Um-)Speicherungen für die Suche des Maximums ω_n in η. Analysen des stochastischen Verhaltens von `max-search` und `straightselection` sind unseres Wissens in der Literatur bisher nur unter Gleichverteilungsannahmen durchgeführt worden (vgl. etwa Knuth (1973), Bd. 3, Abschnitt 5.2.3 und Kemp (1984), Kapitel 3). Wir wollen daher hier ein etwas allgemeineres Verteilungsmodell vorstellen, welches die Gleichverteilung als Spezialfall umfaßt, und das in der Statistik unter dem Stichwort *successive sampling* bekannt ist (vgl. etwa Hájek (1981), Abschnitt 9). Dazu sei $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ ein Wahrscheinlichkeitsvektor, d.h. es gelte $0 \le \alpha_k \le 1$, $1 \le k \le n$, und $\sum_{k=1}^{n} \alpha_k = 1$. Die zufällige Permutation η entstehe nun vermöge der Verteilung $\boldsymbol{\alpha}$ rekursiv wie folgt:

1. Wähle den Index $i_1 \in \{1, 2, \ldots, n\}$ mit Wahrscheinlichkeit α_{i_1}. Setze $\eta_{i_1} = \omega_n$.
2. Sind die Komponenten $\eta_{i_1}, \ldots, \eta_{i_{k-1}}$, $2 \le k \le n$, bereits bestimmt, wähle unabhängig von den vorherigen Schritten den Index $i_k \notin I_{k-1} = \{i_1, \ldots, i_{k-1}\}$

mit Wahrscheinlichkeit

$$\frac{\alpha_{i_k}}{1-\sum_{j=1}^{k-1}\alpha_{i_j}} = \frac{\alpha_{i_k}}{\sum\limits_{\ell\in I_{k-1}^c}\alpha_\ell}. \tag{4.1.19}$$

Setze $\eta_{i_k} = \omega_{n-k+1}$.

Die Permutation η wird also durch sukzessive Auswahl einer Platznummer für das größte, zweitgrößte usw. Element des Feldes gemäß der aus $\boldsymbol{\alpha}$ durch Streichung der vorher bestimmten Indizes entstehenden bedingten Verteilung zusammengesetzt[1]. Für $\boldsymbol{\alpha} = (\frac{1}{n},\ldots,\frac{1}{n})$ ergeben sich damit gerade über $Perm_n^n(\Omega; o.W.)$ gleichverteilte Permutationen η.

Bezüglich des stochastischen Verhaltens von `max-search` unter diesem Verteilungsmodell ist vor allem das folgende Resultat von Bedeutung.

Satz 4.1.2. *Es bezeichne*

$$p_k = \frac{\alpha_k}{\sum_{j=1}^{k}\alpha_j}, \quad 1\le j\le n. \tag{4.1.20}$$

Dann sind die in (4.1.18) definierten Zufallsvariablen $X_1,\ldots,X_n$ stochastisch unabhängig binomialverteilt mit

$$P(X_k=1) = 1-P(X_k=0) = p_k, \quad 1\le j\le n. \tag{4.1.21}$$

Ferner ist die Verteilung der Permutation η gegeben durch

$$P(\eta_1=\omega_{\sigma(1)},\ldots,\eta_n=\omega_{\sigma(n)}) = \prod_{k=1}^{n}\frac{\alpha_{\sigma^{-1}(k)}}{\sum_{j=1}^{k}\alpha_{\sigma^{-1}(j)}} \tag{4.1.22}$$

für jede (feste) Permutation σ der Zahlen $\{1,2,\ldots,n\}$, wobei σ^{-1} die zugehörige inverse Permutation bezeichne.

Beweise dieses Satzes findet man in Pfeifer (1989b, 1990); vgl. auch die Aufgaben 4.8 und 4.9.

Für die Prozedur `max-search` bedeutet dies, daß die interessierende Anzahl S der (Um-)Speicherungen der Referenzelemente hier eine Poisson-Binomial-Verteilung $\mathfrak{PB}(n;p_1,\ldots,p_n)$ besitzt, die — sogar gleichmäßig — durch die Poisson-Verteilung $\mathfrak{P}(\lambda_n)$ mit $\lambda_n = \sum_{k=1}^{n}p_k$ approximiert werden kann, wenn der Quotient $\frac{\sum_{k=1}^{n}p_k^2}{\sum_{k=1}^{n}p_k}$ "klein" ist; dies folgt unmittelbar aus den Konvergenzabschätzungen (2.1.75) und (2.1.78). Im Fall einer Gleichverteilung für η ergibt sich damit insbesondere

$$\rho_0\big(P^S,\mathfrak{P}(\ln n+\gamma)\big) \le \rho\big(P^S,\mathfrak{P}(\ln n+\gamma)\big) \le \frac{\pi^2}{6\ln n}+\frac{1}{n} \tag{4.1.23}$$

---

[1] man kann dieses Verfahren also mit der Prozedur "Ziehen ohne Zurücklegen" nach der Verteilung $\boldsymbol{\alpha}$ identifizieren

für alle $n \in \mathbb{N}$ (wobei γ wieder die Euler'sche Konstante bezeichne), also eine erhebliche Verschärfung einer entsprechenden Aussage in Kemp (1984), S. 23 (vgl. auch Aufgabe 2.4 und Aufgabe 4.10).
Als erwartete Anzahl von (Um–)Speicherungen erhält man nach Satz 4.1.2 weiter

$$E(S) = \sum_{k=1}^{n} p_k \quad \text{mit} \quad \mathrm{Var}(S) = \sum_{k=1}^{n} p_k^2. \tag{4.1.24}$$

Im Fall einer Gleichverteilung für η bedeutet dies:

$$E(S) = \sum_{k=1}^{n} \frac{1}{k} \approx \ln n, \quad \mathrm{Var}(S) = \sum_{k=1}^{n} \frac{1}{k}\Big(1 - \frac{1}{k}\Big) \approx \ln n - \frac{\pi^2}{6}$$

für große n.
Das folgende Rechenbeispiel zeigt eine Anwendung von Satz 4.1.2 bei nicht–gleichverteiltem Input: es sei $n = 4$ und $\boldsymbol{\alpha} = \left(\frac{1}{2}\ \frac{1}{4}\ \frac{1}{8}\ \frac{1}{8}\right)$; dann ist $p_1 = 1$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{7}$, $p_4 = \frac{1}{8}$ und somit

$$E(S) = 1 + \frac{1}{3} + \frac{1}{7} + \frac{1}{8} = 1.6012, \quad \mathrm{Var}(S) = \frac{2}{9} + \frac{6}{49} + \frac{7}{64} = 0.4540.$$

Für die Verteilung von η erhält man gemäß (4.1.22) die 24 möglichen Werte (in lexikographischer Anordnung, $\Omega = \{1,2,3,4\}$):

τ	(1 2 3 4)	(1 2 4 3)	(1 3 2 4)	(1 3 4 2)	(1 4 2 3)	(1 4 3 2)
$P(\eta = \tau)$	0.0060	0.0060	0.0072	0.0072	0.0083	0.0083
τ	(2 1 3 4)	(2 1 4 3)	(2 3 1 4)	(2 3 4 1)	(2 4 1 3)	(2 4 3 1)
$P(\eta = \tau)$	0.0119	0.0119	0.0286	0.0286	0.0333	0.0333
τ	(3 1 2 4)	(3 1 4 2)	(3 2 1 4)	(3 2 4 1)	(3 4 1 2)	(3 4 2 1)
$P(\eta = \tau)$	0.o238	0.0238	0.0476	0.0476	0.0833	0.0833
τ	(4 1 2 3)	(4 1 3 2)	(4 2 1 3)	(4 2 3 1)	(4 3 1 2)	(4 3 2 1)
$P(\eta = \tau)$	0.0417	0.0417	0.0833	0.0833	0.125	0.125

Allgemeiner läßt sich zeigen, daß bei *exponentiell fallenden* Wahrscheinlichkeiten $\alpha_1, \ldots, \alpha_n$, also $\dfrac{\alpha_k}{\alpha_{k-1}} \le L < 1$ für $2 \le k \le n-1$, die $p_1, \ldots, p_n$ ebenfalls exponentiell fallen (vgl. Pfeifer (1990)), also der erwartete (Um–)Speicheraufwand bei `max-search` hier durch eine nur von L abhängige Konstante beschränkt ist, d.h. es gilt $E(S) = O(1)$.
Ist dagegen $\boldsymbol{\alpha}$ etwa gegeben durch

$$\alpha_k = \begin{cases} (1-c)^{n-1} & \text{für } k = 1 \\ c(1-c)^{n-j} & \text{für } 2 \le k \le n, \end{cases} \tag{4.1.25}$$

mit einer Konstanten $c \in (0,1)$, d.h. sind die Wahrscheinlichkeiten $\alpha_1, \ldots, \alpha_n$ *exponentiell wachsend*, so ergibt sich $p_k = \begin{cases} 1 & \text{für } k = 1 \\ c & \text{für } 2 \le k \le n, \end{cases}$ also $E(S) = 1 + (n-1)c = O(n)$ mit $n \to \infty$.

Wir wollen die für `max-search` erhaltenen Ergebnisse jetzt auf die stochastische Analyse von `straightselection` anwenden. Aufgrund der Konstruktion der Verteilung von η mittels *consecutive sampling* liegt bei jedem Durchgang, d.h. nach Streichung des größten, zweitgrößten usw. Elements, stets wieder dieselbe (bedingte) Verteilungsstruktur vor, allerdings mit Wahrscheinlichkeiten, die jetzt von den vorher entfernten Elementen bzw. deren Position innerhalb von η abhängen. Um diese komplexe Situation mathematisch zu erfassen, definieren wir für $m \ge 2$:

$$\alpha_\ell(i_1, \ldots, i_m) = \begin{cases} 0 & \text{für } \ell \in \{i_1, \ldots, i_m\} \\ \dfrac{\alpha_\ell(i_1, \ldots, i_{m-1})}{\sum\limits_{\substack{j=1 \\ j \ne i_m}}^{m} \alpha_j(i_1, \ldots, i_{m-1})} & \text{sonst,} \end{cases} \qquad 1 \le \ell \le n$$

für Permutationen $(i_1, \ldots, i_m)$ ohne Wiederholung aus Elementen der Menge $\{1, 2, \ldots, n\}$. $\alpha_\ell(i_1, \ldots, i_m)$ gibt also die Wahrscheinlichkeit dafür an, daß für η_ℓ das Element ω_{n-m+1} ausgewählt wird, wenn vorher für $\eta_{i_1}, \ldots, \eta_{i_{m-1}}$ die Elemente $\omega_n, \ldots, \omega_{n-m+2}$ ausgewählt wurden. Setzt man entsprechend

$$p_k(i_1, \ldots, i_m) = \frac{\alpha_k(i_1, \ldots, i_m)}{\sum_{j=1}^{k} \alpha_j(i_1, \ldots, i_m)}, \quad 1 \le k \le n,$$

so erhält man unter Berücksichtigung des vorher gezeigten für die Gesamtanzahl aller (Um-)Speicherungen S^* in allen Durchgängen von `straigtselection`:

$$E(S^*) = \sum_{m=1}^{n} \sum_{(i_1,\ldots,i_m) \in \Sigma_m} \alpha_{i_1}\alpha_{i_2}(i_1) \cdots \alpha_{i_m}(i_1, \ldots, i_{m-1}) \sum_{k=1}^{n} p_k(i_1, \ldots, i_{m-1}), \tag{4.1.26}$$

wobei für $m = 1$ in der rechten Summe die p_k aus Beziehung (4.1.20) zu wählen sind und $\Sigma_m = Perm_m^m(\{1, \ldots, m\}; o.W.)$, $1 \le m \le n$, bezeichne.

Der Ausdruck (4.1.26) ist i.a. nicht in einfacher, geschlossener Form darstellbar. Allerdings ist dies möglich bei einer Gleichverteilung als Inputverteilung $\boldsymbol{\alpha}$, da die bedingten Verteilungen in jedem Durchgang von `straightselection` dann selbst wieder Gleichverteilungen bilden. Es ergibt sich dann die — bekannte — vereinfachte Darstellung

$$E(S^*) = \sum_{m=1}^{n} \sum_{j=1}^{m} \frac{1}{j} = \sum_{j=1}^{n} \frac{n-j+1}{j} = (n+1) \sum_{j=1}^{n} \frac{1}{j} - n = O(n \ln n) \tag{4.1.27}$$

(vgl. auch Aufgabe 2.8).

Allerdings läßt sich eine einfache untere Schranke für $E(S^*)$ wie folgt gewinnen. Es ist jedenfalls $S^* \ge \dfrac{T(T+1)}{2}$ mit $T = \sum_{k=1}^{m} X_k$, da die Anordnungen der

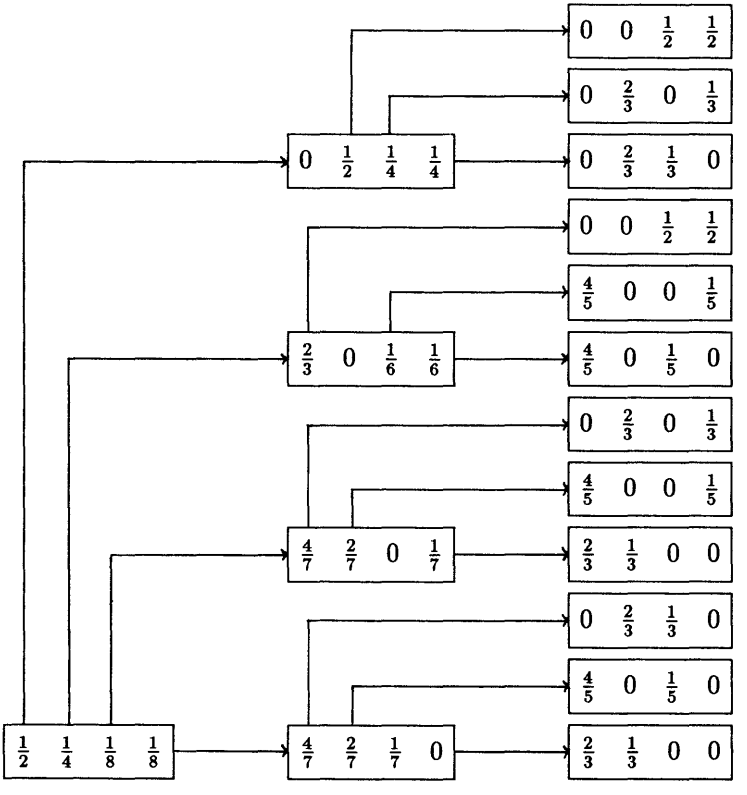
ersten Elemente von η bei jedem Durchgang von `straightselection` unverändert bleiben, bis sie entfernt werden. Daraus ergibt sich aber

$$E(S^*) \geq \frac{1}{2}E(T^2) = \frac{1}{2}\big[\operatorname{Var}(T) + \{E(T)\}^2\big] = \frac{1}{2}\sum_{k=1}^{n} p_k^2 + \frac{1}{2}\Big[\sum_{k=1}^{n} p_k\Big]^2. \quad (4.1.28)$$

Man beachte, daß das Verteilungsbeispiel aus (4.1.25) hier speziell $E(S^*) \geq c^2n^2 = \mathrm{O}(n^2)$ $(n \to \infty)$ liefert, d.h. die durchschnittliche Komplexität in diesem Beispiel genau so schlecht ist wie im Worst–Case.
Die vorigen Ausführungen zu dem erwarteten Verhalten von `max-search` bei exponentiell fallenden bzw. wachsenden Wahrscheinlichkeiten $\alpha_1, \ldots, \alpha_n$ übertragen sich allgemeiner entsprechend auch auf `straightselection`. Man erhält hier:

$\alpha_1, \ldots, \alpha_n$	exponentiell fallend	konstant	exponentiell wachsend
$E(S^*)$	$\mathrm{O}(n)$	$\mathrm{O}(n \ln n)$	$\mathrm{O}(n^2)$

An dem schon bei `max-search` betrachteten Beispiel $\boldsymbol{\alpha} = \left(\frac{1}{2}\ \frac{1}{4}\ \frac{1}{8}\ \frac{1}{8}\right)$ wollen wir die Berechnung von $E(S^*)$ über (4.1.26) abschließend exemplarisch verdeutlichen. Die folgende Graphik enthält die in den jeweiligen Durchgängen von `straightselection` auftretenden Wahrscheinlichkeiten $\alpha_\ell(i_1, \ldots, i_m)$. Die wegführenden Kanten sind dabei mit dem Wahrscheinlichkeiten zu gewichten, von denen sie ausgehen.



Hieraus resultiert die folgende Berechnung von $E(S^*)$:

$$\begin{aligned}E(S^*) = &\left(1+\tfrac{1}{3}+\tfrac{1}{7}+\tfrac{1}{8}\right)+\tfrac{1}{2}\left(1+\tfrac{1}{3}+\tfrac{1}{4}\right)+\tfrac{1}{4}\left(1+\tfrac{1}{5}+\tfrac{1}{6}\right)+\tfrac{1}{8}\left(1+\tfrac{1}{3}+\tfrac{1}{7}\right)+\tfrac{1}{8}\left(1+\tfrac{1}{3}+\tfrac{1}{7}\right)\\ &+\tfrac{1}{2}\cdot\tfrac{1}{2}\left(1+\tfrac{1}{2}\right)+\tfrac{1}{2}\cdot\tfrac{1}{4}\left(1+\tfrac{1}{3}\right)+\tfrac{1}{2}\cdot\tfrac{1}{4}\left(1+\tfrac{1}{3}\right)\tfrac{1}{4}\cdot\tfrac{2}{3}\left(1+\tfrac{1}{2}\right)+\tfrac{1}{4}\cdot\tfrac{1}{6}\left(1+\tfrac{1}{5}\right)+\tfrac{1}{4}\cdot\tfrac{1}{6}\left(1+\tfrac{1}{5}\right)\\ &+\tfrac{1}{8}\cdot\tfrac{4}{7}\left(1+\tfrac{1}{3}\right)+\tfrac{1}{8}\cdot\tfrac{2}{7}\left(1+\tfrac{1}{5}\right)+\tfrac{1}{8}\cdot\tfrac{1}{7}\left(1+\tfrac{1}{3}\right)+\tfrac{1}{8}\cdot\tfrac{4}{7}\left(1+\tfrac{1}{3}\right)+\tfrac{1}{8}\cdot\tfrac{2}{7}\left(1+\tfrac{1}{5}\right)+\tfrac{1}{8}\cdot\tfrac{1}{7}\left(1+\tfrac{1}{3}\right)+1\\ &= 5.4857.\end{aligned}$$

Bei einer Gleichverteilung für $\boldsymbol{\alpha}$ hätte sich hier der erwartungsgemäß höhere Wert $E(S^*) = 6.4167$ ergeben.
Will man bei der stochastischen Analyse von `straightselection` lediglich die reinen *Umspeicherungen* der Referenzelemente berücksichtigen (wie etwa in Knuth (1973), Kemp (1984) oder Mehlhorn (1988)), so hat man nur statt $E(S^*)$ den Ausdruck $E(S^*) - n$ zu betrachten.

4.2. Markoff–Modelle für Algorithmen

In diesem Abschnitt wollen wir zeigen, daß einige der bereits besprochenen Algorithmen sowie auch andere durch Markoff–Modelle beschrieben werden können. Eine Average–Case–Analyse dieser Modelle kann dann leicht unter Verwendung von Lemma 3.2.6 durchgeführt werden, indem man den Abbruch des Algorithmus durch geeignete erste Eintrittszeiten beschreibt.

Beispiel 4.2.1. (`binary search`)
Wir betrachten die in den Beispielen 1.1.1 und 2.2.2 gegebene Situation, ein Schlüsselelement aus einem Feld der Größe 2^n-1, $n \in \mathbb{N}$, zu suchen, wobei die Möglichkeit des Nichtvorhandenseins berücksichtigt sein soll. Als mögliche Zustände $\mathcal{S} = \{s_1, \ldots, s_{n+1}\}$ wählen wir die nach jedem Schritt verbleibenden möglichen Restfeldlängen $s_1 = 2^n - 1, s_2 = 2^{n-1} - 1, \ldots, s_{n-1} = 3, s_n = 1, s_{n+1} = 0$, wobei der Zustand $s_{n+1} = 0$ bedeutet, daß das Schlüsselelement im gerade ausgeführten Suchschritt gefunden bzw. als nicht vorhanden erkannt wurde. Es bezeichne L_k, $k \in \mathbb{N}_0$, die Zufallsvariable, die die verbleibende Restfeldlänge nach dem k–ten Schritt beschreibt, wobei angenommen sei, daß mit dem Erreichen der Zustände 0 oder 1 die Folge mit 0 fortgesetzt werde (absorbierender Zustand). Unter der Annahme einer Gleichverteilung für die Platznummer des Schlüsselelements über der Menge $\{0, 1, \ldots, 2^n - 1\}$ (wobei die Platznummer 0 wieder das Nichtvorhandensein bedeute) bildet dann die Folge $\{L_n\}_{n\in\mathbb{N}_0}$ eine homogene Markoff–Kette mit der Übergangsmatrix

$$\boldsymbol{\Pi} = \begin{array}{c} \\ s_1 \\ s_2 \\ \vdots \\ s_{n-1} \\ s_n \\ s_{n+1} \end{array} \begin{array}{c} \begin{array}{ccccccc} s_1 & s_2 & s_3 & \cdots & s_{n-1} & s_n & s_{n+1} \end{array} \\ \begin{pmatrix} 0 & 1-\alpha_1 & 0 & \cdots & 0 & 0 & \alpha_1 \\ 0 & 0 & 1-\alpha_2 & \cdots & 0 & 0 & \alpha_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1-\alpha_{n-1} & \alpha_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.2.1}$$

mit

$$\alpha_k = \frac{2^{k-1}}{2^n - 2^{k-1} + 1}, \quad 1 \le k \le n-1,$$

sowie der Anfangsverteilung $P(L_0 = s_1) = 1$. Die Stoppzeit

$$S(B) = \min\{k \in \mathbb{N}_0 \mid L_k \in B\}, \quad B = \{0,1\}, \tag{4.2.2}$$

beschreibt dann den Abbruch des Algorithmus; die erwartete Anzahl der Schritte ist damit also gegeben durch $E(S(B))$.
Wir weisen zunächst die Markoff-Eigenschaft der Folge $\{L_k\}_{k\in\mathbb{N}_0}$ nach. Nach Konstruktion der Folge $\{L_k\}_{k\in\mathbb{N}_0}$ ist $P\Big(\bigcup_{k=0}^{n}\{L_k = j_k\}\Big)$, $j_0,\ldots,j_n \in \mathcal{S}$, nur dann positiv, wenn $(j_0,\ldots,j_n)$ ein Element der Menge

$$\mathcal{J}_n = \big\{(s_1,0,0,\ldots,0),\ (s_1,s_2,0,0,\ldots,0),\ \ldots,\ (s_1,s_2,\ldots,s_{n+1})\big\}$$

ist, d.h. es können mit positiver Wahrscheinlichkeit nur Übergänge von s_k nach s_{k+1}, $1 \le k \le n-1$, oder 0 (absorbierender Zustand) erfolgen. Man kann daher für $(j_0,\ldots,j_k) \notin \mathcal{J}_k$ die bedingten Wahrscheinlichkeiten $P(L_k = j_k \mid L_0 = j_0,\ldots,L_{k-1} = j_{k-1})$, $1 \le k \le n$, im Sinne von (3.1.1.) beliebig festlegen. Wählt man etwa für $(j_0,\ldots,j_{k-1}) \in \mathcal{S}^k$, $k \in \mathbb{N}$,

$$P(L_k = s_{k+1} \,|\, L_0 = j_0,\ldots,L_{k-1} = j_{k-1}) = P(L_k = s_{k+1} \,|\, L_0 = s_1,\ldots,L_{k-1} = s_k)$$

für $j_{k-1} = s_k$ und 0 sonst, $k \le n-1$, sowie für $j_{k-1} = 0$

$$P(L_k = 0 \mid L_0 = j_0,\ldots,L_{k-1} = j_{k-1}) = 1, \quad k \ge n,$$

so ergibt sich in der Tat die Markoff-Eigenschaft der Folge $\{L_k\}_{k\in\mathbb{N}_0}$ mit

$$\begin{aligned}
1-\alpha_k &= \frac{P(L_k = s_{k+1} \mid L_0 = s_1,\ldots,L_{k-1} = s_k)}{P(L_{k-1} = s_k \mid L_0 = s_1,\ldots,L_{k-2} = s_{k-1})} \\
&= \frac{P\Big(\bigcap_{j=1}^{k} A_j^c\Big)}{P\Big(\bigcap_{j=1}^{k-1} A_j^c\Big)} = \frac{1-P\Big(\bigcup_{j=1}^{k} A_j\Big)}{1-P\Big(\bigcup_{j=1}^{k-1} A_j\Big)} \\
&= \frac{1-\sum_{j=1}^{k} P(A_j)}{1-\sum_{j=1}^{k-1} P(A_j)} = \frac{1-\sum_{j=1}^{k} 2^{i-n-1}}{1-\sum_{j=1}^{k-1} 2^{i-n-1}} \\
&= \frac{2^n-2^k+1}{2^n-2^{k-1}+1}, \quad 2 \le k \le n-1
\end{aligned}$$

unter Verwendung der Notation aus Beispiel 1.1.1, woraus

$$\alpha_k = 1 - \frac{2^n-2^k+1}{2^n-2^{k-1}+1} = \frac{2^{k-1}}{2^n-2^{k-1}+1}, \quad 2 \le k \le n-1$$

folgt, wie behauptet. Für $k = 1$ und $k = n$ argumentiert man analog.

Kürzt man $\beta_k = 1 - \alpha_k$, $1 \le k \le n$, ab, so erhält man also mit den Bezeichungen aus Lemma 3.2.6

$$I_n - \Pi_{B^c} = \begin{pmatrix} 1 & -\beta_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\beta_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -\beta_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -\beta_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

Offensichlich sind alle Zeilensummen der Matrix Π_{B^c} kleiner als 1; nach der Bemerkung im Anschluß an Lemma 3.2.6 existiert also $E\big(S(B)\big)$. Da die Anfangsverteilung durch

$$p_{B^c}(0) = (1\ 0\ 0\ \ldots\ 0)$$

gegeben ist, wird zur Berechnung von $E\big(S(B)\big)$ gemäß (3.2.39) lediglich die erste Zeile $(z_1\ z_2\ \ldots\ z_n)$ der Inversen $(I_n - \Pi_{B^c})^{-1}$ benötigt; diese ist gegeben durch die Lösung des Gleichungssystems

$$(z_1\ z_2\ \ldots\ z_n)(I_n - \Pi_{B^c})^{-1} = (1\ 0\ \ldots\ 0),$$

d.h.

$$z_1 = 1$$
$$-z_k\beta_k + z_{k+1} = 0, \quad 1 \le k \le n-1,$$

oder in expliziter Form

$$z_1 = 1, \quad z_k = \prod_{j=1}^{k-1} \beta_j = \prod_{j=1}^{k-1} \frac{2^n - 2^j + 1}{2^n - 2^{j-1} + 1}$$
$$= \frac{2^n - 2^{k-1} + 1}{2^n} = 1 - 2^{k-1-n} + 2^{-n}, \quad 2 \le k \le n.$$

Nach (3.2.39) ergibt sich jetzt

$$E\big(S(B)\big) = \sum_{k=1}^{n} z_k = n + n2^{-n} - 2^{-n} \sum_{k=0}^{n-1} 2^k = n - 1 + \frac{n+1}{2^n},$$

was erwartungsgemäß mit dem Ergebnis in Beispiel 2.2.2 übereinstimmt.
Ist die Verteilung der Position des Schlüsselelements beliebig — etwa gegeben durch die Wahrscheinlichkeiten $q_0, \ldots, q_{2^n-1}$ — so zeigt eine analoge Rechnung, daß die Folge $\{L_k\}_{k \in \mathbb{N}_0}$ auch in diesem allgemeinen Fall eine homogene Markoff-Kette bildet; man hat dabei in (4.2.1) lediglich

$$\alpha_1 = q_{2^n-1}, \quad \alpha_k = \frac{\sum_{j=1}^{2^{k-1}} q_{(2j-1)2^{n-k}}}{1 - \sum_{j=1}^{k-1} \sum_{i=1}^{2^{j-1}} q_{(2i-1)2^{n-j}}}, \quad 2 \le k \le n-1,$$

zu wählen (vgl. hierzu auch Aufgabe 1.5). Der Fall, daß das Schlüsselelement in dem Feld vorhanden ist, wird dabei durch die Wahl $q_0 = 0$ miterfaßt. Mit derselben Argumentation wie oben ergibt sich jetzt allgemeiner

$$E(S(B)) = \sum_{k=1}^{n} \prod_{j=1}^{k-1} (1 - \alpha_j). \tag{4.2.3}$$

Insbesondere ergibt sich Beziehung (2.2.5) für die Wahl $q_0 = 0$, $q_k = \dfrac{1}{2^n - 1}$, $1 \leq k \leq 2^n - 1$, denn es ist

$$\alpha_1 = \frac{1}{2^n - 1}, \quad \alpha_k = \frac{\dfrac{2^{k-1}}{2^n - 1}}{1 - \displaystyle\sum_{j=1}^{k-1} \frac{2^{j-1}}{2^n - 1}} = \frac{2^{k-1}}{2^n - 2^{k-1}}, \quad 2 \leq k \leq n-1.$$

Hiermit folgt aber

$$E(S(B)) = \sum_{k=1}^{n} \prod_{j=1}^{k-1} (1-\alpha_j) = \sum_{k=1}^{n} \prod_{j=1}^{k-1} \frac{2^n - 2^j}{2^n - 2^{j-1}} = \sum_{k=1}^{n} \frac{2^n - 2^{k-1}}{2^n - 1} = n - 1 + \frac{n}{2^n - 1};$$

dies ist (2.2.5).
Für die durch

k	0	1	2	3	4	5	6	7
q_k	0.2	0.2	0.1	0.05	0.05	0	0.1	0.1

gegebene Verteilung des Schlüsselelements erhält man beispielsweise

$$\alpha_1 = 0.05, \quad \alpha_2 = \frac{4}{19} = 0.2105...,$$

also nach (4.2.3)

$$E(S(B)) = 1 + 0.95 + 0.95 \cdot 0.7894... = 2.7.$$

Beziehung (4.2.3) erlaubt die folgende Abschätzung für den erwarteten Suchaufwand:

$$n - \sum_{j=1}^{n-1} (n-j)\alpha_j \leq E(S(B)) \leq n. \tag{4.2.4}$$

Dies folgt z.B. aus der leicht mit vollständiger Induktion zu beweisenden Ungleichung

$$1 - \sum_{j=1}^{k-1} \alpha_j \leq \prod_{j=1}^{k-1} (1 - \alpha_j) \leq 1.$$

Man beachte dabei, daß hier α_k, $1 \le k \le n-1$, die bedingte Wahrscheinlichkeit dafür ist, daß nach $k-1$ erfolglosen Suchschritten im nachfolgenden Schritt das Schlüsselelement gefunden wird. In dem betrachteten Rechenbeispiel ergibt sich beispielsweise $n - \sum_{j=1}^{n-1}(n-j)\alpha_j = 3 - 2\alpha_1 - \alpha_2 = 2.6894...$, also eine bereits recht gute Näherung für den wahren Wert von $E(S(B)) = 2.7$.
Optimale Verfahren zur Suche von Schlüsselelementen bei nicht-gleichverteilten Platznummern werden im Abschnitt 5.3 (Binäre Suchbäume) besprochen. ∎

Wir wollen nun eine Variante von `binary search` behandeln, die sich vor allem dann anbietet, wenn die exakte Teilung des gegebenen Feldes bzw. der sich im Laufe des Verfahrens ergebenden Restfelder nicht oder nur mit größerem Aufwand möglich ist. (Eine solche Situation liegt z.B. vor, wenn ein Eintrag in einem Telefonbuch gesucht wird.) Die Idee besteht darin, den Teilungspunkt des (Rest-)Feldes nicht deterministisch, sondern gemäß einer Gleichverteilung zufällig zu bestimmen; der Algorithmus verläuft ansonsten wie bei `binary search`.

Zur Vorbereitung benötigen wir allerdings noch das folgende Hilfsresultat.

Lemma 4.2.1. *(zufällige Suche)*
Es sei $\Omega = \{n, n+1, \dots, m\}$, $n \le m$, $n, m \in \mathbf{N}$. Die Zufallsvariable X sei gleichverteilt über Ω, d.h. es gelte $P^X = \mathfrak{L}(\Omega)$. Zur Suche von X in dem Feld Ω sei Y eine weitere, von X stochastisch unabhängige, ebenfalls $\mathfrak{L}(\Omega)$-verteilte Zufallsvariable. Die Zufallsvariablen N und M seien wie folgt definiert:

$$N = \begin{cases} Y & \text{für } Y \le X \\ n & \text{sonst} \end{cases} \qquad M = \begin{cases} Y & \text{für } Y \ge X \\ m & \text{sonst.} \end{cases}$$

Dann bezeichnet $L = M - N$ die (zufällige) Restfeldlänge, und es gilt:
a) $N \le X \le M$ mit

$$P^X(\cdot \mid N = i, M = j) = \mathfrak{L}(\{i, i+1, \dots, j\}), \quad n \le i \le j \le m;$$

b)
$$P(L = k) = \begin{cases} \dfrac{1}{m-n+1} & \text{für } k = 0 \\ \dfrac{2k}{(m-n+1)^2} & \text{für } 1 \le k \le m-n. \end{cases}$$

Beweis. a) Es genügt wegen (3.1.1) die Fälle $i = n$, $j \in \{n, \dots, m\}$ und $j = m$, $i \in \{n, \dots, m\}$ zu betrachten. Im ersten Fall ergibt sich

$$P(X = k \mid N = n, M = j) = \frac{P(X = k \le Y = j)}{P(X \le Y = j)}$$

mit

$$P(X = k \le Y = j) = \begin{cases} 0 & \text{für } j < k \le m \\ P(X = k)P(Y = j) = \dfrac{1}{(m-n+1)^2} & \text{für } n \le k \le j \end{cases}$$

$$P(X \le Y = j) = \sum_{k=n}^{j} P(X = k \le Y = j) = \frac{j-n+1}{(m-n+1)^2},$$

also

$$P(X=k \mid N=n, M=j)=\begin{cases}\dfrac{1}{j-n+1} & \text{für } n\le k\le j\\ 0 & \text{sonst;}\end{cases}$$

für den zweiten Fall argumentiert man analog.
b) Es ist

$$P(L=0)=P(X=Y)=\sum_{k=n}^{m}P(X=k,Y=k)=\frac{m-n+1}{(m-n+1)^2}=\frac{1}{m-n+1}$$

sowie für $1\le k\le m-n$

$$\begin{aligned}P(L=k)&=P\big(\{0\le X\le k-1, Y=k\}\cup\{m-k+1\le X\le m, Y=m-k\}\big)\\&=P(0\le X\le k-1)P(Y=k)+P(m-k+1\le X\le m)P(Y=m-k)\\&=\frac{2k}{(m-n+1)^2}.\end{aligned}$$

Damit ist das Lemma beweisen. ∎

Beispiel 4.2.2. (`random search`)
Es sei $\Omega=\{1,2,\dots,n\}$, $n\ge 2$, sowie die Zufallsvariable X gleichverteilt über Ω, d.h. $P^X=\mathfrak{L}(\Omega)$. Zur Suche von X in dem Feld Ω werde das folgende "zufällige" Suchverfahren angewandt:
Es sei $\{U_k\}_{k\in\mathbb{N}}$ eine Folge (auch von X) unabhängiger, je $\mathcal{R}((0,1))$-verteilter Zufallsvariablen. Die Zufallsvariablen $\{Y_k\}_{k\in\mathbb{N}}$, $\{N_k\}_{k\in\mathbb{N}}$ und $\{M_k\}_{k\in\mathbb{N}}$ seien rekursiv wie folgt definiert:

$$Y_1=\lceil nU_1\rceil,\qquad Y_{k+1}=N_k+\lceil(M_k-N_k-1)U_{k+1}\rceil$$

$$N_1=0,\qquad N_{k+1}=\begin{cases}Y_{k+1} & \text{für } Y_{k+1}\le X\\ N_k & \text{sonst}\end{cases}$$

$$M_1=n+1,\qquad M_{k+1}=\begin{cases}Y_{k+1} & \text{für } Y_{k+1}\ge X\\ M_k & \text{sonst.}\end{cases}$$

Hierbei bedeutet anschaulich Y_k den unabhängig von den vorherigen Schritten gemäß einer Gleichverteilung bestimmten Teilungspunkt des Restfeldes im k-ten Suchschritt und $L_k=\max\{M_k-N_k-1,0\}$, $k\in\mathbb{N}$, die jeweilige Restfeldlänge. Das Verfahren bricht ab, wenn erstmalig $L_k=0$ ist.
Unter Heranziehung von Lemma 4.2.1 läßt sich zeigen, daß die Folge $\{L_k\}_{k\in\mathbb{N}_0}$ mit $L_0=n$ der Restfeldlängen eine homogene Markoff-Kette bildet mit der Übergangsmatrix

$$\Pi=\begin{array}{c}\\ n\\ n-1\\ n-2\\ \vdots\\ 2\\ 1\\ 0\end{array}\!\!\begin{array}{c}\begin{array}{ccccccc}n & n-1 & n-2 & n-3 & \cdots & 1 & 0\end{array}\\ \begin{pmatrix}0 & \frac{2(n-1)}{n^2} & \frac{2(n-2)}{n^2} & \frac{2(n-3)}{n^2} & \cdots & \frac{2}{n^2} & \frac{1}{n}\\ 0 & 0 & \frac{2(n-2)}{(n-1)^2} & \frac{2(n-3)}{(n-1)^2} & \cdots & \frac{2}{(n-1)^2} & \frac{1}{n-1}\\ 0 & 0 & 0 & \frac{2(n-3)}{(n-2)^2} & \cdots & \frac{2}{(n-2)^2} & \frac{1}{n-2}\\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & 0 & 0 & \cdots & 0 & 1\\ 0 & 0 & 0 & 0 & \cdots & 0 & 1\end{pmatrix}\end{array}\tag{4.2.5}$$

Bezeichnet wieder $B = \{0,1\}$ und $S(B) = \min\{k \in \mathbb{N}_0 \mid L_k \in B\}$, so bricht also wie in Beispiel 4.2.1 das Verfahren nach $S(B)$ Schritten ab. Die Berechnung von $E(S(B))$ läßt sich daher völlig analog wie in Beispiel 4.2.1 durchführen. Allerdings führt die Berechnung der ersten Zeile $(z_1\ z_2\ \dots\ z_n)$ der Inversen Matrix $I_n - \Pi_{B^c}$ hier auf die etwas kompliziertere Rekursion

$$z_1 = 1, \quad z_k = 2(n-k+1)\sum_{j=1}^{k-1} \frac{z_j}{(n-j+1)^2}, \quad 2 \le k \le n. \tag{4.2.6}$$

Da eine explizite Darstellung der z_k nicht in einfacher Form möglich ist, begnügen wir uns mit einer hinreichend guten Abschätzung. Es gilt nämlich:

$$z_k \le \frac{2}{n-k+2}, \quad 2 \le k \le n.$$

Wir zeigen dies durch vollständige Induktion:
Für $k = 2$ ist $z_k = \frac{2(n-1)}{n^2} \le \frac{2}{n-1}$, d.h. die Abschätzung ist für $k = 2$ richtig.
Die Abschätzung gelte für $k \in \mathbb{N}$, $k \ge 2$. Es ist

$$\begin{aligned} z_{k+1} &= 2(n-k)\sum_{j=1}^{k} \frac{z_j}{(n-j+1)^2} \\ &= 2(n-k)\sum_{j=1}^{k-1} \frac{z_j}{(n-j+1)^2} + 2(n-k)\frac{z_k}{(n-k+1)^2} \\ &= \frac{n-k}{n-k+1} z_k + \frac{2(n-k)}{(n-k+1)^2)} z_k = \frac{(n-k)(n-k+3)}{(n-k+1)^2} z_k \\ &\le \frac{(n-k+1)(n-k+2)}{(n-k+1)^2} \frac{2}{n-k+2} = \frac{2}{n-k+1}, \end{aligned}$$

d.h. die Abschätzung gilt dann auch für $k+1$.
Für den erwarteten Suchaufwand erhält man also

$$\begin{aligned} E(S(B)) &= \sum_{k=1}^{n} z_k \le 1 + \sum_{k=2}^{n} \frac{2}{n-k+2} \\ &= 1 + 2\sum_{k=2}^{n} \frac{1}{k} \le 1 + 2\ln n. \end{aligned} \tag{4.2.7}$$

Vergleicht man den erwarteten Suchaufwand von `random search` mit `binary search` für $n = 2^m - 1$, $m \ge 2$, so ergibt sich im ersten Fall

$$E(S(B)) \le 1 + 2\ln(2^m - 1) \le 1 + 2\ln 2\, m \approx 1 + 1.386\, m,$$

im zweiten

$$E(S(B)) = m - 1 + \frac{m}{2^m - 1}.$$

Für $n = 2^m - 1$, $m = 2, 3, \ldots, 9$ sind in der folgenden Tabelle die exakten Werte von $E_R = E(S(B))$ bei `random search` angegeben, darunter zum Vergleich die Werte von $E_B = E(S(B))$ bei `binary search`.

n	3	7	15	31	63	127	255	511
E_R	1.8889	2.9265	4.0789	5.3143	6.6066	7.9361	9.2889	10.6558
E_B	1.6667	2.4286	3.2667	4.1612	5.0952	6.0551	7.0314	8.0176

Der erwartete Aufwand ist bei `random search` gegenüber `binary search` also etwa nur maximal 39% größer bei derselben Komplexität von $O(m)$ für große m; dies zeigt, daß `random search` im Mittel dieselbe Effektivität wie `binary search` besitzt. Allerdings unterscheiden sich beide Verfahren im Worst-Case drastisch: während `binary search` auch hier höchstens m Schritte benötigt, können dies bei `random search` im ungünstigsten Fall $2^m - 1$ Schritte sein.
Betrachtet man statt der ursprünglichen Menge $\Omega = \{1, \ldots, n\}$ die Menge $\Omega^* = \{\frac{1}{n}, \frac{2}{n}, \ldots, 1\} \subset [0,1]$, so läßt sich die diskrete Gleichverteilung über Ω^* durch eine Rechteckverteilung $\mathcal{R}([0,1])$ approximieren. Die Ausführungen in Pfeifer (1985) zeigen dann, daß sogar bei beliebiger Verteilung P^X des Schlüsselelements in dem Feld der Größe n der erwartete Suchaufwand unter `random search` immer noch $O(\ln n)$ beträgt; eine obere Schranke hierfür ist gegeben durch

$$E(S(B)) \leq 1 + 4 \ln n.$$

Allerdings kann das durchschnittliche Verhalten von `random search` unter gewissen Verteilungsannahmen für X sogar besser sein als bei `binary search`, etwa im Fall einer Zweipunktverteilung auf den Enden des Feldes, d.h.

$$P(X = 1) = p, \quad P(X = n) = 1 - p, \quad p \in [0,1]$$

(vgl. hierzu auch Aufgabe 4.12). Man erhält hier

$$E(S(B)) = \sum_{k=1}^{n} \frac{1}{k} \leq 1 + \ln n,$$

d.h. bei Feldgrößen $n = 2^m - 1$, $m \in \mathbb{N}$, ergibt sich

$$E(S(B)) \leq 1 + \ln(2^m - 1) \leq 1 + \ln 2m \approx 1 + 0.6931\, m;$$

`random search` ist hier also im Mittel etwa 30% schneller als `binary search`, da bei dieser Verteilung von X `binary search` stets genau m Suchschritte benötigt. Dies zeigt, daß `random search` im Mittel dann schneller als `binary search` ist, wenn sich die Verteilung P^X weniger auf die Mitte als vielmehr auf die Ränder des Feldes konzentriert. ∎

Zum Abschluß dieses Abschnitts wollen wir zeigen, daß auch die allgemeine Version von `max-search`, die bereits in Abschnitt 4.1 ausführlicher behandelt wurde, als Markoff-Modell formulierbar ist, was im wesentlichen auf das folgende Resultat zurückzuführen ist.

Satz 4.2.1. *(Stoppzeiten bei Binomialverteilungen)*
Es sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge stochastisch unabhängiger, jeweils binomialverteilter Zufallsvariablen mit

$$P(X_n = 1) = 1 - P(X_n = 0) = p_n \in (0,1], \quad n \in \mathbb{N},$$

und

$$\sum_{n=1}^{\infty} p_n = \infty. \tag{4.2.8}$$

Dann bilden die rekursiv über

$$S_0 = \inf\{n \in \mathbb{N} \mid X_n = 1\}, \quad S_{k+1} = \inf\{n > S_k \mid X_n = 1\}, \quad k \in \mathbb{N}_0,$$

definierten Stoppzeiten $\{S_k\}_{k\in\mathbb{N}_0}$ eine homogene Markoff-Kette mit Übergangswahrscheinlichkeiten

$$P(S_{k+1} = j \mid S_k = i) = \begin{cases} p_j \prod_{\ell=i+1}^{j-1}(1-p_\ell) & \text{für } j > i \geq 1 \\ 0 & \text{sonst} \end{cases} \tag{4.2.9}$$

sowie Anfangsverteilung

$$P(S_0 = j) = p_j \prod_{\ell=1}^{j-1}(1-p_\ell), \quad j \in \mathbb{N}. \tag{4.2.10}$$

Beweis. Aufgrund der Beziehung (2.1.16) ist klar, daß Teil a) des Satzes 2.1.2 auch unter der allgemeineren Annahme unabhängiger, aber beliebig verteilter Zufallselemente gültig bleibt. Insbesondere sind also das Ereignis $\{S_k = n\}$ und die Folge $\{X_{n+1}, X_{n+2}, \ldots\}$ für alle $n > k \in \mathbb{N}_0$ stochastisch unabhängig. Nach der Bemerkung im Anschluß an Lemma 3.2.2 bildet also die Folge $\{S_k\}_{k\in\mathbb{N}_0}$ eine Markoff-Kette mit

$$P(S_1 = j) = P\Big(\bigcap_{\ell=1}^{j-1}\{X_\ell = 0\} \cap \{X_j = 1\}\Big) = p_j \prod_{\ell=1}^{j-1}(1-p_\ell)$$

und

$$P(S_{k+1} = j \mid S_k = i) = P\Big(\bigcap_{\ell=i+1}^{j-1}\{X_\ell = 0\} \cap \{X_j = 1\}\Big), \quad k \in \mathbb{N}_0,\ j > i \in \mathbb{N}.$$

Man beachte dabei, daß die Bedingung (4.2.8) nach dem Borel–Cantelli–Lemma Satz 1.1.3 garantiert, daß die Folge $\{S_k\}_{k\in\mathbb{N}_0}$ f.s. nicht degeneriert, d.h. f.s. unendlich oft das Ereignis $\{1\}$ in der Folge $\{X_n\}_{n\in\mathbb{N}}$ eintritt. ∎

Für unsere Zwecke benötigen wir noch die folgende leichte Modifikation von Satz 4.2.1.

Lemma 4.2.2. *Es sei $n \in \mathbb{N}$ fest. Unter den Voraussetzungen von Satz 4.2.1 seien die Zufallsvariablen $\{T_k\}_{k\in\mathbb{N}_0}$ definiert durch*

$$T_k = \begin{cases} S_k & \text{falls } S_k \le n \\ n+1 & \text{sonst} \end{cases}, \quad k \in \mathbb{N}_0.$$

Dann ist auch die Folge $\{T_k\}_{k\in\mathbb{N}_0}$ eine homogene Markoff-Kette mit absorbierendem Zustand $n+1$, Übergangswahrscheinlichkeiten

$$P(T_{k+1} = j \mid T_k = i) = \begin{cases} P(S_{k+1} = j \mid S_k = i) & \text{für } 1 \le i < j \le n \\ \sum_{\ell=n+1}^{\infty} P(S_{k+1} = \ell \mid S_k = i) & \text{für } 1 \le i \le n,\ j = n+1 \\ 1 & \text{für } i = j = n+1 \end{cases} \tag{4.2.11}$$

für $k \in \mathbb{N}_0$ und Anfangsverteilung

$$P(T_0 = j) = \begin{cases} P(S_0 = j) & \text{für } 1 \le j \le n \\ \sum_{\ell=n+1}^{\infty} P(S_{k+1} = \ell) & \text{für } j = n+1. \end{cases}$$

Beweis. Dies folgt unmittelbar aus der Definition der Folge $\{T_k\}_{k\in\mathbb{N}_0}$. ∎

Beispiel 4.2.3. (`max-search`)
Wir betrachten ein angeordnetes Feld $\Omega = \{\omega_1, \ldots, \omega_n\}$, $n \in \mathbb{N}$. Die (zufällige) Permutation $\eta = (\eta_1, \ldots, \eta_n) \in \Omega^{(n,n)} = Perm_n^n(\Omega; o.W.)$ sei über $\Omega^{(n)}$ gleichverteilt. Die Zufallsvariablen $X_1, \ldots, X_n$ seien definiert durch

$$X_1(\eta) = 1, \quad X_j(\eta) = \begin{cases} 1 & \text{falls } \eta_j > \max\{\eta_1, \ldots, \eta_{j-1}\} \\ 0 & \text{sonst} \end{cases}, \quad 2 \le j \le n.$$

Nach den Ausführungen in Abschnitt 4.1 bzw. Aufgabe 1.10 sind die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig mit

$$P(X_j = 1) = 1 - P(X_j = 0) = \frac{1}{j}, \quad 1 \le j \le n. \tag{4.2.12}$$

Die Zufallsvariable $S = \sum_{j=1}^{n} X_j$ gibt also die Anzahl der (Um-)Speicherungen der Referenzelemente bei `max-search` an. Nach Lemma 4.2.2 kann `max-search` alternativ auch über die Markoff-Kette $\{T_j\}_{j\in\mathbb{N}_0}$ der Indizes der Referenzelemente der Permutation η beschrieben werden, wobei das Erreichen des absorbierenden Zustands $n+1$ signalisiert, daß das Maximum der Permutation gefunden wurde. Setzt man die Wahrscheinlichkeiten (4.2.12) in (4.2.11) ein, so ergibt sich die folgende Übergangsmatrix

$$\Pi = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n-1 \\ n \\ n+1 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & n-1 & n & n+1 \end{array} \\ \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{6} & \cdots & \frac{1}{(n-1)(n-2)} & \frac{1}{n(n-1)} & \frac{1}{n} \\ 0 & 0 & \frac{2}{6} & \cdots & \frac{2}{(n-1)(n-2)} & \frac{2}{n(n-1)} & \frac{2}{n} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{n-1}{n(n-1)} & \frac{n-1}{n} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.2.13}$$

mit der Anfangsverteilung $P(T_0 = 1) = 1$. Setzt man hier $B = \{n+1\}$, so gibt wieder die Stoppzeit $S(B)$ die Anzahl der benötigten (Um-)Speicherungen der Referenzelemente an. Wie in den vorangehenden Beispielen kann nun $E\big(S(B)\big)$ über die erste Zeile $(z_1 \ \dots \ z_n)$ der Inversen $I_n - \mathbf{\Pi}_{B^c}$ berechnet werden. Es ist

$$I_n - \mathbf{\Pi}_{B^c} = \begin{pmatrix} 1 & -\beta_2 & -\beta_3 & -\beta_4 & \cdots & -\beta_{n-1} & -\beta_n \\ 0 & 1 & -2\beta_3 & -2\beta_4 & \cdots & -2\beta_{n-1} & -2\beta_n \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & 1 & -(n-1)\beta_n \\ 0 & 0 & 0 & 0 & \vdots & 0 & 1 \end{pmatrix} \tag{4.2.14}$$

mit $\beta_j = \frac{1}{j(j-1)}$, $2 \le j \le n$, woraus sich die Rekursion

$$z_1 = 1, \quad z_j = \beta_j \sum_{i=1}^{j-1} i z_i, \quad 2 \le j \le n$$

ergibt, bzw. explizit

$$z_1 = 1, \quad z_j = \frac{1}{j}, \quad 2 \le j \le n.$$

Man erhält somit erwartungsgemäß

$$E\big(S(B)\big) = \sum_{j=1}^{n} \frac{1}{j} \le 1 + \ln n.$$

Das in Abschnitt 4.1 vorgestellte allgemeine Modell für `max-search` läßt sich völlig analog behandeln, indem man mit den Bezeichnungen aus (4.1.20)

$$p_i = \frac{\alpha_i}{\alpha_1 + \ldots + \alpha_i}, \quad 1 \le i \le n,$$

wählt; für die Matrix $I_n - \mathbf{\Pi}_{B^c}$ ergibt sich dann die Darstellung

$$\begin{pmatrix} 1 & -p_2 & -q_2p_3 & -q_2q_3p_4 & \cdots & -q_2\cdots q_{n-2}p_{n-1} & -q_2\cdots q_{n-1}p_n \\ 0 & 1 & -p_3 & -q_3p_4 & \cdots & -q_3\cdots q_{n-2}p_{n-1} & -q_3\cdots q_{n-1}p_n \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -q_{n-1}p_n \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

wobei $q_j = 1 - p_j$, $2 \le j \le n-1$ zu setzen ist, was wieder zu

$$E\big(S(B)\big) = \sum_{j=1}^{n} p_j$$

führt. ∎

Weitere Markoff-Modelle für Algorithmen findet man z.B. in Barth (1982, 1984), Kemp (1984), Abschnitt 2.2 und Mathar & Mann (1990).

4.3. Konvexe Hüllen von Zufallspunkten

Ist eine große Zahl von Punkten $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in der Ebene gegeben, so interessiert oft das größte Gebiet, das bei gradliniger Verbindung der Punkte umschlossen wird. Präzisiert wird diese Fragestellung durch den Begriff der konvexen Hülle einer Menge $A \subseteq \mathbf{R}^m$, $m \in \mathbf{N}$. Dabei heißt eine Menge $B \subseteq \mathbf{R}^m$ konvex, wenn sie mit je zwei Punkten auch deren Verbindungsgerade enthält, wenn also $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in B$ für alle $\boldsymbol{x}, \boldsymbol{y} \in B$, $\alpha \in [0,1]$.

Definition 4.3.1. *Die konvexe Hülle einer Menge $A \in \mathbf{R}^m$ ist definiert als die kleinste, konvexe Menge B, die A enthält. Sie wird bezeichnet mit $B = \operatorname{conv}(A)$.*

Da der Durchschnitt von beliebig vielen konvexen Mengen wieder konvex ist, existiert eine solche kleinste, konvexe Menge, und $\operatorname{conv}(A)$ ist wohldefiniert. Sie besitzt die Darstellung

$$\operatorname{conv}(A) = \Big\{ \boldsymbol{y} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \mid \boldsymbol{x}_i \in A,\ \lambda_i \geq 0,\ i = 1, \ldots, k,\ \textstyle\sum_{i=1}^{k} \lambda_i = 1,\ k \in \mathbf{N} \Big\}. \tag{4.3.1}$$

Die konvexe Hülle von n Punkten $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbf{R}^2$ läßt sich graphisch als konvexes Polygon veranschaulichen, dessen Eckpunkte von gewissen $\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_\ell}$ gestellt werden und das alle Punkte $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ enthält. Es entsteht, indem man jeden Punkt mit jedem anderen durch eine Gerade verbindet und dann die äußere Hülle aller Geraden nimmt. Der kürzeste Polygonzug, der alle Punkte umgibt, bildet den Rand dieses Gebiets. Die konvexe Hülle läßt sich durch Angabe der sie aufspannenden Eckpunkte eindeutig beschreiben, wobei die Eckpunkte sinnvollerweise nach aufsteigenden Winkeln zwischen der Verbindungsgeraden zu einem Referenzpunkt, der zur Hülle gehört, und der x–Achse sortiert werden.

Die Bestimmung der konvexen Hülle von n Punkten fällt leicht, wenn n klein ist, sie muß jedoch algorithmisch mit Hilfe von Computern durchgeführt werden, wenn die Anzahl der Punkte sehr groß ist. Das oben beschriebene Verfahren, aus allen Verbindungsgeraden die äußeren herauszusuchen, läßt sich zwar relativ leicht implementieren, ist aber mit quadratischem Aufwand nicht effizient. Es gibt jedoch bessere Algorithmen, die diese Aufgabe selbst im schlechtesten Fall (Worst–Case) mit einem Zeitaufwand von $O(n \ln n)$ lösen.

Der von Graham 1972 vorgeschlagene Algorithmus (Graham scan) benötigt höchstens $O(n \ln n)$ Schritte zur Bestimmung der konvexen Hülle von n Punkten $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in der Ebene mit orthogonalen x–y–Koordinaten. Die Idee soll hier kurz skizziert werden, eine vollständige Implementierung findet sich etwa in Sedgewick (1988).

Beginne mit dem Punkt $\boldsymbol{a}_{i_1}$, der die kleinste y– und größte x–Koordinate besitzt, und sortiere alle Punkte nach aufsteigenden Winkeln zwischen der x–Richtung und der Geraden mit den Endpunkten $\boldsymbol{a}_{i_1}$ und $\boldsymbol{a}_\ell$, $\ell = 1, \ldots, n$, $\ell \neq i_1$. Die so sortierten Punkte mit $\boldsymbol{a}_{i_1}$ als erstem Element seien mit $\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(n)}$ bezeichnet. Bestimme nun sukzessive die konvexe Hülle von $\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(r)}$, $4 \leq r \leq n$, nach folgendem Verfahren.

Bilden $\boldsymbol{a}_{(i_1)}, \ldots, \boldsymbol{a}_{(i_{n_r})}$ die Eckpunkte der konvexen Hülle von $\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(r)}$ und ist $\boldsymbol{a}_{(r+1)}$ der nächste in Frage kommende Punkt, so eliminiere so lange Punkte $\boldsymbol{a}_{(i_{n_r-\ell})}$, $\ell = 0, 1, 2, \ldots$, bis der Kantenzug $\boldsymbol{a}_{(i_{n_r-\ell-1})} \to \boldsymbol{a}_{(i_{n_r-\ell})} \to \boldsymbol{a}_{(r)}$ erstmalig einen Linksknick macht. Jeder Punkt wird hierbei höchstens einmal eliminiert,

woraus nach Sortierung höchstens lineare Zeit für die Berechnung der konvexen Hülle resultiert. Der Hauptaufwand wird durch die anfängliche Sortierung aller Punkte nach aufsteigenden Winkeln verursacht, wofür bekanntlich Algorithmen der Komplexität $O(n \ln n)$ zur Verfügung stehen.

Also existiert zur Berechnung der konvexen Hülle von n Punkten $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbf{R}^2$ ein Algorithmus *CH*, dessen Rechenkomplexität höchstens proportional zu $n \ln n$ wächst, wobei die Anzahl der Punkte die Problemgröße bestimmt.

Ein solcher Algorithmus wird typischerweise häufig auf verschiedene Konstellationen von Eingaben angewendet, wobei man natürlich nicht genau weiß, welche Punkte in Zukunft als Input auftreten. In dieser Situation hilft ein stochastisches Modell. Die Punkte $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ können als Realisationen von stochastisch unabhängigen, zweidimensionalen Zufallsvektoren $X_1, \ldots, X_n$ angesehen werden, deren Verteilung P^{X_i} man kennt. Die Rechenkomplexität ist dann eine Funktion des zufälligen Inputs, deren Erwartungswert Auskunft über die mittlere Laufzeit gibt.

Wir nehmen im folgenden an, daß $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ Realisationen von stochastisch unabhängigen, identisch verteilten Zufallsvektoren $X_1, \ldots, X_n$ sind, deren Verteilung jeweils eine Gleichverteilung auf $K_1^a = \{(x_1, x_2) \in \mathbf{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$ ist mit der Dichte

$$f(x_1, x_2) = \frac{1}{\pi} \mathbb{1}_{K_1^a}(x_1, x_2), \quad (x_1, x_2) \in \mathbf{R}^2$$

(vergleiche (1.4.28)). Wir sprechen dann von unabhängigen, im Einheitskreis gleichverteilten Punkten.

Wir werden jetzt sehen, wie mit Hilfe des Algorithmus *CH* ein Verfahren zur Bestimmung der konvexen Hülle von n unabhängigen, in der Einheitskreisscheibe gleichverteilten Punkten konstruiert wird, dessen erwartete Komplexität nur noch linear in n wächst. Es geht auf Borgwardt, Gaffke, Jünger & Reinelt (1989) zurück.

Das Verfahren basiert auf der Idee, daß Punkte innerhalb des einbeschriebenen offenen Kreises $K_\eta = \{(x_1, x_2) \mid x_1^2 + x_2^2 < \eta^2\}$, $0 < \eta < 1$, für nahe bei 1 liegendes, von n unabhängiges η nur mit geringer Wahrscheinlichkeit als Eckpunkte der konvexen Hülle in Frage kommen. In der ersten Phase des Algorithmus werden also Punkte innerhalb der Kreisscheibe K_η ausgesondert, anschließend wird die konvexe Hülle der verbleibenden Punkte im Kreisring $K_1^a \setminus K_\eta$ gebildet und dann überprüft, ob sie K_η enthält. Ist dies der Fall, hat man $\operatorname{conv}(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\})$ schon aus den äußeren Punkten bestimmt, falls nicht, wird in einer zweiten Phase der Algorithmus *CH* auf die gesamte Punktwolke angewendet.

Wählt man $\eta = \eta(n)$ geschickt, lohnt der oben beschriebene Versuch, in einer ersten Phase die konvexe Hülle nur aus wenigen, äußeren Punkten zu konstruieren, wenn nur die Wahrscheinlichkeit gering ist, daß dieser Versuch fehlschlägt. $\eta(n)$ wird so bestimmt, daß der erwartete Gesamtaufwand beim Einsatz des Algorithmus *CH* sublinear in n bleibt.

Wir beschreiben zunächst den zweiphasigen Algorithmus *LDCH* [1] in einer PASCAL-nahen Notation. Bezeichne $A = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$, $n \in \mathbf{N}$, die Menge der unabhängigen, im Einheitskreis gleichverteilten Punkte und $K_\eta^c = K_1^a \setminus K_\eta$ das

[1] *LDCH* ist ein Kürzel aus den charakteristischen Eigenschaften: linear expected running time, unit disk, convex hull.

Komplement von K_η im Einheitskreis.

```
PROCEDURE LDCH;
BEGIN
  {1. Bestimme den Parameter η = η(n) > 0.}
  {2. Setze S = {a_i ∈ A | a_i ∈ K_η^c}.}
  {3. Wende CH auf S an ⟶ conv(S).}
  {4. IF K_η ⊂ conv(S)} THEN {conv(A):=conv(S)}
        ELSE {5. Wende CH auf A an ⟶ conv(A)}
END;
```

(4.3.2)

Die Mächtigkeit der Menge S ist in unserem Modell eine Zufallsvariable mit Träger $\{0,1,\dots,n\} \subset \mathbf{N}$. Sie läßt sich schreiben als

$$\#(S) = \sum_{i=1}^{n} \mathbb{1}_{K_\eta^c}(X_i). \tag{4.3.3}$$

Die konvexe Hülle von S ist eine Funktion der zufälligen Punkte $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$. Es gilt

$$K_\eta \subseteq \operatorname{conv}(S) \quad \text{genau dann, wenn} \quad K_\eta \subseteq \operatorname{conv}(A) \tag{4.3.4}$$

wie man sich leicht mit Hilfe der Eckpunkte beider konvexer Hüllen klarmacht (Aufgabe 4.14). Hieraus folgt mit (4.3.1), daß das Ereignis $K_\eta \not\subseteq \operatorname{conv}(S)$ die Darstellung

$$\bigcup_{\lambda\in[0,1]\cap\mathbf{Q}} \bigcup_{1\le i_1,i_2\le n} \{\|\lambda X_{i_1} + (1-\lambda)X_{i_2}\| \ge \eta\}$$

besitzt, also eine meßbare Menge ist. $\|(x_1,x_2)\| = \sqrt{x_1^2+x_2^2}$ bezeichnet hierbei die euklidische Norm in $\mathbf{R}^2$. Wir werden später die schwierig zu berechnende Wahrscheinlichkeit für $K_\eta \not\subseteq \operatorname{conv}(S)$ nach oben abschätzen.

Der Erwartungswert des Zeitaufwands T, den CH im Algorithmus (4.3.2) insgesamt benötigt, wird in den Schritten 3. und 5. durch die Mächtigkeit der Menge S bestimmt. Er läßt sich folgendermaßen abschätzen.

$$\begin{aligned} E(T) &\le C\Big\{ \sum_{k=1}^{n} k\ln k \cdot P\big(\#(S)=k\big) + n\ln n \cdot P\big(K_\eta \not\subseteq \operatorname{conv}(S)\big)\Big\} \\ &\le C\Big\{ \ln n \cdot \sum_{k=1}^{n} k \cdot P\big(\#(S)=k\big) + n\ln n \cdot P\big(K_\eta \not\subseteq \operatorname{conv}(S)\big)\Big\} \\ &= C\Big\{ \ln n \cdot E\big(\#(S)\big) + n\ln n \cdot P\big(K_\eta \not\subseteq \operatorname{conv}(S)\big)\Big\}. \end{aligned} \tag{4.3.5}$$

C ist hierbei die Konstante, die die obere Schranke $C \cdot n\ln n$ für die Komplexität des Algorithmus CH festlegt.

Der erwartete Aufwand der Prozedur *LDCH* verhält sich nun linear in n, wenn $E(T) = \mathrm{O}(n)$ und die Schritte 2. und 4. höchstens lineare Zeit benötigen.

Für den Schritt 2. ist das klar; man hat für jeden der n Punkte lediglich abzufragen, ob er innerhalb des Kreises K_η liegt oder nicht.

Die Abfrage in 4. wird nun in zwei Schritten realisiert. Das Ergebnis des Algorithmus *CH* in Schritt 3. ist eine Liste von Punkten, von denen $\boldsymbol{a}_{i_1}$ die kleinste y– und größte x–Koordinate besitzt und die nach aufsteigenden Winkeln zwischen der Geraden $[\boldsymbol{a}_{i_1}, \boldsymbol{a}_j]$ und der x–Richtung sortiert sind, $j = 2, \ldots, \ell$. Für diese Punkte wird untersucht,

a) ob der Nullpunkt $(0,0)$ in $\mathrm{conv}(\{\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_\ell}\})$ liegt. Ist dies nicht der Fall, ist das Ergebnis der Abfrage in 4. `false` und 5. wird durchgeführt. Anderenfalls wird untersucht,

b) ob alle Verbindungsgeraden zwischen $\boldsymbol{a}_{i_j}$ und $\boldsymbol{a}_{i_{j+1}}$, $j = 1, \ldots, l-1$, und $\boldsymbol{a}_{i_\ell}$ und $\boldsymbol{a}_{i_1}$ keinen Punkt mit K_η gemeinsam haben. Ist dies richtig, wurde die konvexe Hülle bereits bestimmt, anderenfalls folgt Schritt 5.

Diese Untersuchung ist in linearer Zeit $\mathrm{O}(\ell)$, also wegen $\ell \leq n$ in linearer Zeit $\mathrm{O}(n)$ möglich.

Es bleibt also, einen erwarteten Aufwand $E(T) = \mathrm{O}(n)$ zu realisieren. Wir widmen uns zunächst $E\big(\#(S)\big)$, dem ersten Summanden in (4.3.5).

Lemma 4.3.1. *$\#(S)$ ist binomialverteilt mit Parametern n und $p = 1 - \eta^2$. Folglich gilt*

$$E\big(\#(S)\big) = n(1-\eta^2).$$

Beweis. Für alle $i = 1, \ldots, n$ besitzt $\mathbb{1}_{K_\eta^c}(X_i)$ den Träger $\{0,1\}$. $\mathbb{1}_{K_\eta^c}(X_i)$ ist also $\mathfrak{B}(1,p)$–verteilt, wobei

$$p = P\big(\mathbb{1}_{K_\eta^c}(X_i) = 1\big) = P\big(X_i \in K_\eta^c\big) = \frac{1}{\pi}\int_{K_\eta^c} d(x_1,x_2) = \frac{1}{\pi}\pi(1-\eta^2) = 1-\eta^2. \tag{4.3.6}$$

$\int_{K_\eta^c} d(x_1,x_2)$ ist hierbei die Fläche des Kreisrings mit Innenradius η und Außenradius 1, das ist die Differenz der Flächen der Einheitskreisscheibe und der Kreisscheibe K_η, die wir mit elementargeometrischen Mitteln zu $\pi - \pi\eta^2$ berechnen. Die Zufallsvariablen $\mathbb{1}_{K_\eta^c}(X_i)$, $i = 1, \ldots, n$ sind nach Lemma 2.1.7 stochastisch unabhängig und identisch verteilt. Mit der Darstellung (4.3.3) und (2.1.18) besitzt $\#(S)$ eine $\mathfrak{B}(n, 1-\eta^2)$–Verteilung, deren Erwartungswert bereits in (2.2.88) berechnet wurde. ∎

Die im zweiten Summanden der letzten Zeile von (4.3.5) auftretende Wahrscheinlichkeit $P\big(K_\eta \not\subseteq \mathrm{conv}(S)\big)$ ist schwierig zu berechnen. Wir werden jedoch im weiteren eine obere Schranke hierfür herleiten. Sei hierzu $\xi > 0$ ein Parameter und n so groß, daß $(\frac{1}{3} + \xi)\ln n < n^{2/3}$. Wir setzen

$$\begin{aligned} \delta = \delta(n) &= \sqrt{1 - n^{-\frac{2}{3}}\Big(\frac{1}{3} + \xi\Big)\ln n}, \\ \eta = \eta(n) &= \delta\sqrt{1 - \Big(\frac{2\pi}{\lfloor n^{1/3}\rfloor}\Big)^2}. \end{aligned} \tag{4.3.7}$$

Dann gilt $0 < \eta < \delta < 1$, und K_δ^c bildet einen Kreisring mit Innenradius δ und Außenradius 1, der die Kreisscheibe K_η umschließt. Für große Werte von n liegen $\delta(n)$ und $\eta(n)$ nahe bei 1, und der Kreisring K_δ^c wird sehr schmal. Wir

definieren in diesem Kreisring Segmente $T_1, T_2, \ldots, T_{\lfloor n^{1/3} \rfloor}$ gleicher Größe mit Hilfe von Polarkoordinaten (r, φ), $r > 0, 0 \leq \varphi \leq 2\pi$, durch

$$T_i = \{(r,\varphi) \mid \delta \leq r \leq 1,\ t_{i-1} \leq \varphi \leq t_i\},$$

wobei die Winkel t_i definiert sind durch

$$t_i = i\frac{2\pi}{\lfloor n^{1/3} \rfloor}, \quad i = 0, 1, \ldots, \lfloor n^{1/3} \rfloor.$$

Lemma 4.3.2. *Es seien η und δ wie in (4.3.7). Falls dann jedes der Segmente T_i, $i = 1, \ldots, \lfloor n^{1/3} \rfloor$, mindestens einen Punkt von A enthält, gilt $K_\eta \subseteq \operatorname{conv}(A)$.*

Beweis. Die Behauptung ist richtig, wenn die Verbindungsgerade zwischen beliebigen Punkten benachbarter Segmente $\mathbf{u} = (u_1, u_2) \in T_{i-1}$ und $\mathbf{v} = (v_1, v_2) \in T_i$, $i = 1, \ldots, \lfloor n^{1/3} \rfloor$, ganz in K_η^c verläuft $(T_0 = T_{\lfloor n^{1/3} \rfloor})$. Wegen der Rotationssymmetrie reicht es, die Segmente T_1 und $T_{\lfloor n^{1/3} \rfloor}$ zu betrachten. Sei also $\mathbf{x} = (x_1, x_2) \in \{\lambda\mathbf{u} + (1-\lambda)\mathbf{v} \mid 0 \leq \lambda \leq 1\}$. Geometrisch macht man sich klar, daß $u_1, v_1 \geq \delta \cos \frac{2\pi}{\lfloor n^{1/3} \rfloor}$. Folglich gilt $x_1 \geq \delta \cos \frac{2\pi}{\lfloor n^{1/3} \rfloor}$. Die Ungleichung

$$\cos\Big(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\Big) = \sqrt{1 - \sin^2\Big(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\Big)} \geq \sqrt{1 - \Big(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\Big)^2}$$

liefert

$$\sqrt{x_1^2 + x_2^2} \geq x_1 \geq \delta\sqrt{1 - \Big(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\Big)^2} = \eta,$$

d.h. $\mathbf{x} \in K_\eta^c$, woraus die Behauptung folgt. ∎

Mit Hilfe dieser geometrischen Überlegungen läßt sich nun eine obere Schranke für die gesuchte Wahrscheinlichkeit konstruieren.

Lemma 4.3.3. *Für die in (4.3.7) festgelegten Werte von η und δ gilt für alle $n \in \mathbb{N}$, $\xi > 0$ mit $(\frac{1}{3} + \xi)\ln n < n^{2/3}$:*

$$P\big(K_\eta \not\subseteq \operatorname{conv}(S)\big) \leq n^{-\xi}.$$

Beweis. Wegen (4.3.4) gilt $\{K_\eta \not\subseteq \operatorname{conv}(S)\} = \{K_\eta \not\subseteq \operatorname{conv}(A)\}$, und mit Lemma 4.3.2 folgt

$$\begin{aligned}
P\big(K_\eta \not\subseteq &\operatorname{conv}(S)\big) \leq P\big(A \cap T_i = \emptyset \text{ für mindestens ein } i \in \{1, \ldots, \lfloor n^{1/3} \rfloor\}\big) \\
&= P\Big(\bigcup_{i=1}^{\lfloor n^{1/3} \rfloor} \{A \cap T_i = \emptyset\}\Big) \leq \sum_{i=1}^{\lfloor n^{1/3} \rfloor} P(A \cap T_i = \emptyset) \\
&= \sum_{i=1}^{\lfloor n^{1/3} \rfloor} \Big(1 - \frac{1}{\pi}\int_{T_i} d(x_1, x_2)\Big)^n = \lfloor n^{1/3} \rfloor \Big(1 - \frac{\pi - \pi\delta^2}{\pi\,\lfloor n^{1/3} \rfloor}\Big)^n \\
&= \lfloor n^{1/3} \rfloor \Big(1 - \frac{(\frac{1}{3} + \xi)\ln n}{\lfloor n^{1/3} \rfloor n^{2/3}}\Big)^n \leq n^{1/3}\Big(1 - \frac{(\frac{1}{3} + \xi)\ln n}{n}\Big)^n \\
&\leq n^{1/3} e^{-(\frac{1}{3}+\xi)\ln n} = n^{-\xi}.
\end{aligned}$$

In dieser Formel wurde verwendet, daß $\int_{T_i} d(x_1, x_2)$, der Flächeninhalt des Ringsegments T_i, unabhängig von i den Wert $\frac{\pi - \pi\delta^2}{\lfloor n^{1/3} \rfloor}$ hat.
Die letzte Ungleichung folgt durch äquivalente Umformungen aus der bekannten Ungleichung $\ln z \leq z - 1$ für alle $0 < z < 1$. ∎

Wir sind jetzt in der Lage, den erwarteten Zeitaufwand des Algorithmus *CH* in (4.3.5) abzuschätzen. Hierzu werden die in (4.3.7) definierten Werte $\eta = \eta(n)$ und $\delta = \delta(n)$ mit einem festen Parameter $\xi > 0$ für *LDCH* verwendet.

Lemma 4.3.4. *η, δ und ξ seien wie oben definiert. Dann gilt für den erwarteten, gesamten Zeitaufwand des Algorithmus CH, dessen Komplexität im Worst–Case durch $C \cdot n \ln n$ beschränkt ist, für alle $n \geq 27$:*

$$E(T) \leq C \cdot \left\{ n^{\frac{1}{3}} \ln n \left(\left(\frac{1}{3} + \xi \right) \ln n + (4\pi)^2 \right) + n^{1-\xi} \ln n \right\}.$$

Beweis. Für obiges η und δ folgt mit Lemma 4.3.1

$$\begin{aligned} E(\#(S)) &= n(1-\eta^2) = n\left(1 - \delta^2\left(1 - \left(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\right)^2\right)\right) \\ &= n\left(1 - \delta^2 + \delta^2\left(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\right)^2\right) \leq n\left(1 - \delta^2 + \left(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\right)^2\right) \\ &= n\left(n^{-\frac{2}{3}}\left(\frac{1}{3} + \xi\right)\ln n + \left(\frac{2\pi}{\lfloor n^{1/3} \rfloor}\right)^2\right) = n^{\frac{1}{3}}\left(\frac{1}{3} + \xi\right)\ln n + \frac{n(2\pi)^2}{(\lfloor n^{1/3} \rfloor)^2} \\ &\leq n^{\frac{1}{3}}\left(\left(\frac{1}{3} + \xi\right)\ln n + (4\pi)^2\right), \quad \text{falls } n \geq 27. \end{aligned}$$

Dies eingesetzt in (4.3.5) liefert, wenn man wie in Lemma 4.3.3 die Wahrscheinlichkeit $P(K_\eta \not\subseteq \mathrm{conv}(S))$ abschätzt, sofort die Behauptung. ∎

Betrachten wir noch einmal zusammenfassend die Vorgehensweise: Zur Verfügung steht der Algorithmus *CH*, dessen Worst–Case–Verhalten den Zeitaufwand $O(n \ln n)$ erfordert. Dieser wird in den Algorithmus *LDCH* eingebaut, indem er die konvexe Hülle einer kleinen Zahl außenliegender Punkte von im Einheitskreis gleichverteilten ermittelt.

Wenn die trennende Kreisscheibe K_η mit η aus (4.3.7) eine Teilmenge dieser konvexen Hülle ist, stoppt *LDCH*, im anderen Fall muß *CH* auf alle Punkte angewendet werden. Dieser Fall tritt aber mit so kleiner Wahrscheinlichkeit ein, daß der erwartete Aufwand beim Einsatz von *CH* sublinear bleibt, wie man an der oberen Schranke aus Lemma 4.3.4 erkennt.

Dominierend für den Gesamtaufwand bleiben damit die linearen Laufzeiten von Schritt 2. und Schritt 4. in *LDCH*. Insgesamt haben wir das folgende Ergebnis bewiesen.

Satz 4.3.1. (*Borgwardt u.a. (1989)*)
Sind $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, $n \in \mathbb{N}$, stochastisch unabhängige, aus einer Gleichverteilung auf der Einheitskreisscheibe des $\mathbb{R}^2$ erzeugte Punkte, so ist die erwartete Laufzeit

des Algorithmus LDCH linear in n, wenn η wie in (4.3.7) mit einem beliebigen Parameter $\xi > 0$ gewählt wird.

Mit den gleichen Methoden, allerdings einer anderen, trennenden Menge K_η, läßt sich der Algorithmus *CH* auch im Fall einer Gleichverteilung auf dem Einheitsquadrat

$$Q_1^a = [0,1]^2 = \{(x_1, x_2) \mid 0 \leq x_1, x_2 \leq 1\} \subset \mathbf{R}^2$$

im Mittel auf linearen Aufwand reduzieren.

Seien im folgenden $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, $n \in \mathbf{N}$, Realisationen von stochastisch unabhängigen, je auf Q_1^a gleichverteilten Zufallsvariablen $X_1, \ldots, X_n$ mit der Dichte (vergleiche (1.4.27))

$$f(x_1, x_2) = \mathbb{1}_{Q_1^a}(x_1, x_2).$$

Für diese Situation betrachten wir den Algorithmus *LSCH* [1], der wie *LDCH* aus (4.3.2) arbeitet, in dem jedoch die Menge K_η ersetzt wird durch G_η, $0 < \eta < 1$, mit

$$G_\eta = \{(x_1, x_2) \in Q_1^a \mid \min\{x_1, 1 - x_1\} \cdot \min\{x_2, 1 - x_2\} > \eta\}$$

und $G_\eta^c = Q_1^a \setminus G_\eta$. Wählt man

$$\eta = \eta(n) = \sqrt{e}\,\frac{(2 + \xi)\ln n}{n}$$

für einen Parameter $\xi > 0$, so gilt Satz 4.3.1 unverändert für die Gleichverteilung auf dem Einheitsquadrat bei Einsatz des Algorithmus *LSCH*.

Eine analoge Abschätzung der Wahrscheinlichkeit $P(G_\eta \not\subseteq \mathrm{conv}(S))$ wie in Lemma 4.3.3 ist hier schwieriger, da die Rotationssymmetrie bei Aufteilung von G_δ^c, $\delta > \eta$, in Segmente verlorengeht. Für die technisch aufwendigen, sehr trickreichen Details sei auf die Arbeit von Borgwardt, Gaffke, Jünger & Reinelt (1989) verwiesen.

[1] *LSCH*: linear expected running time, unit square, convex hull.

4.4. Aufgaben

4.1 Zeigen Sie, daß für alle $1 \le k \le n \in \mathbb{N}$

$$\sum_{\ell=0}^{\min\{k-1,n-k\}} \binom{k-1}{\ell}\binom{n-k}{\ell} = \binom{n-1}{k-1}$$

gilt.

4.2 Bestimmen Sie zu festem $n \in \mathbb{N}$ die Anzahl der Felder der Länge n, die nach einem Aufruf der Prozedur **quicksort** schon sortiert sind (ohne daß der Algorithmus dann abbricht).

4.3 Untersuchen Sie die Funktion $f(x) = x + 0.35 + x\ln x - (x+1)\ln(x+1)$, $x > 0$, auf Nullstellen und Konvexität.

4.4 Zeigen Sie, daß die erwartete Anzahl $E(S)$ von Sortierschritten unter **hybridsort** bei paarweisen Vergleichen in den Körben und der Inputverteilung mit der (unbeschränkten) Dichte $f_m(x) = \dfrac{(-\ln x)^m}{m!}$, $0 < x < 1$, $m \in \mathbb{N}_0$, gegeben ist durch

$$E(S) \approx \frac{(2m)!}{(m!)^2}\frac{n}{2\alpha} \quad (n \to \infty).$$

Bemerkung: f_m ist die Dichte der Zufallsvariablen $\prod_{j=1}^{m+1} X_j$, wobei $X_1, \ldots, X_{m+1}$ stochastisch unabhängig und jeweils $\mathcal{R}((0,1])$–verteilt sind.

4.5 Zeigen Sie, daß **hybridsort** im Mittel auch dann mit $O(n)$ Sortierschritten auskommt, wenn die Inputverteilung eine unbeschränkte Dichte f besitzt, für die das uneigentliche Riemann–Integral $\int_0^1 f^2(x)\,dx$ existiert.

4.6 Weisen Sie nach, daß die Komplexität von **hybridsort** bei vergleichsorientierter Sortierung in den Körben unter einer polynmialen Inputverteilung (Beziehung (4.1.15)) gegeben ist durch $O(n)$ für $1/2 < \beta < 1$ und $O(n\ln n)$ für $\beta = 1/2$. Bestimmen Sie jeweils asymptotisch geeignete Faktoren hierfür.

4.7 Führen Sie eine Average–Case–Analyse von **hybridsort** durch für den Fall einer beliebigen, *stetigen* Verteilung über $\mathcal{B}^1$. Wie müssen Sie die Körbe wählen, um ein möglichst günstiges Sortierverhalten zu erzielen? Welches Vorgehen ist bei einer *beliebigen* Verteilung empfehlenswert?
Anleitung: Transformieren Sie die Zufallsvariablen $X_1, \ldots, X_n$ mit Hilfe der gegebenen Verteilungsfunktion F und benutzen Sie Satz 2.1.1.

4.8 Zeigen Sie mit elementaren Überlegungen, daß die in (4.1.20) angegebene Beziehung für die Wahrscheinlichkeiten p_k im Fall $k = n$, $k = n-1$ und $k = 1$ richtig sind. Zeigen Sie weiter durch direktes Nachrechnen, daß

$$P\Big(\bigcap_{k=1}^{n}\{X_k = 1\}\Big) = \prod_{k=1}^{n} p_k$$

gilt. Kann man hieraus schon auf die stochastische Unabhängigkeit der Zufallsvariablen $X_1, \ldots, X_n$ schließen?

4.9 Beweisen Sie die Gültigkeit der Beziehung (4.1.22).
Anleitung: Benutzen Sie die Eigenschaften von *consecutive sampling*.

4.10 Zeigen Sie die Gültigkeit der Fehlerabschätzungen in Beziehung (4.1.23).
Anleitung: Aufgabe 2.4, Dreiecksungleichung für Metriken, Beziehung (2.1.66) und die Beziehung

$$\ln n + \gamma - \frac{1}{n} \leq \sum_{k=1}^{n} \frac{1}{k} \leq \ln n + \gamma, \quad n \in \mathbf{N}$$

(γ =Euler'sche Konstante).

4.11 (**max-search** bei Dreiecksverteilung) Die Verteilungen $\boldsymbol{\alpha}$ und $\overline{\boldsymbol{\alpha}}$ seien gegeben durch

$$\alpha_i = \frac{2i}{n(n+1)}, \quad \overline{\alpha}_i = \alpha_{n-i+1}, \quad 1 \leq i \leq n.$$

Zeigen Sie (Notation wie in Beispiel 4.1.3):

$$p_i = \frac{2}{i+1}, \quad \overline{p}_i = p_{n-i+1}, \quad 1 \leq i \leq n;$$

$$\begin{aligned} E(S_n) &= 2\ln n + 2\gamma - 2 + O\Big(\frac{1}{n}\Big) \\ E(\overline{S}_n) &= \ln n + \gamma - \ln 2 + O\Big(\frac{\ln n}{n}\Big) \end{aligned} \qquad (n \to \infty).$$

Wie beurteilen Sie das Verhalten von **max-search** mit diesen Verteilungen im Vergleich zu $\boldsymbol{\alpha} = (\frac{1}{n}, \ldots, \frac{1}{n})$? Kann man hier sinnvoll eine Poisson–Approximation für die Verteilung von S_n bzw. $\overline{S}_n$ vornehmen (vgl. Pfeifer (1990))?

4.12 (**random search**) Es sei $\Omega = \{1, 2, \ldots, n\}$, $n \in \mathbf{N}$, und die Zufallsvariable X gegeben durch $X \equiv 1$. Zeigen Sie, daß unter der Prozedur **random search** die Folge der Restfeldlängen eine homogene Markoff–Kette bildet. Geben Sie die zugehörige Übergangsmatrix $\boldsymbol{\Pi}$ explizit an und leiten Sie daraus die Beziehung

$$E(S(B)) = \sum_{k=1}^{n} \frac{1}{k} \approx \ln n + \gamma$$

für die erste Eintrittszeit $S(B)$ mit $B = \{0\}$ ab. Wieso bleibt dasselbe Ergebnis auch für beliebige Zweipunktverteilungen auf den Enden des Feldes (d.h. mit Träger $\subseteq \{1, n\}$) richtig?

4.13 Analysieren Sie das Average–Case–Verhalten von **random search** bei einer Dreiecks–Inputverteilung $\boldsymbol{\alpha}$ für die Zufallsvariable X wie in Aufgabe 4.11 im Fall $n = 5$. Läßt sich das Verhalten des Algorithmus hier auch durch ein homogenes Markoff–Modell beschreiben (Begründung)?

4.14 S und A bezeichne die Mengen aus Algorithmus **LDCH** aus Beziehung (4.3.2), K_η den Kreis mit Radius η. Zeigen Sie: $K_\eta \subset \mathrm{conv}(S)$ gilt genau dann, wenn $K_\eta \subset \mathrm{conv}(A)$ ist.

5. Elemente der Informationstheorie

Die Codierung von Signalfolgen spielt in der Informatik eine wichtige Rolle. Man denke etwa an Codierungsprobleme bei der Übertragung von Signalfolgen in Kommunikationssystemen oder das Abspeichern und Wiederauffinden von Schlüsselwörtern einer Programmiersprache in der Symboltafel durch den Compiler. Die Vorschrift, die bei einer bestimmten Speicherstrategie zum Wiederauffinden eines Identifiers führt, kann dabei auch als Codierung des entsprechenden Namens gedeutet werden.

Sind nun Wahrscheinlichkeiten für das Auftreten einzelner Quellsignale oder Schlüssel bekannt, etwa in Form von relativen Häufigkeiten aus Erfahrungswerten, wird man häufig auftretende Symbole durch kurze Wörter codieren, selten auftretende dagegen eher durch lange. Entsprechend steigt man in Suchbäumen zur Identifikation wenig benutzter Schlüssel lieber durch häufige Abfragen tief hinab, wenn man dafür den Vorteil einer schnellen Identifikation durch wenige Abfragen für oft auftretende Identifier genießt.

Wir setzen im folgenden voraus, daß a–priori–Kenntnisse über das Auftreten von zu codierenden Elementen einer endlichen Menge Ω in Form einer vorgegebenen Wahrscheinlichkeitsverteilung P auf $(\Omega, \mathfrak{P}(\Omega))$ vorliegen. Unser weiteres Vorgehen hängt dann davon ab, wie die Wahrscheinlichkeitsverteilung P strukturiert ist. Die erwartete Länge von Codewörtern oder die erwartete Anzahl von Abfragen ist ein Maß dafür, mit welchem Aufwand man bei häufigem Einsatz eines Verfahrens im Mittel rechnen muß. Wir werden bemüht sein, diese Zielgröße zu minimieren, und dabei sehen, daß sie eng mit einem Maß für die Unbestimmtheit der unterliegenden Verteilung verbunden ist.

Folgendes Beispiel, in dem zwei stochastische Vektoren der Länge drei verglichen werden, zeigt, daß Verteilungen einen verschiedenen Grad von "Unbestimmtheit" besitzen.

$$\boldsymbol{p} = (0.3,\ 0.4,\ 0.3) \qquad \boldsymbol{q} = (0.02,\ 0.04,\ 0.94)$$

Im ersten Fall tritt jeder der drei möglichen Ausgänge mit beinahe gleicher Wahrscheinlichkeit auf; es fällt schwer, vor Beobachten des zugehörigen Experimentes den Ausgang richtig vorherzusagen, während im zweiten Fall die letzte Möglichkeit mit hoher Wahrscheinlichkeit eintritt. Nach Beobachten des Ausgangs ist der Informationsgewinn bei der ersten Verteilung groß, bei der zweiten dagegen gering, da man dort den Ausgang mit hoher Trefferquote hätte vorhersagen können.

Ziel der nachfolgenden Betrachtungen ist es, ein vernünftiges, quantitatives Maß für die "Unbestimmtheit" oder den "Informationsgewinn" bei Zufallsexperimenten herzuleiten und dessen Auswirkungen bei Codierungs– und Suchproblemen zu untersuchen. Eine zentrale Rolle spielt hierbei der Begriff der Entropie.

5.1. Information und Entropie

Wir werden im folgenden synonym von der Entropie einer Zufallsvariablen X und der einer Verteilung P sprechen. Beide Sprechweisen lassen sich in einem gemeinsamen Modell begründen, wenn P mit P^X, der Verteilung einer Zufallsvariablen X, identifiziert wird.

Definition 5.1.1. *X sei eine endlich diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Wertebereich $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \in \mathbb{N}$, und Verteilung P^X. Für $1 \leq j \leq m$ bezeichne $p_j = P^X(\{x_j\}) = P(X = x_j)$.*

$$H(X) = -\sum_{j=1}^{m} P(X = x_j) \cdot \log P(X = x_j) = -\sum_{j=1}^{m} p_j \cdot \log p_j \tag{5.1.1}$$

heißt Entropie der Zufallsvariablen X oder der Verteilung P^X.

In (5.1.1), wie auch im folgenden, verwenden wir die Konventionen

$$0 \cdot \log 0 = 0 \quad \text{und} \quad 0/0 = 0. \tag{5.1.2}$$

Als Basis des zugehörigen Logarithmus kann eine beliebige Konstante > 1 gewählt werden, ihr spezieller Wert bestimmt die Skala, auf der $H(X)$ gemessen wird.

Die Entropie hängt nicht von der speziellen Gestalt des Wertebereichs oder der Zufallsvariablen ab, sondern nur von den Wahrscheinlichkeiten $p_j = P(X = x_j)$, $1 \leq j \leq m$. Für eine beliebige endlich diskrete Verteilung, die durch einen stochastischen Vektor $\boldsymbol{p} = (p_1, \ldots, p_m)$ beschrieben wird, können wir die Entropie äquivalent durch $H(p_1, \ldots, p_m)$ notieren. Die Schreibweise mit Hilfe von Zufallsvariablen ist in vielen Fällen jedoch kürzer und prägnanter.

Die oben definierte Entropie besitzt einige charakteristische Eigenschaften, die man von einem Maß für Unbestimmtheit fordert. Mehr noch – wir werden nachweisen, daß die Entropie bis auf skalare Vielfache das einzige Maß für Unbestimmtheit oder Informationsgewinn ist, wenn man im wesentlichen drei einleuchtende Bedingungen für solche Maße fordert.

Zunächst wird der Begriff der Entropie, der für beliebige endlich diskrete Verteilungen definiert ist, speziell im Fall von endlich diskreten Zufallsvektoren und bedingten Verteilungen betrachtet.

Hierzu sei (X, Y) ein zweidimensionaler, endlich diskreter Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Verteilung $P^{(X,Y)}$ und Wertebereich $\mathcal{X} \times \mathcal{Y}$, wobei $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \in \mathbb{N}$, und $\mathcal{Y} = \{y_1, \ldots, y_n\}$, $n \in \mathbb{N}$. Es bezeichne für $1 \leq i \leq m$, $1 \leq j \leq n$

$$\begin{aligned}
&p_{ij} = P^{(X,Y)}\big(\{(x_i, y_j)\}\big) = P(X = x_i, Y = y_j), \\
&p_{i\cdot} = \sum_{j=1}^{n} p_{ij} = P(X = x_i), \quad p_{\cdot j} = \sum_{i=1}^{m} p_{ij} = P(Y = y_j), \\
&p_{i|j} = \frac{p_{ij}}{p_{\cdot j}} = P(X = x_i \mid Y = y_j), \quad p_{j|i} = \frac{p_{ij}}{p_{i\cdot}} = P(Y = y_j \mid X = x_i)
\end{aligned} \tag{5.1.3}$$

die Wahrscheinlichkeiten der Elementarereignisse der gemeinsamen Verteilung, der Randverteilungen und der bedingten Verteilungen von (X,Y). Dann ist nach Definition 5.1.1

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \cdot \log p_{ij} \tag{5.1.4}$$

die Entropie des Zufallsvektors (X,Y) und

$$H(X \mid Y = y_j) = -\sum_{i=1}^{m} p_{i|j} \cdot \log p_{i|j} \tag{5.1.5}$$

die bedingte Entropie von X, gegeben $Y = y_j$.

Die bedingte Entropie von X unter Y erhält man durch Erwartungswertbildung von $H(X \mid Y = \cdot)$ aus (5.1.5) bezüglich der Verteilung P^Y. Es ist

$$\begin{aligned} H(X \mid Y) &= \sum_{j=1}^{n} P(Y = y_j) H(X \mid Y = y_j) \\ &= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{\cdot j} p_{i|j} \log p_{i|j} = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{i|j}. \end{aligned} \tag{5.1.6}$$

Ganz analog erhält man für die bedingte Entropie von Y unter X

$$H(Y \mid X) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{i\cdot} p_{j|i} \log p_{j|i} = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{j|i}. \tag{5.1.7}$$

Einen Zusammenhang zwischen der Entropie der gemeinsamen Verteilung, der der Randverteilungen und der bedingten Entropie stellt das folgende Lemma her.

Lemma 5.1.1. *Unter den Bezeichnungen (5.1.3) – (5.1.7) gilt*

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y). \tag{5.1.8}$$

Beweis. Es gilt

$$\begin{aligned} H(X,Y) &= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \cdot \log p_{ij} = -\sum_{i=1}^{m}\left(\sum_{j=1}^{n} p_{ij} \cdot \log p_{i\cdot} + \sum_{j=1}^{n} p_{ij} \cdot \log \frac{p_{ij}}{p_{i\cdot}}\right) \\ &= -\sum_{i=1}^{m} p_{i\cdot} \cdot \log p_{i\cdot} - \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \cdot \log p_{j|i} = H(X) + H(Y \mid X) \end{aligned}$$

Man beachte, daß für alle Indizes i und j aus $p_{i\cdot} = 0$ die Gleichheit $p_{ij} = 0$ folgt, so daß obige Quotienten und Produkte mit den Konventionen (5.1.2) wohldefiniert sind.

Die zweite Identität wird analog bewiesen. ∎

Induktiv erhält man mit Lemma 5.1.1 die folgende allgemeinere Aussage. Hierbei sind $X_1, \dots, X_L$, $L \in \mathbb{N}$, endlich diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Für jede Auswahl von Indizes $1 \le i_1 \le \cdots \le i_\ell \le L$, $\ell \le L$, bildet $Y = (X_{i_1}, \dots X_{i_\ell})$ wieder eine endlich diskrete Zufallsvariable. Wir können damit die sukzessive bedingten Entropien bilden, für die gilt

$$H(X_1, \dots, X_L) = H(X_1) + \sum_{k=2}^{L} H\big(X_k \mid (X_1, \dots, X_{k-1})\big). \tag{5.1.9}$$

Für $L = 2$ ist dies Gleichung (5.1.8). Im Induktionsschritt zeigen wir mit (5.1.8)

$$\begin{aligned} H(X_1, \dots, X_{L+1}) &= H(X_1, \dots, X_L) + H\big(X_{L+1} \mid (X_1, \dots, X_L)\big) \\ &= H(X_1) + \sum_{k=2}^{L+1} H\big(X_k \mid (X_1, \dots, X_{k-1})\big). \end{aligned}$$

Die Eigenschaft (5.1.8) kann als einleuchtende Eigenschaft eines Maßes für Unbestimmtheit oder Information interpretiert werden. Die Unbestimmtheit des zusammengesetzten Experimentes, das durch den Zufallsvektor (X, Y) beschrieben wird, setzt sich additiv zusammen aus der Unbestimmtheit der einen Komponente und der der anderen, gegeben die erste. Information kann in diesem Sinn also stufenweise gewonnen werden.

Wir wenden uns nun einigen wichtigen Ungleichungen der Entropie zu.

Satz 5.1.1. *Unter den Bezeichnungen (5.1.3) – (5.1.7) gilt:*

a) $$0 \overset{\text{(i)}}{\le} H(X) \overset{\text{(ii)}}{\le} \log m,$$

wobei Gleichheit in (i) *genau dann gilt, wenn* $P(X = x_i) = 1$ *für ein* $i \in \{1, \dots, m\}$, *und Gleichheit in* (ii) *genau dann, wenn* $P(X = x_i) = \frac{1}{m}$ *für alle* $i \in \{1, \dots, m\}$ *(Gleichverteilung).*

b) $$0 \overset{\text{(i)}}{\le} H(X \mid Y) \overset{\text{(ii)}}{\le} H(X)$$

(Shannon'sche Ungleichung), wobei Gleichheit in (i) *genau dann gilt, wenn* $P(X = x_i \mid Y = y_j) \in \{0, 1\}$ *für alle* i, j, *und Gleichheit in* (ii) *genau dann, wenn* X *und* Y *stochastisch unabhängig sind.*

c) $$H(X) \overset{\text{(i)}}{\le} H(X, Y) \overset{\text{(ii)}}{\le} H(X) + H(Y),$$

wobei Gleichheit in (i) *genau dann gilt, wenn* $P(Y = y_j \mid X = x_i) \in \{0, 1\}$ *für alle* i, j, *und Gleichheit in* (ii) *genau dann, wenn* X *und* Y *stochastisch unabhängig sind.*

Beweis. a) (i) ist leicht einzusehen, wobei $-\sum_{i=1}^m p_i \cdot \log p_i = 0$ genau dann, wenn $p_i \in \{0,1\}$ für alle $i = 1,\dots,m$. Da $\sum_{i=1}^m p_i = 1$, gilt dies genau dann, wenn $p_i = 1$ für ein $i \in \{1,\dots,m\}$. Zum Beweis von (ii) benutzen wir die Ungleichung

$$\log z = \log e \cdot \ln z \;\le\; (\log e)\cdot(z-1), \quad z > 0, \tag{5.1.10}$$

mit Gleichheit genau dann, wenn $z = 1$. Es folgt

$$H(X) - \log m = \sum_{i=1}^m p_i \cdot \log \frac{1}{p_i} - \sum_{i=1}^m p_i \cdot \log m = \sum_{i=1}^m p_i \cdot \log \frac{1}{mp_i}$$

$$= (\log e) \sum_{\substack{i=1\\ p_i\neq 0}}^m p_i \cdot \ln \frac{1}{mp_i} \le (\log e) \sum_{\substack{i=1\\ p_i\neq 0}}^m p_i\Big(\frac{1}{mp_i} - 1\Big) = (\log e)\Big(\sum_{\substack{i=1\\ p_i\neq 0}}^m \frac{1}{m} - 1\Big) \le 0,$$

wobei Gleichheit genau dann gilt, wenn $1/(mp_i) = 1$ für alle $i = 1,\dots,m$.
b) (i) folgt unmittelbar aus der Definition (5.1.6). Die notwendige und hinreichende Bedingung für Gleichheit ergibt sich aus den Äquivalenzen

$$-\sum_{i=1}^m \sum_{j=1}^n p_{ij} \cdot \log p_{i|j} = 0 \quad\Longleftrightarrow\quad p_{ij} \cdot \log p_{i|j} = 0 \quad \text{für alle } i,j$$

$$\Longleftrightarrow \quad p_{ij} = 0 \quad \text{oder} \quad p_{i|j} = 1 \quad \text{für alle } i,j$$

$$\Longleftrightarrow \quad p_{i|j} = 1 \quad \text{für alle } i,j \text{ mit } p_{ij} > 0,$$

da unter (5.1.2) $p_{ij} = 0$ dann und nur dann gilt, wenn $p_{i|j} = 0$. (ii) ergibt sich aus der Ungleichungskette

$$H(X \mid Y) - H(X) = \sum_{i=1}^m p_{i\cdot} \cdot \log p_{i\cdot} - \sum_{i=1}^m \sum_{j=1}^n p_{ij} \cdot \log p_{i|j}$$

$$= \sum_{i=1}^m \sum_{j=1}^n \Big(p_{ij} \cdot \log p_{i\cdot} - p_{ij} \cdot \log p_{i|j}\Big) = (\log e) \sum_{\substack{i=1\\ p_{ij}\neq 0}}^m \sum_{j=1}^n p_{ij} \cdot \ln \frac{p_{i\cdot}}{p_{i|j}}$$

$$\le (\log e)\Big(\sum_{\substack{i=1\\ p_{ij}\neq 0}}^m \sum_{j=1}^n p_{i\cdot}p_{\cdot j} - 1\Big) \le 0,$$

wobei die vorletzte Ungleichung wegen (5.1.10) folgt. Gleichheit erhält man genau dann, wenn für alle i,j im Fall $p_{ij} > 0$ die Identität $p_{i\cdot} = p_{i|j}$ gilt und im Fall $p_{ij} = 0 \quad p_{i\cdot}p_{\cdot j} = 0$. Diese beiden Bedingungen charakterisieren wegen (1.1.22) und (1.1.23) gerade die stochastische Unabhängigkeit von X und Y.
c) Zum Beweis von (i) schließen wir mit Lemma 5.1.1

$$H(X,Y) - H(X) = H(Y \mid X) \ge 0,$$

wobei Gleichhheit analog zu b)(i) genau dann gilt, wenn $p_{j|i} \in \{0,1\}$ für alle i,j. (ii) folgt schließlich aus Lemma 5.1.1 und b(ii), indem man dort die Rollen von X und Y vertauscht, und zwar

$$H(X,Y) - H(X) = H(Y \mid X) \le H(Y).$$

Gleichheit gilt hierbei dann und nur dann, wenn X und Y stochastisch unabhängig sind. ∎

Bezeichne $\mathcal{P}_m$ die Menge der Wahrscheinlichkeitsvektoren der Länge m und $\mathcal{P}$ die Menge aller Wahrscheinlichkeitsvektoren, also

$$\mathcal{P}_m = \{(p_1,\dots,p_m) \mid p_i \geq 0, \sum_{i=1}^{m} p_i = 1\}, \quad m \in \mathbf{N}, \quad \mathcal{P} = \bigcup_{m=1}^{\infty} \mathcal{P}_m. \tag{5.1.11}$$

Die Entropie kann aufgefaßt werden als reellwertige Abbildung auf $\mathcal{P}$,

$$H : \mathcal{P} \to \mathbf{R} : (p_1,\dots,p_m) \mapsto -\sum_{i=1}^{m} p_i \cdot \log p_i. \tag{5.1.12}$$

Ungleichung a)(ii) aus Satz 5.1.1 mit der notwendigen und hinreichenden Bedingung für Gleichheit lautet dann:

$$H(p_1,\dots,p_m) \leq H(\frac{1}{m},\dots,\frac{1}{m}) \quad \text{für alle } m \in \mathbf{N},\ (p_1,\dots,p_m) \in \mathcal{P}_m \quad . \tag{E1}$$

Ist $(p_{11},\dots,p_{1n},p_{21},\dots,p_{2n},\dots,p_{m1},\dots,p_{mn}) \in \mathcal{P}_{mn}$ ein stochastischer Vektor, der die Verteilung eines Zufallsvektors (X,Y) repräsentiert, so besagt Lemma 5.1.1 mit den Abkürzungen (5.1.3) und (5.1.7):

$$\begin{aligned} H(p_{11},\dots,p_{1n},p_{21},\dots,p_{2n},\dots,p_{m1},\dots,p_{mn}) \\ = H(p_{1\cdot},\dots,p_{m\cdot}) + \sum_{i=1}^{m} p_{i\cdot} H(p_{1|i},\dots,p_{n|i}) \\ \text{für alle } m,n \in \mathbf{N},\ (p_{11},\dots,p_{mn}) \in \mathcal{P}_{mn} \quad . \end{aligned} \tag{E2}$$

Folgende Eigenschaft der Entropie ist mit Definition 5.1.1 leicht einzusehen:

$$\begin{aligned} H(p_1,\dots,p_l,0,p_{l+1},\dots,p_m) = H(p_1,\dots,p_l,p_{l+1},\dots,p_m) \\ \text{für alle } m \in \mathbf{N},\ 1 \leq l \leq m,\ (p_1,\dots,p_m) \in \mathcal{P}_m \quad . \end{aligned} \tag{E3}$$

(E3) besagt, daß die Unbestimmtheit eines Zufallsexperiments sich nicht verändert, wenn ein Ereignis mit in Betracht gezogen wird, das mit Wahrscheinlichkeit null (also nie) eintritt.

Betrachten wir nun umgekehrt eine reellwertige Funktion H auf $\mathcal{P}$, die die Eigenschaften (E1), (E2) und (E3) besitzt. Der folgende Satz zeigt, daß eine solche Funktion, wenn sie stetig ist, mit der Entropie (5.1.1) übereinstimmt. Die Entropie wird also im wesentlichen durch die Bedingungen (E1) – (E3) bis auf skalare Vielfache eindeutig bestimmt.

Satz 5.1.2. *(Eindeutigkeitssatz der Entropie, Chinchin (1953))*
Sei $H : \mathcal{P} \to \mathbf{R}$, H nicht identisch 0, eine Funktion, die den Bedingungen (E1), (E2) und (E3) genügt. Ist ferner $H|_{\mathcal{P}_m}$ (die Restriktion von H auf $\mathcal{P}_m$) stetig für alle $m \in \mathbf{N}$, so gilt

$$H(p_1,\dots,p_m) = -c\sum_{i=1}^{m} p_i \cdot \log p_i$$

für eine Konstante $c > 0$.

Beweis. Wir setzen $f(m) = H(1/m,\dots,1/m)$ und zeigen zunächst, daß $f(m) = c \cdot \log m$ für eine Konstante $c > 0$. Wegen (E3) und (E1) gilt

$$f(m) = H(1/m,\dots,1/m,0) \leq H\big(1/(m+1),\dots,1/(m+1)\big) = f(m+1),$$

also ist $f(m)$ monoton steigend in m.

Setzt man für beliebiges $r, s \in \mathbb{N}$ in (E2) $m = r$, $n = r^{s-1}$ und $p_{ij} = p_{i\cdot}p_{\cdot j} = 1/r^s$, wobei $p_{i\cdot} = 1/r$, $p_{\cdot j} = 1/r^{s-1}$, $i = 1, \ldots, r$, $j = 1, \ldots, r^{s-1}$, so gilt $p_{j|i} = p_{ij}/p_{i\cdot} = r/r^s = 1/r^{s-1}$, wobei $(p_{11}, \ldots, p_{1r^{s-1}}, \ldots, p_{r1}, \ldots, p_{rr^{s-1}}) \in \mathcal{P}_{r^s}$. (Ausgedrückt durch Zufallsvariable X, Y bedeutet dies mit der obigen Bezeichnungsweise: X und Y sind unabhängig und jeweils gleichverteilt auf den Wertebereichen $\mathcal{X} = \{1, \ldots, r\}$ bzw. $\mathcal{Y} = \{1, \ldots, r^{s-1}\}$, d.h. der Zufallsvektor (X, Y) ist auf $\mathcal{X} \times \mathcal{Y}$ gleichverteilt.)
Diese spezielle Verteilung in (E2) eingesetzt liefert

$$\begin{aligned} H(1/r^s, \ldots, 1/r^s) &= H(1/r, \ldots, 1/r) + \frac{1}{r}\sum_{i=1}^{r} H(1/r^{s-1}, \ldots, 1/r^{s-1}) \\ &= H(1/r, \ldots, 1/r) + H(1/r^{s-1}, \ldots, 1/r^{s-1}). \end{aligned}$$

Mit vollständiger Induktion folgt

$$f(r^s) = H(1/r^s, \ldots, 1/r^s) = sH(1/r, \ldots, 1/r) = sf(r). \tag{5.1.13}$$

Insbesondere erhält man hieraus

$$f(1) = f(1^s) = sf(1)$$

für alle $s \in \mathbb{N}$, also $f(1) = 0$. Nach Voraussetzung ist dann $f(2) > 0$, da anderenfalls $f(2^s) = sf(2) = 0$ für alle $s \in \mathbb{N}$ folgen würde, also die Monotonie von f zwingend $f \equiv 0$ ergäbe. Damit ist aber sogar $f(m) > 0$ für alle $m \geq 2$. Für alle $r, s, n \in \mathbb{N}$, $r \geq 2$, existiert nun ein $m \in \mathbb{N}_0$ mit

$$r^m \leq s^n < r^{m+1} \quad \Longleftrightarrow \quad \frac{m}{n} \leq \frac{\log s}{\log r} < \frac{m}{n} + \frac{1}{n}. \tag{5.1.14}$$

Aus der Monotonie von f folgt mit (5.1.13)

$$mf(r) \leq nf(s) \leq (m+1)f(r) \quad \Longleftrightarrow \quad \frac{m}{n} \leq \frac{f(s)}{f(r)} < \frac{m}{n} + \frac{1}{n}$$

Dies zusammen mit (5.1.14) gibt $|f(s)/f(r) - \log s/\log r| \leq 1/n$ für alle $n \in \mathbb{N}$, also $f(s)/\log s = f(r)/\log r$ für alle $r, s \in \mathbb{N}$, $r, s \geq 2$. Mit der Monotonie von f folgt, daß für alle $m \in \mathbb{N}$ eine Konstante $c > 0$ existiert derart, daß

$$f(m) = H(1/m, \ldots, 1/m) = c \cdot \log m \tag{5.1.15}$$

Seien jetzt $(p_1^*, \ldots, p_m^*) \in \mathcal{P}_m$, p_i^* rational, mit $p_i^* = g_i/g$, $g_i \in \mathbb{N}$, $i = 1, \ldots, m$, $\sum_{i=1}^m g_i = g$. Setze p_{ij} wie in folgender Matrix angegeben:

p_{ij}										$g_{i\cdot}/g = p_{i\cdot}$
$1/g$	$\ldots$	$1/g$	0	$\ldots$	0	$\ldots$	0	$\ldots$	0	$g_1/g = p_1^*$
0	$\ldots$	0	$1/g$	$\ldots$	$1/g$	$\ldots$	0	$\ldots$	0	$g_2/g = p_2^*$
$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
0	$\ldots$	0	0	$\ldots$	0	$\ldots$	$1/g$	$\ldots$	$1/g$	$g_m/g = p_m^*$
$\underbrace{\qquad}_{g_1}$			$\underbrace{\qquad}_{g_2}$				$\underbrace{\qquad}_{g_m}$			

Es gilt

$$p_{j|i} = \begin{cases} 1/g_i, & \text{falls } j \text{ Index im } i\text{-ten Block} \\ 0, & \text{sonst} \end{cases}, \quad i = 1,\dots,m,\ j = 1,\dots,n.$$

Mit (E2) folgt

$$H(\underbrace{1/g,\dots,1/g}_{g_1},0,\dots,0,\underbrace{1/g,\dots,1/g}_{g_2},0,\dots,0,\underbrace{1/g,\dots,1/g}_{g_m})$$
$$= H(p_1^*,\dots,p_m^*) + \sum_{i=1}^{m} p_i^* H(0,\dots,0,\underbrace{1/g_i,\dots,1/g_i}_{g_i},0,\dots,0).$$

Mit (E3) und (5.1.15) folgt hieraus $c \cdot \log g = H(p_1^*,\dots,p_m^*) + \sum_{i=1}^m p_i^*(c \cdot \log g_i)$ für eine positive Konstante c, so daß

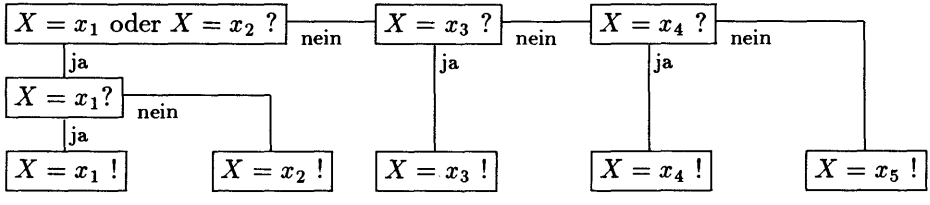
$$H(p_1^*,\dots,p_m^*) = c\Big(\log g - \sum_{i=1}^{m} p_i^* \cdot \log g_i\Big) = -c\sum_{i=1}^{m} p_i^* \cdot \log p_i^*.$$

Für beliebige $p_i \in \mathbf{R}$, $p_i \geq 0$, $\sum_{i=1}^m p_i = 1$, folgt die Behauptung mit der geforderten Stetigkeit von H. ∎

5.2. Optimale Codierung

Wir beginnen mit einem Beispiel, in dem anschaulich ein Codierungsverfahren beschrieben wird. Es sei vorausgesetzt, daß Kenntnisse über das Auftreten zu codierender Elemente in Form eines stochastischen Modells vorliegen. Wir betrachten ein Zufallsexperiment mit fünf Ausgängen $\mathcal{X} = \{x_1,\dots,x_5\}$, beschrieben durch eine endlich diskrete Zufallsvariable X mit Wertebereich $\mathcal{X}$ und Wahrscheinlichkeitsverteilung $p_j = P(X = x_j)$, $j = 1,\dots,5$, wobei $p_1 = 0.3$, $p_2 = p_3 = 0.2$, $p_4 = p_5 = 0.15$.

Jemand beobachtet den Ausgang des Experiments, und wir wollen versuchen, durch Fragen, die nur mit "ja" oder "nein" beantwortet werden, herauszubekommen, welches Ergebnis vorgelegen hat. Eine mögliche Fragestrategie kann etwa folgendermaßen skizziert werden:



Mögliche Antwortfolgen unter dieser Fragestrategie können unter der Kurzschreibweise "ja $\cong$ 1" und "nein $\cong$ 0" als Codierung von $\mathcal{X}$ aufgefaßt werden, wobei sich hier

$$(11) \leftrightarrow x_1,\ (10) \leftrightarrow x_2,\ (01) \leftrightarrow x_3,\ (001) \leftrightarrow x_4,\ (000) \leftrightarrow x_5$$

identifizieren lassen. Zur Beurteilung, ob unsere Strategie gut ist, wird die erwartete Anzahl von Fragen N zur Bestimmung des Experimentausgangs herangezogen. Offensichtlich ist hier

$$N(x_1) = N(x_2) = N(x_3) = 2, \quad N(x_4) = N(x_5) = 3$$

eine meßbare Abbildung $N: \mathcal{X} \to \mathbf{R}$, also eine Zufallsvariable. Ihr Erwartungswert beträgt $E(N) = 2 \cdot 0.3 + 4 \cdot 0.2 + 6 \cdot 0.15 = 2.3$. Wir werden später einsehen, daß dieser Wert mit keiner anderen ja/nein–Fragestrategie unterboten werden kann.

Die Entropie des Zufallsexperimentes beträgt $H(X) = -0.3 \cdot \log_2 0.3 - 0.4 \cdot \log_2 0.2 - 0.3 \cdot \log_2 0.15 \approx 2.271$. Es gilt $E(N) \geq H(X)$, und dieser Sachverhalt wird sich auch als allgemeingültig herausstellen. Stets ist die Entropie nicht größer als die erwartete Anzahl von Fragen zur Bestimmung des Ausgangs, egal welche Fragestrategie eingesetzt wird. Wir werden sogar zeigen, daß man mit geschickten "Blockfragestrategien" mit der erwarteten Anzahl von Fragen beliebig nahe an diese untere Grenze herankommen kann.

Daß man sich um eine im Mittel möglichst kurze Codierung (in Form von Fragestrategien, Suchbäumen, Signalfolgen zur digitalen Übertragung etc.) bemüht, ist unmittelbar klar. Eine ungünstige Codierung verschlingt unnötig Speicherplatz und Rechenzeit.

Wir stellen zunächst ein stochastisches Modell für folgenden allgemeinen Rahmen auf. Eine Quelle sendet Folgen von Buchstaben aus einem endlichen Alphabet $\mathcal{X} = \{x_1, \ldots, x_m\}$. Jeder Buchstabe tritt hierbei mit einer gewissen Wahrscheinlichkeit auf. In einer ersten Näherung wird angenommen, daß das Auftreten der einzelnen Buchstaben unabhängig voneinander geschieht. Später wird dieses für viele praktische Situationen sicher zu einfache Modell erweitert, indem auch Abhängigkeiten zwischen den einzelnen Komponenten zugelassen werden.

Definition 5.2.1. *$\{X_n\}_{n \in \mathbf{N}}$ sei eine Folge von stochastisch unabhängigen, identisch verteilten Zufallsvariablen mit Verteilung $P^{X_n} = P^X$, die den Wertebereich $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \in \mathbf{N}$, besitzt. $\{X_n\}_{n \in \mathbf{N}}$ (oder auch X) heißt diskrete, gedächtnislose Quelle. $\mathcal{X}$ heißt Quellalphabet.*

Ziel ist nun, Folgen von Buchstaben des Quellalphabets einer diskreten gedächtnislosen Quelle optimal zu codieren. Wir betrachten dabei eine Codierung über einem beliebigen Codealphabet $\mathcal{Y} = \{y_1, \ldots, y_d\}$, $d \in \mathbf{N}$. Zunächst wird jeder einzelne Buchstabe der Quelle mit einem Codewort, einer Folge von Buchstaben des Codealphabets, codiert. Mathematisch läßt sich dieser Vorgang in folgender Definition fassen. A^k, $k \in \mathbf{N}_0$, bedeutet hierbei das k–fache kartesische Produkt einer Menge A. Ist A ein endliches Alphabet, so läßt sich A^k gerade mit den Wörtern der Länge k über A identifizieren, A^0 enthält dann als einziges Element das leere Wort.

Definition 5.2.2. *Seien $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \in \mathbf{N}$, $\mathcal{Y} = \{y_1, \ldots, y_d\}$, $d \in \mathbf{N}$, und*

$$g : \mathcal{X} \to \bigcup_{l=1}^{\infty} \mathcal{Y}^l : x_j \mapsto (w_{j1}, \ldots, w_{jn_j}),$$

$w_{jk} \in \mathcal{Y}$, $j = 1, \ldots, m$, $k = 1, \ldots, n_j$, $n_j \in \mathbf{N}$, eine injektive Abbildung. $g(x_j) = (w_{j1}, \ldots, w_{jn_j})$ heißt Codewort des Buchstabens x_j. n_j heißt Länge des Codewortes. g heißt Code mit Codewortmenge $\mathcal{K} = g(\mathcal{X}) = \{g(x_1), \ldots, g(x_m)\}$.

Damit eine Codierung überhaupt brauchbar ist, müssen auch ohne Trennzeichen aneinandergereihte Codewörter (Konkatenation) die ursprünglichen Quellbuchstaben wieder identifizieren lassen, also eine eindeutige Decodierung ermöglichen.

Definition 5.2.3. *Unter den Bezeichnungen von Definition 5.2.2 heißt ein Code g eindeutig decodierbar (kurz: e.d.), wenn die Abbildung*

$$G : \bigcup_{\ell=1}^{\infty} \mathcal{X}^{\ell} \to \bigcup_{\ell=1}^{\infty} \mathcal{Y}^{\ell} : (u_1, \ldots, u_r) \mapsto \big(g(u_1), \ldots, g(u_r)\big),$$

$r \in \mathbb{N}$, injektiv ist.

Im allgemeinen ist es schwierig festzustellen, ob ein Code g eindeutig decodierbar ist. Nützlich sind Konstruktionsprinzipien, die zu eindeutig decodierbaren Codes führen. Dies leisten präfixfreie Codes, die in folgender Definition vorgestellt werden.

Definition 5.2.4. *$\boldsymbol{a} = (a_1, \ldots, a_r)$, $\boldsymbol{b} = (b_1, \ldots, b_s) \in \bigcup_{\ell=1}^{\infty} \mathcal{Y}^{\ell}$, $r \leq s$, seien Codewörter. $\boldsymbol{a}$ heißt Präfix von $\boldsymbol{b}$, wenn $\boldsymbol{c} = (c_1, \ldots, c_{s-r}) \in \bigcup_{\ell=0}^{\infty} \mathcal{Y}^{\ell}$ existiert mit $\boldsymbol{b} = (a_1, \ldots, a_r, c_1, \ldots, c_{s-r})$. Ein Code g mit Codewortmenge $\mathcal{K}$ heißt präfixfrei (kurz: PF–Code), wenn kein Codewort Präfix eines anderen ist.*

Lemma 5.2.1. *Präfixfreie Codes sind eindeutig decodierbar.*

Ein formaler Beweis von Lemma 5.2.1 ist recht aufwendig, die Beweisidee läßt sich jedoch durch den folgenden Algorithmus einfach skizzieren. Man suche die Folge der Codebuchstaben aus verketteten Wörtern von links beginnend ab, bis das erste Codewort (und damit der erste Quellbuchstabe) identifiziert ist. Dies ist eindeutig möglich, da der Code präfixfrei ist. Mit dem Rest der Folge fahre man genauso fort.

Das Morse–Alphabet, welches Codewortlängen zwischen eins und sechs sowie Trennzeichen (Pausen) benutzt, ist nicht präfixfrei, wenn man die Trennzeichen wegläßt. Identifiziert man etwa "kurz" mit dem Symbol 0 und "lang" mit dem Symbol 1, so liefern die Quellwörter (kode) und (tamms) beide das Codewort (1011111000), wenn keine Trennzeichen verwendet werden.

Eine typische Anwendung von präfixfreien Codes finden wir beim Fernkopieren (Telefax). Hierbei bilden Sequenzen aus schwarzen bzw. weißen Pixeln zeilenweise die Quellwörter $\{s_1, s_2, \ldots, s_{\max}, w_1, w_2, \ldots, w_{\max}\}$. Diese werden mit einem optimalen PF–Code binär codiert und dann übertragen.

Eine wichtige Frage ist daher, wann zu einem gegebenen Quellalphabet $\mathcal{X} = \{x_1, \ldots, x_m\}$, einem Codealphabet $\mathcal{Y} = \{y_1, \ldots, y_d\}$ und gegebenen Codewortlängen $n_1, \ldots, n_m$ ein präfixfreier Code mit eben diesen Codewortlängen existiert. Eine Antwort hierauf gibt der folgende Satz. Die Bezeichnungen stimmen mit denen aus Definition 5.2.2 überein.

Satz 5.2.1. *(Kraft (1949), McMillan (1959))*
Für jeden eindeutig decodierbaren Code mit Codewortlängen $n_1, \ldots, n_m$ gilt

$$\sum_{j=1}^{m} d^{-n_j} \leq 1. \tag{5.2.1}$$

Gilt umgekehrt (5.2.1) für Zahlen $n_1, \ldots, n_m \in \mathbb{N}$, so existiert ein präfixfreier (insbesondere also e.d.) Code mit Codewortlängen $n_1, \ldots, n_m$.

Beweis. Sei g ein e.d. Code mit Wortlängen $n_1, \ldots, n_m$. Für $\ell \in \mathbb{N}$ bezeichne $\beta_\ell = \#(\{j \mid n_j = \ell\})$ die Anzahl der Codewörter der Länge ℓ und $r = \max_{1 \le j \le m} n_j$ die maximale Codewortlänge. Für alle $k \in \mathbb{N}$ ist dann

$$\left(\sum_{j=1}^{m} d^{-n_j}\right)^k = \left(\sum_{\ell=1}^{r} \beta_\ell \cdot d^{-\ell}\right)^k = \sum_{\ell=k}^{k \cdot r} \gamma_\ell \cdot d^{-\ell} ,$$

wobei

$$\gamma_\ell = \sum_{\substack{i_1, \ldots, i_k \in \{1, \ldots, r\} \\ i_1 + \ldots + i_k = \ell}} \beta_{i_1} \cdots \beta_{i_k} , \quad \ell = k, \ldots, k \cdot r .$$

γ_ℓ ist gerade die Anzahl der Quellwörter aus $\bigcup_{j=1}^{\infty} \mathcal{X}^j$, deren Codewort die Länge ℓ hat und die aus k einzelnen Codewörtern zusammengesetzt sind, $\ell = k, \ldots, k \cdot r$. Wegen der eindeutigen Decodierbarkeit von g besitzt jedes Codewort höchstens ein Quellwort. d^ℓ ist die maximale Anzahl der Codewörter der Länge ℓ, so daß $\gamma_\ell \le d^\ell$ für alle $\ell = k, \ldots, k \cdot r$. Es folgt für alle $k \in \mathbb{N}$

$$\left(\sum_{j=1}^{m} d^{-n_j}\right)^k \le \sum_{\ell=k}^{k \cdot r} 1 = k \cdot r - k + 1 \le k \cdot r,$$

also

$$\sum_{j=1}^{m} d^{-n_j} \le (k \cdot r)^{1/k} \quad \longrightarrow 1 \quad (k \to \infty),$$

woraus (5.2.1) folgt.
Umgekehrt seien ohne Einschränkung der Allgemeinheit $n_1 \le \cdots \le n_m \in \mathbb{N}$ mit $\sum_{j=1}^{m} d^{-n_j} \le 1$. Wir konstruieren Wörter $\boldsymbol{g}_j \in \mathcal{Y}^{n_j}$, $j = 1, \ldots, m$, die einen PF–Code $\mathcal{K} = \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m\}$ bilden. Für $\boldsymbol{a} \in \mathcal{Y}^\ell$, $\ell \le n_m$, bezeichne

$$\mathcal{N}(\boldsymbol{a}) = \{\boldsymbol{b} \in \mathcal{Y}^{n_m} \mid \boldsymbol{a} \text{ ist Präfix von } \boldsymbol{b}\}.$$

Wähle nun beliebig $\boldsymbol{g}_1 \in \mathcal{Y}^{n_1}$. Es gilt $\#(\mathcal{N}(\boldsymbol{g}_1)) = d^{n_m - n_1}$. (5.2.1) impliziert $\sum_{j=1}^{m} d^{n_m - n_j} \le d^{n_m}$, also $d^{n_m - n_1} < d^{n_m}$, falls $m \ge 2$. Somit gilt $\mathcal{Y}^{n_m} \setminus \mathcal{N}(\boldsymbol{g}_1) \ne \emptyset$ und folglich existiert $\boldsymbol{g}_2 \in \mathcal{Y}^{n_2}$, das $\boldsymbol{g}_1$ nicht als Präfix enthält.
Dieses Verfahren läßt sich fortsetzen, falls $m \ge 3$. Es gilt $\#(\mathcal{N}(\boldsymbol{g}_2)) = d^{n_m - n_2}$ und wegen (5.2.1) $d^{n_m - n_1} + d^{n_m - n_2} < d^{n_m}$. Da $\mathcal{Y}^{n_m} \setminus (\mathcal{N}(\boldsymbol{g}_1) \cup \mathcal{N}(\boldsymbol{g}_2)) \ne \emptyset$, existiert ein Wort $\boldsymbol{g}_3 \in \mathcal{Y}^{n_3}$, das weder $\boldsymbol{g}_1$ noch $\boldsymbol{g}_2$ als Präfix enthält.
Die m–fache Anwendung dieses Verfahrens führt zu der Menge $\mathcal{K} = \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m\}$, die mit $g(x_j) = \boldsymbol{g}_j$, $j = 1, \ldots, m$, einen präfixfreien Code für $\mathcal{X}$ liefert. ∎

Der zweite Teil des Beweises läßt sich in ein konstruktives Verfahren zur Bestimmung eines präfixfreien Codes mit vorgegebenen Wortlängen $n_1 \le \cdots \le n_m$ umsetzen. Wir betrachten hierzu einen vollständigen Baum der Tiefe n_m mit Grad

d. Die d direkten Nachfolger jedes Knotens werden hierbei jeweils aufsteigend geordnet. Jedes Codewort aus $\bigcup_{j=1}^{n_m} \mathcal{Y}^j$ kann dann mit einem Knoten des Baums identifiziert werden, indem seine Komponenten eindeutig den Weg beschreiben, der von der Wurzel zu dem entsprechenden Knoten führt. Ist etwa y_j die k–te Komponente eines Codeworts und befindet man sich zur Zeit $k-1$ in einem bestimmten Knoten, so gehe man beim nächsten Schritt in den j–ten direkten Nachfolger dieses Knotens. Zur Zeit 0 startet man in der Wurzel des Baums.

Einen Code mit vorgeschriebenen Codewortlängen erhält man nun folgendermaßen. Wähle einen beliebigen Knoten auf der n_1–ten Stufe. Alle nachfolgenden Knoten fallen wegen der Präfix–Forderung als mögliche Codewörter aus. Die Bedingung $\sum_{j=1}^{m} d^{-n_j} \leq 1$ stellt jedoch sicher, daß mindestens ein Knoten auf der n_2–ten Stufe als Codewort verbleibt. Wähle diesen als zweites Codewort. Wieder verbleibt mindestens ein Knoten auf der n_3–ten Stufe. Dieses Verfahren läßt sich fortsetzen, bis m Codewörter mit den vorgeschriebenen Längen entstanden sind.

Die Güte einer Codierung $g : \mathcal{X} \to \bigcup_{l=1}^{\infty} \mathcal{Y}^l$ wird nun durch ihre erwartete Codewortlänge bewertet. Hierzu ordne unter den Bezeichnungen von Definition 5.2.2 die Abbildung

$$N_g : \mathcal{X} \to \mathbb{N} : x_j \mapsto n_j$$

jedem Quellbuchstaben x_j, $j = 1, \ldots, m$, die Länge seines Codewortes zu. Ist X nun wie in Definition 5.2.1 eine diskrete gedächtnislose Quelle, so bezeichnet

$$\bar{n} = \bar{n}(g) = E(N_g) = \sum_{j=1}^{m} n_j \cdot P(X = x_j)$$

die erwartete Codewortlänge. Der Erwartungswert $\bar{n} = \bar{n}(g)$ hängt über die Zahlen n_j von der verwendeten Codierung g ab; er ist eng mit der Entropie $H(X)$ der Quelle verbunden.

Satz 5.2.2. *(Noiseless coding theorem, Shannon (1948))*
$\{X_n\}_{n \in \mathbb{N}}$ sei eine diskrete, gedächtnislose Quelle mit Alphabet $\mathcal{X} = \{x_1, \ldots, x_m\}$ und Entropie $H(X)$. $\mathcal{Y} = \{y_1, \ldots, y_d\}$, $d \geq 2$, sei ein Codealphabet.

a) Für alle eindeutig decodierbaren Codes g mit Wortlängen $n_1, \ldots, n_m \in \mathbb{N}$ gilt

$$\frac{H(X)}{\log d} \leq \bar{n}(g). \tag{5.2.2}$$

b) Es existiert ein präfixfreier Code g mit Wortlängen $n_1, \ldots, n_m$ derart, daß

$$\bar{n}(g) < \frac{H(X)}{\log d} + 1 \tag{5.2.3}$$

Man beachte, daß in Satz 5.2.2 die Gedächtnislosigkeit der Quelle — im Modell die stochastische Unabhängigkeit der Zufallsvariablen X_n — keine Rolle spielt, von Bedeutung ist lediglich die identische Randverteilung $P^{X_n} = P^X$. Der Faktor $\log d$ verschwindet, wenn von vorneherein Logarithmen zur Basis d gewählt werden.

Beweis. a) Mit der Abkürzung $p_j = P(X = x_j)$, $j = 1, \ldots, m$ gilt

$$\begin{aligned} H(X) - \bar{n} \log d &= \sum_{j=1}^{m} p_j \cdot \log \frac{1}{p_j} - \sum_{j=1}^{m} p_j \cdot n_j \cdot \log d = \sum_{\substack{j=1 \\ p_j \neq 0}}^{m} p_j \log \frac{d^{-n_j}}{p_j} \\ &\leq \log e \Big(\sum_{\substack{j=1 \\ p_j \neq 0}}^{m} p_j \big(\frac{d^{-n_j}}{p_j} - 1 \big) \Big) = \log e \Big(\sum_{\substack{j=1 \\ p_j \neq 0}}^{m} d^{-n_j} - 1 \Big) \end{aligned} \tag{5.2.4}$$

In (5.2.4) wurden die Konventionen (5.1.2) und die Ungleichung (5.1.10) eingesetzt. Satz 5.2.1 lehrt, daß die rechte Seite für alle e.d. Codes nicht positiv ist, woraus a) folgt.

b) Wir weisen die Existenz eines PF–Codes mit Wortlängen $n_1, \ldots, n_m$ nach. Wähle n_j hierzu so, daß

$$d^{-n_j} \leq p_j < d^{-n_j+1}, \; 1 \leq j \leq m. \tag{5.2.5}$$

Dann gilt $\sum_{j=1}^{m} d^{-n_j} \leq \sum_{j=1}^{m} p_j = 1$ und nach Satz 5.2.1 existiert ein PF–Code mit Wortlängen $n_1, \ldots, n_m$. Logarithmiert man die rechte Seite von (5.2.5) und formt äquivalent um, erhält man $n_j < -\frac{\log p_j}{\log d} + 1$. Es folgt $\sum_{j=1}^{m} p_j \cdot n_j < H(X)/\log d + 1$. Das ist die Behauptung. ∎

Wählt man also $n_j = \lceil -\frac{\log p_j}{\log d} \rceil$, $j = 1, \ldots, m$, so existiert ein PF–Code g mit Wortlängen n_j, dessen erwartete Codewortlänge um höchstens 1 schlechter ist als die bestmögliche. Nach dem Beweis von Satz 5.2.1 haben wir gesehen, wie man einen solchen Code mit Hilfe von vollständigen Bäumen konstruiert.

Satz 5.2.2 a) präzisiert die Aussage des einführenden Beispiels, wenn ja/nein–Fragestrategien als binäre Codierung interpretiert und Logarithmen zur Basis zwei gewählt werden. (5.2.2) besagt dann $\bar{n} = E(N) \geq H(X)$ für alle Fragestrategien.

Codiert man nicht jeden einzelnen Buchstaben des Quellalphabets sondern "Blöcke" aus Quellbuchstaben, so kann man mit der mittleren erwarteten Codewortlänge beliebig nahe an die untere Schranke $H(X)/\log d$ herankommen. Wir werden jetzt wesentlich die stochastische Unabhängigkeit der Zufallsvariablen X_n ausnutzen.

Für eine gedächtnislose Quelle $\{X_n\}_{n \in \mathbb{N}}$ bezeichne $\boldsymbol{X}^{(L)} = (X_1, \ldots, X_L)$ den Zufallsvektor aus den ersten L Komponenten. $\boldsymbol{X}^{(L)}$ ist wieder eine endlich diskrete Zufallsvariable mit Wertebereich $\mathcal{X}^L$, für die mit Satz 5.1.1 c)

$$H(\boldsymbol{X}^{(L)}) = H(X_1) + \cdots + H(X_L) = L \cdot H(X) \tag{5.2.6}$$

folgt.

Codiert werden jetzt Wörter der Länge L des Alphabets $\mathcal{X} = \{x_1, \ldots, x_m\}$. Diese werden aufgefaßt als Buchstaben des Alphabets $\tilde{\mathcal{X}} = \mathcal{X}^L$ und über $\mathcal{Y} = \{y_1, \ldots, y_d\}$, $d \geq 2$, mit einer injektiven Abbildung $g^{(L)}: \tilde{\mathcal{X}} \to \bigcup_{l=1}^{\infty} \mathcal{Y}^l$ codiert. Die Abbildung $g^{(L)}$ heißt Blockcodierung. $n(u_1, \ldots, u_L)$ bezeichne die Codewortlänge von $(u_1, \ldots, u_L) \in \tilde{\mathcal{X}}$ bei Verwendung von $g^{(L)}$. Die erwartete Codewortlänge ist dann

$$\bar{n}^{(L)} = \bar{n}(g^{(L)}) = \sum_{(u_1,\ldots,u_L)\in\mathcal{X}^L} P(X_1 = u_1, \ldots, X_L = u_L) \cdot n(u_1, \ldots, u_L). \quad (5.2.7)$$

$\bar{n}^{(L)}/L$ ist ein Maß für die mittlere Codewortlänge pro Quellbuchstabe. Durch Anwendung von Satz 5.2.2 ergibt sich für die spezielle Situation der Blockcodierung bei diskreten gedächtnislosen Quellen das folgende

Korollar 5.2.1. *(Blockcodierung)*
$\{X_n\}_{n\in\mathbb{N}}$ sei eine diskrete, gedächtnislose Quelle, $L \in \mathbb{N}$ und $\mathcal{Y} = \{y_1, \ldots, y_d\}$ ein Codealphabet. Bei Blockcodierung von Wörtern der Länge L gilt unter den Bezeichnungen (5.2.6) und (5.2.7):

a) Für alle eindeutig decodierbaren Block–Codes $g^{(L)}$ ist

$$\frac{H(X)}{\log d} \leq \frac{\bar{n}(g^{(L)})}{L}. \quad (5.2.8)$$

b) Es existiert ein präfixfreier Block–Code $g^{(L)}$ mit

$$\frac{\bar{n}(g^{(L)})}{L} < \frac{H(X)}{\log d} + \frac{1}{L}. \quad (5.2.9)$$

Mit wachsender Blocklänge L konvergiert die obere Schranke in (5.2.9) gegen die untere in (5.2.8), so daß für genügend großes L die mittlere erwartete Codewortlänge beliebig nahe an den optimalen Wert $H(X)/\log d$ herangebracht werden kann.

Nach Satz 5.2.2 a) kann die erwartete Codewortlänge für keinen eindeutig decodierbaren Code kürzer als die mit $\log d$ normierte Entropie der Quelle sein. Eindeutig decodierbare Codes, die die untere Schranke in (5.2.2) erreichen, wollen wir *absolut optimal* nennen.

Gilt für die Quelle X, daß $p_j = P(X = x_j) > 0$ für alle $j = 1, \ldots, m$, so existiert ein absolut optimaler, eindeutig decodierbarer Code genau dann, wenn Zahlen $n_1, \ldots, n_m \in \mathbb{N}$ existieren mit $p_j = d^{-n_j}$, $j = 1, \ldots, m$. Dies folgt aus der Tatsache, daß Gleichheit in (5.2.4) genau dann gilt, wenn $d^{-n_j}/p_j = 1$ für alle $j = 1, \ldots, m$.

Ist p_j irrational für ein $j \in \{1, \ldots, m\}$ für eine Quelle X, kann damit kein absolut optimaler Code existieren.

In der Regel existieren also keine absolut optimalen Codes, und es stellt sich die Frage nach der Existenz und Konstruktion von optimalen Codes, das sind solche Codes, die kleinste erwartete Codewortlänge unter allen eindeutig decodierbaren besitzen.

Definititon 5.2.5. *X sei eine diskrete, gedächtnislose Quelle und $p_j = P(X = x_j)$, $j = 1, \ldots, m$. Ein eindeutig decodierbarer Code g^* mit Codewortlängen $n_1^*, \ldots, n_m^*$ heißt optimal, wenn*

$$\bar{n}(g^*) = \sum_{j=1}^{m} p_j n_j^* \leq \sum_{j=1}^{m} p_j n_j = \bar{n}(g)$$

für alle eindeutig decodierbaren Codes g mit Wortlängen $n_1, \ldots, n_m$.

Die Frage nach der Existenz von optimalen, eindeutig decodierbaren Codes wird durch das folgende Lemma beantwortet.

Lemma 5.2.2. *Stets existiert ein optimaler präfixfreier Code.*

Beweis. Wegen Satz 5.2.1 muß nachgewiesen werden, daß

$$\inf \Big\{ \sum_{j=1}^{m} p_j n_j \mid (n_1, \ldots, n_m) \in \mathcal{Z} \Big\} \tag{5.2.10}$$

für ein Element des Bereichs $\mathcal{Z} = \{\boldsymbol{n} = (n_1, \ldots, n_m) \in \mathbb{N}^m \mid \sum_{j=1}^{m} d^{-n_j} \leq 1\}$ zulässiger Punkte angenommen wird. Bezeichne hierzu

$$\mathcal{M} = \{\boldsymbol{n}' = (n_1', \ldots, n_m') \in \mathcal{Z} \mid \text{es existiert kein } \boldsymbol{n} = (n_1, \ldots, n_m) \in \mathcal{Z} \text{ mit } \boldsymbol{n} \neq \boldsymbol{n}' \text{ und } n_j \leq n_j' \text{ für alle } j = 1, \ldots, m\}.$$

die Menge der minimalen Elemente in $\mathcal{Z}$ bzgl. der komponentenweise Halbordnung in $\mathbb{N}^m$. Es gilt $\mathcal{M} \neq \emptyset$. Denn angenommen, $\mathcal{Z}$ besitzt keine minimalen Elemente. Dann gibt es für alle $\boldsymbol{n} = (n_1, \ldots, n_m) \in \mathcal{Z}$ ein $\boldsymbol{n}' = (n_1', \ldots, n_m') \in \mathcal{Z}$, $\boldsymbol{n} \neq \boldsymbol{n}'$, mit $n_j' \leq n_j$ für alle $j = 1, \ldots, m$, im Widerspruch dazu, daß für jedes $\boldsymbol{n} \in \mathcal{Z}$ die Menge $\{\boldsymbol{n}' \mid n_j' \leq n_j$ für alle $j = 1, \ldots, m$ und $n_{j_0}' < n_{j_0}$ für ein $j_0\}$ endliche Mächtigkeit besitzt.
Da die Zielfunktion $\sum_{j=1}^{m} p_j n_j$ monoton steigend in der komponentenweisen Halbordnung ist, ändert sich das Infimum in (5.2.10) nicht, wenn dort $\mathcal{Z}$ durch $\mathcal{M} \subset \mathcal{Z}$ ersetzt wird.
Angenommen in $\mathcal{M}$ gibt es eine unendliche Folge von verschiedenen Elementen $\{\boldsymbol{n}_k\}_{k \in \mathbb{N}}$, $\boldsymbol{n}_k = (n_{k,1}, \ldots, n_{k,m})$. Dann gilt wegen der Minimalität der $\boldsymbol{n}_k$, daß für alle k eine Komponente $\ell(k)$ existiert mit $n_{k+1,\ell(k)} < n_{k,\ell(k)}$. Somit existiert eine Komponente ℓ_0 mit $n_{k+1,\ell_0} < n_{k,\ell_0}$ für unendlich viele $k \in \mathbb{N}$. Dies führt zum Widerspruch, da $n_{k,j} \geq 1$ für alle $j = 1, \ldots, m$, $k \in \mathbb{N}$.
Damit ist $\mathcal{M}$ endlich und das Infimum in (5.2.10) wird in einem Punkt $\boldsymbol{n}^* \in \mathcal{M} \subset \mathcal{Z}$ angenommen. Wegen Satz 5.2.1 existiert ein zugehöriger präfixfreier Code. ∎

$\bar{n}(g^*)$ aus Definition 5.2.5 heißt auch *reale Entropie* der Quelle X. Korollar 5.2.1 gibt einen Zusammenhang zwischen realer Entropie und der Entropie $H(X)$. Da bei Blockcodierung für jeden optimalen Block–Code $g^{(L)*}$

$$H(X)/\log d \leq \bar{n}(g^{(L)*})/L < H(X)/\log d + 1/L$$

gilt, konvergiert die Folge der normierten realen Entropien $\{\bar{n}(g^{(L)*})/L\}_{L\in\mathbb{N}}$ gegen die Entropie $H(X)$.

Der Beweis der Existenzaussage von Lemma 5.2.2 gibt keinen Hinweis zur Konstruktion optimaler Codes. Dies gelingt jedoch mit dem im folgenden beschriebenen Huffman–Verfahren, dessen Kernstück das folgende Lemma bildet. Wir werden hier nicht auf den Beweis eingehen, er findet sich etwa in Ash (1965), Gallager (1968), Oberschelp & Wille (1976) und Topsøe (1974). Ferner werden wir nur binäre Codierung über dem Alphabet $\mathcal{Y} = \{0,1\}$ betrachten.

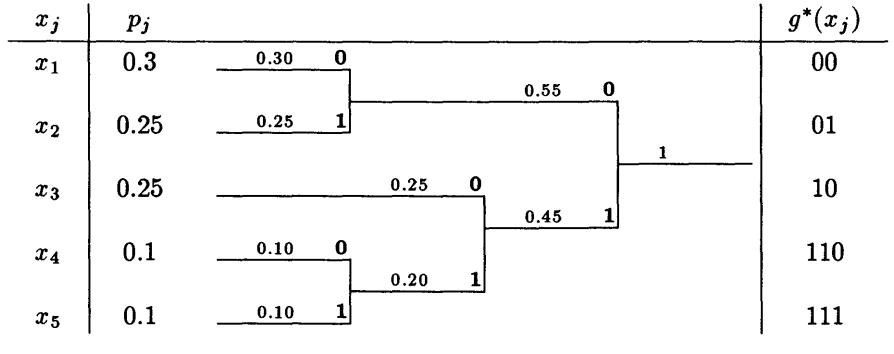
Lemma 5.2.3. *X sei eine diskrete Quelle mit Alphabet $\mathcal{X} = \{x_1,\dots,x_m\}$ und Wahrscheinlichkeiten $p_j = P(X = x_j)$, $j = 1,\dots,m$, wobei $p_1 \geq \dots \geq p_m$, $p_m > 0$. X' sei eine Quelle mit Alphabet $\mathcal{X}' = \{x'_1,\dots,x'_{m-1}\}$ und Wahrscheinlichkeiten $p'_j = p_j$, $j = 1,\dots,m-2$, und $p'_{m-1} = p_{m-1} + p_m$. Ist g' mit Codewortmenge $\mathcal{K}' = \{g'(x'_1),\dots,g'(x'_{m-1})\} \subset \bigcup_{l=0}^{\infty}\{0,1\}^l$ ein optimaler, präfixfreier Code für X', so bildet g, definiert durch*

$$g(x_j) = g'(x'_j), j = 1,\dots,m-2,$$
$$g(x_{m-1}) = \big(g'(x'_{m-1}),0\big), \quad g(x_m) = \big(g'(x'_{m-1}),1\big),$$

einen optimalen, präfixfreien Code für X.

Beim Huffman–Verfahren wird Lemma 5.2.3 rekursiv angewendet. Es wird eine Folge von Quellen konstruiert, indem jede aus der Vorgängerquelle durch Addition der beiden kleinsten Wahrscheinlichkeiten wie in Lemma 5.2.3 verkürzt wird. Die Folge bricht bei einer Quelle mit Alphabet der Mächtigkeit zwei ab, welches leicht optimal binär mit 0 bzw. 1 codiert werden kann. Rückwärtsgehend werden dann mit der Konstruktion aus Lemma 5.2.3 optimale, präfixfreie Codes bestimmt, bis man einen optimalen, präfixfreien Code für die vorgegebene Quelle X erhält.

Das folgende Beispiel mag die Vorgehensweise verdeutlichen. Die sukzessive verwendeten Codebuchstaben sind in der Graphik fettgedruckt eingetragen, links daneben befinden sich die Summen der jeweils kleinsten Wahrscheinlichkeiten. Die Ausgangswahrscheinlichkeiten $p_1 \geq \dots \geq p_5$ sind absteigend geordnet; dies kann stets durch Umnumerierung der Quellbuchstaben erreicht werden. In der Spalte rechts finden sich die Codewörter des optimalen Codes. Die Entropie der Quelle beträgt (bei Verwendung von $\log_2$) $H(X) = 2.1855$, die minimale erwartete Codewortlänge $\bar{n}(g^*) = 2.2$.

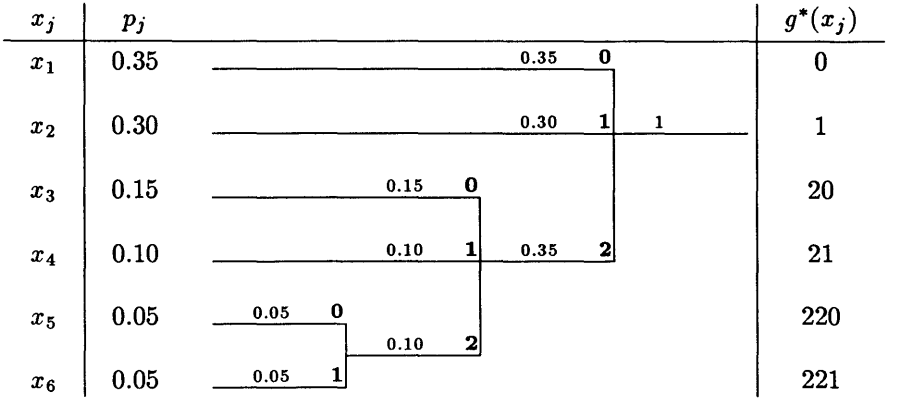


Wir betrachten noch einmal das eingangs von Kapitel 5.2 vorgestellte Beispiel eines Experiments mit fünf Ausgängen $\mathcal{X} = \{x_1, \ldots, x_5\}$ und Wahrscheinlichkeitsverteilung (0.3, 0.2, 0.2, 0.15, 0.15). Die ja/nein-Antworten auf die gewählte Fragestrategie wurden bereits dort als binäre Codierung interpretiert. Wendet man nun das Huffman-Verfahren zur Codierung dieses Experiments an, so erhält man als einen Code g^* kürzester, erwarteter Codewortlänge unter allen e.d. Codes

$$x_1 \mapsto (10),\ x_2 \mapsto (00),\ x_3 \mapsto (01),\ x_4 \mapsto (110),\ x_5 \mapsto (111)$$

mit $\bar{n}(g^*) = 2.3$. Die eingangs betrachtete Fragestrategie führte zu einer Codierung mit der gleichen minimalen, erwarteten Codewortlänge. Dies beweist jetzt die behauptete Optimalität dieser Fragestrategie.

Ein analoges Verfahren liefert auch bei beliebigem $d \geq 2$ mit Codealphabet $\mathcal{Y} = \{y_1, \ldots, y_d\}$ einen optimalen, präfixfreien Codes (s. Gallager (1968)). Besondere Aufmerksamkeit muß man lediglich dem ersten Schritt der Addition der kleinsten der Wahrscheinlichkeiten $p_1 \geq \cdots \geq p_m > 0$ schenken, und zwar werden lediglich die k kleinsten $p_m + \cdots + p_{m-k+1}$ addiert, wobei $k = 2 + (m-2) \pmod{(d-1)}$. In allen nachfolgenden Schritten werden dann die jeweils d kleinsten Wahrscheinlichkeiten addiert. Zur Verdeutlichung dient folgendes Beispiel mit $m = 6$, $d = 3$, $\mathcal{Y} = \{0, 1, 2\}$, $k = 2 + 4 \pmod 2 = 2$. Es gilt $H(X)/\log 3 = 1.2587$, $\bar{n}(g^*) = 1.35$.



Ein interessantes Codierungsverfahren, das unter Umständen auch dann noch angewendet werden kann, wenn die Menge der zugelassenen Codes Restriktionen unterliegt, wurde von Fano vorgeschlagen. Allerdings liefert es nicht notwendig optimale Codes. In der Regel jedoch liegt die erwartete Codewortlänge in der Nähe des Minimalwertes. Der Einsatz des Verfahrens kann dann sinnvoll sein, wenn nicht jeder beliebige Code von einem Codierer realisiert werden kann. Wir werden in Übungsaufgabe 5.7 ein Beispiel hierfür kennenlernen.

Ausgangspunkt unserer Betrachtungen ist eine diskrete Quelle X mit Quellalphabet $\mathcal{X} = \{x_1, \ldots, x_m\}$, das über dem Codealphabet $\mathcal{Y} = \{y_1, \ldots, y_d\}$, $d \geq 2$, mit einem präfixfreien Code $g : \mathcal{X} \to \bigcup_{\ell=1}^{n} \mathcal{Y}^\ell$ maximaler Codewortlänge $n \in \mathbb{N}$ codiert wird. Wir ergänzen die Codewörter $y_i = g(x_i)$, deren Länge n_i kleiner als n ist, durch Konkatenation eines beliebigen Wortes der Länge $n - n_i$ über $\mathcal{Y}$ zu Wörtern $\tilde{y}_i$der Länge n, $i = 1, \ldots, m$. Die eindeutige Decodierbarkeit des

hieraus entstehenden Codes mit Wörtern gleicher Länge bleibt wegen der Präfix–Eigenschaft von g erhalten, da $\tilde{g} : \mathcal{X} \to \mathcal{Y}^n : x_j \mapsto \tilde{y}_j$ ebenfalls ein PF–Code ist. Die Zufallsvariable $\boldsymbol{Y} = (Y_1, \ldots, Y_n) = \tilde{g}(X)$ besitzt wegen der Injektivität von $\tilde{g}$ die gleiche Entropie wie X, die sich mit (5.1.9) iteriert berechnen läßt zu

$$H(Y_1, \ldots, Y_n) = H(Y_1) + H(Y_2 \mid Y_1) + \cdots + H\big(Y_n \mid (Y_1, \ldots, Y_{n-1})\big). \quad (5.2.12)$$

Die Idee der Fano–Codierung basiert darauf, die gesamte Unbestimmtheit $H(Y_1, \ldots, Y_n) = H(X)$ durch möglichst wenige Komponenten $Y_1, \ldots, Y_n$ auszuschöpfen. Der folgende Greedy–Algorithmus steuert dieses Ziel mit einer kurzsichtigen Strategie an, er löst das Problem jedoch nicht immer optimal.

Wähle $n \in \mathbb{N}$ und die Verteilung von $(Y_1, \ldots, Y_n)$ so, daß von links nach rechts in jedem Schritt die einzelnen Summanden auf der rechten Seite von (5.2.12) maximiert werden, der gesamte Wert H(X) also durch möglichst wenige Summanden ausgeschöpft wird. Dies muß natürlich so geschehen, daß ein zugehöriger präfixfreier Code $\tilde{g}$ angegeben werden kann, für den $\big(\tilde{g}(X)\big)_j = Y_j$, $j = 1, \ldots, n$, gilt. $\big(\tilde{g}(X)\big)_j$ bezeichnet hierbei die j–te Komponente von $\tilde{g}(X)$.

Wir wollen diese Konstruktion im Fall einer binären Codierung $\mathcal{Y} = \{0, 1\}$ durchführen. Zunächst gilt mit der Abkürzung $p_i = P(X = x_i)$, $i = 1, \ldots, m$,

$$P(Y_1 = 0) = P\big((\tilde{g}(X))_1 = 0\big) = \sum_{i:(\tilde{g}(x_i))_1 = 0} p_i = q_0$$

$$\text{und} \quad P(Y_1 = 1) = q_1 = 1 - q_0,$$

also für die Entropie

$$H(Y_1) = -q_0 \log q_0 - (1 - q_0) \log(1 - q_0). \quad (5.2.13)$$

Dieser Wert ist zu maximieren über alle Indexteilmengen $T \subseteq \{1, \ldots, m\}$, wobei $q_0 = \sum_{i \in T} p_i$. (5.2.13) ist maximal für $q_0 = 1/2$, symmetrisch um den Punkt $1/2$ und konkav, so daß obiges Maximierungsproblem mit einer Menge T_0 gelöst wird, für die

$$\Big| \sum_{i \in T_0} p_i - \frac{1}{2} \Big| = \min \quad \text{über alle } T \subseteq \{1, \ldots, n\}. \quad (5.2.14)$$

T_0 bestimmt gleichzeitig den Code $\tilde{g}$ bzw. g. Wir setzen nämlich $\big(\tilde{g}(x_i)\big)_1 = 0$ für alle $i \in T_0$ und mit $T_1 = \{1, \ldots, m\} \setminus T_0$ entsprechend $\big(\tilde{g}(x_i)\big)_1 = 1$ für alle $i \in T_1$.

Analog verfährt man nun in den beiden Mengen T_0 und T_1 mit den bedingten Wahrscheinlichkeiten. Wir verfolgen dies für T_0. Es gilt, falls $P(Y_1 = 0) > 0$,

$$P(Y_2 = 0 \mid Y_1 = 0) = \frac{P(Y_2 = 0, Y_1 = 0)}{P(Y_1 = 0)} = \sum_{\substack{i:(\tilde{g}(x_i))_1 = 0, \\ (\tilde{g}(x_i))_2 = 0}} \frac{p_i}{q_0} = q_{0|0}$$

$$\text{und} \quad P(Y_2 = 1 \mid Y_1 = 0) = 1 - q_{0|0} = q_{1|0}.$$

Entsprechend seien $q_{0|1}$ und $q_{1|1}$ definiert. Für die bedingte Entropie $H(Y_2 \mid Y_1)$ gilt

$$H(Y_2 \mid Y_1) = P(Y_1 = 0) H(Y_2 \mid Y_1 = 0) + P(Y_1 = 1) H(Y_2 \mid Y_1 = 1).$$

Sie wird maximal, wenn jede der beiden bedingten Entropien $H(Y_2 \mid Y_1 = 0)$ und $H(Y_2 \mid Y_1 = 1)$ maximal ist. Wir lösen dies für die erste bedingte Entropie.

$$\text{maximiere } \{H(Y_2 \mid Y_1 = 0) = -q_{0|0} \log q_{0|0} - (1 - q_{0|0}) \log(1 - q_{0|0})\}$$

über alle Indexteilmengen $T \subseteq T_0$, wobei $q_{0|0} = \sum_{i \in T} p_i/q_0$. Eine Lösung ergibt sich wie in (5.2.14): Teile die bedingten Wahrscheinlichkeiten p_i/q_0, $i \in T_0$, so auf, daß $\sum_{i \in T_{00}} p_i/q_0$ möglichst nahe bei $1/2$ liegt. $T_{00} \subseteq T_0$ bezeichnet hierbei die optimale Indexteilmenge und T_{01} entsprechend $T_0 \setminus T_{00}$. Für den zugehörigen Code $\tilde{g}$ setzen wir $\big(\tilde{g}(x_i)\big)_2 = 0$ für alle $i \in T_{00}$ bzw. $\big(\tilde{g}(x_i)\big)_2 = 1$ für alle $i \in T_{01}$. Analog verfahren wir mit T_1 und spalten in T_{10} bzw. T_{11} auf.

Mit jeder dieser Teilmengen und allen jeweils neu entstehenden Teilmengen verfährt man entsprechend.

Einelementige Teilmengen werden nicht mehr weiter aufgeteilt — der Algorithmus stoppt an dieser Stelle. Es ist klar, daß das Verfahren nach höchstens $n = m$ Schritten abbricht. Spätestens dann gilt $P\big(Y_{n+1} = y_{n+1} \mid (Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)\big) = 1$, falls $P\big((Y_1, \ldots, Y_{n+1}) = (y_1, \ldots, y_{n+1})\big) > 0$. Wegen der notwendigen und hinreichenden Bedingung für Gleichheit in Satz 3.1.1. b)(i) gilt für die bedingte Entropie $H\big(Y_{n+1} \mid (Y_1, \ldots, Y_n)\big) = 0$, die dann keinen Beitrag mehr zur Summe (5.2.12) leistet.

Diese Situation kann auch schon früher eintreten. Gilt $P\big((Y_1, \ldots, Y_{k+1}) = (y_1, \ldots, y_{k+1})\big) > 0$ und $P\big(Y_{k+1} = y_{k+1} \mid (Y_1, \ldots, Y_k) = (y_1, \ldots, y_k)\big) = 1$, für $(y_1, \ldots y_{k+1}) \in \{0,1\}^k$, so hat man bereits ein eindeutiges Codewort $(y_1, \ldots, y_k)$ zugeordnet. Dieser Fall tritt genau dann auf, wenn die entsprechende Indexmenge $T_{y_1,\ldots,y_k}$ einelementig ist. Die zugehörige Codewortlänge beträgt dann k, $1 \leq k \leq m$.

Wir führen das Verfahren an folgendem Beispiel vor.

$p_i = P(X = x_i)$	$g(x_i)$				
0.3	0	0			
0.2	0	1			
0.1	1	0	0		
0.1	1	0	1	0	
0.05	1	0	1	1	
0.07	1	1	0	0	
0.05	1	1	0	1	
0.08	1	1	1	0	
0.03	1	1	1	1	0
0.02	1	1	1	1	1

Die waagerechten Linien zeigen an, wo die Aufteilung in zwei Gruppen mit möglichst gleichen, bedingten Wahrscheinlichkeiten vorgenommen wurde. Für die zugehörige erwartete Codewortlänge gilt

$$\bar{n}(g) = 2 \cdot 0.5 + 3 \cdot 0.1 + 4 \cdot 0.35 + 5 \cdot 0.05 = 2.95.$$

Die Entropie von X beträgt $H(X) = 2.8728$ bei Verwendung von Logarithmen zur Basis 2. Der Fano–Code liefert in diesem Beispiel die kürzeste erwartete Codewortlänge. Folgender Code g^* ist ein Huffman–Code mit der gleichen erwarteten Codewortlänge $\bar{n}(g^*) = 2.95$. Die Codewörter sind nach fallenden Wahrscheinlichkeiten der Quellbuchstaben sortiert.

$$g^*(\mathcal{X}) = \{00,\ 10,\ 110,\ 0100,\ 0101,\ 0110,\ 0111,\ 1110,\ 11110,\ 11111\}$$

5.3. Binäre Suchbäume

Die gleichen Prinzipien, die bei der Konstruktion optimaler Codes verwendet wurden, können auch bei der Bestimmung effizienter binärer Suchbäume eingesetzt werden. Hierbei sollen Elemente einer endlichen Menge S so als Knoten eines binären Baumes angeordnet werden, daß das Wiederfinden durch einen festen Suchalgorithmus in möglichst kurzer Zeit geschehen kann.

Die im vorigen Abschnitt bestimmten binären Codes lassen sich folgendermaßen mit den Knoten eines binären Baumes identifizieren. Wenn man in der Wurzel eines vollständigen binären Baumes genügend großer Tiefe startet und jede "1" mit der Anweisung "Besuche den rechten Nachfolger", jede "0" mit "Besuche den linken Nachfolger" identifiziert, so findet sich das j-te Codewort $(w_{j1}, \ldots, w_{jn_j})$, $w_{ji} \in \{0,1\}$, gerade bei dem Knoten, in dem man durch sukzessives Abarbeiten des Codewortes $(w_{j1}, \ldots, w_{jn_j})$ nach obiger Vorschrift endet. Präfixfreie Codes haben sich als besonders wichtig erwiesen. Ihre Besonderheit in diesem Zusammenhang ist, daß nach Identifizierung aller Codewörter mit den Knoten eines binären Baumes kein Knoten Nachfolger eines anderen sein kann. Wäre nämlich Knoten l Nachfolger von Knoten k, so müßte das zu k gehörige Codewort Präfix des zu l gehörigen sein. Dieses "Nachfolgerverbot" wird bei der jetzt folgenden Konstruktion von Suchbäumen aufgegeben und verhindert eine direkte Anwendung der im vorigen Abschnitt erhaltenen Ergebnisse, verwandte Methoden jedoch werden auch hier zum Ziel führen.

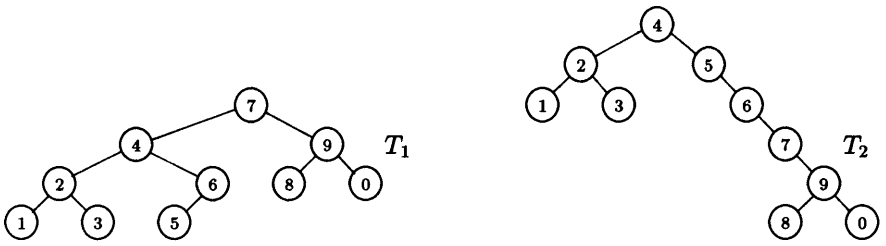
Wir werden zunächst den Begriff des binären Suchbaums definieren, wobei vorausgesetzt wird, daß der Leser binäre Bäume und eine Implementation durch eine Programmiersprache kennt, etwa in rekursiver PASCAL–Notation, wenn man die Knoten mit einer Teilmenge der natürlichen Zahlen identifiziert:

```
TYPE node = RECORD no:    INTEGER;
                left,right: ↑node
            END.
```
(5.3.1)

Definition 5.3.1. *$S = \{x_1, \ldots, x_m\}$, $m \in \mathbb{N}$, $m \geq 2$, sei eine bezüglich "$<$" vollständig geordnete, endliche Menge mit $x_1 < x_2 < \ldots < x_m$. Ein binärer Baum T mit Knoten $\{x_1, \ldots, x_m\}$ heißt binärer Suchbaum über S, wenn für alle Knoten $x_j \in S$, für alle Knoten x_i im linken bzw. alle Knoten x_k im rechten Teilbaum mit Wurzel x_j gilt, daß $x_i < x_j < x_k$.*

Die in (5.3.2) dargestellten Bäume T_1 und T_2 sind binäre Suchbäume über der Menge $S = \{1,2,3,\ldots,9,0\} \subset \mathbb{N}_0$ mit der Ordnung $1 < 2 < \ldots < 0$.

Ein geeigneter Algorithmus, um ein Element $x \in S$ in einem binären Suchbaum T über S wiederzufinden, ist der folgende:



Binäre Suchbäume für $\{1, 2, \ldots, 9, 0\}$ (5.3.2)

Beginne mit der Wurzel des Baumes x^*. Ist $x < x^*$, setze x^* = Wurzel des linken Teilbaums, sonst x^* = Wurzel des rechten Teilbaums, bis $x = x^*$ oder ein Nachfolgerbaum leer ist.

Ein solchermaßen strukturierter Algorithmus benötigt bis zum Zugriff auf Knoten x_i im Baum T genau $Z_T(x_i) = t_i + 1$ Vergleiche, wobei t_i die Tiefe (= Anzahl der Vorgänger) des Knotens x_i ist. $Z_T(x_i)$ heißt Zugriffszeit des Knotens x_i; sie ist eine meßbare Abbildung von S, der Knotenmenge des Baumes, in die natürlichen Zahlen, $Z_T : S \to \mathbb{N}$, und hängt natürlich von der speziellen Gestalt des Baumes T ab.

Wir nehmen jetzt an, daß Vorkenntnisse über die Häufigkeit, mit der auf ein Element von S zugegriffen wird, in Form einer Wahrscheinlichkeitsverteilung über $S = \{x_1, \ldots, x_m\}$ vorliegen, charakterisiert durch $p_i = P(\{x_i\})$, $i = 1, \ldots, m$. P heißt *Zugriffsverteilung*. In dieser Situation kann die Qualität eines binären Suchbaumes durch die erwartete Zugriffszeit beurteilt werden.

$$E(Z_T) = \sum_{i=1}^{m} (t_i + 1) P(Z_T = t_i + 1) = \sum_{i=1}^{m} p_i (t_i + 1). \tag{5.3.3}$$

Für die in (5.3.2) betrachteten Bäume T_1 bzw. T_2 beträgt bei Vorliegen einer Gleichverteilung $P(\{i\}) = 0.1$, $i = 1, \ldots, 0$,

$$E(Z_{T_1}) = 2.9 \qquad \text{bzw.} \qquad E(Z_{T_2}) = 3.5.$$

Der erste Suchbaum ist daher effizienter als der zweite. Bei einer Zugriffsverteilung $P(\{i\}) = p_i$, $i = 1, \ldots, 10$, mit

$$(p_1, \ldots, p_{10}) = (0.2,\ 0.2,\ 0.2,\ 0.1,\ 0.05,\ 0.05,\ 0.05,\ 0.05,\ 0.05,\ 0.05)$$

beträgt jedoch

$$E(Z_{T_1}) = 3.2 \qquad \text{bzw.} \qquad E(Z_{T_2}) = 2.55,$$

so daß in diesem Fall der zweite Suchbaum die höhere Effizienz besitzt, die damit offensichtlich von der speziellen Gestalt der Zugriffsverteilung abhängt. Bei fester Zugriffsverteilung werden solche Suchbäume günstig sein, die den Knoten mit geringer Wahrscheinlichkeitsmasse eine große Zugriffszeit geben und denen, die mit großer Wahrscheinlichkeit auftreten, eine kleine Zugriffszeit.

Unser Ziel lautet nun, bei gegebener Menge S und Zugriffsverteilung P einen binären Suchbaum über S so zu finden, daß die erwartete Zugriffszeit $E(Z_T)$ klein ist. Wir werden hier keine Algorithmen für optimale Suchbäume angeben (vgl. hierzu Mehlhorn S. 148ff.), sondern durch Vergleich mit der Entropie der Zugriffsverteilung lediglich fast optimale Suchbäume konstruieren.

Der im folgenden betrachtete "Median–Algorithmus" läßt in schöner Weise die stochastischen Argumente bei der Untersuchung seiner Eigenschaften hervortreten, ohne daß seine Formulierung von der algorithmischen Seite her große Mühe bereitet.

Analog zu Satz 5.2.2 (Noiseless coding theorem) werden wir untere und obere Schranken für $E(Z_T)$ aus (5.3.3) angeben. Auf diesem Weg wird das folgende Lemma benötigt.

Lemma 5.3.1. *T sei ein binärer Suchbaum über der Menge $S = \{x_1, \ldots, x_m\}$, $m \geq 2$, t_i bezeichne die Tiefe des Knotens x_i. Dann gilt für alle $0 \leq c \leq 1$ die Ungleichung* $c \sum_{i=1}^{m} \left((1-c)/2\right)^{t_i} \leq 1$.

Beweis. Für $m(k) = \#\big(\{x_i | t_i = k\}\big)$, die Anzahl der Knoten mit Tiefe k, $k \in \mathbb{N}_0$, gilt $m(k) \leq 2^k$, da T ein binärer Baum ist. Es folgt

$$c \sum_{i=1}^{m} \Big(\frac{1-c}{2}\Big)^{t_i} = c \sum_{k=0}^{m} m(k) \Big(\frac{1-c}{2}\Big)^{k} \leq c \sum_{k=0}^{m} (1-c)^k$$
$$= 1 - (1-c)^{m+1} \leq 1. \qquad \blacksquare$$

Im folgenden sei $(p_1, \ldots, p_m)$, $p_i \geq 0$, $\sum_{i=1}^{m} p_i = 1$, eine gegebene Zugriffsverteilung auf der Menge $S = \{x_1, \ldots, x_m\}$.

Satz 5.3.1. *$H = H(p_1, \ldots, p_m)$ sei die Entropie der Zugriffsverteilung. Dann gilt für alle binären Suchbäume T über $S = \{x_1, \ldots, x_m\}$*

$$\max_{y \in \mathbf{R}} \left\{ \frac{H - y}{\log(2 + b^{-y})} \right\} \leq E(Z_T), \tag{5.3.4}$$

wobei $b > 0$ die Basis von "log" ist, die auch bei der Berechnung von H verwendet wird.

Beweis. Für $0 \leq c < 1$ bezeichne $c' = \frac{1-c}{2}$ und $q_i = c\big(\frac{1-c}{2}\big)^{t_i}$, $i = 1, \ldots, m$, wobei t_i wie oben die Tiefe des Knotens x_i bedeutet. Dann gilt $t_i + 1 = (\log q_i - \log c)/\log c' + 1$ und

$$E(Z_T) = \sum_{i=1}^{m} p_i (t_i + 1) = 1 - \frac{\log c}{\log c'} + \frac{1}{\log c'} \sum_{i=1}^{m} p_i \log q_i \tag{5.3.5}$$

Mit Hilfe von (5.1.10) und Lemma 5.3.1 erhalten wir die Abschätzung

$$H(p_1, \ldots, p_m) + \sum_{i=1}^{m} p_i \log q_i = \sum_{i=1}^{m} p_i \log \frac{q_i}{p_i}$$
$$= \log e \sum_{\substack{i=1 \\ p_i \neq 0}}^{m} p_i \ln \frac{q_i}{p_i} = \log e \Big(\sum_{\substack{i=1 \\ p_i \neq 0}}^{m} p_i \frac{q_i}{p_i} - 1 \Big) \leq 0,$$

die mit (5.3.5), da $\log c' \le 0$,

$$E(Z_T) \ge 1 - \frac{\log c}{\log c'} - \frac{H}{\log c'} = \frac{1}{\log c'}\left(\log \frac{c'}{c} - H\right) = \frac{H - \log(c'/c)}{\log(1/c')}$$

liefert. $y = \log \frac{c'}{c} = \log \frac{1-c}{2c}$ durchläuft mit $0 < c < 1$ alle reellen Zahlen, so daß bei Exponenten zur Basis $b > 0$ gilt: $b^{-y} = c/c'$, bzw. $2 + b^{-y} = 1/c'$. Insgesamt erhält man $E(Z_T) \ge \frac{H-y}{\log(2+b^{-y})}$ für alle $y \in \mathbf{R}$, woraus die Behauptung folgt. ∎

Aus (5.3.4) folgt für den speziellen Wert $y = 0$ die untere Schranke

$$H/\log 3 \le E(Z_T) \tag{5.3.6}$$

für alle binären Suchbäume T. Eine bessere, implizite Schranke läßt sich aus Satz 5.3.1 ableiten, indem man dort $y = \log(E(Z_T)/2)$ einsetzt und die entstehende Ungleichung nach H auflöst.

Korollar 5.3.1. *Unter den Bezeichnung von Satz 5.3.1 gilt für alle binären Suchbäume T:*

$$H \le E(Z_T) + \log E(Z_T) + \log e - 1.$$

Ein binärer Suchbaum T, der die untere Schranke in (5.3.4) erreicht, ist bezüglich der erwarteten Zugriffszeit sicher optimal, "absolut optimal" im Sinn von Kapitel 5.2. Im allgemeinen wird ein solcher Suchbaum jedoch nicht existieren. Analog zu Satz 5.2.2 b) geben wir jetzt konstruktiv einen binären Suchbaum T^* an, für den $E(Z_{T^*})$ nicht weit entfernt von der unteren Schranke aus (5.3.4) bzw. (5.3.6) liegt.

Satz 5.3.2. *$H = H(p_1, \ldots, p_m)$ sei die Entropie der Zugriffsverteilung. Dann existiert ein binärer Suchbaum T^* für $S = \{x_1, \ldots, x_m\}$ mit*

$$E(Z_{T^*}) \le \frac{H}{\log 2} + 1. \tag{5.3.7}$$

Die Konstruktionsidee für einen solchen Baum T^* beruht darauf, die Summe der Wahrscheinlichkeiten im rechten und linken Teilbaum jedes Knotens möglichst gleich zu machen. Genauer, sind $x_\ell, \ldots, x_r \in S$, $1 \le \ell < r \le m$, als Knoten eines binären Teilbaumes anzuordnen, so wähle x_k als Wurzel dieses Teilbaumes, wenn

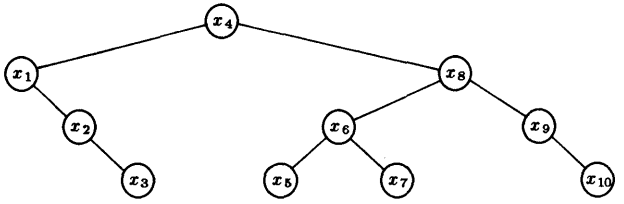
$$\sum_{j=\ell}^{k-1} p_j < \frac{1}{2}\sum_{j=\ell}^{r} p_j \quad \text{und} \quad \sum_{j=\ell}^{k} p_j \ge \frac{1}{2}\sum_{j=\ell}^{r} p_j. \tag{5.3.8}$$

Verfahre mit den Knoten $x_\ell, \ldots, x_{k-1}$ bzw. $x_{k+1}, \ldots, x_r$ genauso bis $\ell > k-1$ bzw. $r < k+1$.

Ein Beispiel mag die Vorgehensweise verdeutlichen. Sei $S = \{x_1, \ldots, x_{10}\}$, $x_1 < x_2 < \ldots < x_{10}$ und die Zugriffsverteilung

$$(p_1, \ldots, p_{10}) = (0.2,\ 0.1,\ 0.05,\ 0.3,\ 0.05,\ 0.03,\ 0.08,\ 0.03,\ 0.1,\ 0.06).$$

Dann ist $\sum_{j=1}^{4} p_j \ge \frac{1}{2}$ erstmalig, und x_4 bildet die Wurzel des Baumes. Im linken Teilbaum verbleiben die Knoten x_1, x_2, x_3. Es gilt $p_1 = 0.2 \ge \frac{1}{2}\sum_{j=1}^{3} p_j = 0.35/2$, so daß x_1 Wurzel des linken Teilbaums ist, im rechten Teilbaum gilt $\sum_{j=5}^{8} p_j = 0.18 \ge \frac{1}{2}\sum_{j=5}^{10} p_j = 0.35/2$ erstmalig, so daß x_8 Wurzel des rechten Teilbaums wird. Führt man das Verfahren vollständig durch, entsteht der folgende binäre Suchbaum.



Der Algorithmus zur allgemeinen Konstruktion solcher ausgewogenen binären Suchbäume wird durch die folgende rekursive Prozedur in PASCAL-Notation mit dem Datentyp `node` aus (5.3.1) beschrieben.

```
FUNCTION bintree(l,r:  INTEGER): ↑node;
  VAR z:  ↑node;
  m:  INTEGER;
BEGIN
  { bestimme Index m zwischen l und r mit ausgeglichenen
    Wahrscheinlichkeiten, d.h.
    ∑_{j=l}^{m-1} p_j < ½ ∑_{j=l}^{r} p_j, ∑_{j=l}^{m} p_j ≥ ½ ∑_{j=l}^{r} p_j  };
  new(z); z↑.no:=m;
  IF l<=m-1 THEN z↑.left:=bintree(l,m-1)
    ELSE z↑.left:=NUL;
  IF m+1<=r THEN z↑.right:=bintree(m+1,r)
    ELSE z↑.right:=NUL;
  bintree:=z;                                              (5.3.9)
END;
```

Zum Beweis von Satz 5.3.2 benötigen wir noch

Lemma 5.3.2. *Für jeden Teilbaum mit Wurzel x_k und Knoten $x_\ell, \ldots, x_k, \ldots, x_r$ eines ausgewogenen Suchaums T^* aus (5.3.9) gilt $\sum_{j=\ell}^{r} p_j \leq 2^{-t_k^*}$, insbesondere also $p_k \leq 2^{-t_k^*}$, wobei t_k^* die Tiefe des Knotens x_k bezeichnet.*

Beweis. Wir führen den Beweis durch Induktion über die Tiefe der Wurzeln. Ist x_{k_0} Wurzel von T^*, so gilt $t_{k_0}^* = 0$, also $\sum_{j=1}^{m} p_j \leq 2^{-t_{k_0}^*} = 1$.
x_k sei Wurzel des Teilbaums mit Knoten $x_\ell, \ldots, x_k, \ldots, x_r$ und gelte $\sum_{j=\ell}^{r} p_j \leq 2^{-t_k^*}$. Für den linken und rechten Teilbaum gilt dann wegen (5.3.8) $\sum_{j=\ell}^{k-1} p_j \leq 2^{-t_k^*-1}$ und $\sum_{j=k+1}^{r} p_j \leq 2^{-t_k^*-1}$, wobei Summen mit leerem Indexbereich zu 0 gesetzt werden. Die Tiefe der Wurzel des linken bzw. rechten Teilbaums beträgt gerade $t_k + 1$, falls dieser nichtleer ist, woraus die Behauptung folgt. ∎

Der Beweis von Satz 5.3.2 fällt jetzt leicht. Mit Lemma 5.3.2 gilt für einen nach (5.3.9) konstruierten Baum T^*, daß $p_j \leq 2^{-t_j^*}$ für alle $j = 1, \ldots, m$, also $\log p_j \leq -t_j^* \log 2$. Es folgt

$$E(Z_{T^*}) = \sum_{j=1}^{m} p_j(t_j^* + 1) \leq -\frac{1}{\log 2} \sum_{j=1}^{m} p_j \log p_j + 1 = \frac{H(p_1, \ldots, p_m)}{\log 2} + 1.$$

Wir sind jetzt in der Lage, einen binären Suchbaum T^* zu konstruieren, dessen erwartete Zugriffszeit mit Hilfe von (5.3.6) und (5.3.7) beurteilt werden kann, und zwar

$$\frac{H}{\log 3} \leq E(Z_{T^*}) \leq \frac{H}{\log 2} + 1. \tag{5.3.10}$$

Ist die Zugriffsverteilung eine diskrete Gleichverteilung auf $S = \{x_1, \ldots, x_m\}$, bedeutet (5.3.10) wegen Satz 5.1.1 a)(ii) unter Verwendung von Logarithmen zur Basis 2

$$\log_2 m / \log_2 3 \leq E(Z_{T^*}) \leq \log_2 m + 1.$$

Mit den gleichen stochastischen Hilfsmitteln können analoge Aussagen zu Satz 5.3.1 und 5.3.2 für den Fall bewiesen werden, daß nicht alle Elemente des relevanten Universums im Baum abgespeichert sind und die Wahrscheinlichkeiten q_i dafür bekannt sind, daß Elemente des Universums zwischen jeweils abgespeicherten Elementen x_i und x_{i+1} liegen (vgl. Mehlhorn S. 164 ff).

Der Algorithmus zur Konstruktion des entsprechenden fast optimalen Suchbaums, der Blätter mit vorgegebenen Wahrscheinlichkeiten an einer geeigneten Stelle im Baum hinzufügen muß, basiert auf dem Prinzip "strikter Halbierung". Seine algorithmische Darstellung ist recht aufwendig.

5.4. Stationäre Quellen und Markoff–Quellen

In Definition 5.2.1 haben wir diskrete, gedächtnislose Quellen als Folge stochastisch unabhängiger, identisch verteilter Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ mit diskreter Randverteilung modelliert. Diese spezielle Struktur der Quelle kam dann bei Blockcodierung in Korollar 5.2.1 zum Tragen: die Entropie einer Signalfolge der Länge L beträgt gerade das L–fache der Entropie von X_1. Mit jeder weiteren betrachteten Stelle in der Signalfolge erhöht sich die Unbestimmtheit um den gleichen konstanten Betrag.

Allerdings ist das Modell der stochastischen Unabhängigkeit der Komponenten in vielen Fällen zu einfach. In der deutschen Sprache etwa taucht keine Sequenz von fünf Konsonanten hintereinander auf, die Wahrscheinlichkeit für das Auftreten eines bestimmten Buchstabens ist sicher nicht unabhängig davon, welche Buchstaben vorher aufgetreten sind. Bei der Modellierung vieler praktisch relevanter Situationen sollten daher auch Abhängigkeiten zwischen den Komponenten zugelassen sein.

Die Frage stellt sich, wie man sinnvoll die Entropie solcher Quellen definiert. Im Fall abhängiger Zufallsvariablen verhält sich wegen Satz 5.1.1 c) die Entropie nicht additiv, da —anschaulich gesprochen— jedes Signal auch Information über andere trägt, der Zuwachs an Information bei jedem neu beobachteten Signal damit kleiner ist als im stochastisch unabhängigen Fall.

Für stationäre Verteilungen $P^{\{X_n\}}$ werden wir im folgenden eine vernünftige Definition der Entropie der zugehörigen diskreten Quelle $\{X_n\}_{n\in\mathbb{N}}$ angeben. Erinnert sei an Definition 3.2.3.

Eine Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißt stationär, wenn für alle Auswahlen $s, k, i_1, \ldots, i_k \in \mathbb{N}$ gilt, daß $P^{(X_{i_1}, \ldots, X_{i_k})} = P^{(X_{i_1+s}, \ldots, X_{i_k+s})}$.

Definition 5.4.1. *Eine stationäre Folge* $\{X_n\}_{n\in\mathbb{N}}$ *von Zufallsvariablen, deren Randverteilungen* $P^{X_n} = P^X$ *den endlichen Wertebereich* $\mathcal{X} = \{x_1, \dots x_m\}$, $m \in \mathbb{N}$, *besitzt, heißt diskrete, stationäre Quelle.* $\mathcal{X}$ *heißt Quellalphabet.*

Wie in (5.2.6) bezeichne $\boldsymbol{X}^{(L)} = (X_1, \dots, X_L)$ den Zufallsvektor aus den ersten L Komponenten von $\{X_n\}_{n\in\mathbb{N}}$.

$$H\big(\boldsymbol{X}^{(L)}\big) = -\sum_{u_1,\dots,u_L\in\mathcal{X}} P(X_1 = u_1, \dots, X_L = u_L) \cdot \log P(X_1 = u_1, \dots, X_L = u_L)$$

ist dann die Entropie von $\boldsymbol{X}^{(L)}$, sie ist wegen Lemma 5.1.1 monoton nicht fallend in L, und es gilt mit Satz 5.1.1 c), daß $H\big(\boldsymbol{X}^{(L)}\big) \le \sum_{i=1}^{L} H(X_i)$.

Die Mittelung $\frac{1}{L}H\big(\boldsymbol{X}^{(L)}\big)$ ist ein Maß für die Unbestimmtheit pro Komponente, in das die Auswirkung von Abhängigkeiten bis zur Weite L mit einbezogen wird. Das folgende Lemma zeigt, daß $\lim_{L\to\infty} H\big(\boldsymbol{X}^{(L)}\big)/L$ für stationäre Quellen existiert und mit der bedingten Entropie von X_L gegeben alle Vorgänger $X_1, \dots, X_{L-1}$, im Limes übereinstimmt.

Lemma 5.4.1. $\{X_n\}_{n\in\mathbb{N}}$ *sei eine diskrete, stationäre Quelle. Dann gilt*

a) $\big\{H\big(X_L \mid (X_1, \dots, X_{L-1})\big)\big\}_{L\in\mathbb{N}}$ *ist monoton nicht wachsend.*

b) $H\big(X_L \mid (X_1, \dots, X_{L-1})\big) \le \frac{1}{L}H\big(\boldsymbol{X}^{(L)}\big)$ *für alle* $L \in \mathbb{N}$.

c) $\big\{\frac{1}{L}H\big(\boldsymbol{X}^{(L)}\big)\big\}_{L\in\mathbb{N}}$ *ist monoton nicht wachsend.*

d) $\lim\limits_{L\to\infty} \frac{1}{L}H\big(\boldsymbol{X}^{(L)}\big) = \lim\limits_{L\to\infty} H\big(X_L \mid (X_1, \dots, X_{L-1})\big)$,

wobei die in d) auftretenden Limites existieren.

Beweis. Ist (X, Y, Z) ein diskreter Zufallsvektor, so gilt allgemein

$$H\big(X \mid (Y, Z)\big) \le H\big(X \mid Y\big) \tag{5.4.1}$$

Dies beweist man nach Einsetzten der Definition (5.1.6) völlig analog zur Ungleichung (ii) aus Satz 5.1.1 b). Ungleichung (5.4.1) zusammen mit der Stationarität von $\{X_n\}_{n\in\mathbb{N}}$ liefert

$$\begin{aligned} H\big(X_L \mid (X_1, \dots, X_{L-1})\big) &\le H\big(X_L \mid (X_2, \dots, X_{L-1})\big) \\ &= H\big(X_{L-1} \mid (X_1, \dots, X_{L-2})\big), \end{aligned}$$

also Behauptung a).

b) Die Entropie von $X^{(L)}$ läßt sich wie in (5.1.9) iteriert, bedingt berechnen, so daß mit (5.4.1)

$$\begin{aligned} \frac{1}{L}H\big(\boldsymbol{X}^{(L)}\big) &= \frac{1}{L}\Big\{H(X_1, \dots, X_{L-1}) + H\big(X_L \mid (X_1, \dots, X_{L-1})\big)\Big\} \\ &= \frac{1}{L}\Big\{H(X_1) + H(X_2 \mid X_1) + \cdots + H\big(X_L \mid (X_1, \dots, X_{L-1})\big)\Big\} \\ &\ge H\big(X_L \mid (X_1, \dots, X_{L-1})\big), \end{aligned}$$

also die behauptete Ungleichung folgt.
c) Mit Hilfe von b) erhalten wir

$$H(\boldsymbol{X}^{(L)}) = H(X_1,\dots,X_{L-1}) + H(X_L \mid (X_1,\dots,X_{L-1}))$$
$$\leq H(\boldsymbol{X}^{(L-1)}) + \frac{1}{L}H(\boldsymbol{X}^{(L)}),$$

so daß nach Zusammenfassen der linken und rechten Seite dieser Ungleichung

$$\frac{1}{L}H(\boldsymbol{X}^{(L)}) \leq \frac{1}{L-1}H(\boldsymbol{X}^{(L-1)})$$

für alle $L \in \mathbb{N}$ folgt.
d) Die Existenz der Limiten ergibt sich aus der Monotonie der Folgen in a) und c), die beide nach unten durch 0 beschränkt sind.
Daß $\lim_{L\to\infty} H(X_L \mid (X_1,\dots,X_{L-1})) \leq \lim_{L\to\infty} \frac{1}{L}H(\boldsymbol{X}^{(L)})$ ist eine Konsequenz von b). Die umgekehrte Ungleichung leiten wir folgendermaßen her.
Für alle $k, L \in \mathbb{N}$ gilt mit Lemma 5.1.1, (5.4.1) und b):

$$\frac{1}{L+k}H(\boldsymbol{X}^{(L+k)})$$
$$= \frac{1}{L+k}\Big\{H(X_{L+k} \mid (X_1,\dots,X_{L+k-1})) + \cdots + H(X_{L+1} \mid (X_1,\dots,X_L))$$
$$+ H(X_L \mid (X_1,\dots,X_{L-1})) + H(X_1,\dots,X_{L-1})\Big\}$$
$$\leq \frac{1}{L+k}H(X_1,\dots,X_{L-1}) + \frac{k+1}{L+k}H(X_L \mid (X_1,\dots,X_{L-1}))$$

Mit $k \to \infty$ konvergiert der erste Summand gegen 0, der zweite Summand gegen $H(X_L \mid (X_1,\dots,X_{L-1}))$, so daß für alle $L \in \mathbb{N}$ gilt

$$\lim_{L'\to\infty} \frac{1}{L'}H(\boldsymbol{X}^{(L')}) \leq H(X_L \mid (X_1,\dots,X_{L-1})).$$

Der Grenzübergang $L \to \infty$ liefert dann Behauptung d). ∎

Definition 5.4.2. *$\{X_n\}_{n\in\mathbb{N}}$ sei eine diskrete, stationäre Quelle. Dann heißt $\bar{H}(\{X_n\}) = \lim_{L\to\infty} \frac{1}{L}H(\boldsymbol{X}^{(L)})$ die Entropie der Quelle.*

Bei Blockcodierung von Wörtern der Länge L über dem Quellalphabet $\mathcal{X} = \{x_1,\dots,x_m\}$ mit Hilfe von Codes $g^{(L)} : \mathcal{X}^L \to \bigcup_{l=1}^{\infty} \mathcal{Y}^l$, $\mathcal{Y} = \{y_1,\dots,y_d\}$, $d \geq 2$, ein Codealphabet, erhalten wir als Konsequenz aus Satz 5.2.2 auch bei stationären Quellen das folgende Korollar. $\bar{n}(g^{(L)})$ bezeichnet wie in (5.2.7) die erwartete Codewortlänge bei Verwendung des Codes $g^{(L)}$.

Korollar 5.4.1. *$\{X_n\}_{n\in\mathbb{N}}$ sei eine diskrete, stationäre Quelle. Für alle $\varepsilon > 0$ existieren $L \in \mathbb{N}$ und ein Blockcode $g^{(L)}$ derart, daß*

$$\frac{\bar{H}(\{X_n\})}{\log d} \leq \frac{\bar{n}(g^{(L)})}{L} < \frac{\bar{H}(\{X_n\})}{\log d} + \varepsilon.$$

Beweis. Wegen Satz 5.2.2 existiert ein präfixfreier Blockcode $g^{(L)}$ mit der Eigenschaft

$$\frac{H(\boldsymbol{X}^{(L)})}{\log d} \leq \bar{n}(g^{(L)}) < \frac{H(\boldsymbol{X}^{(L)})}{\log d} + 1.$$

Nach Division durch L erhält man mit Lemma 5.4.1 b)

$$\frac{\bar{H}(\{X_n\})}{\log d} \leq \frac{H(\boldsymbol{X}^{(L)})}{L \log d} \leq \frac{\bar{n}(g^{(L)})}{L} < \frac{H(\boldsymbol{X}^{(L)})}{L \log d} + \frac{1}{L}.$$

Die rechte Seite konvergiert mit $L \to \infty$ gegen $\bar{H}(\{X_n\})/\log d$, woraus die Behauptung folgt. ∎

Im folgenden Beispiel werden wir für eine stationäre Quelle $\{X_n\}_{n\in\mathbf{N}}$ mit dem Huffman–Verfahren einen optimalen Blockcode der Länge L bestimmen und dann über den Zusammenhang in Korollar 5.4.1 $\bar{H}(\{X_n\})$ bestimmen.

Sei $\mathcal{X} = \{x_1, x_2, x_3\}$ ein dreielementiges Quellalphabet und $\{Z_n\}_{n\in\mathbf{N}}$ eine Folge stochastisch unabhängiger, identisch verteilter Zufallsvariablen mit Verteilung $P(Z_n = x_2) = \frac{1}{2} = P(Z_n = x_3), n \in \mathbf{N}$. Ferner sei U eine $\mathfrak{B}(1,\frac{1}{2})$–verteilte Zufallsvariable (d.h. $P(U = 0) = \frac{1}{2} = P(U = 1)$) so, daß $\{Z_n\}_{n\in\mathbf{N}}$ und U stochastisch unabhängig sind. Setze

$$X_n = x_1 U + Z_n(1 - U), \quad n \in \mathbf{N}. \tag{5.4.2}$$

Dann ist $\{X_n\}_{n\in\mathbf{N}}$ eine stationäre Folge, denn

$$\begin{aligned} P\Big(\bigcap_{j=1}^{k}\{X_{i_j} \leq z_j\}\Big) &= \int P\Big(\bigcap_{j=1}^{k}\{x_1 u + Z_{i_j}(1-u) \leq z_j\}\Big) dP^U(u) \\ &= \int P\Big(\bigcap_{j=1}^{k}\{x_1 u + Z_{i_j+s}(1-u) \leq z_j\}\Big) dP^U(u) \\ &= P\Big(\bigcap_{j=1}^{k}\{x_1 U + Z_{i_j+s}(1-U) \leq z_j\}\Big) = P\Big(\bigcap_{j=1}^{k}\{X_{i_j+s} \leq z_j\Big) \end{aligned}$$

für alle $s, i_1, \ldots, i_k \in \mathbf{N}$, $z_1, \ldots, z_k \in \mathbf{R}$, nach Lemma 3.1.4 über das Rechnen mit bedingten Verteilungen.

Eine mögliche Interpretation der stationären Quelle (5.4.2) ist, daß mit gleicher Wahrscheinlichkeit die Signale x_2 oder x_3 ausgesendet werden oder die konstante Folge $\{x_1\}$, etwa wenn die Quelle defekt ist. In beide Zustände gelangt die Quelle mit gleicher Wahrscheinlichkeit, U hat die Funktion eines "zufälligen Schalters".

Bei optimaler binärer Blockcodierung von Wörtern der Länge L über dem Codealphabet $\mathcal{Y} = \{0, 1\}$ erhalten wir mit dem Huffman–Verfahren aus Lemma 5.2.3 :

	$\boldsymbol{u}^{(L)}$	$P(\boldsymbol{X}^{(L)} = \boldsymbol{u}^{(L)})$	$n_j,\ j = 1,\ldots,2^L+1$
1	$(x_1,\ldots,x_1)$	$1/2$	1
2	$(u_1,\ldots,u_L)$,	$1/2 \cdot 1/2^L = 1/2^{L+1}$	$L+1$
$\vdots$	$u_i \in \{x_2, x_3\}$,	$\vdots$	$\vdots$
2^L+1	$i = 1,\ldots,L$	$1/2 \cdot 1/2^l = 1/2^{L+1}$	$L+1$

Dies liefert für den zugehörigen Code $g^{*(L)}$

$$\bar{n}(g^{*(L)}) = 1/2 + (L+1)(2^L/2^{L+1}) = (L+2)/2.$$

Bei Verwendung von Logarithmen zur Basis 2 folgt dann mit Korollar 5.4.1

$$\bar{H}(\{X_n\}) = \lim_{L\to\infty} \frac{\bar{n}(g^{*(L)})}{L} = \lim_{L\to\infty} \frac{L+2}{2L} = \frac{1}{2}.$$

Wir werden im weiteren solche stationären Quellen untersuchen, bei denen Abhängigkeiten zwischen dem Auftreten einzelner Buchstaben nicht beliebig weit reichen (gemessen durch die Differenz der Indizes) sondern nur über endlich viele Stufen bestehen. Die Buchstabenfolgen von natürlichen Sprachen können in ihrem stochastischen Verhalten gut durch solche Modelle approximiert werden. In Kapitel 3.2 wurden Markoff-Ketten behandelt, das sind Folgen von Zufallsvariablen, deren bedingte Verteilungen $P^{(X_n|X_{n-1},\ldots,X_1)}$ für alle $n \in \mathbb{N}$ nur vom unmittelbaren Vorgänger X_{n-1}, nicht aber von $X_{n-2},\ldots,X_1$ abhängen. Mit Hilfe von Markoff-Ketten werden wir obige Vorstellung von nur endlich viele Indizes weit reichenden Abhängigkeiten präzisieren.

Definition 5.4.3. *$\{Z_n\}_{n\in\mathbb{N}}$ sei eine homogene Markoff-Kette mit endlichem Zustandsraum $\mathcal{S} = \{s_1,\ldots,s_r\}$, $r \in \mathbb{N}$, $\mathcal{X} = \{x_1,\ldots,x_m\}$, $m \in \mathbb{N}$, ein Alphabet und $f : \mathcal{S} \to \mathcal{X}$ eine Abbildung. Die Folge $\{X_n\}_{n\in\mathbb{N}}$ der Zufallsvariablen $X_n = f(Z_n)$, $n \in \mathbb{N}$, heißt diskrete Markoff-Quelle mit Quellalphabet $\mathcal{X}$. f heißt assoziierte Funktion.*

Im folgenden werden die Übergangsmatrizen von Markoff-Ketten $\{Z_n\}_{n\in\mathbb{N}}$ mit $\Pi = (p_{ij})_{1\le i,j\le r}$ bezeichnet.

Markoff-Ketten mit endlichem Zustandsraum werden nicht direkt als Markoff-Quellen definiert, sondern erst nach Dazwischenschalten einer assoziierten Funktion f, weil auch weiterreichende Abhängigkeiten als nur einstufige miterfaßt werden sollen. Folgendes Beispiel verdeutlicht den Einsatz einer assoziierten Funktion f.

Betrachtet wird eine diskrete Quelle $\{X_n\}_{n\in\mathbb{N}}$ mit binärem Quellalphabet $\mathcal{X} = \{0,1\}$, bei der die Verteilung des Buchstabens zu einem bestimmten Zeitpunkt n homogen lediglich von den Buchstaben zu drei Vorgängerzeitpunkten abhängen soll. Wir wählen dann

$$\begin{aligned}\mathcal{S} &= \{0,1\}^3 = \{s_1,\ldots,s_8\}\\ &= \{(000),(001),(010),(011),(100),(101),(110),(111)\}\end{aligned} \tag{5.4.3}$$

und $\{Z_n\}_{n\in\mathbf{N}}$ als homogene Markoff-Kette mit Zustandsraum $\mathcal{S}$. Die Zustände $s_1,\ldots,s_8 \in \mathcal{S}$ werden als "Ausschnitte" der Länge drei aus 0-1-Zahlenfolgen gedeutet, so daß die Übergangsmatrix $\Pi = (p_{ij})_{1\le i,j\le 8}$ folgende Struktur besitzen muß

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{23} & p_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{35} & p_{36} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{47} & p_{48} \\ p_{51} & p_{52} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{63} & p_{64} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{75} & p_{76} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{87} & p_{88} \end{pmatrix} \tag{5.4.4}$$

Lediglich die mit "p_{ij}" notierten Elemente dieser Matrix können von 0 verschieden sein. Die assoziierte Funktion zeigt dann sukzessive die neu auftretenden Buchstaben, also wählt man f als Projektion auf die dritte Komponente.

$$f : \mathcal{S} \to \mathcal{X} : (u_1,u_2,u_3) \mapsto u_3. \tag{5.4.5}$$

Insgesamt sind dann Abhängigkeiten über drei Indizes hinweg durch eine mit f transformierte Markoff-Kette modelliert.

Ist die Anfangsverteilung der Markoff-Kette $\{Z_n\}$ einer Markoff-Quelle stationär, so bildet $\{Z_n\}_{n\in\mathbf{N}}$ und damit $\{X_n\}_{n\in\mathbf{N}}$ eine stationäre Folge von Zufallsvariablen. Die Entropie $\bar{H}(\{X_n\})$ existiert in diesem Fall als Grenzwert (vgl. Definition 5.4.2), der im allgemeinen schwierig zu bestimmen ist. Für einen bestimmten Typ von Markoff-Quellen läßt sich die Entropie jedoch relativ leicht berechnen.

Definition 5.4.4. *Unter den Bezeichnungen von Definition 5.4.3 sei $\{X_n\}_{n\in\mathbf{N}}$ eine Markoff-Quelle mit zugehöriger stationärer Markoff-Kette $\{Z_n\}_{n\in\mathbf{N}}$ und assoziierter Funktion f. Für $s_k \in \mathcal{S}$ bezeichne $R(s_k) = \{s_\ell \in \mathcal{S} \mid p_{k\ell} > 0\}$ die Menge der in einem Schritt von s_k aus erreichbaren Zustände. $\{X_n\}_{n\in\mathbf{N}}$ heißt dann unifilar, wenn für alle $s_k \in \mathcal{S}$, $s_i \neq s_j \in R(s_k)$ gilt, daß $f(s_i) \neq f(s_j)$.*

Unifilar bedeutet, daß die Restriktion von f auf $R(s_k)$ für alle $s_k \in \mathcal{S}$ injektiv ist. Damit erzeugt jeder Zustand, der von s_k aus in einem Schritt erreicht werden kann, über f einen anderen Buchstaben des Quellalphabets. Offensichtlich bestimmen dann der Zustand zur Zeit $n \in \mathbf{N}$ und der Buchstabe zur Zeit $n+1$ eindeutig den Zustand zur Zeit $n+1$.

Die Markoff-Quelle aus (5.4.3)-(5.4.5) ist unifilar, wie man leicht anhand der Definition nachprüft.

Satz 5.4.1. *Für jede unifilare (und damit stationäre) Markoff-Quelle $\{X_n\}_{n\in\mathbf{N}}$ gilt*

$$\bar{H}(\{X_n\}) = \sum_{j=1}^{r} w_j H(p_{j1},\ldots,p_{jr}),$$

wobei $\boldsymbol{w} = (w_1,\ldots,w_r)$ die stationäre Anfangsverteilung und $(p_{j1},\ldots,p_{jr})$ die j-te Zeile der Übergangsmatrix Π ist, und somit die homogene Übergangsverteilung $P^{(Z_n \mid Z_{n-1}=j)}$ bestimmt.

Beweis. $\{X_n\}_{n\in\mathbf{N}}$ ist als Transformation $X_n = f(Z_n)$ einer stationären Folge $\{Z_n\}_{n\in\mathbf{N}}$ wieder stationär, so daß $\bar{H}(\{X_n\})$ nach Lemma 5.4.1 wohldefiniert ist.

Sind $z_1,\dots,z_n \in \mathcal{S}$ mit

$$0 < P(Z_1 = z_1,\dots,Z_n = z_n) = P(Z_1 = z_1)P(Z_2 = z_2 \mid Z_1 = z_1) \cdots P(Z_n = z_n \mid Z_{n-1} = z_{n-1}),$$

so gilt $z_j \in R(z_{j-1})$ für alle $j = 2,\dots,n$. Die assoziierte Funktion ist nach Voraussetzung injektiv auf $R(z_{j-1})$ für alle $j = 2,\dots,n$, so daß

$$P(Z_j = z_j \mid Z_{j-1} = z_{j-1}) = P\big(X_j = f(z_j) \mid X_{j-1} = f(z_{j-1})\big),\ j = 2,\dots,n.$$

Insgesamt folgt mit $u_j = f(z_j)$, $j = 2,\dots,n$,

$$\begin{aligned} P(Z_1 = z_1,\dots,Z_n = z_n) &= P(Z_1 = z_1)P(X_2 = u_2 \mid Z_1 = z_1) \\ &\quad \cdot P(X_3 = u_3 \mid X_2 = u_2) \cdots P(X_n = u_n \mid X_{n-1} = u_{n-1}) \\ &= P(Z_1 = z_1, X_2 = u_2,\dots,X_n = u_n), \end{aligned}$$

falls $P(Z_1 = z_1,\dots,Z_n = z_n) > 0$. Wir erhalten

$$\begin{aligned} &\frac{1}{n}H(Z_1,\dots,Z_n) \\ &\quad = -\frac{1}{n}\sum_{z_1,\dots,z_n \in \mathcal{S}} P(Z_1 = z_1,\dots,Z_n = z_n)\log P(Z_1 = z_1,\dots,Z_n = z_n) \\ &\quad = -\frac{1}{n}\sum_{\substack{u_2,\dots,u_n \in \mathcal{X} \\ z_1 \in \mathcal{S}}} P(X_2 = u_2,\dots,X_n = u_n \mid Z_1 = z_1)P(Z_1 = z_1) \\ &\qquad\quad \cdot \big(\log P(X_2 = u_2,\dots,X_n = u_n \mid Z_1 = z_1) + \log P(Z_1 = z_1)\big) \\ &\quad = \frac{1}{n}\Big(H(Z_1) + H\big((X_2,\dots,X_n) \mid Z_1\big)\Big), \end{aligned} \tag{5.4.6}$$

wobei die letzte Gleichheit durch Summation der bedingten Wahrscheinlichkeiten zu 1 und aus (5.1.6) folgt. Lemma 5.1.1 liefert

$$\begin{aligned} H\big((X_2,\dots,X_n) \mid Z_1\big) &= H(Z_1, X_2,\dots,X_n) - H(Z_1) \\ &= H(X_2,\dots,X_n) + H\big(Z_1 \mid (X_2,\dots,X_n)\big) - H(Z_1). \end{aligned}$$

Wegen Satz 5.1.1 b) ist $H\big(Z_1 \mid (X_2,\dots,X_n)\big) \le H(Z_1) \le \log r$, so daß nach Einsetzen in (5.4.6)

$$\begin{aligned} \bar{H}\big(\{Z_n\}\big) &= \lim_{n\to\infty}\frac{1}{n}H(Z_1,\dots,Z_n) = \lim_{n\to\infty}\frac{1}{n}H(X_2,\dots,X_n) \\ &= \lim_{n\to\infty}\frac{1}{n-1}H(X_1,\dots,X_{n-1}) = \bar{H}\big(\{X_n\}\big) \end{aligned}$$

folgt. $\bar{H}\big(\{Z_n\}\big)$ läßt sich mit Lemma 5.4.1 d) berechnen als

$$\begin{aligned} \bar{H}\big(\{Z_n\}\big) &= \lim_{n\to\infty} H\big(Z_n \mid (Z_1,\dots,Z_{n-1})\big) = \lim_{n\to\infty} H(Z_n \mid Z_{n-1}) = H(Z_2 \mid Z_1) \\ &= \sum_{j=1}^{r} P(Z_1 = s_j)H(Z_2 \mid Z_1 = s_j) = \sum_{j=1}^{r} w_j H(p_{j1},\dots,p_{jr}). \end{aligned}$$

Obige Identitäten folgen aus der Markoff-Eigenschaft, der Stationarität und Formel (5.1.6). ∎

5.5. Aufgaben

5.1 Für $n \in \mathbf{N}$ und $\alpha > 0$ sei die Wahrscheinlichkeitsverteilung $P(n,\alpha)$ über dem Meßraum $(\mathcal{X} = \{x_1,\dots,x_n\}, \mathfrak{P}(\mathcal{X}))$ definiert durch

$$p_j(n,\alpha) = P(n,\alpha)(\{x_j\}) = \frac{j^\alpha}{s(n,\alpha)},\ 1 \le j \le n,$$

mit $s(n,\alpha) = \sum_{k=1}^n k^\alpha$. Zeigen Sie für die zugehörige Entropie $H(n,\alpha)$ bei Verwendung des natürlichen Logarithmus: $H(n,\alpha) = \ln n + \frac{\alpha}{\alpha+1} - \ln(\alpha+1) + O\left(\frac{\ln n}{n}\right)$.

Hinweis: Benutzen Sie die Ungleichung $\int_0^n f(x)dx \le \sum_{k=1}^n f(k) \le \int_1^{n+1} f(x)dx$ für monoton wachsende Funktionen f.

5.2 In einer unabhängigen Versuchsserie mit den Ausgängen 0 und 1 mit Eintrittswahrscheinlichkeiten $1-p$ und p, $0 < p < 1$, bezeichne X_n die Anzahl der Versuche unter den ersten n, die der ersten 1 vorausgehen. Berechnen Sie $H(X_n)$. Was ergibt sich hier für $n \to \infty$?

5.3 X, Y und Z seien endlich diskrete Zufallsvariable. Man beweise $H((X,Y) \mid Z) \le H(X \mid Z) + H(Y \mid Z)$.

5.4 $(p_1,\dots,p_m)$, $(q_1,\dots,q_m) \in \mathcal{P}_m$ seien stochastische Vektoren der Länge m, $m \in \mathbf{N}$. Zeigen Sie: $-\sum_{i=1}^m p_i \log p_i \le -\sum_{i=1}^m p_i \log q_i$ mit Gleichheit genau dann, wenn $p_i = q_i$ für alle $i = 1,\dots,m$.

5.5 $X : \Omega \to \mathcal{X}$, $\mathcal{X} = \{x_1,\dots,x_m\}$ sei eine endlich diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $g : \mathcal{X} \to \mathcal{Y}$, $\mathcal{Y} = \{y_1,\dots,y_n\}$ eine Abbildung. Man zeige, daß $H(g(X)) \le H(X)$ gilt.

5.6 $\mathcal{P} = \bigcup_{m\in\mathbf{N}} \mathcal{P}_m$ bezeichne die Menge der Wahrscheinlichkeitsvektoren. $H : \mathcal{P} \to \mathbf{R}$ sei eine Funktion mit folgenden Eigenschaften. Für alle $m \in \mathbf{N}$, $(p_1,\dots,p_m) \in \mathfrak{P}_m$ gilt:

(1) $H|_{\mathcal{P}_m}$ ist stetig und symmetrisch, d.h. $H(p_1,\dots,p_m) = H(p_{\pi(1)},\dots,p_{\pi(m)})$ für alle Permutationen π,

(2) $H(p_1,\dots,p_m) \le H(\frac{1}{m},\dots,\frac{1}{m})$,

(3) $H(p_1,\dots,p_m,0) = H(p_1,\dots,p_m)$,

(4) $H(p_1,\dots,p_m) = H(p_1+p_2,p_3,\dots,p_m) + (p_1+p_2)H\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right)$.

Zeigen Sie, daß dann $H(p_1,\dots,p_m) = -\sum_{i=1}^m p_i \log_b p_i$ für eine Konstante $b > 0$ ist.

Hinweis: Verifizieren Sie die Voraussetzungen des Charakterisierungssatzes 5.1.2.

5.7 Von 12 äußerlich gleichen Kugeln besitzen 11 gleiches Gewicht. Eine der Kugeln hat abweichendes Gewicht, jedoch ist nicht bekannt, ob sie leichter oder schwerer als die übrigen ist. Mit Hilfe einer Balkenwaage soll durch Vergleichswägungen herausgefunden werden, welche der Kugeln abweichendes Gewicht besitzt, und gleichzeitig, ob diese leichter oder schwerer ist. Zeigen Sie: Mit drei Wägungen kann man obige Aufgabe lösen, mit weniger Wägungen jedoch nicht.

5.8 Das Quellalphabet $\mathcal{X} = \{x_1, x_2, x_3\}$ werde durch den binären Kode g kodiert, wobei $g(x_1) = (0)$, $g(x_2) = (1,0)$, $g(x_3) = (1,1)$. Die Abbildung

$$G : \bigcup_{j=1}^\infty \mathcal{X}^j \to \bigcup_{j=1}^\infty \{0,1\}^j : (x_1,\dots,x_k) \mapsto (g(x_1),\dots,g(x_k))$$

beschreibe die Kodierung von endlichen Wörtern über dem Quellalphabet.
Bestimmen Sie eine allgemeine Formel, die für gegebenes $\ell \in \mathbf{N}$ die Mächtigkeit der Menge

$$\mathcal{M} = \{(x_1,\dots,x_k) \in \textstyle\bigcup_{j=1}^\infty \mathcal{X}^j \mid G(x_1,\dots,x_k) \in \mathcal{Y}^\ell,\ k \in \mathbf{N}\}$$

angibt. $\mathcal{M}$ ist die Menge der Nachrichten, die bei Verwendung von g mit Kodewörtern der Länge ℓ dargestellt werden können. Welche Anzahlen ergeben sich für $\ell = 1, \ldots, 5$?

5.9 Untersuchen Sie, ob die folgenden Kodes g_1 und g_2 eindeutig dekodierbar sind. Das Quellalphabet sei $\mathcal{X} = \{x_1, \ldots, x_8\}$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$g_1(x_i)$	010	0001	0110	1100	00011	00110	11110	101011
$g_2(x_i)$	abc	abcd	e	dba	bace	ceac	ceab	eabd

5.10 $\boldsymbol{p} = (p_1, p_2, p_3) \in \mathcal{P}_3$ sei ein stochastischer Vektor. Bestimmen Sie eine Lösung des Minimierungsproblems $\min\left\{ \sum_{i=1}^3 p_i n_i \mid n_1, n_2, n_3 \in \mathbf{N},\ \sum_{i=1}^3 2^{-n_i} \leq 1 \right\}$.

5.11 Sei $m \in \mathbf{N}$, $m \geq 2$, und $\boldsymbol{p} = (p_1, \ldots, p_m) \in \mathcal{P}_m$ ein stochastischer Vektor mit $p_i > 0$ für alle $i = 1, \ldots, m$. $\boldsymbol{n}^* = (n_1^*, \ldots, n_m^*)$, $n_1^* \leq \cdots \leq n_m^*$ sei eine Lösung von $\min \sum_{j=1}^m p_j n_j$ über $n_1, \ldots, n_m \in \mathbf{N}$ mit $\sum_{j=1}^m 2^{-n_j} \leq 1$.
Zeigen Sie: a) $\sum_{j=1}^m 2^{-n_j^*} = 1$ b) $n_{m-1}^* = n_m^*$.

5.12 $\{X\}_{n \in \mathbf{N}}$ sei eine diskrete Quelle mit Alphabet $\mathcal{X} = \{x_1, \ldots, x_m\}$, $m \geq 2$, und zugehöriger Verteilung $\boldsymbol{p} = (\frac{1}{m}, \ldots, \frac{1}{m}) \in \mathcal{P}_m$. Zeigen Sie: Für jeden optimalen binären Kode g gilt $\bar{n}(g) = \lfloor \log_2 m \rfloor + \frac{2}{m}\left(m - 2^{\lfloor \log_2 m \rfloor}\right)$.

5.13 $(\mathcal{X}, \mathfrak{P}(\mathcal{X}), P)$ sei ein Wahrscheinlichkeitsraum mit endlicher Grundmenge $\mathcal{X}$. $\mathcal{Y}$ sei ebenfalls eine endliche Menge und $Y : \mathcal{X} \to \mathcal{Y}$ eine Zufallsvariable. $\mathcal{G} = \{g \mid g : \mathcal{X} \to \{0,1\}\}$ bezeichne die Menge der $\{0,1\}$–wertigen Funktionen mit Definitionsbereich $\mathcal{X}$. Bestimmen Sie eine Lösung des Maximierungsproblems $\max_{g \in \mathcal{G}} H(g \mid Y)$.

5.14 $H = H(p_1, \ldots, p_m)$ bezeichne die Entropie einer Zugriffsverteilung $(p_1, \ldots, p_m) \in \mathcal{P}_m$, $d > 1$ eine reelle Zahl und $f : \mathbf{R} \to \mathbf{R} : y \mapsto \dfrac{H - y}{\log_d(2 + d^{-y})}$. Existiert stets $y^* \in \mathbf{R}$ mit $f(y^*) \geq f(y)$ für alle $y \in \mathbf{R}$?

5.15 $\mathcal{X} = \{x_1, \ldots, x_{m+1}\}$ sei eine total geordnete Menge und $(p_1, \ldots, p_{m+1}) \in \mathcal{P}_{m+1}$ eine Zugriffsverteilung mit $p_i = 2^{-i}$, $i = 1, \ldots, m$. $H = H(p_1, \ldots, p_{m+1})$ bezeichne die zugehörige Entropie.

a) Zeigen Sie, daß ein binärer Suchbaum T^* existiert mit $E(Z_{T^*}) < H/\log 2$.

b) Gibt es einen binären Suchbaum T^{**} mit $E(Z_{T^{**}}) < H/\log 3$?

Beachten Sie, daß alle auftretenden Logarithmen zur gleichen Basis $d > 1$ gewählt werden.

5.16 $\{X_n\}_{n \in \mathbf{N}}$ sei eine Markoff–Quelle mit Alphabet $\mathcal{X} = \{a, b\}$, wobei $X_n = f(Z_n)$ für alle $n \in \mathbf{N}$ gilt und $\{Z_n\}_{n \in \mathbf{N}}$ eine homogene Markoff–Kette mit Zustandsraum $\mathcal{S} = \{s_1, s_2, s_3, s_4\}$, Übergangsmatrix

$$\Pi = \begin{pmatrix} 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0.2 & 0.8 \\ 0.7 & 0.3 & 0 & 0 \end{pmatrix}$$

und assoziierter Funktion $f : \mathcal{S} \to \mathcal{X}$ mit $f(s_1) = f(s_4) = a$, $f(s_2) = f(s_3) = b$ ist. Die Anfangsverteilung der Markoff–Kette sei stationär.

a) Berechnen Sie die Entropie $H_\infty(\{X_n\})$ und zeigen Sie, daß $P(Z_n = z_n \mid U_n = u_n, U_{n-1} = u_{n-1}) \in \{0,1\}$ für alle $n \geq 2$, $z_n \in \mathcal{S}$, $u_n, u_{n-1} \in \mathcal{X}$ ist.

Es sollen Wörter der Länge L dieser Quelle blockweise mit Wörtern über einem Kodealphabet der Mächtigkeit $d = 5$ kodiert werden.

b) Bestimmen Sie eine möglichst kleine Blocklänge L, für die ein präfixfreier Kode $g^{(L)}$ mit einer erwarteten Kodewortlänge pro Quellbuchstabe $\frac{1}{L}\bar{n}(g^{(L)}) < 0.3$ existiert.

6. Simulationsverfahren

Wir haben in vorhergehenden Kapiteln gesehen, welchen Nutzen Zahlen haben, die nach einem Zufallsprinzip ausgesucht werden. Erinnert sei an stochastische Algorithmen (Simulated Annealing) oder auch an Zufallspermutationen der Zahlen $\{1,\ldots,n\}$, mit denen Such- und Sortierverfahren empirisch getestet werden können. Weitere Anwendungen finden sich im weiten Gebiet der Simulation, wo Zufallszahlen benutzt werden, um natürliche Phänomene realistisch nachzubilden und deren Verhalten durch häufige Wiederholung zu studieren.

Systeme, die durch eine Menge von Zuständen beschrieben werden können und bei denen die Übergänge von einem Zustand in einen anderen zufällig geschehen oder von einem zufälligen Input abhängen, können häufig durch geeignete Datenstrukturen abstrahiert und zufällige Übergänge oder zufälliger Input durch entsprechend ausgewählte Zahlen gesteuert werden. Die Konstruktion der Zufallszahlen sollte aus naheliegenden Gründen innerhalb einer Programmumgebung rechnerintern erfolgen. Weil meist sehr große Mengen solcher Zahlen benötigt werden, sollte sie durch einfache, schnelle Algorithmen bewerkstelligt werden. Dies schließt in der Regel "echte" Zufallsexperimente wie Münzwurf, Würfelwurf oder physikalische Versuchsaufbauten als Erzeuger von Zufallszahlen aus, deren Ergebnisse auf umständliche Weise in ein Programm eingeschleust werden müßten. Besser geeignet sind Mechanismen, die die gewünschten Zahlen intern algorithmisch produzieren.

Folgen von Zufallszahlen müssen natürlich Eigenschaften aufweisen, die mit unserem intuitiven Verständnis von Zufälligkeit übereinstimmen. Sie sollten ferner den Gesetzen genügen, die wir in Kapitel 1 auf axiomatischer Grundlage für zufällige Ereignisse hergeleitet haben.

Einige interessante Bemerkungen sind hier angebracht, die sich um die zentrale Frage "Was ist eine zufällige Zahlenfolge?" ranken. Eine Antwort hierauf können wir nicht aus den in Kapitel 1 entwickelten Methoden der Wahrscheinlichkeitstheorie oder mathematischen Statistik erwarten, die dieses Problem durch ihren mengentheoretischen Zugang vermeiden.

Wir wollen uns auf den einfachsten Fall von Zahlenfolgen zurückziehen, die 0–1–Folgen $\mathcal{X} = \{\{x_n\}_{n\in\mathbb{N}} \mid x_i \in \{0,1\},\ i \in \mathbb{N}\}$. Die Menge $\mathcal{X}$ ist überabzählbar, da sich jedes Element in natürlicher Weise mit der Binärdarstellung einer reellen Zahl im Intervall $[0,1]$ identifizieren läßt.

$\mathcal{X}$ ist das abzählbare kartesische Produkt der Menge $\{0,1\}$, also $\mathcal{X} = \{0,1\}^n$. Als σ–Algebra wählen wir die Produkt–σ–Algebra $\mathcal{A}' = \bigotimes_{i=1}^{\infty} \mathfrak{P}(\{0,1\})$ und als Wahrscheinlichkeitsmaß hierauf das Produktmaß $\mu = \bigotimes_{i=1}^{\infty} \mathfrak{B}(1,p)$, $0 < p < 1$. Diese Konstruktion haben wir in Kapitel 1.4 vor Definition 1.4.6 als geeignet zur Beschreibung der unendlichen, unabhängigen Wiederholung des Münzwurfs mit Trefferwahrscheinlichkeit p erkannt. Meßbare Mengen sind hierdurch als Elemente der Produkt–σ–Algebra definiert. Auch die einelementigen Mengen $\{\{x_n\}_{n\in\mathbb{N}}\}$, $\{x_n\}_{n\in\mathbb{N}} \in \mathcal{X}$, gehören zur Produkt–$\sigma$–Algebra und besitzen jede für sich die Wahrscheinlichkeit 0.

Dieses Modell kümmert sich nicht darum, ob Folgen so aussehen, als seien

sie durch unendlichen Münzwurf entstanden, *jede* Folge als einelementige Menge ist zufällig. Will man das Modell anschaulich begreifen, so muß man sich vor die Ausführung des unendlichen Münzwurfs stellen. Für jede nicht allzu irreguläre Menge — solche nämlich, die in der σ–Algebra liegen — weiß man, mit welcher Wahrscheinlichkeit das spätere Ergebnis hier hinein fällt. Wird die Menge im Verhältnis zum riesigen Raum aller möglichen Ergebnisse zu klein, so erhält sie die Wahrscheinlichkeit 0. Will man sogar den genauen Ausgang vorhersagen, so hat man mit Sicherheit Unrecht. Es gibt zu viele Ereignisse, als daß man eines hiervon mit positiver Wahrscheinlichkeit richtig vorhersagen könnte.

All dies hilft natürlich nicht, für eine gegebene Folge zu entscheiden, ob sie *zufällig* ist oder nicht. Jede einzelne Folge hat, wie bemerkt, gleiche Wahrscheinlichkeit 0, und auch jedes Anfangsstück gleicher Länge n besitzt die identische Wahrscheinlichkeit $1/2^n$, wenn die Trefferwahrscheinlichkeit $p = 1/2$ beträgt. Die beiden in obigem Modell völlig gleichwertigen Folgen

$$\ldots,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,\ldots$$
$$\text{und } \ldots,1,1,0,1,1,0,0,1,0,0,0,0,1,0,1,0,0,1,1,1,\ldots$$

begreifen wir bezüglich ihrer Entstehung dennoch als sehr unterschiedlich, obwohl nur ein Ausschnitt der Länge 20 zu sehen ist. Das (scheinbare) Bildungsgesetz der ersten Folge ist eben leicht zu erklären: sie besteht nur aus Nullen. Allerdings besitzen fast alle Folgen lokal regelmäßige Struktur. Das folgende Beispiel macht dies klar.

Die Menge B der 0–1–Folgen, bei denen unendlich oft ein fest vorgegebenes Teilstück $x_1^*,\ldots,x_k^*$ der Länge k (zum Beispiel $k = 10^6$ Nullen hintereinander) vorkommt, besitzt Wahrscheinlichkeit 1. Dies sieht man mit Hilfe des Borel–Cantelli–Lemmas ein. Wir bedienen uns der bequemeren Schreibweise mit Zufallsvariablen und betrachten eine Folge von stochastisch unabhängigen, je $\mathfrak{B}(1,\frac{1}{2})$–verteilten Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$, also

$$\{X_n\}_{n\in\mathbb{N}} : (\Omega,\mathcal{A},P) \to (\{0,1\}^{\mathbb{N}},\mathcal{A}',\mu). \tag{6.1}$$

Die Ereignisse $A_n = \{X_{(n-k)k+1} = x_1^*,\ldots,X_{nk} = x_k^*\}$, $n \in \mathbb{N}$, sind für jedes $k \in \mathbb{N}$ stochastisch unabhängig, und es gilt $P(A_n) = 1/2^k$ für alle $n \in \mathbb{N}$, also $\sum_{n=1}^{\infty} P(A_n) = \infty$. Mit Satz 1.1.3 folgt, daß $P(B) \geq P(\limsup_{n\to\infty} A_n) = 1$.

Dies ist auf den ersten Blick erstaunlich: Fast alle 0–1–Folgen haben unter obiger Gleichverteilung unendlich viele Teilstücke von 10^6 aufeinanderfolgenden Nullen. Wenn man allerdings ein solches Teilstück als Repräsentanten für Münzwürfe anbietet, werden einem sicher Zweifel an der Zufälligkeit entgegengebracht.

Wie man sieht, ist es im bisherigen axiomatischen Konzept der Wahrscheinlichkeitstheorie unmöglich zu beurteilen, ob eine gegebene Folge zufällig — zum Beispiel durch Münzwurf — entstanden ist oder durch ein determiniertes Verfahren. Um als zufällig zu gelten, muß sie genügend unregelmäßig sein. Dies können wir an Eigenschaften beurteilen, die wir für Zufallsfolgen in dem bisherigen Modell hergeleitet haben. Eine hiervon ist Häufigkeitsstabilität.

Eine 0–1–Folge $\{x_n\}_{n\in\mathbb{N}}$ heißt häufigkeitsstabil, wenn die relative Häufigkeit der Einsen bis zur n–ten Stelle mit $n \to \infty$ gegen p konvergiert, wenn also

$\lim_{n\to\infty}(\nu_n/n) = \frac{1}{2}$, wobei $\nu_n = \sum_{i=1}^{n} \mathbf{1}_{\{1\}}(x_i)$ die Anzahl der Einsen unter den ersten n Stellen ist. Diese Forderung allein reicht aber zur Definition von zufälligen 0–1–Folgen nicht aus. Die Folge $x_n = \frac{1}{2}\left((-1)^n + 1\right)$, $n \in \mathbb{N}$, ist häufigkeitsstabil mit $p = 1/2$ aber bei weitem nicht zufällig.

Nun besitzt jede unendliche Teilfolge einer Folge von unabhängig wiederholten Münzwürfen die gleiche Verteilung wie die ursprüngliche Folge. Dies fordern wir auch für die Häufigkeitsstabilität einer gegebenen Folge.

Definition 6.0.1. *Eine 0–1–Folge heißt zufällig, wenn jede zulässige, unendliche Teilfolge häufigkeitsstabil mit dem gleichen p ist.*

Die Crux in obiger Definition ist die Spezifikation zulässiger Teilfolgen. Wenn wir jede Teilfolge zulassen, gibt es keine zufälligen 0–1–Folgen mehr. Jede zufällige Folge muß, um häufigkeitsstabil mit $0 < p < 1$ zu sein, unendliche viele Nullen und Einsen enthalten. Die Regel: "Wähle die nächste 0." führt immer zur konstanten Nullfolge als Teilfolge, die selbst nicht p–häufigkeitsstabil ist. Es bleibt, geeignete Auswahlregeln zu definieren, die bei der Auswahl des n–ten Glieds x_n nicht von x_n abhängen, sondern a–priori die in die Teilfolge aufzunehmenden Elemente festlegen.

Bezeichne hierzu $\mathcal{X}^* = \bigcup_{\ell=0}^{\infty}\{0,1\}^\ell$ die Menge der endlichen Wörter über dem Alphabet $\{0,1\}$ (alle endlichen 0–1–Ketten), wobei $\{0,1\}^0$ das leere Wort repräsentiert. $\{\varphi_n\}_{n\in\mathbb{N}_0}$ sei eine Folge von 0–1–wertigen Funktionen,

$$\varphi_n : \{0,1\}^n \longrightarrow \{0,1\},\ n \in \mathbb{N},$$

φ_0 ist hierbei eine Konstante.

$\{\varphi_n\}_{n\in\mathbb{N}}$ heißt berechenbar, wenn es für alle $n \in \mathbb{N}$ einen Algorithmus gibt, der den Wert von φ_n bei Input $n \in \mathbb{N}$ und $x_1, \ldots, x_n \in \{0,1\}$ bestimmt. Die Folge $\{\varphi_n\}_{n\in\mathbb{N}}$ liefert eine Auswahlregel durch: "Nehme x_n in die Teilfolge auf, wenn $\varphi_{n-1}(x_1, \ldots, x_{n-1}) = 1$".

Von berechenbaren $\{\varphi_n\}_{n\in\mathbb{N}}$ erzeugte Teilfolgen haben notwendig monoton steigende Indexfolgen. Dies trifft noch nicht ganz unsere Vorstellung von Zufälligkeit. Vertauscht man in (6.1) die Indizes der Zufallsvariablen X_n, so besitzt die so gebildete Folge die gleiche Verteilung $\mu = P^{\{X_n\}}$. Diese Eigenschaft sollte sich in der Häufigkeitsstabilität vorgegebener , zufälliger 0–1–Folgen wiederspiegeln.

In Definition 6.0.1 heißen nun solche Teilfolgen bzw. Auswahlregeln zulässig, die mit Hilfe einer unendlichen, berechenbaren Folge von verschiedenen natürlichen Zahlen $\{i_n\}_{n\in\mathbb{N}}$, $i_n \in \mathbb{N}$, $i_k \neq i_\ell$, falls $k \neq \ell$, zunächst eine Folge $\{x_{i_n}\}_{n\in\mathbb{N}}$ bilden und hieraus unter einer berechenbaren Folge $\{\varphi_n\}_{n\in\mathbb{N}}$ wie oben eine Subteilfolge auswählen.

Definition 6.0.1 legt eine Teilmenge aller 0–1–Folgen fest, die genügend chaotisch sind, um einer unendlichen Serie von unabhängigen Münzwürfen entstammen zu können. Die so definierten Folgen genügen einer Reihe von Grenzwertsätzen, die μ–fast alle Folgen in dem axiomatischen Modell aus Kapitel 1 erfüllen. Für die praktische Überprüfung einer gegebenen Folge auf Zufälligkeit ist Definititon 6.1 jedoch nicht geeignet, da wir nur einen höchstens endlichen Abschnitt dieser Folge untersuchen können.

Für unsere Zwecke ist entscheidend, Zahlenfolgen zu erhalten, die typische Eigenschaften von zufälligen Zahlen unter dem abstrakten Modell aus Kapitel 1 besitzen, die sich also in wesentlichen Punkten so verhalten, als seien sie aus einem

echten Zufallsexperiment entstanden. Auf die Erzeugung solcher Zahlen und die Überprüfung ihrer Eigenschaften werden wir in den folgenden Abschnitten eingehen.

6.1. Erzeugung von Zufallszahlen

Die Basis zur Erzeugung von zufälligen Ergebnissen aus beliebigen Verteilungen und auch stochastischen Prozessen bilden sogenannte *Standardzufallszahlen*. Das sind Zahlen, die durch einen Zufallsmechanismus generiert werden, der mit stochastisch unabhängigen, je $\mathcal{R}([0,1])$–verteilten Zufallsvariablen (rechteckverteilt auf dem Intervall $[0,1]$, vgl. (1.2.21)) modelliert werden kann. Ob Zufallszahlen einer solchen Forderung genügen, kann mit Hilfe statistischer Tests als nicht zutreffend nachgewiesen werden. Überlebt nun eine gegebene Zahlenfolge eine Reihe statistischer Tests auf Erzeugung aus $\mathcal{R}([0,1])$–verteilten, stochastisch unabhängigen Zufallsvariablen, so bleibt sie ein möglicher Kandidat für Standardzufallszahlen, auch wenn sie nach einem bekannten Bildungsgesetz entstanden ist. Wir werden im folgenden alle Kandidaten für Standardzufallszahlen als *Pseudozufallszahlen* bezeichnen.

Eine nach einem bekannten Mechanismus rechnerintern erzeugte Menge von Zufallszahlen kann nach den zuvor gemachten Bemerkungen nicht echt zufällig im Sinn von Definition 6.1 sein. Dies ist aber nicht weiter tragisch. Ein Beobachter, der das Bildungsgesetz nicht kennt und nicht weiß, wie diese Zahlen entstanden sind, wird seinen unsicheren Kenntnisstand in ein stochastisches Modell fassen und überprüfen, ob die Zahlen diesem Modell widersprechen. Ist dies nicht der Fall, wird er die Zufallszahlen als echt akzeptieren und bei diesem Urteil bleiben, bis er durch einen weiteren Test oder tiefergehende Informationen vom Gegenteil überzeugt ist. Dies ist aber genau die Position stochastischer Modelle in der Realität. Sie werden dann benutzt, wenn ein Mechanismus zur Erzeugung von Ergebnissen so komplex ist, daß man ihn nicht genau beschreiben kann oder eine genaue Beschreibung zwar möglich aber unübersichtlich ist und auch keine tiefere Einsicht vermittelt als ein stochastisches Modell. Geeignete Pseudozufallszahlen bilden den Entstehungsprozeß wirklicher Zufallsergebnisse nach, und wir können berechtigt hoffen, mit ihnen ein adäquates Abbild realer Vorgänge nachvollziehen zu können.

Eine weitere Schwierigkeit bei der Erzeugung von Standardzufallszahlen bildet die approximative Darstellung reller Zahlen in Computern mit nur endlich vielen binären Stellen. Dies bedeutet, daß nur eine endliche Menge von rationalen Zahlen rechnerintern verarbeitet werden kann. Das Modell einer $\mathcal{R}([0,1])$–verteilten Zufallsvariablen kann mit Hilfe eines digitalen Rechners also nur diskret approximiert werden.

Als brauchbare Methode zur Erzeugung von Pseudozufallszahlen haben sich die 1948 von D.H. Lehmer vorgeschlagenen linearen Kongruenzgeneratoren herauskristallisiert. Sie sind wegen ihrer Einfachheit noch theoretisch zu überblicken und leicht zu implementieren. Damit lassen sich auftretende Parameter so wählen, daß die erzeugten Zahlenfolgen möglichst Standardzufallszahlen entsprechen.

Definition 6.1.1. *(Lineare Kongruenzmethode)*
Wähle ganze Zahlen $m > 0$ (Modul), $a > 0$ (Faktor), $r \geq 0$ (Inkrement) und $z_0 \geq 0$ (Startwert). Das Verfahren einer rekursiven Erzeugung von Pseudozufallszahlen

$\{u_n\}_{n\in\mathbb{N}_0}$ *durch*

$$z_{n+1} = (az_n + r) \bmod m \quad \text{und} \quad u_n = z_n/m,\ n \in \mathbb{N}_0, \tag{6.1.1}$$

heißt lineare Kongruenzmethode oder linearer Kongruenzgenerator (im Fall $r = 0$*: multiplikative Kongruenzmethode).*

Für ganze Zahlen $u, v \in \mathbb{Z}$ bezeichnet hierbei $u = v \bmod m$ den Rest bei Division von v durch m, also $v = km+u$, $k \in \mathbb{Z}$, $0 \le u \le m-1$ bzw. $u = v - \lfloor \frac{v}{m} \rfloor \cdot m$.

Bei falscher Wahl von m, a, r und z_0 sind die durch (6.1.1) erzeugten Folgen ungeeignet, ein zufälliges Entstehungsgesetz nachzubilden. Man betrachte etwa $m = 21$, $a = 5$, $r = 5$, $z_0 = 3$. Die ersten Glieder lauten dann $3, 18, 3, 18, \ldots$, und die Folge setzt sich periodisch mit $3, 18$ und der Periodenlänge 2 fort. Dieses Beispiel zeigt einen generellen Nachteil von Rekurrenzgeneratoren; sie besitzen stets eine Periode, deren Länge kleiner oder gleich m ist. Dies gilt, da die Werte von z_n in der Menge $\{0, 1, \ldots, m-1\}$ liegen und nach spätestens $m+1$ Iterationen wieder der Startwert z_0 errreicht ist, von dem aus die gleichen Zahlenwerte erzeugt werden.

Als Pseudozufallszahlen kommen nur die Werte der ersten Periode in Frage, alle weitere sind ein Duplikat der vorhergehenden und damit nutzlos. Um möglichst viele Zufallszahlen zur Verfügung zu haben, sollte m von vorneherein sehr groß gewählt werden, möglichst in der Nähe der größten, auf dem zur Verfügung stehenden Rechner darstellbaren Integer–Zahl.

Wir werden im folgenden untersuchen, welche Werte für m, a, r, und z_0 zu brauchbaren Pseudozufallszahlen führen, und unsere Überlegungen zunächst auf die Periodenlänge konzentrieren. Die Periodenlänge eines Kongruenzgenerators ist definiert durch

$$L = L(m, a, r, z_0) = \min\{k \in \mathbb{N} \mid \text{ es existiert ein } n \in \mathbb{N}_0 \text{ mit } z_{n+k} = z_n\}.$$

Da alle Rekurrenzgeneratoren wegen der endlichen Zahlendarstellung in digitalen Rechnern periodische Zahlenfolgen erzeugen, ist L als endliche natürliche Zahl wohldefiniert.

Unser Ziel ist eine Wahl von m, a und r, mit der die volle Periodenlänge erreicht wird. In diesem Fall tritt in jedem Zyklus des Kongruenzgenerators (6.1.1) jede der Zahlen $\{0, \ldots, m-1\}$ genau einmal auf, so daß die Periodenlänge unabhängig vom speziellen Startwert z_0 ist. Der folgende Satz gibt einfache notwendige und hinreichende Bedingungen für maximale Periodenlänge m an. Sein Beweis basiert auf zahlentheoretischen Resultaten, die vorbereitend bewiesen werden, aber auch für sich von Interesse sind. Für natürliche Zahlen u, $v \in \mathbb{N}$ bezeichne im folgenden $\mathrm{GGT}(u, v)$ den größten gemeinsamen Teiler und $\mathrm{KGV}(u, v)$ das kleinste gemeinsame Vielfache von u und v.

Lemma 6.1.1. *Der Modul* m *des linearen Kongruenzgenerators (6.1.1) besitze die Primfaktorzerlegung* $m = p_1^{t_1} \cdots p_k^{t_k}$, k, $t_1, \ldots, t_k \in \mathbb{N}$. *Dann gilt*

$$L(m, a, r, z_0) = \mathrm{KGV}\left(L(p_1^{t_1}, a, r, z_o), \ldots, L(p_k^{t_k}, a, r, z_0)\right). \tag{6.1.2}$$

Beweis. (6.1.2) kann leicht durch vollständige Induktion aus folgendem Sachverhalt gefolgert werden: Ist $m = u \cdot v$ für teilerfremdes $u, v \in \mathbb{N}$ $\left(\mathrm{GGT}(u, v) = 1\right)$, so gilt

$$L(m, a, r, z_0) = \mathrm{KGV}\left(L(u, a, r, z_0), L(v, a, r, z_0)\right) \tag{6.1.3}$$

Wir kürzen im folgenden die Periodenlängen mit $L(m)$, $L(u)$ bzw. $L(v)$ ab. Zum Beweis von (6.1.3) bezeichne $z_n^{(m)}$, $z_n^{(u)}$, bzw. $z_n^{(v)}$, $n \in \mathbb{N}$, die durch die Kongruenzgeneratoren mit Moduln m, u, bzw. v und den gemeinsamen Werten a, r, z_0 erzeugten Zahlen. Dann gilt für alle $n \in \mathbb{N}_0$

$$z_n^{(u)} = z_n^{(m)} \bmod u \quad \text{und} \quad z_n^{(v)} = z_n^{(m)} \bmod v. \tag{6.1.4}$$

Für $n = 0$ ist (6.1.4) wegen der identischen Startwerte trivialerweise erfüllt. Die obigen Identitäten folgen für beliebiges n mit vollständiger Induktion. Wir führen dies für die erste Aussage durch. Es gilt

$$z_{n+1}^{(m)} = \left(az_n^{(m)} + r\right) \bmod m = az_n^{(m)} + r - \ell m \text{ für ein } \ell \in \mathbb{N}.$$

u ist Teiler von m, so daß mit der Induktionsvoraussetzung

$$\begin{aligned} z_{n+1}^{(m)} &= \left(az_n^{(m)} + r\right) \bmod u = \left(a(z_n^{(m)} \bmod u) + r \bmod u\right) \bmod u \\ &= \left(az_n^{(u)} + r\right) \bmod u = z_{n+1}^{(u)} \bmod u \end{aligned}$$

folgt. Analog erhält man den zweiten Teil von (6.1.4).
Für alle $\ell, n \in \mathbb{N}$ folgt mit (6.1.4) und Übungsaufgabe 6.1

$$z_n^{(m)} = z_\ell^{(m)} \text{ genau dann, wenn } z_n^{(u)} = z_\ell^{(u)} \text{ und } z_n^{(v)} = z_\ell^{(v)}. \tag{6.1.5}$$

Ist hiermit $z_n^{(m)} = z_{n+L(m)}^{(m)}$, so gilt $z_n^{(u)} = z_{n+L(m)}^{(u)}$ und $z_n^{(v)} = z_{n+L(m)}^{(v)}$. Also ist $L(m)$ ein Vielfaches von $L(u)$ und $L(v)$, folglich $L(m) \geq \mathrm{KGV}\left(L(u), L(v)\right) = L'$. Andererseits gilt $z_n^{(u)} = z_{n+L'}^{(u)}$ und $z_n^{(v)} = z_{n+L'}^{(v)}$, also mit (6.1.5) $z_n^{(m)} = z_{n+L'}^{(m)}$. Damit ist $L' \geq L(m)$ gezeigt und insgesamt $L' = L(m)$. ∎

Für $u, v \in \mathbb{Z}$ und $m \in \mathbb{N}$ notieren wir im folgenden mit $u \equiv v$ (modulo m), daß u und v bei Division durch m den gleichen Rest besitzen, falls also $u \bmod m = v \bmod m$.

Lemma 6.1.2. *p sei eine Primzahl und $t \in \mathbb{N}$ so, daß $p^t > 2$. Ist dann für $x \in \mathbb{Z}$*

$$x \equiv 1 \text{ (modulo } p^t) \quad \textit{und} \quad x \not\equiv 1 \text{ (modulo } p^{t+1}),$$

so gilt

$$x^p \equiv 1 \text{ (modulo } p^{t+1}) \quad \textit{und} \quad x^p \not\equiv 1 \text{ (modulo } p^{t+2}).$$

Beweis. x besitzt die Darstellung $x = 1 + qp^t$ für eine Zahl $q \in \mathbb{N}$, die nicht Vielfaches von p ist. Es gilt

$$\begin{aligned} x^p = (1 + qp^t)^p &= \sum_{i=0}^{p} \binom{p}{i} q^i p^{ti} = 1 + qp^{t+1} + \sum_{i=2}^{p} \binom{p}{i} q^i p^{ti} \\ &= 1 + qp^{t+1}\left(1 + \sum_{i=2}^{p} \frac{1}{p}\binom{p}{i} q^{i-1} p^{t(i-1)}\right). \end{aligned}$$

Alle Terme unter obigem Summenzeichen sind Vielfache von p, da bei den Binomialkoeffizienten $\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{i!}$ für $2 \le i < p$ die Primzahl $p > i$ nur im Zähler auftritt und für $i = p$ nach Voraussetzung $t(p-1) > 1$ gilt. Ansonsten wäre $t = 1$ und $p = 2$, also $p^t = 2$.
x^p läßt sich daher in der Form $x^p = 1 + qp^{t+1}(1+u)$ mit $u \bmod p = 0$ schreiben. Hieraus folgt unmittelbar $x^p \equiv 1$ (modulo p^{t+1}). Auch gilt $x^p \not\equiv 1$ (modulo p^{t+2}), da q und $1+u$ beides keine Vielfachen von p sind. ∎

Satz 6.1.1. *Der lineare Kongruenzgenerator (6.1.1) mit $r > 0$ besitzt die volle Periodenlänge m genau dann, wenn*

$$\begin{aligned} &\text{(i)} && \mathrm{GGT}(r,m) = 1, \\ &\text{(ii)} && a \bmod p = 1 \text{ für alle Primfaktoren } p \text{ von } m \text{ und} \\ &\text{(iii)} && a \bmod 4 = 1, \text{ falls } m \text{ Vielfaches von 4 ist.} \end{aligned} \tag{6.1.6}$$

Beweis. Die Primfaktorzerlegung von m sei $m = p_1^{t_1} \cdots p_k^{t_k}$. Dann gilt mit Lemma 6.1.1

$$L(m,a,r,z_0) = \mathrm{KGV}\left(L(p_1^{t_1},a,r,z_0),\ldots,L(p_k^{t_k},a,r,z_0)\right) \le p_1^{t_1}\cdots p_k^{t_k} = m$$

mit Gleichheit genau dann, wenn $L(p_i^{t_i},a,r,z_0) = p_i^{t_i}$ für alle $i = 1,\ldots,k$. Damit reicht es aus, die Bedingungen (i) – (iii) im Fall einer Primzahlpotenz $m = p^t$ als notwendig und hinreichend nachzuweisen, $p \in \mathbb{N}$ eine Primzahl, $t \in \mathbb{N}$.
Ist $a = 1$, so sind (ii) und (iii) stets erfüllt. (6.1.1) reduziert sich in diesem Fall für alle $n \in \mathbb{N}$ auf

$$z_n = (z_0 + nr) \bmod m = \left(z_0 \bmod m + (nr) \bmod m\right) \bmod m. \tag{6.1.7}$$

Es gilt $(nr) \bmod m = nr - qm$ für ein $q \in \mathbb{N}_0$, so daß $(nr) \bmod m$ Vielfaches von $\mathrm{GGT}(r,m)$ ist. Die Darstellung (6.1.7) zeigt, daß die volle Periodenlänge m genau dann erreicht wird, wenn $\mathrm{GGT}(r,m) = 1$. Damit ist der Fall $a = 1$ bewiesen, und wir nehmen im folgenden an, daß $a > 1$.

Zunächst wird gezeigt, daß die Bedingungen (i)–(iii) notwendig sind. Hat ein Kongruenzgenerator die Periodenlänge m, so kommen alle Reste aus der Menge $\{0,1,\ldots,m-1\}$ in jeder Periode genau einmal vor. Insbesondere muß hierbei die 0 auftreten, die man ohne Einschränkung der Allgemeinheit als Startwert z_0 wählen kann. Dann gilt

$$\begin{aligned} z_n &= \left(a^n z_0 + (a^{n-1}r + \cdots + ar + r)\right) \bmod m \\ &= \left(\frac{a^n - 1}{a-1}\, r\right) \bmod m = \left(\frac{a^n-1}{a-1}\, r\right) \bmod p^t, \end{aligned} \tag{6.1.8}$$

und auch hier werden nur Vielfache von $\mathrm{GGT}(r,m)$ erzeugt. Wie oben folgt bei maximaler Periodenlänge m die Bedingung (i). Ferner erhalten wir wegen $z_m = z_0 = 0$ aus $r(a^{p^t}-1)/(a-1) \equiv 0$ (modulo p^t) und $\mathrm{GGT}(r,m) = 1$, daß

$$\frac{a^{p^t}-1}{a-1} \equiv 0 \text{ (modulo } p^t). \tag{6.1.9}$$

Angenommen (ii) gilt nicht, d.h. $a \not\equiv 1$ (modulo p). In diesem Fall ist (6.1.9) äquivalent mit $(a^{p^t} - 1) \bmod p^t = 0$, so daß $(a^{p^t} - 1) \bmod p = 0$, also $a^{p^t} \equiv 1$ (modulo p) folgt.
Aus dem Satz von Fermat (vgl. Übungsaufgabe 6.3) folgt $a^p \equiv a$ (modulo p), also $a^{p^2} = (a^p)^p \equiv a^p \equiv a$ (modulo p) und allgemein mit Induktion $a^{p^t} \equiv a$ (modulo p). Insgesamt folgt der Widerspruch $(a-1) \equiv 0$ (modulo p) und damit die Gültigkeit von (ii).
Wenn $m = p^t$ Vielfaches von 4 ist, haben wir den Fall $p = 2$, $t \geq 2$. Gilt dann $a \not\equiv 1$ (modulo 4), so folgt

$$\frac{a^{2^{t-1}} - 1}{a - 1} \equiv 0 \text{ (modulo } 2^t). \tag{6.1.10}$$

Dies sieht man folgendermaßen ein. Nach dem gerade Bewiesenen muß $a \bmod 2 = 1$, also nach obiger Annahme $a \bmod 4 = 3$ gelten.
Da $a^2 \equiv 1$ (modulo 2^3) und $a^2 \equiv 4$ (modulo 2^4), schließen wir mit Lemma 6.1.2, daß $(a^2)^2 = a^4 \equiv 1$ (modulo 2^4) und $a^4 \not\equiv 1$ (modulo 2^5), sowie durch iterierte Anwendung dieses Arguments für $t \geq 2$

$$a^{2^{t-1}} \equiv 1 \text{ (modulo } 2^{t+1}).$$

Wegen $a - 1 \equiv 2$ (modulo 4) haben wir die Darstellung $a - 1 = 4\ell + 2 = 2(2\ell + 1)$ für ein $\ell \in \mathbb{N}$. $2\ell + 1$ enthält nicht den Primfaktor 2. Daher ist 2^t Teiler der natürlichen Zahl $(a^{2^t} - 1)/(a - 1)$, woraus (6.1.10) folgt.
(6.1.10) liefert, eingesetzt in (6.1.8), $z_{2^{t-1}} = 0 = z_0$, also eine Periodenlänge $L \leq 2^{t-1}$, im Widerspruch zur Voraussetzung $L = 2^t$. Also gilt Bedingung (iii).
Wir beweisen nun die umgekehrte Richtung und nehmen die Gültigkeit der Bedingungen (i)–(iii) an.
Der Fall $p = 2$, $t = 1$ (also $m = p^t = 2$) liefert die volle Periodenlänge. Wegen (i) und (ii) sind nämlich a und m beide ungerade, so daß in (6.1.1) $z_{n+1} = r \bmod 2 = 1$, falls $z_n = 0$, und $z_{n+1} = (a + r) \bmod 2 = 0$, falls $z_n = 1$, für alle $n \in \mathbb{N}_0$, also eine alternierende Folge aus 0 und 1 mit Periodenlänge 2 erzeugt wird.
Sei also $p^t > 2$, d.h. $p > 2$ oder $m = 2^t$ mit $t \geq 2$. Ist $p > 2$, so folgt aus $a > 1$ und $a \bmod p = 1$, daß $a > p$. Ist $p = 2$, so teilt 4 den Modul m, und wegen $a \bmod 4 = 1$ gilt $a > 4$. In jedem Fall existieren $k, q \in \mathbb{N}$, $\mathrm{GGT}(q, p) = 1$, $p^k > 2$ mit $a = 1 + qp^k$, so daß

$$a \equiv 1 \text{ (modulo } p^k) \quad \text{und} \quad a \not\equiv 1 \text{ (modulo } p^{k+1}).$$

Durch iterierte Anwendung von Lemma 6.1.2 folgt $a^{p^\ell} \equiv 1$ (modulo $p^{k+\ell}$) und $a^{p^\ell} \not\equiv 1$ (modulo $p^{k+\ell+1}$) für alle $\ell \in \mathbb{N}$, also $a^{p^\ell} = 1 + q'p^{k+\ell}$ für ein mit p teilerfremdes $q' \in \mathbb{N}$. Obige Darstellungen von a und a^{p^ℓ} liefern für alle $\ell \in \mathbb{N}$

$$\frac{a^{p^\ell} - 1}{a - 1} \equiv 0 \text{ (modulo } p^\ell) \quad \text{und} \quad \frac{a^{p^\ell} - 1}{a - 1} \not\equiv 0 \text{ (modulo } p^{\ell+1}). \tag{6.1.11}$$

Insbesondere gilt $(a^{p^t} - 1)/(a - 1) \equiv 0$ (modulo p^t). Wegen (6.1.8) ist damit p^t ein Vielfaches der Periodenlänge $L(p^t, a, r, 0)$.

Angenommen p^{t-1} ist ebenfalls Vielfaches der Periodenlänge $L(p^t, a, r, 0)$. Dann gilt $\left(r\,(a^{p^{t-1}} - 1)/(a-1)\right) \equiv 0 \pmod{p^t}$, und wegen $\mathrm{GGT}(r, p^t) = 1$ folgt

$$\frac{a^{p^{t-1}} - 1}{a - 1} \equiv 0 \pmod{p^t},$$

im Widerspruch zur zweiten Aussage in (6.1.11) mit $\ell = t - 1$.
Insgesamt folgt $L(p^t, a, r, 0) = p^t = m$ und damit die Behauptung auch für beliebigen Startwert z_0. ∎

Die Bedingungen (i),(ii) und (iii) aus (6.1.6) beschreiben vollständig, wann die Periodenlänge eines linearen Kongruenzgenerators mit positivem Inkrement maximal wird, eine Approximation an die stetige Gleichverteilung auf $[0,1]$ also bestmöglichst gelingt.

Ein wichtiger Fall für die Realisation auf Computern liegt vor, wenn $m = 2^t$ die größte darstellbare Integerzahl ist, $t \in \mathbb{N}$. Die Berechnung der Reste $(az_n + r) \bmod m$ fällt dann automatisch ab, da das Register `y ← a*z+r` bei Overflow gerade den Rest modulo m enthält.

Für $m = 2^t$, $t \geq 2$, bleiben immer noch große Freiheiten in der Wahl von a und r. Ist r ungerade und $a = 4\ell + 1$ für ein $\ell \in \mathbb{N}$, so sind (i)–(iii) aus Satz 6.1.1 erfüllt und der zugehörige Kongruenzgenerator hat maximale Periodenlänge. Das gleiche Ergebnis liefert im Fall $m = 10^t$, $t \geq 2$, die Wahl "r nicht durch 2 oder 5 teilbar" sowie $a = 20\ell + 1$ für ein $\ell \in \mathbb{N}$.

Wir behandeln jetzt die multiplikative Kongruenzmethode

$$z_{n+1} = (az_n) \bmod m \quad \text{und} \quad u_n = z_n/m, \quad n \in \mathbb{N}_0, \tag{6.1.12}$$

die aus (6.1.1) mit Inkrement $r = 0$ entsteht. $a \equiv 0 \pmod{m}$ wird im folgenden ausgeschlossen. Der Wert 0 würde sich in diesem Fall bei jedem Iterationsschritt in (6.1.12) reproduzieren. Die maximale Periodenlänge kann für multiplikative Kongruenzgeneratoren also höchstens $m - 1$ betragen. Im folgenden wird für den Startwert $0 \leq z_0 \leq m - 1$ vorausgesetzt.

Eine geschlossenen Darstellung der z_n in (6.1.12) lautet für alle $n \in \mathbb{N}$

$$z_n = (a^n z_0) \bmod m. \tag{6.1.13}$$

Zur Bestimmung der Periodenlänge spielt offensichtlich das kleinste $n \in \mathbb{N}$ mit $a^n \bmod m = 1$ eine wichtige Rolle. Wir bemerken zunächst

Lemma 6.1.3. *Gilt* $\mathrm{GGT}(a, m) = 1$, *so existiert* $k \in \mathbb{N}$ *mit* $a^k \bmod m = 1$.

Beweis. Die Folge $\left\{(a^n \bmod m)\right\}_{n \in \mathbb{N}}$ ist periodisch, da ihre Glieder nur Werte in $\{0, 1, \ldots, m-1\}$ annehmen können. Also existieren $n, k \in \mathbb{N}$ mit $a^{n+k} \equiv a^n \pmod{m}$. Damit gilt $a^{n+k} - a^n \equiv a^n(a^k - 1) \equiv 0 \pmod{m}$. Nach Voraussetzung sind a^n und m teilerfremd für alle $n \in \mathbb{N}$, so daß $a^k - 1 \equiv 0 \pmod{m}$. Also existiert $k \in \mathbb{N}$ mit $a^k \equiv 1 \pmod{m}$. ∎

Lemma 6.1.4. *Für* $a, t \in \mathbb{N}$, $t \geq 3$ *und* $a > 1$ *ungerade, gilt*

$$L = \min\{n \in \mathbb{N} \mid a^n \bmod 2^t = 1\} \leq 2^{t-2} \tag{6.1.14}$$

mit Gleichheit genau dann, wenn $a \bmod 8 \in \{3, 5\}$.

Beweis. Für ungerades a ist $a+1$ oder $a-1$ durch 4 teilbar. $a \pm 1$ hat also eine eindeutige Darstellung $a \pm 1 = (2r+1)2^k$, $k \geq 2$, $r \in \mathbb{N}$, also $a \pm 1 = r2^{k+1} + 2^k$, d.h.

$$a \pm 1 \equiv 2^k \text{ (modulo } 2^{k+1}). \tag{6.1.15}$$

"$\pm$" bedeutet hierbei, daß die Aussage für beide Vorzeichen $+$ und $-$ gilt.
Sei nun $1 < a < 2^t - 1$. Aus der Darstellung (6.1.15) folgt $k < t$. Es gilt

$$\pm a \equiv 1 \text{ (modulo } 2^k) \quad \text{und} \quad \pm a \not\equiv 1 \text{ (modulo } 2^{k+1}),$$

und durch iterierte Anwendung von Lemma 6.1.2 erhalten wir

$$(\pm a)^{2^{t-k-1}} \equiv a^{2^{t-k+1}} \not\equiv 1 \text{ (modulo } 2^{t-1}) \quad \text{und}$$
$$(\pm a)^{2^{t-k}} \equiv a^{2^{t-k}} \equiv 1 \text{ (modulo } 2^t).$$

L ist also Teiler von 2^{t-k}, teilt aber nicht 2^{t-k-1}, woraus $L = 2^{t-k}$ mit $k \geq 2$ folgt. Für $a = 2^t - 1$ gilt $L = 2 \leq 2^{t-2}$, da $a^2 \bmod 2^t = 1$. Insgesamt folgt die Ungleichung (6.1.14). Gleichheit gilt hierin genau dann, wenn $k = 2$, was wegen (6.1.15) äquivalent ist zu $a \equiv 4 \pm 1$ (modulo 8). ∎

Wir behandeln im weiteren den für die Praxis wichtigen Fall $m = 2^t$, $t \in \mathbb{N}$.

Satz 6.1.2. *Seien $m = 2^t$, $t \geq 3$, $1 \leq z_0 \leq 2^t - 1$ ungerade und $a \bmod 8 \in \{3,5\}$. Dann besitzt der multiplikative Kongruenzgenerator (6.1.12) die Periode 2^{t-2}. Diese ist maximal für alle Startwerte $z_0 \in \{0,1,\ldots,m-1\}$ und Faktoren $a \in \mathbb{N}$.*

Beweis. Für jeden Startwert $z_0 \in \{1,\ldots,m-1\}$ und ungerades a wird wegen Lemma 6.1.3 die Periodenlänge des multiplikativen Kongruenzgenerators bestimmt durch

$$L(m,a,0,z_0) = \min\{n \in \mathbb{N} \mid z_0 = a^n z_0 \text{ (modulo } 2^t)\}.$$

$a^i z_0 \bmod 2^t = a^j z_0 \bmod 2^t$ für zwei Indizes $i,j \in \mathbb{N}$ ist äquivalent mit $z_0(a^i - a^j) \equiv 0$ (modulo 2^t), also $a^i - a^j \equiv 0$ (modulo 2^t), d.h. $a^i \bmod 2^t = a^j \bmod 2^t$, da z_0 ungerade. Die Folge $\{z_n\}_{n\in\mathbb{N}}$ besitzt daher die gleiche Periodenlänge wie die Folge $\{(a^n \bmod 2^t)\}_{n\in\mathbb{N}}$, und es gilt

$$L(m,a,0,z_0) = \min\{n \in \mathbb{N} \mid a^n \equiv 1 \text{ (modulo } 2^t)\}.$$

Dieses Minimum wird nach Lemma 6.1.3 als endliche natürliche Zahl angenommen. Lemma 6.1.4 besagt, daß $L(m,a,0,z_0) \leq 2^{t-2}$ mit Gleichheit genau dann, wenn a den angegebenen Bedingungen genügt.
Enthält z_0 den Primfaktor 2, etwa $z_0 = q2^s$ für $s \in \mathbb{N}$ und q ungerade, so ist $a^n q\, 2^s \bmod 2^t = q2^s$ für ein $n \in \mathbb{N}$ äquivalent zu $a^n q \bmod 2^{t-s} = q$. In diesem Fall ist wie gerade bewiesen mit $z_0 = q$ und $m = 2^{t-s}$ die Periodenlänge kleiner oder gleich 2^{t-s-2}.
Für gerades a mit der Darstellung $a = q\,2^s$ beträgt die Periodenlänge 1. Denn für alle $n \in \mathbb{N}$ mit $n > \lfloor \frac{t}{s} \rfloor$ gilt, daß $a^n = q^n 2^{sn} = 0 \bmod 2^t$ also $z_n = 0$. Insgesamt folgt die behauptete Maximalität. ∎

Auch bei den multiplikativen Kongruenzgeneratoren mit Modul $m = 2^t$ bleiben noch große Freiheiten bei der Wahl des Faktors a, um die größtmögliche Periodenlänge 2^{t-2} und damit eine gute Approximation an die stetige Gleichverteilung auf dem Intervall $[0,1]$ zu erreichen.

Für Spezialfälle läßt sich die Menge der mit der multiplikativen Kongruenzmethode erzeugten Zahlen vollständig überblicken.

Lemma 6.1.5. *Seien* $t \geq 3$, $m = 2^t$ *und* $a \bmod 8 = 5$.

a) *Im Fall* $z_0 \bmod 4 = 1$ *werden durch den multiplikativen Kongruenzgenerator* $z_{n+1} = (az_n) \bmod m$ *alle Zahlen der Form* $4k+1$, $0 \leq k \leq 2^{t-2}-1$, *erzeugt.*

b) *Gilt* $z_0 \bmod 4 = 3$, *so werden alle Zahlen der Form* $4k+3$, $0 \leq k \leq 2^{t-2}-1$, *erzeugt.*

Beweis. Die Voraussetzungen von Satz 6.1.2 sind erfüllt, so daß die Periodenlänge 2^{t-2} beträgt, also mindestens 2^{t-2} verschiedene Zahlen erzeugt werden. Es bleibt zu zeigen, daß $z_n \bmod 4 = 1$ für alle $n \in \mathbb{N}$, falls $z_0 \bmod 4 = 1$. Wir weisen dies mit vollständiger Induktion nach. Der Induktionsanfang $n = 0$ ist klar.
z_n besitze nun eine Darstellung $z_n = 4k_n + 1$, $k_n \in \mathbb{N}_0$. Nach Voraussetzung existiert $q \in \mathbb{N}_0$ mit $a = 8q + 5$. Hieraus folgt

$$az_n = (8q+5)(4k_n+1) = 4(8qk_n + 5k_n + 2q + 1) + 1 = 4\ell_n + 1,$$

wobei $\ell_n = 8qk_n + 5k_n + 2q + 1 \in \mathbb{N}$. Mit $\alpha_n = \lfloor (4\ell_n+1)/2^t \rfloor$ gilt weiter $4\ell_n + 1 - \alpha_n 2^t = 4(\ell_n - \alpha_n 2^{t-2}) + 1 \equiv 1$ (modulo 4), so daß

$$z_{n+1} = (az_n) \bmod 2^t = 4\ell_n + 1 - \alpha_n 2^t \equiv 1 \text{ (modulo 4)}.$$

Die Argumentation für den Fall $z_0 \bmod 4 = 3$ verläuft analog. ∎

In der Praxis stellt sich oft die Aufgabe, auch höherdimensionale zufällige Ereignisse durch Simulation mit entsprechenden Zufallszahlen nachzuvollziehen. Eine einfache Methode, aus Pseudozufallszahlen auf $[0,1]$ für eine vorgegebene Dimension $p \geq 2$ solche auf $[0,1]^p$ zu gewinnen, ist, p-Tupel aus aufeinanderfolgenden Zufallszahlen zu bilden und diese als Vektoren in $\mathbb{R}^p$ zu interpretieren. Ist also $\{u_n\}_{n \in \mathbb{N}_0}$ eine Folge von Zufallszahlen, $0 \leq u_n < 1$, so definiere eine Folge $\{\boldsymbol{u}_m\}_{m \in \mathbb{N}_0}$, $\boldsymbol{u}_m \in \mathbb{R}^p$, von p-dimensionalen Zufallszahlen durch

$$\boldsymbol{u}_m = (u_m, u_{m+1}, \ldots, u_{m+p-1})', \quad m \in \mathbb{N}_0. \tag{6.1.16}$$

Für den Fall des multiplikativen Kongruenzgenerators (6.1.12), der den Bedingungen

$$z_{n+1} = az_n \bmod 2^t,\ u_n = z_n/2^t,\ t \in \mathbb{N},\ a \bmod 8 = 5 \text{ und } z_0 \bmod 4 = 1 \tag{6.1.17}$$

aus Lemma 6.1.5 genügt, werden wir im folgenden Eigenschaften der hierdurch erzeugten p-dimensionalen Pseudozufallszahlen untersuchen. Es wird sich zeigen, daß die durch (6.1.16) konstruierten Zufallszahlen auf einem Gitter des p-dimensionalen Einheitskubus auf Hyperebenen liegen.

Nach Lemma 6.1.5 a) werden für z_n aus (6.1.17) alle Zahlen der Form $4k+1$, $k = 0, 1, \ldots, 2^{t-2}-1$, erzeugt. Folglich findet man wegen (6.1.13) zu jedem $m \in \mathbb{N}_0$ eine natürliche Zahl ℓ, $0 \le \ell \le 2^{t-2}-1$, mit

$$\boldsymbol{u}_m = \frac{1}{2^t}\big((4\ell+1) \bmod 2^t, ((4\ell+1)a) \bmod 2^t, \ldots, ((4\ell+1)a^{p-1}) \bmod 2^t\big)'. \quad (6.1.18)$$

Mit der folgenden Basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ von $\mathbb{R}^p$

$$\boldsymbol{b}_1 = \frac{1}{2^{t-2}}\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{p-1} \end{pmatrix}, \quad \boldsymbol{b}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{b}_p = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad (6.1.19)$$

läßt sich nun das Gitter beschreiben, auf dem alle erzeugten p–dimensionalen Vektoren liegen.

Lemma 6.1.6. *Wenn Faktor a, Modul m und Startwert z_0 des multiplikativen Kongruenzgenerators (6.1.17) erfüllen, stimmt die Menge der durch (6.1.16) erzeugten p–dimensionalen Pseudozufallsvektoren $\boldsymbol{u}_m$, $m \in \mathbb{N}_0$, überein mit dem verschobenen Gitter*

$$\Big(\frac{1}{4}\boldsymbol{b}_1 + \mathcal{G}\Big) \cap [0,1]^p, \quad \textit{wobei} \quad \mathcal{G} = \Big\{ \sum_{i=1}^{p} q_i \boldsymbol{b}_i \mid q_1, \ldots, q_p \in \mathbb{Z} \Big\}. \quad (6.1.20)$$

Beweis. Jedes $\boldsymbol{u}_m$ besitzt wegen (6.1.18) eine Darstellung

$$\boldsymbol{u}_m = \Big(\frac{1+4\ell}{2^t} - r_1, \frac{1+4\ell}{2^t}a - r_2, \ldots, \frac{1+4\ell}{2^t}a^{p-1} - r_p\Big)' \in [0,1]^p,$$

wobei $r_j = \lfloor (1+4\ell)a^{j-1}/2^t \rfloor \in \mathbb{N}$, $j = 1, \ldots, p$. Offensichtlich gilt $r_1 = 0$, so daß

$$\begin{aligned} \boldsymbol{u}_m &= \frac{1}{2^t}(1, a, \ldots, a^{p-1})' + \frac{\ell}{2^{t-2}}(1, a, \ldots, a^{p-1})' - (r_1, r_2, \ldots, r_p)' \\ &= \frac{1}{4}\boldsymbol{b}_1 + \ell \boldsymbol{b}_1 - \sum_{i=2}^{p} r_i \boldsymbol{b}_i. \end{aligned}$$

Also gilt $\boldsymbol{u}_m \in \mathcal{G}$ für alle $m \in \mathbb{N}_0$. Existieren umgekehrt $q_1, \ldots q_p \in \mathbb{Z}$ mit

$$(x_1, \ldots, x_p)' = \frac{1}{4}\boldsymbol{b}_1 + \sum_{i=1}^{p} q_i \boldsymbol{b}_i = \frac{4q_1+1}{2^t}(1, a, \ldots, a^{p-1})' + (0, q_2, \ldots, q_p)' \in [0,1]^p,$$

so folgt $q_1 \ge 0$ und $(4q_1+1)a^{i-1}/2^t + q_i \in [0,1]$ für alle $i = 2, \ldots, p$. Mit Lemma 6.1.5 a) existieren $u_m, \ldots, u_{m+p-1}$ aus (6.1.17), für die $u_{m+i-1} = x_i$ für alle $i = 1, \ldots, p$ gilt. ∎

Die Gitterpunkte (6.1.20), auf denen alle erzeugten Punkte liegen, lassen sich auch als Schnittpunkte von Hyperebenenscharen darstellen. Wir betrachten hierzu die Gitterbasis $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p\}$ aus (6.1.19). Bezeichnet

$$\mathcal{L}_i = \Big\{ \sum_{\substack{j=1 \\ j \neq i}}^{p} \lambda_j \boldsymbol{b}_j \mid \lambda_j \in \mathbf{R} \Big\}, \quad i = 1, \ldots, p,$$

den durch $\boldsymbol{b}_1 \ldots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \ldots, \boldsymbol{b}_p$ aufgespannten linearen Unterraum in $\mathbf{R}^p$ und $\mathcal{S}_i$ die entsprechende Schar aus mit $q_i \boldsymbol{b}_i$, $q_i \in \mathbf{Z}$, verschobenen, parallelen Hyperebenen, $\mathcal{S}_i = \bigcup_{q \in \mathbf{Z}} (q\boldsymbol{b}_i + \mathcal{L}_i)$, so gilt

$$\mathcal{G} = \mathcal{S}_1 \cap \ldots \cap \mathcal{S}_p \cap [0,1]^p.$$

Allerdings ist diese Darstellung nicht eindeutig. Es gibt weitere Gitterbasen $\bar{\boldsymbol{b}}_1, \ldots, \bar{\boldsymbol{b}}_p$, die das gleiche Gitter (6.1.20) aber eine andere Repräsentation durch Schnittpunkte von Hyperebenenscharen liefern.

Ein Zufallszahlengenerator ist zur Erzeugung p–dimensionaler Vektoren unbrauchbar, wenn eine Schar paralleler Hyperebenen großen Abstands untereinander existiert, die alle Punkte enthält. Dann liegen alle Vektoren auf wenigen Hyperebenen in $[0,1]^p$, eine Eigenschaft, die unserer Forderung nach gleichmäßiger Verteilung in $[0,1]^p$ widerspricht. Ein wichtiger Punkt ist also die Bestimmung des Abstands solcher paralleler Hyperebenen.

Wir beschreiben die $(p-1)$–dimensionalen Unterräume $\mathcal{L}_i$ durch ihre *Normalvektoren* $\boldsymbol{n}_i$ mit den Eigenschaften $\boldsymbol{n}_i \perp \mathcal{L}_i$, d.h. $\boldsymbol{n}_i' \boldsymbol{b}_j = 0$ für alle $j = 1, \ldots, p$, $j \neq i$, und $\boldsymbol{n}_i' \boldsymbol{b}_i = 1$.

Im Fall der Basis (6.1.19) läßt sich sofort nachrechnen, daß

$$\boldsymbol{n}_1 = \begin{pmatrix} 2^{t-2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{n}_2 = \begin{pmatrix} -a \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{n}_3 = \begin{pmatrix} -a^2 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{n}_p = \begin{pmatrix} -a^{p-1} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{6.1.21}$$

die entsprechenden Normalvektoren sind. Jedes $\mathcal{L}_i$ besitzt die Darstellung $\mathcal{L}_i = \{\boldsymbol{y} \in \mathbf{R}^p \mid \boldsymbol{n}_i' \boldsymbol{y} = 0\}$. Ist $\mathcal{H}_i = q\boldsymbol{b}_i + \mathcal{L}_i$ für ein $q \in \mathbf{Z}$ eine parallele Hyperebene, so gilt $(\mathcal{H}_i - q\boldsymbol{b}_i) \perp \boldsymbol{n}_i$, d.h. $(\boldsymbol{y} - q\boldsymbol{b}_i)'\boldsymbol{n}_i = 0$ für alle $y \in \mathcal{H}_i$, also $\boldsymbol{y}'\boldsymbol{n}_i = q$, da $\boldsymbol{b}_i'\boldsymbol{n}_i = 1$.

Zur Bestimmung des Abstands verwenden wir die Cauchy–Schwarz Ungleichung: $|\boldsymbol{a}'\boldsymbol{b}| \leq \|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|$ für alle Vektoren $\boldsymbol{a}, \boldsymbol{b} \in \mathbf{R}^p$, wobei Gleichheit genau dann gilt, wenn $\boldsymbol{a} = \lambda \boldsymbol{b}$ für ein $\lambda \in \mathbf{R}$. $\|\boldsymbol{a}\| = \big(\sum_{i=1}^p a_i^2\big)^{1/2}$, $\boldsymbol{a} = (a_1, \ldots, a_p)' \in \mathbf{R}^p$ bezeichnet hierbei die euklidische Norm in R^p.

Benachbarte Hyperebenen (etwa $\mathcal{L}_i$ und $\boldsymbol{b}_i + \mathcal{L}_i$) haben den Abstand $1/\|\boldsymbol{n}_i\|$, da $1 = \boldsymbol{y}'\boldsymbol{n}_i \leq \|\boldsymbol{y}\| \cdot \|\boldsymbol{n}_i\|$ für alle $\boldsymbol{y} \in (\boldsymbol{b}_i + \mathcal{L}_i)$, also $1/\|\boldsymbol{n}_i\| \leq \|\boldsymbol{y}\|$ mit Gleichheit genau dann, wenn $\boldsymbol{y} = \boldsymbol{n}_i/\|\boldsymbol{n}_i\|^2$.

Bei Gitterbasis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ erhält man also den maximalen Abstand benachbarter, paralleler Hyperebenen aus der Lösung von

$$\max \big\{1/\|\boldsymbol{n}_i\| \mid \boldsymbol{n}_i \text{ ist Normalenvektor zu } \mathcal{L}_i,\ i = 1, \ldots, p\big\}. \tag{6.1.22}$$

Möglicherweise vergrößert sich dieser Abstand, wenn man zu einer anderen Gitterbasis $\bar{b}_1, \ldots, \bar{b}_p$ übergeht. Solche Basiswechsel von Gitterbasen können durch unimodulare Matrizen beschrieben werden. Es läßt sich zeigen, daß der maximale Abstand benachbarter, paralleler Hyperebenen, die alle Punkte des Gitters $\mathcal{G}$ aus (6.1.20) enthalten, $1/d_p$ beträgt, wobei

$$d_p = \min\left\{\|\bar{n}\| \mid \bar{n} = \sum_{i=1}^{p} q_i n_i,\ q_1, \ldots, q_p \in \mathbf{Z},\ \bar{n} \neq \mathbf{0}\right\}. \tag{6.1.23}$$

Dieses Problem entsteht aus (6.1.22), indem man zusätzlich alle Scharen paralleler Hyperebenen $\bar{\mathcal{L}}_i + q\bar{b}_i$ bei der Minimierung zuläßt, die aus möglichen Gitterbasen $\bar{b}_1, \ldots, \bar{b}_p$ für $\mathcal{G}$ entstehen.

Die Minimierungsaufgabe (6.1.23) ist ein diskretes Optimierungsproblem, zu dessen Lösung U. Dieter einen effizienten Suchalgorithmus vorgeschlagen hat (vgl. Afflerbach & Lehn (1986), S. 15ff).

Werden aus einem Pseudozufallszahlengenerator wie in (6.1.16) Vektoren in $[0,1]^p$ zusammengesetzt und entsteht dabei ein Gitter in $[0,1]^p$, so ist ein wichtiges Kriterium für die Beurteilung der Güte, daß der maximale Abstand von parallelen Hyperebenen, die alle Gitterpunkte enthalten, nicht zu groß wird. Wir haben die hierbei auftretenden Probleme an der speziellen Klasse der multiplikativen Kongruenzgeneratoren (6.1.17) vorgeführt. Ähnliche Gitterstrukturen gelten für alle linearen Kongruenzgeneratoren. Eine Übersicht hierüber findet sich in Afflerbach & Lehn (1986).

Eine ungeeignete Pseudozufallszahlenfolge wird durch folgende Rekursion erzeugt.

Beispiel 6.1.1. Wir betrachten den multiplikativen Kongruenzgenerator

$$z_{n+1} = (2^{16} + 3) z_n \bmod 2^{32}, \quad u_n = z_n / 2^{32}. \tag{6.1.24}$$

Es gilt $a = 2^{16} + 3 \bmod 8 = 3$, so daß dieser Kongruenzgenerator nach Satz 6.1.2 für ungeraden Startwert z_0 die maximale Periodenlänge 2^{30} besitzt.
Für die dreidimensionalen Vektoren $\boldsymbol{u}_m = (u_m, u_{m+1}, u_{m+2})$, $m \in \mathbf{N}_0$, gilt $z_{m+2} = (2^{16} + 3)^2 z_m \bmod 32 = (6 \cdot 2^{16} + 9) z_m \bmod 2^{32}$ und daher

$$z_{m+2} - 6z_{m+1} + 9z_m \bmod 2^{32} = \left(6 \cdot 2^{16} + 9 - 6(2^{16} + 3) + 9\right) z_m \bmod 2^{32} = 0.$$

Es folgt $u_{m+2} - 6u_{m+1} + 9u_m = q \in \mathbf{Z}$ mit $-6 < q < 10$ für alle $m \in \mathbf{N}_0$. Alle dreidimensionalen Vektoren aus dem multiplikativen Generator (6.1.24) liegen damit auf höchstens 15 Ebenen in $\mathbf{R}^3$. Zur Simulation von im Einheitswürfel gleichverteilten Zufallsexperimenten ist dieser Generator trotz maximaler Periodenlänge ungeeignet, da starke Korrelationen zwischen je drei aufeinanderfolgenden Pseudozufallszahlen bestehen. ∎

Um Unabhängigkeit von Pseudozufallsvektoren zu erreichen, wird man eine Überlappung der erzeugenden u_n durch mehrfache Verwendung in aufeinanderfolgenden Vektoren wie in (6.1.16) vermeiden und jeden Vektor $\boldsymbol{u}_m \in \mathbf{R}^p$ aus neuen Zufallszahlen zusammensetzen. Die Folge

$$\boldsymbol{u}_m = (u_{mp}, u_{mp+1}, \ldots, u_{mp+p-1}), \quad m \in \mathbf{N}_0, \tag{6.1.25}$$

erfüllt diese Forderung. Ist L die Periodenlänge des zugrundeliegenden Generators, erhält man in dieser Weise allerdings nur $\lfloor L/p \rfloor$ nicht überlappende Vektoren der Dimension p.

Das Problem der Gitterstruktur wird durch (6.1.25) jedoch nicht aufgelöst. Die hierdurch erzeugten Vektoren liegen lediglich auf einer gewissen Teilmenge des vollen Gitters $\mathcal{G}$. Wenn sich dieses nur auf wenige Hyperebenen konzentriert, gilt dieser Nachteil auch für die durch (6.1.25) erzeugten Punkte.

Lineare Kongruenzgeneratoren mit anderen Parametern als den hier speziell untersuchten können mit verwandten Methoden analysiert werden und liefern für geeignete Werte von a, r, und m brauchbare Pseudozufallszahlen. Eine Übersicht über die umfangreiche Theorie mit einigen praktischen Anwendungsfällen findet sich in Afflerbach & Lehn (1986) und Schmitz & Lehmann (1985).

Als brauchbare Verallgemeinerung haben sich gewisse *mehrfach rekursive Kongruenzgeneratoren* erwiesen

$$z_{n+1} = (a_1 z_{n-1} + a_2 z_{n-2} + \cdots + a_k z_{n-k} + r) \bmod m, \; u_n = z_n/m, \; n \geq k \in \mathbb{N},$$

mit $a_1, a_2, \ldots, a_k$, $r \in \mathbb{Z}$ und Startwerten $z_0, \ldots, z_k \in \mathbb{Z}$. Das gleiche gilt für Kombinationen von linearen Kongruenzgeneratoren. Die von solchen Generatoren erzeugten Zahlenfolgen sind natürlich schwerer zu überblicken, was einem intuitiven Verständnis von "Zufälligkeit" entgegenkommt. Man kann nicht mehr relativ leicht vorherbestimmen, welche Zahl als nächste von einem solchen Generator erzeugt wird. Allerdings ergeben sich auch hier für ungeeignete Parameterwerte starke Abweichungen von einer gleichmäßigen Verteilung sowie deutliche Abhängigkeiten. Die erzeugten Zahlenfolgen unterliegen ebenfalls einem deterministischen Bildungsgesetz. Warum soll man also nicht gleich Pseudozufallszahlen verwenden, von denen man eine gute Approximation des Modells stochastisch unabhängiger, im Intervall $[0, 1]$ gleichverteilter Zufallsvariablen theoretisch nachweisen kann, auch wenn das deterministische Bildungsgesetz leicht zu beschreiben und analysieren ist ? Für den Zweck einer Nachbildung zufallsgesteuerter Vorgänge auf Computern bildet das keinen Nachteil.

Zum Abschluß dieses Kapitels werden noch kurz einige andere Verfahren zur Erzeugung von Pseudozufallszahlen vorgestellt, die aber schwerwiegende Mängel aufweisen.

Als unbrauchbar haben sich *Fibonacci-Generatoren* herausgestellt. Man startet hierbei mit zwei Werten $u_0, u_1 \in [0, 1)$. Nachfolgende Zahlen werden durch die Rekursion

$$u_{n+1} = (u_n + u_{n-1}) - \lfloor u_n + u_{n-1} \rfloor, \quad n \in \mathbb{N},$$

erzeugt. Für die so definierte Zahlenfolge $\{u_n\}_{n \in \mathbb{N}_0}$ treten die Permutationen $u_{n-1} < u_{n+1} < u_n$ und $u_n < u_{n+1} < u_{n-1}$ allerdings nie auf (Übungsaufgabe 6.6). Bei stochastisch unabhängigen, gleichverteilten Zufallsvariablen beträgt die Wahrscheinlichkeit für solche Anordnungen jedoch 1/6. Fibonacci-Generatoren widersprechen also unserer Modellvorstellung.

Die *Mid-Square-Methode* arbeitet folgendermaßen. $u_n = (0, b_1, \ldots, b_{2k})$, $k \in \mathbb{N}$ repräsentiere eine $2k$-stellige Zahl in $[0, 1)$ mit Nachkommastellen $b_1, \ldots, b_k$. Besitzt dann u_n^2 die Darstellung $u_n^2 = (0, c_1, \ldots, c_{4k})$, so verwende als nächste Pseudozufallszahl $u_{n+1} = (0, c_{k+1}, \ldots, c_{3k})$. Diese Generatoren weisen erhebliche

Abweichungen von der Gleichverteilungsannahme in $[0,1]$ auf, wie statistische Tests zeigen. Häufig degenerieren sie sehr schnell in Zyklen kurzer Periodenlänge. Solche Generatoren liefern in der Tat keine brauchbaren Pseudozufallszahlen.

Die bereits erwähnten physikalischen Generatoren, bei denen die Zufallszahlen in einem Versuchsaufbau bestimmt werden (etwa durch Messung des Pegels eines Rauschsignals zu genügend weit auseinanderliegenden Zeitpunkten), erfordern in der Regel komplizierte Geräte und sind nicht leicht verfügbar. Eine umfangreiche Analyse, ob solche Zahlen auch wirklich dem erhofften Zufallsgesetz gehorchen, ist auch hier nötig. Die Einflußfaktoren sind nicht leicht einzustellen, so daß eine ständige Kontrolle des erzeugenden Prozesses durchgeführt werden muß.

Als Pseudozufallszahlen mit Werten in $\{0,1,\ldots,9\}$ bieten sich auch Dezimalentwicklungen von irrationalen Zahlen an. Die Entstehungsprozesse aufeinanderfolgender Ziffern sind zur Zeit nicht zu überblicken — insofern bieten sie sich als Zufallszahlen an. Man weiß allerdings sehr wenig über die zugrundeliegenden Verteilungen, was der Verwendung für einen bestimmten Zweck mit vorgeschriebener Verteilung im Wege steht. Die Dezimalentwicklung der Zahl π wurde auf einige Milliarden Stellen durchgeführt; in keinem Test auf "Zufälligkeit" wurden die so erhaltenen Ziffern bisher als "unechte" Zufallszahlen abgelehnt.

6.2. Testen von Zufallszahlen

Nicht jede Wahl von a, r, m und z_0, selbst wenn sie für einen linearen Kongruenzgenerator $z_{n+1} = (az_n + r) \bmod m$ maximale Periodenlänge liefert, ist gleich gut. Maximale Periodenlänge sichert nur eine optimale Approximation der stetigen Gleichverteilung im Intervall $[0,1]$, wenn man den zugehörigen Generator eine volle Periodenlänge durchlaufen läßt. Sie sagt aber nichts über das Verhältnis der Zufallszahlen untereinander aus. Der Generator $u_n = n/m - \lfloor n/m \rfloor$, $n \in \mathbb{N}_0$, hat volle Periodenlänge, die hierdurch erzeugte Zahlenfolge ist aber nicht geeignet, Realisationen von stochastisch unabhängigen, auf $[0,1]$ rechteckverteilten Zufallsvariablen zu simulieren.

Welche weiteren Forderungen sollte man an "gute" Zufallszahlengeneratoren stellen? Wie schon zu Anfang des Kapitels bemerkt, werden wir keinen Einwand gegen die "Echtheit" von Pseudozufallszahlen haben, wenn sie eine Reihe statistischer Tests überleben, die Eigenschaften wirklich zufälliger Zahlen abprüfen.

Einen solchen Test haben wir bereits kennengelernt: der Maximalabstand d_p von parallelen Hyperebenen, die alle aus aufeinanderfolgenden Zahlen zusammengesetzten Pseudozufallsvektoren der Dimension p enthalten. Er berechnet sich als Kehrwert einer Lösung des Minimierungsproblems (6.1.23), in der $\boldsymbol{n}_1,\ldots,\boldsymbol{n}_p$ Normalvektoren der p Hyperebenen sind, die von je $p-1$ Vektoren einer Gitterbasis aufgespannt werden. Die normierten Größen

$$c_p = \frac{\pi^{p/2} {d_p}^p}{(p/2)!\, m}, \quad p \in \mathbb{N}, \tag{6.2.1}$$

geben Auskunft über den maximalen Hyperebenenabstand im Verhältnis zum Modul m. Je größer c_p ist, desto kleiner ist der Abstand $1/d_p$. Für ungerades p gilt hierbei $(\frac{p}{2})! = \frac{p}{2}(\frac{p}{2}-1)\cdots\frac{1}{2}\sqrt{\pi}$.

Umfangreiche empirische Untersuchungen zeigen, daß ein Zufallszahlengenerator akzeptabel ist, wenn die zugehörigen Werte

$$c_2 = \frac{\pi {d_2}^2}{m}, \quad c_3 = \frac{4\pi {d_3}^3}{3m}, \quad \text{und} \quad c_4 = \frac{\pi^2 {d_4}^4}{2m}$$

alle nicht kleiner als 0.1 sind. Sind c_2, c_3 und c_4 alle ≥ 1, so gilt der zugehörige Generator als hervorragend bezüglich einer gleichmäßigen Verteilung in den entsprechenden Dimensionen (vgl. Knuth (1969), Bd. 2, Kapitel 3.3.4).

Für den Generator $z_{n+1} = (2^{16} + 5)z_n \bmod 2^{32}$, $u_n = z_n/2^{32}$, der nach Satz 6.1.2 für ungeraden Startwert z_0 volle Periodenlänge 2^{30} besitzt, gilt

$$c_2 \leq 0.3926, \quad c_3 \leq 0.00001908, \quad \text{und} \quad c_4 \leq 0.1550$$

(vgl. Aufgabe 6.5). Zur Erzeugung von dreidimensionalen Pseudozufallsvektoren ist er nicht akzeptabel und muß damit verworfen werden.

Dieses Verfahren einer Beurteilung von Zufallszahlen entspricht dem *Spektraltest*, dessen theoretische Fundierung und praktische Erprobung im Buch von Knuth (1969), Band 2, ausführlich dargestellt werden. Bemerkenswert ist, daß alle Zufallszahlengeneratoren, die im Laufe der Zeit bei speziellen Anwendungen Mängel gezeigt haben, durch den Spektraltest verworfen werden, obwohl man aufgrund anderer Testverfahren keinen Einwand gegen sie haben konnte.

Weitere Tests auf Zufälligkeit vergleichen, ob Eigenschaften von Pseudozufallszahlen dem unterstellten Modell von stochastisch unabhängigen, je $\mathcal{R}([0,1])$–verteilten Zufallsvariablen widersprechen. Im Vordergrund steht hierbei der Gedanke, daß der Entstehungsprozeß der Pseudozufallszahlen gar nicht so sehr interessiert, sondern vielmehr, ob die durch erzeugten Zahlen sich so verhalten, als sei der Entstehungsprozeß echt zufällig. Von den vielen Verfahren, mit denen Eigenschaften von Zufallszahlen abgeprüft werden können, wollen wir lediglich zwei Typen herausgreifen: 1.) Tests auf Anpassung an die Recheckverteilung und 2.) Tests auf stochastische Unabhängigkeit der erzeugten Zahlen.

1. Anpassungstests. Sind vorliegende Zahlen im Einheitsintervall Realisationen von stochastisch unabhängigen, $\mathcal{R}([0,1])$–verteilten Zufallsvariablen $\{U_n\}_{n\in\mathbb{N}}$ und teilt man das Einheitsintervall in k gleichlange Teilintervalle $I_j = [\frac{j-1}{k}, \frac{j}{k})$, $j = 1, \ldots, k$ der Länge $1/k$, so sollte die relative Häufigkeit der in jedem dieser Teilintervalle liegenden Zahlen ungefähr $1/k$ betragen. Nach dem starken Gesetz großer Zahlen (Satz 2.3.3) gilt nämlich für die Zufallsvariablen $X_n^{(j)} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{I_j}(U_i)$ unabhängig von j für alle $j = 1, \ldots, k$, daß $\lim_{n\to\infty} X_n^{(j)} = E\big(\mathbb{1}_{I_j}(U_1)\big) = P\big(\frac{j-1}{k} \leq U_1 \leq \frac{j}{k}\big) = \frac{1}{k}$ P–fast sicher.

Mit einer solchen Intervalleinteilung von $[0,1]$ berechnen wir für eine gegebene Menge von Zufallszahlen $u_1, \ldots, u_n$ die Anzahlen $n_j = \sum_{i=1}^n \mathbb{1}_{I_j}(u_i) = \#\{i \mid \frac{j-1}{k} \leq u_i \leq \frac{j}{k}\}$, $j = 1, \ldots, k$. Die folgende Prüfgröße mißt die Abweichung von den hypothetischen relativen Häufigkeiten $p_j^* = \frac{1}{k}$.

$$T_n = \sum_{j=1}^k \frac{(n_j - np_j^*)^2}{np_j^*} = \frac{1}{n}\sum_{j=1}^k \frac{n_j^2}{p_j^*} - n. \tag{6.2.2}$$

Entstehen die n_j tatsächlich aus stochastisch unabhängigen, $\mathcal{R}([0,1])$–verteilten Zufallsvariablen, so konvergiert die Verteilung P^{T_n} schwach (im Sinn von Definition 2.3.2) gegen eine χ^2_{k-1}–Verteilung mit $k-1$ Freiheitsgraden. Dies ist eine spezielle Γ–Verteilung (vgl. (2.1.84)) mit Parametern $\alpha = \frac{k-1}{2}$, $\lambda = \frac{1}{2}$ und Dichte

$$f_{\chi^2_{k-1}}(y) = \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})}\, y^{\frac{k-3}{2}} e^{-\frac{y}{2}}, \quad y \geq 0.$$

$F_{\chi^2_{k-1}}$ bezeichne die zugehörige Verteilungsfunktion. Die Approximation durch eine χ^2_{k-1}–Verteilung ist bereits brauchbar, wenn $n \geq 5k$ beträgt.

Ist T_n zu groß, werden wir die vorliegenden Zahlen nicht als gleichverteilt anerkennen. Der *kritische Wert* $c > 0$, dessen Überschreitung durch T_n zur Ablehnung führt, wird durch folgende Überlegungen bestimmt.

Wir wollen die Hypothese H_0, daß die Zahlen von stochastisch unabhängigen, $\mathcal{R}([0,1])$–verteilten Zufallsvariablen stammen, nur mit kleiner Wahrscheinlichkeit α fälschlich verwerfen, z.B. zum Niveau $\alpha = 0.05$ oder $\alpha = 0.01$. Das bedeutet, eine Zahl $c > 0$ so zu bestimmen, daß $P(T_n > c) \leq \alpha$. Da bei Gültigkeit der Hypothese H_0 die Prüfgröße T_n für große n approximativ χ^2_{k-1}–verteilt ist, muß

$$P(T_n > c) \approx 1 - \int_0^c f_{\chi^2_{k-1}}(y)\, dy = 1 - F_{\chi^2_{k-1}}(c) \leq \alpha$$

erfüllt werden. Als kritischer Wert wird daher das kleinste c verwendet, für das $F_{\chi^2_{k-1}}(c) \geq 1-\alpha$ gilt. Diese Lösung c_α heißt das $(1-\alpha)$–Quantil der χ^2_{k-1}–Verteilung. Einige Werte für c_α sind in folgender Tabelle angegeben.

$k-1$	3	5	10	20	30	40	50
$\alpha = 0.05$	7.815	11.071	18.307	31.410	43.773	55.758	67.505
$\alpha = 0.1$	6.251	9.236	15.987	28.412	40.256	51.805	63.167

Durch die Zusammenfassung der Zahlen in Klassen geht allerdings ein Teil der Information verloren. Dies wird beim *Kolmogoroff–Smirnov–Test* vermieden, der ebenfalls die Anpassung an eine vorgegebene Verteilung prüft. Man betrachtet hierzu für gegebene Zahlen $u_1, \ldots, u_n \in [0,1]$ die *empirische Verteilungsfunktion*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(u_i) \tag{6.2.3}$$

und bildet deren maximalen Abstand zur hypothetischen Verteilungsfunktion, hier der Verteilungsfunktion der $\mathcal{R}([0,1])$–Verteilung, $F_{\mathcal{R}([0,1])}(x) = x$, falls $0 \leq x \leq 1$,

$$D_n = \sup_{x \in [0,1]} |F_n(x) - x|. \tag{6.2.4}$$

Auch hier wird man die Hypothese H_0 ablehnen, wenn D_n zu groß ist. Unter der Hypothese H_0 besitzt die Zufallsvariable P^{D_n} eine Kolmogoroff–Smirnov–Verteilung mit Verteilungsfunktion F_{D_n}. Mit wachsendem n konvergiert die Verteilungsfunktion von $\sqrt{n}D_n$ gegen $F_{KS}(y) = 1 + 2\sum_{i=1}^{\infty}(-1)^i e^{-2i^2y^2}$, falls $y > 0$, und $F_{KS}(y) = 0$, falls $y \leq 0$.

Wie beim χ^2–Anpassungstest wählt man den kritischen Wert c_α, dessen Überschreitung durch D_n zur Ablehnung der Hypothese H_0 führt, als $(1-\alpha)$–Quantil der Kolmogoroff–Smirnov–Verteilung, also $c_\alpha = \min\{c > 0 \mid F_{D_n}(c) \geq 1-\alpha\}$. Einige Werte für c_α sind in der folgenden Tabelle angegeben.

n	10	30	50	60	70	80	100
$\alpha = 0.05$	0.409	0.242	0.188	0.172	0.160	0.150	0.134
$\alpha = 0.1$	0.369	0.218	0.170	0.155	0.144	0.135	0.121

Allerdings ist die Berechnung der Statistik D_n für große n recht aufwendig. Bezeichnen $u_{(1)}, \ldots, u_{(n)}$ die der Größe nach aufsteigend geordneten Zufallszahlen, so ergibt sich wegen der Monotonie der Wert des Supremums in den Sprungstellen der empirischen Verteilungsfunktion zu

$$D_n = \max\Big\{ \max_{1\leq i\leq n} |u_{(i)} - \tfrac{i}{n}|,\ \max_{1\leq i\leq n} |u_{(i)} - \tfrac{i-1}{n}| \Big\}.$$

Zur Berechnung von D_n sind die Werte $u_1, \ldots, u_n$ also der Größe nach zu sortieren, der Absolutbetrag der Differenzen $u_{(i)} - \frac{i}{n}$ bzw. $u_{(i)} - \frac{i-1}{n}$ muß gebildet und anschließend das Maximum der $2n$ so erhaltenen Zahlen bestimmt werden.

Der χ^2– und Kolmogoroff–Smirnov–Test funktionieren analog mit den gleichen Quantilen, wenn man die Anpassung auf andere hypothetische Verteilungen $\tilde{F}$ überprüfen will. Die Klassenhäufigkeiten p_j^* sind dann aus $\tilde{F}$ zu berechnen (durch Wahl der Klassengrenzen möglichst in gleicher Größe) bzw. $F_{\mathcal{R}([0,1])}$ durch $\tilde{F}$ zu ersetzen. Im allgemeinen besitzt der Kolmogoroff–Smirnov–Test eine größere Schärfe, d.h. er verwirft die Hypothese H_0 mit größerer Wahrscheinlichkeit, wenn sie nicht zutrifft. Die oben angegebenen Fraktile sind hierbei jedoch nur dann gültig, wenn die hypothetische Verteilungsfunktion $\tilde{F}$ stetig ist.

Die auf (6.2.2) und (6.2.4) basierenden Tests sind hauptsächlich auf die Überprüfung der Gleichverteilungsannahme zugeschnitten, obwohl bei der Bestimmung der kritischen Schranken die Annahme der stochastischen Unabhängigkeit indirekt eingeht und eine Abweichung hiervon das Ergebnis beeinflussen kann. Spezielle Tests auf stochastische Unabhängigkeit werden im folgenden behandelt, wobei wir uns auf zwei typische Vertreter beschränken.

2. Tests auf Unabhängigkeit. Die Prüfgrößen solcher Tests basieren auf Eigenschaften, die stochastisch unabhängige Zufallsvariablen auszeichnen. Eine hiervon ist die Verteilung von Sequenzen aufsteigend geordneter Zahlen, sogenannte Runs. In der Ziffernfolge

2 | 1 7 9 | 4 5 7 | 2 8 | 1 8 | 1 6 | 2 | 1 4 5 6 9

finden sich zwei Runs der Länge 1, drei der Länge 2, zwei der Länge 3 und einer der Länge 5. Der *Runtest* basiert auf der Verteilung solcher Sequenzen.

Für stochastisch unabhängige Zufallsvariable läßt sich die Wahrscheinlichkeit für das Auftreten eines Runs bestimmter Länge explizit bestimmen. Diese Verteilung reagiert empfindlich auf Störungen der stochastischen Unabhängigkeit und eignet sich daher gut als Testgröße dieser Eigenschaft.

Lemma 6.2.1. *$\{U_n\}_{n\in\mathbb{N}}$ sei eine Folge von stochastisch unabhängigen, identisch $\mathcal{R}([0,1])$-verteilten Zufallsvariablen. Dann beträgt die Wahrscheinlichkeit für einen Run R der Länge r, $r \in \mathbb{N}$,*

$$P(R=r) = P(U_1 \le \cdots \le U_r > U_{r+1}) = \frac{1}{r!} - \frac{1}{(r+1)!} = \frac{r}{(r+1)!}. \tag{6.2.5}$$

Beweis. Wegen der stochastischen Unabhängigkeit und identischen Verteilung reicht es, die Zufallsvariablen $U_1, \ldots, U_{r+1}$ zu betrachten. Für alle $u \in [0,1]$, $k \in \mathbb{N}$ gilt

$$P(U_1 \le \cdots \le U_r \le u) = \frac{u^r}{r!}. \tag{6.2.6}$$

Wir zeigen diese Behauptung mit vollständiger Induktion. Für $r = 1$ ist (6.2.6) richtig. Ferner gilt mit der Induktionsvoraussetzung und Lemma 3.1.4

$$\begin{aligned} P(U_1 \le \cdots \le U_{r+1} \le u) &= \int_0^1 P(U_1 \le \cdots \le U_r \le t \le u)dt \\ &= \int_0^u P(U_1 \le \cdots \le U_r \le t)dt = \int_0^u \frac{t^r}{r!}dt = \frac{u^{r+1}}{(r+1)!}, \end{aligned}$$

woraus (6.2.6) folgt. Da

$$P(U_1 \le \cdots \le U_r) = P(U_1 \le \cdots \le U_r, U_r \le u) + P(U_1 \le \cdots \le U_r, U_r > u),$$

erhalten wir mit (6.2.6) $P(U_1 \le \cdots \le U_r > u) = \frac{1}{r!} - \frac{u^r}{r!}$ und insgesamt

$$\begin{aligned} P(U_1 \le \cdots \le U_r, U_r > U_{r+1}) &= \int_0^1 P(U_1 \le \cdots \le U_r, U_r > u)du \\ &= \int_0^1 \frac{1}{r!} - \frac{u^r}{r!}du = \frac{1}{r!} - \frac{1}{(r+1)!} \end{aligned}$$

∎

Sind die Runs $R_1, \ldots, R_n$ aus einer Folge von Zufallszahlen stochastisch unabhängig, so ist ein Anpassungstest auf die durch Lemma 6.2.1 gegebene Verteilung

$$P(R=r) = \frac{r}{(r+1)!}, \quad r \in \mathbb{N}, \tag{6.2.7}$$

ein brauchbarer Test auf Unabhängigkeit von vorliegenden Zufallszahlen. Eine Schwierigkeit bei der Anwendung besteht jedoch darin, daß aufeinanderfolgende Runs in einer Folge stochastisch unabhängiger Zufallsvariablen $\{U_n\}_{n\in\mathbb{N}}$ nicht unabhängig sind. Streicht man jedoch die auf einen Run folgende Zufallszahl und beginnt die Zählung der Länge des nächsten Run erst mit der übernächsten Zahl, so sind die so gewonnenen Runlängen stochastisch unabhängig. Wir definieren formal die Zufallsvariablen

$$\begin{aligned} S_1 = \inf\{n \in \mathbb{N} \mid U_n > U_{n+1}\}, \quad S_{k+1} = \inf\{n > S_k + 1 \mid U_n > U_{n+1}\}, \\ R_1 = S_1, \quad R_{k+1} = S_{k+1} - S_k - 1, \quad k \in \mathbb{N}. \end{aligned} \tag{6.2.8}$$

Man beachte, daß die Zufallsvariablen S_k in folgendem Sinn Stoppzeiten bezüglich $\{U_n\}_{n\in\mathbb{N}}$ sind. (vgl. Definition 2.1.3) Für alle $k, \ell \in \mathbb{N}$ besitzt das Ereignis $\{S_k = \ell\}$ eine Darstellung

$$\{S_k = \ell\} = \{(U_1, \ldots, U_{\ell+1}) \in B_k^{(\ell+1)}\} \tag{6.2.9}$$

für eine Menge $B_k^{(\ell+1)} \in \mathcal{B}^{\ell+1} \cap [0,1]^{\ell+1}$. Der Beweis des folgenden Lemmas benutzt Methoden, die bereits in Satz 2.1.2 eingeführt wurden.

Lemma 6.2.2. *$\{U_n\}_{n\in\mathbb{N}}$ sei eine Folge von stochastisch unabhängigen, identisch $\mathcal{R}([0,1])$-verteilten Zufallsvariablen. Dann ist die durch (6.2.8) definierte Folge der Längen sukzessiver Runs $\{R_n\}_{n\in\mathbb{N}}$ stochastisch unabhängig. Alle R_n sind ferner identisch verteilt mit Zähldichten (6.2.7).*

Beweis. Die Folge $\{U_{S_k+n}\}_{n\geq 2}$ besitzt für alle $k\in\mathbb{N}$ stochastisch unabhängige, identisch nach P^{U_1} verteilte Glieder. Denn für alle $k,m\in\mathbb{N}$, $m\geq 2$, $B_1,\ldots,B_m\in\mathcal{B}\cap[0,1]$ gilt wegen (6.2.9) für alle $\ell\in\mathbb{N}$

$$P\Big(S_k=\ell,\bigcap_{j=2}^{m}\{U_{S_k+j}\in B_j\}\Big)=P\Big((U_1,\ldots,U_{\ell+1})\in B_k^{(\ell+1)},\bigcap_{j=2}^{m}\{U_{\ell+j}\in B_j\}\Big)$$
$$=P(S_k=\ell)\prod_{j=2}^{m}P(U_1\in B_j).$$

Durch Summation über ℓ folgt

$$P\Big(\bigcap_{j=2}^{m}\{U_{S_k+j}\in B_j\}\Big)=\sum_{\ell=1}^{\infty}P\Big(S_k=\ell,\bigcap_{j=2}^{m}\{U_{S_k+j}\in B_j\}\Big)$$
$$=\prod_{j=2}^{m}P(U_1\in B_j),\tag{6.2.10}$$

also auch $P\big(U_{S_k+m}\in B_m\big)=P(U_1\in B_m)$ für alle $m\geq 2$.
Ferner sind $(S_1,\ldots,S_k)$ und $\{U_{S_k+n}\}_{n\geq 2}$ stochastisch unabhängig, da wegen $S_1\leq\cdots\leq S_k$ und $\{S_1=\ell_1,\ldots,S_k=\ell_k\}=\{(U_1,\ldots,U_{\ell_k+1})\in B_k^{(\ell_k+1)}\}$ für eine Menge $B_k^{(\ell_k+1)}\in\mathcal{B}^{\ell_k+1}\cap[0,1]^{\ell_k+1}$ wie oben

$$P\Big(S_1=\ell_1,\ldots,S_k=\ell_k,\bigcap_{j=2}^{m}\{U_{S_k+j}\in B_j\}\Big)$$
$$=P(S_1=\ell_1,\ldots,S_k=\ell_k)\prod_{j=2}^{m}P(U_1\in B_j)$$

für alle $\ell_1,\ldots,\ell_k\in\mathbb{N}$, $B_1,\ldots,B_m\in\mathcal{B}\cap[0,1]$ folgt. Weiterhin gilt

$$R_{k+1}=\inf\{n\in\mathbb{N}\mid U_{S_k+n+1}>U_{S_k+n+2}\},\tag{6.2.11}$$

so daß R_{k+1} durch eine meßbare Transformation aus $\{U_{S_k+n}\}_{n\geq 2}$ gewonnen wird. Nach Lemma 2.1.7 sind also R_{k+1} und $(S_1,\ldots,S_k)$ stochastisch unabhängig.
$R_1,\ldots,R_k$ sind wegen (6.2.8) Funktionen von $S_1,\ldots,S_k$. Wieder mit Lemma 2.1.7 sind dann $(R_1,\ldots,R_k)$ und R_{k+1} stochastisch unabhängig. Lemma 2.1.8 zeigt, daß dann die ganze Folge $\{R_n\}_{n\in\mathbb{N}}$ stochastisch unabhängig ist.
In (6.2.10) hängt die Verteilung von $\{U_{S_k+n}\}_{n\geq 2}$ nicht von k ab. Aus der Darstellung (6.2.11) sieht man, daß alle R_n die gleiche Verteilung wie R_1 besitzen, die in Lemma 6.2.1 berechnet wurde. ∎

Die Verteilung der Runlängenfolge $\{R_n\}_{n\in\mathbb{N}}$ bei Überschlagen einer Zufallszahl nach Ende eines Runs ist durch Lemma 6.2.1 und 6.2.2 vollständig bestimmt. Beim Runtest führt man jetzt eine Prüfung auf Anpassung dieser Verteilung mit Hilfe des χ^2–Tests durch. $r_1,\dots,r_n$ bezeichne die aus einer Zufallszahlenfolge wie oben gewonnenen Runlängen.

Der Träger $\mathbb{N}$ der Verteilung (6.2.7) wird in k disjunkte Klassen $\{i_1+1,\dots,i_2\}$, $\{i_2+1,\dots,i_3\},\dots,\{i_k+1,\dots,\infty\}$, $0=i_1<i_2<\dots<i_k\in\mathbb{N}$, mit möglichst gleichgroßen Wahrscheinlichkeiten $p_j^*=\sum_{\ell=i_j+1}^{i_{j+1}}P(R=\ell)$, $j=1,\dots k$, $i_{k+1}=\infty$, eingeteilt. Wegen der Approximationsgenauigkeit sollte $np_j^*\geq 30$ für alle $j=1,\dots k$ gelten. Mit diesen p_j^* und den aus $r_1,\dots,r_n$ gewonnenen relativen Häufigkeiten n_j wird die Prüfgröße T_n aus (6.2.2) des χ^2–Anpassungstests berechnet. Die Unabhängigkeitshypothese wird verworfen, wenn $T_n>c_\alpha$, wobei c_α das $(1-\alpha)$–Quantil der χ^2_{k-1}–Verteilung ist. Um eine genügende Schärfe des Runtests zu erzielen, sollte die Anzahl n der zugrundeliegenden Zufallszahlen sehr groß sein $(n\geq 4000)$.

Auf der *Autokorrelation* basierende Prüfgrößen nutzen die Tatsache aus, daß aus der stochastischen Unabhängigkeit von Zufallsvariablen deren Unkorreliertheit folgt (vgl. Definition 2.2.5). Trifft die Hypothese von stochastisch unabhängigen, identisch verteilten Zufallsvariablen $\{U_n\}_{n\in\mathbb{N}}$ zu, so gilt für alle $k,m\in\mathbb{N}$

$$\rho(k)=\mathrm{Korr}(U_m,U_{m+k})=\frac{E(U_mU_{m+k})-\big(E(U_m)\big)^2}{E(U_m^2)-\big(E(U_m)\big)^2}=0.$$

Es gilt nun, aus vorliegenden Zufallszahlen $u_1,\dots,u_n$ eine Schätzgröße für $\rho(k)$ zu gewinnen. Ausgenommen im Trivialfall, daß alle u_i gleich sind, leistet dies der Autokorrelationskoeffizient

$$\hat\rho(k)=\frac{n\sum_{i=1}^n u_iu_{i+k}-\big(\sum_{i=1}^n u_i\big)^2}{n\sum_{i=1}^n u_i^2-\big(\sum_{i=1}^n u_i^2\big)},\quad 1\leq k\leq n/2,$$

wobei $u_{i+k}=u_{i+k-n}$, falls $i+k>n$. Bei einem guten Zufallszahlengenerator sollte $\hat\rho(k)$ für alle k nahe bei Null liegen, da sonst ein Indiz für lineare Abhängigkeit zwischen k–ten Nachfolgern in der Zufallszahlenfolge vorliegt. Die im Anschluß an Definition 2.2.5 durchgeführten Überlegungen zeigen, daß $-1\leq\rho(k)\leq 1$, wobei die Werte $\rho(k)=-1$ und $\rho(k)=+1$ auf exakte lineare Abhängigkeit zwischen den entsprechenden Zufallsvariablen schließen lassen.

6.3. Transformationsverfahren

Wir gehen im folgenden davon aus, daß ein guter Generator zur Erzeugung von stochastisch unabhängigen, rechteckverteilten Zufallszahlen zur Verfügung steht. Wie solche Generatoren konstruiert werden und wie man ihre Eigenschaften testet, haben wir in den vorhergehenden Kapiteln kennengelernt. In den meisten Anwendungen werden allerdings Zufallszahlen aus anderen Verteilungen benötigt. Bei der Simulation von Warteschlangen benötigt man etwa Ankunftszahlen in festen Zeitintervallen, die sich durch Poisson–Verteilungen beschreiben lassen. Zwischenankunftszeiten genügen hierbei in der Regel Exponentialverteilungen. Normalverteilungen sind nach dem Zentralen Grenzwertsatz immer dann angebracht, wenn sich die Wirkung eines Einflußfaktors als Summe vieler gleichgewichtiger, unabhängiger Einzelfaktoren darstellen läßt. Bei der Simulation von Zufallspermutationen werden Tupel von natürlichen Zahlen benötigt. Solche Anwendungen setzen die Existenz von Zufallszahlen voraus, die als Realisationen aus den entsprechenden Verteilungen interpretiert werden können.

Wir werden uns in diesem Abschnitt mit einigen Verfahren beschäftigen, die die Transformation von rechteckverteilten Zufallszahlen auf gewisse vorgegebene Verteilungen bewirken. Die stochastische Unabhängigkeit der transformierten Zufallszahlen bleibt dabei erhalten, wenn sie aus verschiedenen stochastisch unabhängigen Standardzufallszahlen erzeugt werden. Dies folgt unmittelbar aus den Lemmata 2.1.5 und 2.1.6.

Unterwirft man Zufallszahlen $\{u_n\}_{n\in\mathbf{N}}$ einer Transformation h, so berechnet sich die Verteilung der transformierten Zufallsvariablen $X_n = h(U_n)$ als Bildmaß unter der Transformation h. Wir werden damit für einige Verfahren die in Kapitel 6.1 vorgestellten Methoden einsetzen, wobei oft mehrere Ansätze zum Ziel führen. Die Auswahl einer speziellen Methode hängt dann davon ab, ob die Implementation des zugehörigen Algorithmus genügend schnell arbeitet.

Ein allgemeines Verfahren zur Gewinnung von Zufallszahlen mit beliebiger vorgegebener Verteilungsfunktion F erhält man aus Satz 2.1.1: "Ist U eine $\mathcal{R}([0,1])$–verteilte Zufallsvariable und bezeichnet $F^{-1}(t) = \inf\{x \in \mathbf{R} \mid F(x) \geq t\}$, $t \in (0,1)$, die Pseudoinverse von F, so besitzt die Zufallsvariable $X = F^{-1}(U)$ die Verteilungsfunktion F."

Lemma 6.3.1. *$T = \{t_i \mid i \in M\}$, $M = \{0,1,\ldots,m\}, m \in \mathbf{N}$, oder $M = \mathbf{N}_0$ sei eine abzählbare Menge und $f : T \to [0,1]$ mit $\sum_{i\in M} f(t_i) = 1$ eine diskrete Zähldichte. Ist U eine $\mathcal{R}([0,1])$–verteilte Zufallsvariable, so besitzt die diskrete Zufallsvariable X mit*

$$X = t_k, \quad \text{falls } \sum_{i=0}^{k-1} f(t_i) < U \leq \sum_{i=0}^{k} f(t_i),\ k \in M, \tag{6.3.1}$$

die durch f gegebene Verteilung. Die Summe mit leerem Indexbereich wird hierbei als Null definiert.

Beweis. Bezeichne $F(x) = \sum_{i\in M} f(t_i)\mathbb{1}_{(-\infty,x]}(i)$, $x \in \mathbf{R}$, die Verteilungsfunktion der diskreten Verteilung mit Träger M und Wahrscheinlichkeiten $f(t_i)$. Für

die Pseudo-Inverse F^{-1} der Treppenfunktion F gilt

$$F^{-1}(u) = \inf\{x \in \mathbf{R} \mid F(x) \geq u\} = \inf\{j \in M \mid F(j) \geq u\}$$
$$= \inf\{j \in M \mid \sum_{i=1}^{j} f(t_i) \geq u\} = k, \text{ falls } \sum_{i=0}^{k-1} f(t_i) < u \leq \sum_{i=0}^{k} f(t_i),$$

für alle $u \in (0,1)$. $X' = F^{-1}(U)$ besitzt nach Satz 2.1.1 die Verteilungsfunktion F mit Träger M und $X = h(X')$ mit der injektiven Abbildung $h(i) = t_i$, $i \in M$, die gewünschte Verteilung mit Zähldichte f. Eine geometrische Veranschaulichung dieser Argumente findet sich nach (2.1.11). ∎

Zur Erzeugung von Zufallszahlen $\{x_n\}_{n\in\mathbf{N}}$ aus einer beliebigen diskreten Verteilung hat man also für Standardzufallszahlen $\{u_n\}_{n\in\mathbf{N}}$ Beziehung (6.3.1) auszuwerten, d.h. setze $x_n = t_k$, falls $u_n \in \left(\sum_{i=0}^{k-1} f(t_i), \sum_{i=0}^{k} f(t_i)\right]$, $k \in M$. Dies läßt sich effektiv bei Verteilungen mit endlichem Träger durchführen.

Beispiel 6.3.1. (Laplace-Verteilung auf $\{0,1,\ldots,m-1\}$)
Für die zugehörige Zähldichte gilt $f(i) = \frac{1}{m}$ für alle $i = 0,\ldots,m-1$. (6.3.1) vereinfacht sich in diesem Fall zu $X = k$, falls $\frac{k}{m} < U \leq \frac{k+1}{m}$, $k = 0,\ldots,m-1$. Da $P(U = \frac{j}{m}) = 0$ für alle $j = 0,\ldots,m$, genügt die Zufallsvariable $X = \lfloor mU \rfloor$ einer diskreten Gleichverteilung mit Träger $\{0,1,\ldots,m\}$. ∎

Beispiel 6.3.2. (Binomialverteilung)
Zu den Parametern $0 \leq p \leq 1$ und $m \in \mathbf{N}$ lautet die Zähldichte $p_i = f(i) = \binom{m}{i} p^i (1-p)^{m-i}$, $i = 0,1,\ldots,m$ (vgl. (2.1.18)). Zur Erzeugung von stochastisch unabhängigen $\mathfrak{B}(m,p)$-verteilten Zufallszahlen werden zunächst die kumulierten Wahrscheinlichkeiten $F_k = \sum_{i=0}^{k} \binom{m}{i} p^i (1-p)^{m-i}$, $k = 0,1,\ldots,m$, berechnet.
Sind u_n stochastisch unabhängige Standardzufallszahlen, so entscheidet man, in welchem der Intervalle $(F_{k-1}, F_k]$, $k = 0,1,\ldots,m$, $(F_{-1} = 0)$, die Zufallszahl u_n liegt. $u_n \in (F_{k-1}, F_k]$ für ein k führt zu $x_n = k$. Mit (6.3.1) genügen die so gewonnenen Zufallszahlen x_n einer $\mathfrak{B}(m,p)$-Verteilung.
Zur effizienten Bestimmung des zugehörigen Intervalls $I_k = (F_{k-1}, F_k]$, in dem u_n liegt, werden die Elemente I_k in der natürlichen Ordnung $I_1 < \cdots < I_m$ mit zugehörigen Wahrscheinlichkeiten $p_k = F_k - F_{k-1}$ wie in (5.3.9) in einem binären Suchbaum abgespeichert. Die erwartete Zugriffszeit beträgt nach Satz 4.3.2 höchstens $H\big(\mathfrak{B}(m,p)\big) + 1 \leq \log_2 m + 1$, wobei H die Entropie mit Logarithmen zur Basis 2 bezeichnet. Nach Berechnung und Abspeichern der F_k wächst der Aufwand höchstens logarithmisch in m.

Eine andere Möglichkeit zur Erzeugung binomialverteilter Zufallszahlen ergibt sich aus der Tatsache, daß jede $\mathfrak{B}(m,p)$-verteilte Zufallsvariable X eine Darstellung $X = \sum_{i=1}^{m} X_i$ mit stochastisch unabhängigen, $\mathfrak{B}(1,p)$-verteilten Zufallsvariablen X_i besitzt (vgl. (2.1.18)). $\mathfrak{B}(1,p)$-verteilte Zufallsvariable Y erhält man aus einer $\mathcal{R}([0,1])$-verteilten Zufallsvariablen U durch

$$Y = \mathbf{1}_{[0,p)}(U) \quad \text{oder auch} \quad Y = \lfloor U + p \rfloor. \tag{6.3.2}$$

Dies folgt aus (6.3.1), da mit U auch $1 - U$ rechteckverteilt auf $[0,1]$ ist.

Je m aufeinanderfolgende Standardzufallszahlen $u_{(j-1)m+1}, \ldots, u_{jm}$, $j \in \mathbb{N}$, werden jetzt zu $\mathfrak{B}(1,p)$–verteilten Zufallszahlen $y_{(j-1)m+1}, \ldots, y_{jm}$ mit Hilfe der Abbildung $y_i = \mathbb{1}_{(0,p]}(u_i)$, $i = (j-1)m+1, \ldots, jm$, transformiert. Die Zufallszahlen $x_j = \sum_{i=(j-1)m+1}^{jm} y_i$ sind dann Realisationen von stochastisch unabhängigen, $\mathfrak{B}(m,p)$–verteilten Zufallsvariablen. Die Nachteile dieses Verfahrens sind, daß für jede binomialverteilte Zufallszahl m Standardzufallszahlen gebraucht werden und daß der Rechenaufwand linear in m wächst. ∎

Auch zur Erzeugung diskreter Verteilungen mit unendlichem Träger läßt sich die Methode (6.3.1) verwerten. Allerdings können die unendlich vielen Werte $\sum_{i=1}^{k} f(t_i)$, $k \in \mathbb{N}_0$, dann nicht a priori abgespeichert werden. Sie müssen wiederholt neu berechnet werden, wodurch das Transformationsverfahren erheblich an Schnelligkeit verliert. Für Poisson– und negative Binomialverteilungen werden wir später Verfahren kennenlernen, die besser arbeiten.

Ist die Verteilungsfunktion invertierbar und ist die Inverse F^{-1} leicht zu berechnen, liefert Satz 2.1.1 eine besonders einfache Transformationsmöglichkeit. Für eine Standardzufallszahl u kann $x = F^{-1}(u)$ als Zufallszahl aus der Verteilung F interpretiert werden. Ein Beispiel hierfür ist die Exponentialverteilung mit Parameter $\lambda > 0$ und Verteilungsfunktion $F(x) = \left(1 - e^{-\lambda x}\right)\mathbb{1}_{[0,\infty)}(x)$. Bereits in (2.1.9) wurde gezeigt, daß man aus einer $\mathcal{R}([0,1])$–verteilten Zufallsvariablen U durch die Transformation

$$X = -\frac{1}{\lambda} \ln U, \quad \lambda > 0, \tag{6.3.3}$$

eine $\mathcal{E}(\lambda)$–verteilte Zufallsvariable X erhält. Die Berechnung von exponentialverteilten Zufallszahlen x_n gestaltet sich mit der Transformation $x_n = -\frac{1}{\lambda} \ln u_n$ sehr einfach, wobei der Zeitaufwand hauptsächlich von dem Berechnungsalgorithmus für die benötigten Logarithmen bestimmt wird.

Durch einfaches Abrunden einer $\mathcal{E}(\lambda)$–verteilten Zufallsvariablen wird eine geometrisch verteilte erzeugt (vgl. Übungsaufgabe 6.7), genauer:
Ist X $\mathcal{E}(\lambda)$–verteilt, $\lambda > 0$, so besitzt die Zufallsvariable $Y = \lfloor X \rfloor$ eine $\mathfrak{G}(p)$–Verteilung, wobei $p = 1 - e^{-\lambda}$.

Für gegebenes $0 < p < 1$ kann man also $\mathfrak{G}(p)$–verteilte Zufallszahlen y_n effizient aus $\mathcal{E}(\lambda)$–verteilten x_n durch $y_n = \lfloor x_n \rfloor$ gewinnen, wobei $\lambda = -\ln(1-p)$ gewählt wird.

Eine Methode zur Erzeugung von Poisson–verteilten Zufallszahlen wird in folgendem Lemma aufgezeigt.

Lemma 6.3.2. *(Poisson–Verteilung)*
$\{X_n\}_{n \in \mathbb{N}}$ sei eine Folge von stochastisch unabhängigen, mit Parameter $\lambda > 0$ exponentialverteilten Zufallsvariablen. Dann besitzt

$$S = \inf\left\{n \in \mathbb{N}_0 \mid \sum_{i=1}^{n+1} X_i > 1\right\}$$

eine Poisson–Verteilung mit Parameter λ.

Beweis. Für alle $k \in \mathbb{N}_0$ gilt

$$P(S=k) = P\Big(\sum_{i=1}^{k} X_i \le 1, \sum_{i=1}^{k+1} X_i > 1\Big) = P\Big(\sum_{i=1}^{k} X_i \le 1, X_{k+1} > 1 - \sum_{i=1}^{k} X_i\Big).$$

$T = \sum_{i=1}^{k} X_i$ besitzt nach (2.1.82) eine Erlang–Verteilung mit Dichte $f_T(t) = \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t} \mathbb{1}_{[0,\infty)}(t)$, so daß

$$\begin{aligned} P(S=k) &= \int_0^1 P(X_{k+1} > 1-t) f_T(t)dt = \int_0^1 e^{-\lambda(1-t)} \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t} dt \\ &= e^{-\lambda} \lambda^k \int_0^1 \frac{t^{k-1}}{(k-1)!} dt = e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

■

Lemma 6.3.2 wird folgendermaßen angewendet. Erzeuge sukzessive unabhängige, exponentialverteilte Zufallszahlen x_n. Ist j der erste Index, bei dem die Summe dieser Zahlen 1 überschreitet, setzt man $y = j - 1$ als Zufallszahl aus einer Poisson–Verteilung.

Die exponentialverteilten Zufallszahlen erhält man durch die Transformation (6.3.3) aus Standardzufallszahlen $\{u_n\}$. Rechnerisch zu bestimmen ist dann

$$\inf\Big\{n \in \mathbb{N}_0 \mid \sum_{i=1}^{n+1} -\frac{1}{\lambda} \ln u_i > 1\Big\} = \inf\big\{n \in \mathbb{N}_0 \mid u_1 \cdots u_{n+1} < e^{-\lambda}\big\}.$$

Beispiel 6.3.3. (geometrische und negative Binomialverteilung)
Mit ähnlichen Methoden der ersten bzw. m–ten Eintrittszeit lassen sich aus $\mathfrak{B}(1,p)$–verteilten Zufallszahlen, die mit Hilfe von (6.3.2) erzeugt werden, geometrisch bzw. negativ binomialverteilte gewinnen. Man zählt dabei die Anzahl der Zufallszahlen bis zur ersten Eins bzw. bis zur m-ten Eins einschließlich, $m \in \mathbb{N}$. Ist nämlich $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen, $\mathfrak{B}(1,p)$–verteilten Zufallsvariablen, so folgt aus (2.1.34), daß

$$S = \inf\{n \in \mathbb{N} \mid X_n = 1\}$$

geometrisch verteilt ist mit Zähldichte $f_{\mathfrak{G}(p)}(k) = p(1-p)^{k-1}$, $k \in \mathbb{N}$.
Durch m–fache Addition solcher $\mathfrak{G}(p)$ verteilter Zufallszahlen erhält man eine Zufallszahl aus einer negativen Binomialverteilung $\overline{\mathfrak{B}}(m,p)$, wie in (2.1.50) gezeigt wurde. Die zugehörige Zähldichte lautet $f_{\overline{\mathfrak{B}}(m,p)}(k) = \binom{k-1}{m-1} k^m (1-p)^{k-m}$, $k \ge m$ (vgl. (2.1.51)). ■

Ein weiteres allgemeines Verfahren zur Erzeugung von Zufallszahlen aus einer beliebigen, vorgegebenen Verteilung mit Dichte f basiert auf folgendem Lemma.

Lemma 6.3.3. *(Verwerfungsmethode)*
$f \geq 0$ und $g \geq 0$ seien Verteilungsdichten derart, daß $c\frac{f(x)}{g(x)} \leq 1$ für eine Konstante $c > 0$ und alle $x \in \mathbf{R}$ mit $f(x) > 0$. $\{X_n\}_{n\in\mathbf{N}}$ sei eine Folge von identisch verteilten Zufallsvariablen mit Verteilung $P^{X_n} = P^X$ mit Dichte g und $\{U_n\}_{n\in\mathbf{N}}$ eine unabhängige Folge von $\mathcal{R}([0,1])$-verteilten Zufallsvariablen so, daß $\{X_n\}$ und $\{U_n\}$ stochastisch unabhängig sind. Es bezeichne

$$S = \inf\left\{n \in \mathbf{N} \mid U_n \leq c\,\frac{f(X_n)}{g(X_n)}\right\}.$$

Dann besitzt die Verteilung der gestoppten Zufallsvariablen X_S die Dichte f. Ferner gilt für den Erwartungswert $E(S) = 1/c$.

Beweis. U sei $\mathcal{R}([0,1])$-verteilt. Die Bedingung $c\frac{f(x)}{g(x)} \leq 1$ für alle $x \in \mathbf{R}$ mit $f(x) > 0$ impliziert (mit den Konventionen $a/0 = \infty$, falls $a \neq 0$, und $0/0 = 0$), daß $g(x) > 0$, falls $f(x) > 0$. $f(X_n)/g(X_n)$ ist also P-f.s. endlich und wohldefiniert. Für alle $A \in \mathcal{B}^1$ gilt

$$\begin{aligned} P\Big(U \leq c\,\frac{f(X)}{g(X)}, X \in A\Big) &= \int_A P\Big(U \leq c\,\frac{f(x)}{g(x)}\Big) g(x)dx \\ &= \int_A c\,\frac{f(x)}{g(x)} g(x)dx = c\int_A f(x)dx, \end{aligned} \tag{6.3.4}$$

da $P(U \leq u) = u$ für alle $u \in [0,1]$. Mit $A = \mathcal{B}^1$ folgt insbesondere

$$1 \geq P\Big(U \leq c\,\frac{f(X)}{g(X)}\Big) = c > 0\,. \tag{6.3.5}$$

$S = \inf\left\{n \in \mathbf{N} \mid U_n - c\frac{f(X_n)}{g(X_n)} \leq 0\right\}$ ist eine Stoppzeit bezüglich der unabhängigen Folge $\left\{U_n - c\,\frac{f(X_n)}{g(X_n)}\right\}_{n\in\mathbf{N}}$, insbesondere die erste Eintrittszeit dieser Folge in die Menge $B = (-\infty, 0] \in \mathcal{B}^1$. Wie in Lemma 2.1.5 folgt für alle $A \in \mathcal{B}^1$ wegen der stochastischen Unabhängigkeit, daß

$$\begin{aligned} P(S = n, X_S \in A) &= P\left(\bigcap_{i=1}^{n-1}\Big\{U_i > c\,\frac{f(X_i)}{g(X_i)}\Big\}, U_n \leq c\,\frac{f(X_n)}{g(X_n)}, X_n \in A\right) \\ &= \left[P\Big(U > c\,\frac{f(X)}{g(X)}\Big)\right]^{n-1} P\Big(U \leq c\,\frac{f(X)}{g(X)}, X \in A\Big)\,. \end{aligned}$$

Mit der Abkürzung $P\big(U > c\,\frac{f(X)}{g(X)}\big) = q < 1$ folgt nach Summation über $n \in \mathbf{N}$ mit $\sum_{n=1}^{\infty} q^{n-1} = \frac{1}{1-q}$ und (6.3.4)

$$\begin{aligned} P(X_S \in A) &= \Big(\sum_{n=1}^{\infty} q^{n-1}\Big) \cdot P\Big(U \leq c\,\frac{f(X)}{g(X)}, X \in A\Big) \\ &= \frac{P\big(U \leq c\,\frac{f(X)}{g(X)}, X \in A\big)}{P\big(U \leq c\,\frac{f(X)}{g(X)}\big)} = \int_A f(x)dx \end{aligned}$$

für alle $A \in \mathcal{B}^1$. Die letzte Gleichung besagt, daß f Verteilungsdichte von P^{X_S} ist.
S ist nach Lemma 2.1.5 geometrisch verteilt mit Parameter $p = P\left(U \leq c\frac{f(X)}{g(X)}\right) = c \leq 1$, so daß $E(S) = 1/c$ folgt. Dies zeigt die Behauptung. ∎

Bei der Anwendung von Lemma 6.3.3 zur Erzeugung von Zufallszahlen aus einer Verteilung P^Y mit Dichte f geht man folgendermaßen vor: Wähle g als Verteilungsdichte, aus der Zufallszahlen einfach und effizient generiert werden können, und $c \in \mathbf{R}$ so, daß

$$c\,\frac{f(x)}{g(x)} \leq 1 \quad \text{für alle } x \in \mathbf{R} \text{ mit } f(x) > 0. \tag{6.3.6}$$

Erzeuge dann sukzessive unabhängige Zufallszahlen $x_1, u_1, x_2, u_2, x_3, u_3, \ldots$ aus der Verteilung $P^{X_n} = P^X$ mit Dichte g bzw. einer $\mathcal{R}([0,1])$–Verteilung. Sei $k \in \mathbf{N}$ der erste Index, für den $u_k \leq c\frac{f(x_k)}{g(x_k)}$ gilt. Dann ist $y_k = x_k$ eine Zufallszahl aus der Verteilung P^Y mit Dichte f. Zur Erzeugung weiterer Zufallszahlen fahre mit Elementen der Restfolge $\{(x_{k+n}, u_{k+n})\}_{n\in\mathbf{N}}$ fort.

Dieses Verfahren heißt Verwerfungsmethode. Man beachte, daß es an keinen speziellen Typ von Verteilung gebunden ist. Es kann in der gleichen Form eingesetzt werden, wenn f und g Zähldichten sind (vgl. Beispiel 6.3.5).

Die erwartete Anzahl von Zufallszahlen (x_n, u_n), die für eine Zufallszahl y_k gebraucht werden, beträgt $1/c$. c sollte also unter Berücksichtigung der Nebenbedingung (6.3.6) möglichst groß gewählt werden.

Beispiel 6.3.4. (Beta–Verteilungen)
Die Familie der Beta–Verteilungen mit Parameter $\alpha, \beta > 0$ besitzt den Träger $(0,1)$ und Dichten der Form

$$f_{\alpha,\beta}(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}_{(0,1)}(x),$$

wobei $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Für $\alpha = \beta = 1$ erhält man gerade eine $\mathcal{R}([0,1])$–Verteilung.
Gilt $\alpha, \beta > 1$, so sind die Dichten $f_{\alpha,\beta}$ beschränkt. Das Maximum wird für $x_0 = \frac{\alpha-1}{\alpha+\beta-2}$ mit Wert $\delta_0 = f(x_0) = \frac{1}{B(\alpha,\beta)}\frac{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}{(\alpha+\beta-2)^{\alpha+\beta-2}}$ angenommen, wie man durch Differenzieren von $f_{\alpha,\beta}$ sieht. Zur Erzeugung von Beta–verteilten Zufallszahlen mit Hilfe der Verwerfungsmethode setzen wir $P^X = \mathcal{R}([0,1])$ mit $g(x) = \mathbb{1}_{[0,1]}(x)$ und $c = 1/\delta_0$. Dann ist Bedingung (6.3.6) erfüllt. Mit stochastisch unabhängigen $\mathcal{R}([0,1])$–verteilten Zufallszahlen u_n, v_n liefert

$$y_k = v_k, \quad \text{falls } k = \inf\left\{n \in \mathbf{N} \mid u_n \leq c f_{\alpha,\beta}(v_n)\right\}$$

eine mit Parametern $\alpha, \beta > 1$ Beta–verteilte Zufallszahl y_k. ∎

Beispiel 6.3.5. (Poisson–Verteilungen)
Wir bestimmen zunächst das Maximum der Folge $\{a_k\}_{k\in\mathbf{N}_0}$ mit $a_k = b^k/k!$, $b > 0$. Es gilt

$$\frac{a_{k+1}}{a_k} = \frac{b}{k+1} \quad \begin{cases} \geq 1, & \text{falls } k+1 \leq b \\ \leq 1, & \text{falls } k+1 \geq b \end{cases},$$

d.h. $\{a_k\}$ steigt monoton, falls $k \leq b$, und ist monoton fallend für $k \geq b$. Das Maximum wird also für $k^* = \lfloor b \rfloor$ mit Wert $b^{k^*}/k^*!$ angenommen.
Für den Quotienten der Zähldichten einer $\mathfrak{G}(p)$- und $\mathfrak{P}(\lambda)$-Verteilung erhalten wir hiermit für alle $p \in (0,1)$, $\lambda > 0$ und $k \in \mathbf{N}_0$

$$\frac{e^{-\lambda}\frac{\lambda^k}{k!}}{(1-p)^k p} = \frac{e^{-\lambda}}{p} \cdot \frac{\left(\frac{\lambda}{1-p}\right)^k}{k!} \leq \frac{e^{-\lambda}}{p} \cdot \frac{\left(\frac{\lambda}{1-p}\right)^{k^*}}{k^*!},$$

wobei $k^* = \lfloor \frac{\lambda}{1-p} \rfloor$. Die Konstante $c = c(\lambda, p) = \dfrac{p(1-p)^{k^*}k^*!}{e^{-\lambda}\lambda^{k^*}}$ stellt also sicher, daß

$$c \cdot \frac{\mathfrak{P}(\lambda)(\{k\})}{\mathfrak{G}(p)(\{k\})} \leq 1 \text{ für alle } k \in \mathbf{N}_0.$$

Sind $\{u_n\}_{n\in\mathbf{N}}$ Standardzufallszahlen und $\{k_n\}_{n\in\mathbf{N}}$ $\mathfrak{G}(p)$-verteilte Zufallszahlen und bezeichnet $s \in \mathbf{N}$ den kleinsten Index n, für den

$$u_n \leq c \cdot \frac{e^{-\lambda}\lambda^{k_n}/k_n!}{(1-p)^{k_n}p},$$

so stammt nach Lemma 6.3.3 die Zufallszahl k_s aus einer $\mathfrak{P}(\lambda)$-Verteilung. Man beachte, daß mit der Verwerfungsmethode zur Erzeugung einer Poisson-verteilten Zufallszahl im Mittel $E(S) = 1/c = (e^{-\lambda}\lambda^{k^*})/(p(1-p)^{k^*}k^*!)$ Paare von Zufallszahlen benötigt werden. (vgl. Übungsaufgabe 6.8) ∎

Im folgenden behandeln wir eine Transformationsmethode, die speziell auf normalverteilte Zufallsvariable zugeschnitten ist.

Lemma 6.3.4. *(Transformation auf Normalverteilung)*
U_1 und U_2 seien stochastisch unabhängige, $\mathcal{R}([0,1])$-verteilte Zufallsvariable. Dann sind die Zufallsvariablen

$$X_1 = \sqrt{-2\ln U_1} \cdot \cos(2\pi U_2) \quad \text{und} \quad X_2 = \sqrt{-2\ln U_1} \cdot \sin(2\pi U_2)$$

stochastisch unabhängig, jeweils $\mathcal{N}(0,1)$-verteilt.

Beweis. Wir wenden Satz 2.1.10 auf die Transformation $G : (0,1)^2 \to \mathbf{R}^2$ mit

$$G(u_1,u_2) = \left(\sqrt{-2\ln u_1} \cdot \cos(2\pi u_2), \sqrt{-2\ln u_1} \cdot \sin(2\pi u_2)\right), \ (u_1,u_2) \in (0,1)^2, \tag{6.3.7}$$

an. Es gilt

$$\det\left(\frac{\partial G_i}{\partial u_j}\right)_{1\leq i,j\leq 2} = \det\begin{pmatrix} -\dfrac{\cos(2\pi u_2)}{u_1\sqrt{-2\ln u_1}} & 2\pi\sqrt{-2\ln u_1}\cos(2\pi u_2) \\ -\dfrac{\sin(2\pi u_2)}{u_1\sqrt{-2\ln u_1}} & -2\pi\sqrt{-2\ln u_1}\sin(2\pi u_2) \end{pmatrix} = \frac{2\pi}{u_1}.$$

Die erste Komponente der Umkehrfunktion berechnet sich zu

$$u_1 = g_1^{-1}(x_1, x_2) = e^{-\frac{x_1^2+x_2^2}{2}}, \quad (x_1, x_2) \in \mathbf{R}^2.$$

Mit (2.1.93) folgt

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2+x_2^2}{2}}, \quad (x_1, x_2) \in \mathbf{R}^2,$$

das Produkt der Dichten von zwei stochastisch unabhängigen, je $\mathcal{N}(0,1)$–verteilten Zufallsvariablen. Die Behauptung folgt mit (1.4.52). ∎

Die Anwendung von Lemma 6.3.4 ist klar. Erzeuge zwei unabhängige Standardzufallszahlen u_1, u_2 und transformiere diese mit der Abbildung (6.3.7). Die resultierenden Zufallszahlen x_1, x_2 entstammen einer Standardnormalverteilung und sind stochastisch unabhängig.

Um den in Kapitel 3.4 vorgestellten Algorithmus zur Bestimmung der konvexen Hülle von Punkten im Einheitskreis mit Simulationsmethoden zu testen, werden zufällige, im Einheitskreis $\bar{K}_1$ gleichverteilte Punkte benötigt. Die zugrundeliegende Verteilung besitzt die Dichte (2.1.96)

$$f(x_1, x_2) = \frac{1}{\pi} \mathbb{1}_{\bar{K}_1}(x_1, x_2).$$

Wir erzeugen die verlangten Zufallsvektoren mit der Transformation (2.1.94). Sind u_1, u_2 unabhängige Standardzufallszahlen, so stammt die zweidimensionale Zufallszahl (x_1, x_2) mit

$$x_1 = \sqrt{u_1} \cdot \cos(2\pi u_2), \quad x_2 = \sqrt{u_1} \cdot \sin(2\pi u_2)$$

aus einer Gleichverteilung auf dem Einheitskreis mit obiger Dichte. Dies wurde am Ende von Kapitel 2.1 bewiesen.

Zum Abschluß diese Kapitels werden wir uns noch mit der Generierung höherdimensionaler Verteilungen mit nicht stochastisch unabhängigen Komponenten beschäftigen, wobei wir exemplarisch die Erzeugung von zufälligen Permutationen ohne Wiederholung und die Simulation von Markoff–Ketten herausgreifen. Eine Anwendung für den letzteren Fall haben wir bereits in Kapitel 3.3 bei der Untersuchung des Simulated Annealing–Algorithmus kennengelernt. Die Erzeugung von Permutationen einer bestimmten Grundmenge spielt beim Austesten von Such– und Sortierverfahren mit Simulationsmethoden eine wichtige Rolle. Wir benötigen zunächst ein vorbereitendes Lemma.

Lemma 6.3.5. *$X_1, \ldots, X_n$ seien Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbf{N}$, deren gemeinsame Verteilung für alle $i_1, \ldots, i_j \in \mathbf{N}$ mit $P(X_1 = i_1, \ldots, X_j = i_j) > 0$, $j = 0, 1, \ldots, n-1$, den Bedingungen*

$$P(X_{j+1} = \ell \mid X_1 = i_1, \ldots, X_j = i_j) = \begin{cases} \frac{1}{n-j}, & \text{falls } \ell \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_j\} \\ 0, & \text{sonst} \end{cases} \tag{6.3.8}$$

genügt. Dann gilt $P\big((X_1,\dots X_n) = (i_1,\dots i_n)\big) = 1/n!$ *für alle* $(i_1,\dots,i_n) \in Perm_n^n(\{1,\dots,n\}; o.W.)$, *d.h.* $P^{(X_1,\dots,X_n)}$ *repräsentiert eine Gleichverteilung auf der Menge* $Perm_n^n(\{1,\dots,n\}; o.W.)$.

Beweis. Für alle $(i_1,\dots,i_n) \in Perm_n^n(\{1,\dots,n\}; o.W.)$ gilt mit Formel (3.1.7), daß

$$\begin{aligned} P(X_1 = i_1,\dots,X_n = i_n) &= P(X_n = i_n \mid X_1 = i_1,\dots,X_{n-1} = i_{n-1})\cdot \\ &\quad \cdot P(X_{n-1} = i_{n-1} \mid X_1 = i_1,\dots,X_{n-2} = i_{n-2}) \cdots P(X_1 = i_1) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{2} \cdot 1 = \frac{1}{n!}. \end{aligned}$$

Man beachte, daß die bedingenden Ereignisse in jedem der obigen Faktoren positive Wahrscheinlichkeit besitzen. ∎

Zur Erzeugung einer Gleichverteilung auf $Perm_n^n(\{1,\dots,n\}; o.W.)$ wird Lemma 6.3.5 folgendermaßen eingesetzt. Speichere in einem Feld `a: ARRAY[1..n] OF INTEGER` die natürlichen Zahlen $1,\dots,n$ durch `a[j]:=j`, $j = 1,\dots,n$ aufsteigend ab. Erzeuge dann unabhängige Standardzufallszahlen $u_1,\dots,u_{n-1} \in (0,1)$, berechne im j-ten Schritt $i_j = j + \lfloor (n-j+1)u_j \rfloor$ und tausche den Inhalt der Zellen `a[j]` und `a[i_j]` aus, $j = 1,\dots,n-1$.

Bezeichnet X_j den Inhalt der Zelle `a[j]`, $j = 1,\dots,n$, nach Ablauf des Verfahrens, so genügen die bedingten Verteilungen der Bedingung (6.3.8), da im j-ten Iterationsschritt durch $j + \lfloor (n-j+1)u_j \rfloor$ eine Gleichverteilung auf den noch nicht ausgewählten Elementen in den Zellen `a[j],...,a[n]` erzeugt wird, $j = 1,\dots,n$. Obiges Verfahren liefert also insgesamt eine zufällige Permutation der Zahlen $\{1,\dots,n\}$.

Bei der Nachbildung des Verhaltens von Markoff-Ketten $\{X_n\}_{n\in\mathbb{N}_0}$ (vgl. Definition 3.2.1) in einem Simulationsmodell sind in der Regel die Anfangsverteilung $\boldsymbol{p}(0)$ und die Übergangsmatrizen Π_n oder ein direktes Konstruktionsprinzip für die Zufallsvariablen X_n bekannt. Damit bieten sich Lemma 3.2.1 oder Lemma 3.2.2 als Basis für die Erzeugung entsprechender Zufallszahlen an. Wir behandeln zunächst den ersten Fall und erläutern den zweiten anschließend an einem Beispiel.

Die Erzeugung einer Zufallszahl $x_0 \in \mathcal{S} = \{1,2,\dots\}$ aus der Anfangsverteilung $\boldsymbol{p}(0) = (p_1(0), p_2(0)\dots)$ auf $(\mathcal{S}, \mathfrak{P}(\mathcal{S}))$ wird mit Hilfe von Lemma 6.3.1 durchgeführt. Bezeichnet u_0 eine Standardzufallszahl, so ist

$$x_0 = k, \quad \text{falls } \sum_{i=1}^{k-1} p_i(0) < u_0 \le \sum_{i=1}^{k} p_i(0),\ k \in \mathcal{S}, \tag{6.3.9}$$

ein Anfangszustand, der der Startverteilung genügt.

Die Erzeugung von Folgezuständen erfolgt nach dem gleichem Prinzip mit der ℓ-ten Zeile $(p_{\ell 1}(n), p_{\ell 2}(n),\dots)$ der Übergangsmatrix Π_n, falls $x_{n-1} = \ell$ der erzeugte Vorgängerzustand war, $n \in \mathbb{N}$. Ist u_n eine von $u_0,\dots,u_{n-1}$ unabhängige Standardzufallszahl, so setze

$$x_n = k, \quad \text{falls } \sum_{i=1}^{k-1} p_{\ell i}(n) < u_n \le \sum_{i=1}^{k} p_{\ell i}(n),\ k \in \mathcal{S}. \tag{6.3.10}$$

Daß die durch (6.3.9) und (6.3.10) erzeugte Folge von Zufallszahlen der Verteilung $P^{\{X_n\}}$ genügt, sieht man folgendermaßen ein. $x_0, x_1, x_2, \ldots$ entstammen der Verteilung stochastisch unabhängiger Zufallsvariablen $Y_0, Y_1, Y_2, \ldots$ mit $P^{Y_0} = P^{X_0}$ und $P^{Y_n} = P^{X_n|X_{n-1}=x_{n-1}}$, $n \in \mathbb{N}$. Für alle $x_0, \ldots, x_n \in \mathcal{S}$, $n \in \mathbb{N}$ gilt

$$P(Y_0 = x_0, \ldots, Y_n = x_n) = P(Y_0 = x_0) \cdot \prod_{j=1}^{n} P(Y_j = x_j)$$

$$p_{x_0}(0) \cdot \prod_{j=1}^{n} p_{x_{j-1}x_j}(j) = P(X_0 = x_0, \ldots X_n = x_n),$$

wobei die letzte Gleichheit aus Lemma 3.2.1 folgt. Verfahren (6.3.9) und (6.3.10) liefert also Realisationen aus einer Markoff–Kette mit der gewünschten Verteilung.

Oft liegen explizite Informationen über die Entstehungsprinzipien einer Markoff–Kette vor. Wir erläutern dies am Beispiel der Irrfahrt auf dem Gitter $\mathcal{G} = \{0, 1, \ldots, r\}$ mit absorbierenden Barrieren (vgl. Beispiel 3.2.1). Die zugehörige Markoff–Kette kann hier für alle $n \in \mathbb{N}$ durch die Rekursion

$$X_n = \begin{cases} \max\{0, \min\{X_{n-1} + Z_n, r\}\}, & \text{falls } 1 \le X_{n-1} \le r-1 \\ X_{n-1}, & \text{falls } X_{n-1} \in \{0, r\} \end{cases}$$

mit $P(Z_n = 1) = p$, $P(Z_n = -1) = 1 - p$, $0 < p < 1$, für unabhängige Zufallsvariable $\{Z_n\}_{n\in\mathbb{N}}$ und $P(X_0 = k) = 1$ für ein $k \in \{0, \ldots, r\}$ rekursiv wie in Lemma 3.2.2 konstruiert werden. Zur Erzeugung von Realisationen aus $\{X_n\}_{n\in\mathbb{N}_0}$ benutzen wir die Funktion

$$f(i, z) = \begin{cases} \max\{0, \min\{i + z, r\}\}, & \text{falls } 1 \le i \le r-1 \\ i, & \text{falls } i \in \{0, r\} \end{cases}$$

aus Beispiel 3.2.1 a) und setzen

$$x_0 = k, \quad x_n = f(x_{n-1}, z_n), \; n \in \mathbb{N},$$

wobei z_n unabhängige Zufallszahlen aus einer Zweipunktverteilung auf $\{-1, 1\}$ sind, die wie in (6.3.3) durch $z_n = 2y_n - 1$ mit $y_n = \mathbb{1}_{[0,p)}(u_n)$ aus unabhängigen Standardzufallszahlen u_n generiert werden können.

Ähnliches gilt für die Simulation stochastischer Prozesse: Markoff–Prozesse lassen sich etwa mit Hilfe des Struktursatzes 3.4.5 erzeugen, indem man zuerst die eingebettete Markoff–Kette mit den Übergangswahrscheinlichkeiten aus (3.4.25) nach dem gerade beschriebenen Verfahren generiert und anschließend die Zwischenankunftszeiten in Abhängigkeit von der eingebetteten Markoff–Kette gemäß (3.4.26). Poisson'sche Punktprozesse lassen sich dagegen nach dem in Lemma 3.4.7 beschriebenen Verfahren leicht erzeugen, indem man zunächst die voneinander unabhängigen Poisson–verteilten *Anzahlen* der Punkte in den einzelnen Partitionsmengen B_n, $n \in I$, erzeugt (etwa nach dem in Beispiel 6.3.5 vorgeschlagenen Verfahren) und anschließend unabhängig von dieser Vorlaufphase gemäß den in (3.4.48) spezifizierten Verteilungen entsprechend viele *Realisationen* der Punkte.

In der Praxis werden dabei in der Regel vor allem die Dimensionen eins, zwei und drei (letztere insbesondere bei der Bildverarbeitung) von Interesse sein.

Ist die Intensität des Prozesses hoch, d.h. sind im Mittel "viele" Punkte je Partitionsmenge zu erzeugen, kann man sich auch der Normalapproximation der Poisson–Verteilungen bedienen, um den Simulationsaufwand zu begrenzen, indem man — etwa gemäß Lemma 6.3.4 — unabhängige normalverteilte Zufallszahlen Z_n, $n \in I$, erzeugt und

$$N_n = \lfloor \sqrt{\mu(B_n)} Z_n + \mu(B_n) \rfloor^+$$

(Positivteil) setzt. Die Abschätzung (2.3.47) ergibt dann einen Approximationsfehler von maximal ca. 5% je Partitionsmenge (bezüglich der Metrik ρ), wenn jeweils $\mu(B_n) \geq 8000$, $n \in I$, ist. Dies zeigt die folgende Überlegung:

Ist $\lambda > 0$ und sind $X_1, \ldots, X_n$ unabhängige, je $\mathfrak{P}(\lambda/n)$–verteilte Zufallsvariablen mit $n = \lceil \lambda \rceil$, so ist auch $X = \sum_{k=1}^{n} X_k$ $\mathfrak{P}(\lambda)$–verteilt. Man hat jetzt nur noch zu beachten, daß für eine $\mathfrak{P}(\mu)$–verteilte Zufallsvariable Y mit $0 < \mu \leq 1$

$$E\big[(Y-\mu)^4\big] = 8\mu^2 + 2\mu \leq 10$$

gilt (Berechnung über erzeugende Funktionen), also eine Anwendung der Jensen'schen Ungleichung, Lemma 2.2.1,

$$E\big[|Y-\mu|^3\big] \leq \big\{E\big[(Y-\mu)^4\big]\big\}^{3/4} \leq 10^{3/4} \leq 5.624$$

ergibt. Einsetzen aller Werte in (2.3.47) liefert dann die Aussage.

6.4. Aufgaben

6.1 Seien $a, b \in \mathbb{N}_0$, $a, b \leq m-1$, $u, v \in \mathbb{N}$, u und v teilerfremd und $m = u \cdot v$. Gilt $a \bmod u = b \bmod u$ und $a \bmod v = b \bmod v$, so ist notwendig $a = b$.

6.2 p sei eine Primzahl und $a \in \mathbb{N}$, $a \not\equiv 1$ (modulo p). Zeigen Sie die Gültigkeit der folgenden Äquivalenz für alle $t, n \in \mathbb{N}$, $n \geq 2$. $\dfrac{a^n - 1}{a - 1} \equiv 0$ (modulo p^t) gilt genau dann, wenn $a^n - 1 \equiv 0$ (modulo p^t).

6.3 Der "Kleine Satz von Fermat" besagt: Ist p eine Primzahl und $a \in \mathbb{N}$, a und p teilerfremd, so folgt $a^{p-1} \equiv 1$ (modulo p). Folgern Sie hieraus, daß $a^p \equiv a$ (modulo p) und $a^p \equiv a$ (modulo pa) gilt.

6.4 Für den multiplikativen Kongruenzgenerator (6.1.12) $z_{n+1} = (az_n) \bmod 2^t$ verwende man $a = 65541$ und $t = 32$. Man zeige, daß dieser Generator für ungerade Startwerte z_0 maximale Periodenlänge besitzt, daß aus je drei aufeinanderfolgenden Zufallszahlen gebildete dreidimensionale Vektoren (u_n, u_{n+1}, u_{n+2}) aber auf höchstens 34 Ebenen in $[0,1]^3$ liegen. Wie lautet die maximale Anzahl von Hyperebenen, wenn die Werte $a = 2^{t+3}$ und $m = 4^t$, $t \in \mathbb{N}$ verwendet werden?
Hinweis: Verwenden Sie die Methode aus Beispiel 6.1.1.

6.5 Finden Sie mit Hilfe von Simulated Annealing eine obere Schranke für die Lösung des Minimierungsproblems (5.1.23) bei Verwendung der Gitterbasis (5.1.21) für Dimension p=2,3,4. Untersuchen Sie mit der erstellten Prozedur für einige Werte von a und $m = 2^t$, die nach Satz 5.1.2 maximale Periodenlänge sicherstellen, die normierten, maximalen Hyperebenenabstände (5.2.1).

6.6 Man betrachte den Fibonacci-Generator $u_{n+1} = (u_n + u_{n-1}) - \lfloor u_n + u_{n-1} \rfloor$, $n \in \mathbb{N}$, mit zwei Startwerten $0 \leq u_0, u_1 < 1$. Zeigen Sie, daß in der so definierten Folge die Anordnungen $u_{n-1} < u_{n+1} < u_n$ und $u_n < u_{n+1} < u_{n-1}$ für kein $n \in \mathbb{N}$ auftreten. Mit welcher Wahrscheinlichkeit tritt die Anordnung $u_{n-1} < u_{n+1} < u_n$ in einer Folge von stochastisch unabhängigen Standardzufallszahlen $\{u_n\}_{n \in \mathbb{N}}$ unendlich oft auf?

6.7 X sei eine $\mathcal{E}(\lambda)$-verteilte Zufallsvariable, $\lambda > 0$. Man zeige, daß $Y = \lfloor X \rfloor$ $\mathcal{G}(p)$-verteilt ist mit $p = 1 - e^{-\lambda}$.

6.8 Sei $0 < p < 1$ und $\lambda_n = n \cdot p$, $n \in \mathbb{N}$. Wie in Beispiel 5.3.5 werden mit der Verwerfungsmethode $\mathfrak{P}(\lambda_n)$-verteilte Zufallszahlen erzeugt. Man zeige, daß für den Erwartungswert $E(S_n)$ der Wartezeit bis zur Erzeugung einer $\mathfrak{P}(\lambda_n)$-verteilten Zufallszahl gilt

$$\lim_{n \to \infty} \frac{\sqrt{n}}{e^{np}} E(S_n) = \frac{1}{p\sqrt{2\pi}}.$$

6.9 Erzeugen Sie mit der Verwerfungsmethode stochastisch unabhängige Zufallszahlen aus einer Dreiecksverteilung mit der Dichte

$$f(x) = \begin{cases} x, & \text{falls } 0 \leq x \leq 1, \\ 2 - x, & \text{falls } 1 \leq x \leq 2, \\ 0, & \text{sonst.} \end{cases}$$

Wie groß ist bei Ihrem Verfahren die erwartete Anzahl von Standardzufallszahlen bis zur ersten erzeugten dreiecksverteilten Zufallszahl? Bestimmen Sie $F^{-1}(u)$, $0 \leq u \leq 1$, wenn F die Verteilungsfunktion zu obiger Dichte bezeichnet. Wie erzeugt man mit Hilfe von F^{-1} dreiecksverteilte Zufallszahlen?

Literatur

Aarts, E.H.L. und **van Laarhoven, P.J.M.** (1985): Statistical cooling: a general approach to combinatorial optimization problems. Philips J. Res. 40, 193 – 226.

Aarts, E.H.L. und **van Laarhoven, P.J.M.** (1987): Simulated annealing: a pedestrian review of the theory and some applications. In: Pattern Recognition Theory and Applications, eds.: P.A. Devijer und J. Kittler, NATO ASI Series, Vol. F30, Springer, N.Y.

Abramowitz, M. und **Stegun, I.A.** (1984): *Pocketbook of Mathematical Functions.* Harri Deutsch, Thun und Frankfurt/Main.

Aigner, M. (1988): *Combinatorial Search. Wiley–Teubner Series in Computer Science.* Teubner, Stuttgart und Wiley, Chichester.

Afflerbach, L. und **Lehn, J.** (1986): *Kolloquium über Zufallszahlen und Simulationen.* Darmstadt, 21. März 1986. Teubner, Stuttgart.

Ash, R. (1965): *Information Theory.* Wiley, N.Y.

Barbour, A.D. und **Hall, P.** (1984): On the rate of Poisson convergence. Math. Proc. Cambridge Philos. Soc. 95, 473 – 480.

Barth, G. (1982): Analyzing algorithms by Markov chains. In: Meth. Operat. Res. 45, 405 – 418.

Barth, G. (1984): An analytical comparison of two string searching algorithms. Inform. Proc. Letters 18, 249 – 256.

Bauer, H. (1978): *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie.* 3. Aufl. de Gruyter, Berlin.

Benedetto, J.J. (1976): *Real Variable and Integration.* Teubner, Stuttgart.

Billingsley, P. (1986): *Probability and Measure.* 2nd ed., Wiley, N.Y.

Bolch, G. (1989): *Leistungsbewertung von Rechensystemen mittels analytischer Warteschlangenmodelle.* Leitfäden und Monographien der Informatik. Teubner, Stuttgart.

Borgwardt, K.H., Gaffke, N., Jünger, M. und **Reinelt, G.** (1989): Computing the convex hull in the Euclidean plane in linear expected time. Arbeitsber. Nr. 139, DFG–Schwerpunktprogramm *Anwendungsbezogene Optimierung und Steuerung*, Universität Augsburg.

Breiman, L. (1968): *Probability.* Addison–Wesley, Reading, Mass.

Çinlar, E. (1975): *Introduction to Stochastic Processes.* Prentice–Hall, Englewood Cliffs.

Daley, D.J. und **Vere–Jones, D.** (1988): *An Introduction to the Theory of Point Processes.* Springer Series in Statistics, Springer, N.Y.

Deheuvels, P. und **Pfeifer, D.** (1988a): On a relationship between Uspenky's theorem and Poisson approximations. Ann. Inst. Stat. Math. 40, 671 – 681.

Deheuvels, P. und **Pfeifer, D.** (1988b): Poisson approximations of multinomial distributions and point processes. J. Multivar. Analysis 25, 65 – 89.

Encarnação, J. und **Straßer, W.** (1986): *Computer Graphics.* 2. Aufl., Oldenbourg, München.

Floret, K. (1981): *Maß – und Integrationstheorie.* Teubner Studienbücher Mathematik. Teubner, Stuttgart.

Gallager, R.G. (1968): *Information Theory and Reliable Communication.* Wiley, N.Y.

Gänssler, P. und **Stute, W.** (1977): *Wahrscheinlichkeitstheorie.* Springer Hochschultext, Springer, Berlin.

Geman, S. und **McClure, D.E.** (1987): Statistical methods for tomographic image reconstruction. In: Bulletin of the 46th Session of the International Statistical Institute, Tokyo 1987, Vol. 4, 1 – 17.

Gläßer, L. (1987): Berechnung der Parameter von diskreten Hidden–Markov–Modellen mit Gradientenprojektionsmethoden. Anwendung auf die automatische Spracherkennung. Siemens Forsch.–u. Entwickl.–Ber. 16, 147 – 151.

Graham, R.L. (1972): An efficient algorithm for determining the convex hull of a finite planar set. Inform. Proc. Letters 1, 132 – 133.

Hájek, J. (1981): *Sampling from a Finite Population.* M. Dekker, N.Y.

Heiss, W.–D., Herholz, K., Pawlik, G., Szelies, B., Wagner, R. und **Wienhard, K.** (1988): Positronen–Emissions–Tomographie. Messung des regionalen zerebralen Glukosestoffwechsels. Neurologie Psychiatrie, Sonderheft 2, 47 – 55.

Heuser, H. (1989): *Lehrbuch der Analysis. Teil 1,2.* Mathematische Leitfäden. Teubner, Stuttgart.

Hunter, J.J. (1983): *Mathematical Techniques of Applied Probability. Vol. 1: Discrete Time Models: Basic Theory, Vol. 2: Discrete Time Models: Techniques and Applications.* Ac. Press, N.Y.

Kall, P. (1976): Mathematische Methoden des Operations Research. Teubner Studienbücher Mathematik. Teubner, Stuttgart.

Kallenberg, O. (1986): *Random Measures.* 4th ed., Ac. Press, London.

Karr, A.F. (1986): *Point Processes and their Statistical Inference.* M. Dekker, N.Y.

Kemp, R. (1984): *Fundamentals of the Average Case Analysis of Particular Algorithms.* Wiley–Teubner Series in Computer Science. Teubner, Stuttgart und Wiley, Chichester.

Kiyek, K. und **Schwarz, F.** (1989): *Mathematik für Informatiker 1,2.* Leitfäden und Monographien der Informatik. Teubner, Stuttgart.

Knuth, D.E. (1968,1969,1973): *The Art of Computer Programming. Vol. 1: Fundamental Algorithms,* 2nd ed., *Vol. 2: Seminumerical Algorithms, Vol. 3: Sorting and Searching.* Addison–Wesley, Reading, Mass.

Lauritzen, S.I. und **Spiegelhalter, D.J.** (1988): Local computations with probabilities on graphical structures and their application to expert systems. J. Roy. Statist. Soc. B, 50, 157 – 224.

Mathar, R. und **Mann, A.** (1990): Analyzing the CSA–protocol by Markov chains. Erscheint in: IEEE Trans. Inf. Th.

Mehlhorn, K. (1988): *Datenstrukturen und effiziente Algorithmen.* Leitfäden und Monographien der Informatik. Teubner, Stuttgart.

Oberschelp, W. und **Wille, D.** (1976): *Mathematischer Einführungskurs für Informatiker.* Teubner Studienbücher Informatik. Teubner, Stuttgart.

Pfeifer, D. (1985): An average–case analysis for a continuous random search algorithm. Adv. Appl. Prob. 17, 231 – 233.

Pfeifer, D. (1989a): *Einführung in die Extremwertstatistik.* Teubner Skripten zur Mathematischen Stochastik. Teubner, Stuttgart.

Pfeifer, D. (1989b): Extremal processes, secretary problems and the $1/e$ law. J. Appl. Prob. 26, 722 – 733.

Pfeifer, D. (1990): Some remarks on Nevzorov's record model. Erscheint in: J. Appl. Prob.

Pflug, G. (1986): *Stochastische Modelle in der Informatik.* Leitfäden und Monographien der Informatik. Teubner, Stuttgart.

Ripley, B.D. (1988): *Statistical Inference for Spatial Processes.* Camb. University Press, Cambridge.

Ritter, H., Martinez, T. und **Schulten, K.** (1990): *Neuronale Netze.* Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke. Addison–Wesley, Bonn, München.

Rösler, U. (1988): A limit theorem for quicksort. Technical Report, Universität Göttingen.

Ross, S.M. (1983): *Stochastic Processes.* Wiley, N.Y.

Sedgewick, R. (1988): *Algorithms.* 2nd ed., Addison–Wesley, Reading, Mass.

Schmitz, N. und **Lehmann, F.**(1985): *Monte–Carlo–Methoden I: Erzeugen und Testen von Zufallszahlen.* Skripten zur Math. Statistik Nr. 2, Münster, 3. Aufl.

Topsøe, F. (1974): *Informationstheorie.* Teubner Studienbücher Mathematik. Teubner, Stuttgart.

Valentine, F.A. (1968): *Konvexe Mengen.* BI Hochschultaschenbücher 402/402a. Bibliographisches Institut, Mannheim.

Vardi, Y., Shepp, L.A. und **Kaufman, L.** (1985): A statistical model for positron emission tomography. JASA 80, 8 – 20 und 34 – 37.

Weiß, P. (1987): *Stochastische Modelle für Anwender.* Mathematische Methoden in der Technik. Teubner, Stuttgart.

Symbolverzeichnis

Symbol	Erklärung
$\mathbf{N}$	natürliche Zahlen
$\mathbf{N}_0$	$\mathbf{N} \cup \{0\}$
$\overline{\mathbf{N}}_0$	$\mathbf{N}_0 \cup \{\infty\}$
$\mathbf{Z}$	ganze Zahlen
$\mathbf{Z}^+$	nicht–negative ganze Zahlen
$\mathbf{R}$	reelle Zahlen
$\mathbf{R}^+$	nicht–negative reelle Zahlen
(a,b)	offenes Intervall
$(a,b]$	links–offenes, rechts–abgeschlossenes Intervall
$[a,b)$	links–abgeschlossenes, rechts–offenes Intervall
$[a,b]$	abgeschlossenes Intervall
$A \times B$, $\bigtimes_{i \in I} A_i$	kartesisches Produkt der Mengen A und B, A_i
A^n	n–faches kartesisches Produkt der Menge A
$A^{\mathbf{N}}$	Menge der Folgen mit Gliedern aus A
$\#(A)$	Anzahl der Elemente der Menge A
$\mathfrak{P}(\Omega)$	Potenzmenge der Menge Ω
P	Wahrscheinlichkeitsverteilung
$(\Omega, \mathcal{A}, P)$	Wahrscheinlichkeitsraum
$(\mathcal{X}, \mathcal{B})$	Meßraum
$P(\cdot \mid B)$, $P(\cdot \mid Y = y)$	bedingte Wahrscheinlichkeit unter der Hypothse B, $Y = y$
$\mathcal{B}^m$	Borel'sche σ–Algebra über $\mathbf{R}^m$
$\sigma(\mathcal{E})$	vom Mengensystem $\mathcal{E}$ erzeugte σ–Algebra
$\delta(\mathcal{E})$	vom Mengensystem $\mathcal{E}$ erzeugtes Dynkin–System
$\mathcal{A} \otimes \mathcal{B}$, $\bigotimes_{i \in I} \mathcal{A}_i$	Produkt–σ–Algebra der σ–Algebren $\mathcal{A}$ und $\mathcal{B}$, $\{\mathcal{A}_i\}$
X	Zufallsvariable
$A \cap \mathcal{B}$	Spur–σ–Algebra der Menge A in $\mathcal{B}$
P^X	Verteilung der Zufallsvariablen X
F	Verteilungsfunktion
F^{-1}	Pseudo–Inverse von F
$E(X)$	Erwartungswert der Zufallsvariablen X
$E(X \mid B)$, $E(X \mid Y = y)$	bedingter Erwartungswert unter B, $Y = y$
$\mathrm{Var}(X)$	Varianz der Zufallsvariablen X
$\mathrm{Var}(X \mid Y = y)$	bedingte Varianz von X unter $Y = y$
$\mathrm{Kov}(X,Y)$	Kovarianz der Zufallsvariablen X und Y
$\mathrm{Korr}(X,Y)$	Korrelation der Zufallsvariablen X und Y
$\mathbf{1}_A$	Indikatorfunktion der Menge A
$P * Q$, $\mathop{*}_{i \in I} P_i$	Faltung der Verteilungen P und Q, $\{P_i\}$
$P \otimes Q$, $\bigotimes_{i \in I} P_i$	Produktmaß der Verteilungen P und Q, $\{P_i\}$
$X_n \xrightarrow{P} X$	stochastische Konvergenz der Zufallsvariablen $\{X_n\}$ gegen X

$Q_n \xrightarrow{w} Q$	schwache Konvergenz der Verteilungen $\{Q_n\}$ gegen Q
$\mathfrak{L}(\Omega)$	diskrete Gleichverteilung (Laplace–Verteilung) über Ω
$\mathfrak{B}(n,p)$	Binomialverteilung mit Parametern n und p
$\overline{\mathfrak{B}}(n,p)$, $\overline{\mathfrak{B}}^+(n,p)$	Negative Binomialverteilung mit Parametern n und p
$\mathfrak{G}(p)$, $\mathfrak{G}^+(p)$	Geometrische Verteilung mit Parameter p
$\mathfrak{P}(\lambda)$	Poisson–Verteilung mit Parameter λ
$\mathfrak{PB}(n;p_1,\ldots,p_n)$	Poisson–Binomialverteilung mit Parametern n und $p_1,\ldots,p_n$
$\mathfrak{M}(n;p_1,\ldots,p_n)$	Multinomialverteilung mit Parametern n und $p_1,\ldots,p_n$
$\mathcal{R}(A)$	stetige Gleichverteilung (Rechteck–Verteilung) über A
$\mathcal{E}(\lambda)$	Exponentialverteilung mit Parameter λ
$\mathcal{E}(n,\lambda)$	Erlang–Verteilung mit Parametern n und λ
$\mathcal{N}(\mu,\sigma^2)$	Normalverteilung mit Parametern μ und σ^2
$\Gamma(\alpha,\lambda)$	Gamma–Verteilung mit Parametern α und λ
χ^2_k	χ^2–Verteilung mit k Freiheitsgraden
$\Gamma(\alpha)$	Gamma–Funktion
$B(\alpha,\beta)$	Beta–Funktion
F_P, F_X	Verteilungsfunktion von P, der Verteilung von X
$H(X)$	Entropie der Zufallsvariablen X
$H(X \mid Y)$	bedingte Entropie von X unter Y
$\bar{H}(\{X_n\})$	Entropie der stationären Folge $\{X_n\}$
$\lfloor x \rfloor$	größte ganze Zahl kleiner gleich x
$\lceil x \rceil$	kleinste ganze Zahl größer gleich x
∂B	Rand der Menge B
$\rho(P,Q)$	Kolmogoroff–Metrik; Abstand der Verteilungen P und Q
$\|\ \|$	Euklidische Norm
$\perp$	senkrecht zu
A^{tr}	Transponierte der Matrix A
$\det(A)$	Determinante der Matrix A
$\int X\,dP$	Lebesgue–Integral der Zufallsvariablen X nach der Verteilung P
$\Psi(t)$	momenterzeugende Funktion
$\psi(s)$	wahrscheinlichkeitserzeugende Funktion
ln	natürlicher Logarithmus
$\log_a$	Logarithmus zur Basis a
$O(n)$	Landau–Symbol
ξ	Punktprozeß
$E\xi$	Intensitätsmaß von ξ
τ_B	Evolutionsabbildung
γ	Euler'sche Konstante: $\gamma = 0.5772156649...$
$\mathrm{conv}(A)$	konvexe Hülle der Menge A
$a \bmod m = b$	b ist der Rest bei Teilen von a durch m
$a \equiv b (\mathrm{modulo}\ m)$	a und b haben den gleichen Rest bei Teilen durch m
$\sum_{x \in \mathbb{R}} f(x)$	Summe über die höchstens abzählbar vielen von Null verschiedenen Werte der Zähldichte $f(x)$

Index