PROBABILISTIC CONCEPTS OF APPROXIMATION THEORY IN CONNEXION WITH OPERATOR SEMIGROUPS

Dietmar Pfeifer

Institut für Statistik und Wirtschaftsmathematik
RWTH Aachen, Wüllnerstr. 3
D-5100 Aachen, West Germany

Received Jan. 16, 1985

Introduction

It was in the middle of the 1930's that E. Hille started to develop the theory of operator semigroups ending up with his famous monograph entitled "functional Analysis and Semi-Groups" [15] published in 1948. A second enlarged edition co-written with R.S. Philips appeared about a decade later [16]; here a first connexion of semigroups with probability theory was pointed out, originating from a personal communication of the authors with M. Riesz. He as well as Kendall [19] was certainly inspired by Bernstein's [1] famous proof of Weierstrass' [32] approximation theorem for continuous functions, given completely in probabilistic terms; however, their approach to semigroup theory was not consequently pursued except for a paper by K.L. Chung [8] in 1962. Unfortunately, the basic formula [11, p.157] which could have been used to develop a more general probabilistic representation theory for operator semigroups was only given in a heuristic framework here, without formal proof.

Besides the investigation of problems concerning representation theory for operator semigroups as such, attention was early drawn also to approximation - theoretic aspects in this field (for instance by Butzer [4], Hsu [17], Butzer and Berens [5], Ditzian [10 -12], Shaw [30], Butzer and Hahn [6], and most recently by the author [26], [29]. Interestingly enough, in the papers before 1980 no explicit reference to probability theory was made, although probabilistic arguments were used here throughout (such as moment calculations and Markov-type inequalities or other estimations of what we call tail-probabilities). In fact, it is possible to describe representation theory for operator semigroups completely in probabilistic terms as has been shown by the author [21,23,25,26,29]; under mild assumptions on the form of representation theorems, this approach is even exhaustive [24,26]. The full power of these probabilistic methods, however, turns out when approximation-theoretic questions in this area are investigated [26,27,29]. In fact, probability theory enables us to simplify to a great extend most of the relevant estimations being involved here, and even to extend them - thanks to the unified setting - to the widest possible generality. This gives a deep insight into the structure of such approximation problems, providing at the same time simple and illustrative answers to questions that had remained open for almost ten years (cf. Hsu [17], Butzer and Berens [5], and Ditzian [10,11]).

In this paper, we shall mainly make use of some elementary probability theory, although for rigorous proofs of some theorems, a more advanced level is necessary. A reader feeling less familiar with this subject may have a look at one of the basic textbooks on probability theory, for instance Billingsley [3].

1. Preliminaries: basic definitions and properties

Throughout this paper, we consider a strongly continuous one-parameter operator semigroup $\{T(t);\ t\geq 0\}$ being a subset of the Banach algebra $\mathcal{E}(\mathcal{X})$ of bounded endomorphisms of a Banach space $\mathcal{X}$ with norm $\|\cdot\|$, characterized by the three conditions

$$\begin{aligned} &T(s+t) = T(s)T(t), \quad s,t\geq 0 \quad \text{(the semigroup property);}\\ &T(0) = I \quad \text{(the identity operator)}\ ; \\ &\lim_{t\downarrow 0}\|T(t)f-f\| = 0\ , \quad f\in\mathcal{X}. \end{aligned} \tag{1.1}$$

Although the last condition is a local continuity condition only, strong continuity is readily obtained by means of the semigroup property; further, there exists constants $M\geq 1$ and $\omega\geq 0$ such that

$$\|T(t)\| \leq M\, e^{\omega t}\ , \quad t\geq 0\ . \tag{1.2}$$

An important tool in semigroup theory is the infinitesimal generator. A with domain $D(A)\subseteq\mathcal{X}$, given by

$$Af = \lim_{h\downarrow 0}\frac{1}{h}(T(h)-I)f = \lim_{h\downarrow 0} A_h f\ , \quad f\in D(A) \tag{1.3}$$

where D(A) is the set of all $f \in \mathfrak{X}$ fulfilling the limit relation (1.3). Obviously, A plays the role of a differential operator, and it can be shown that D(A) is a dense subspace of $\mathfrak{X}$ (correspondingly for the powers A^r, $r \geq 1$, of A). With respect to the structure of semigroups, we have to distinguish between two different cases, i.e. whether A is a bounded or unbounded operator. In the first case, the semigroup has a unique representation

$$T(t) = e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k , \quad t \geq 0 , \tag{1.4}$$

implying also that the semigroup is even uniformly continuous. Conversely, every uniformly continuous operator semigroup has a bounded generator, hence is of the form (1.4). In the second case, no such formula is available; however, several limit relations can be established in this situation, even with estimations for the rates of convergence. For instance, Hille's first exponential formula [15] states

$$T(t)f = \lim_{h \downarrow 0} \exp(A_h t)f , \qquad f \in \mathfrak{X} , \quad t \geq 0 \tag{1.5}$$

which means that although A might not be bounded, the semigroup can be approximated by uniformly continuous ones, even uniformly in t in every bounded interval. In 1960, Hsu [17] gave an estimation for the rate of convergence for Hille's formula in terms of the rectified modulus of continuity

$$\omega^b(\delta,f) = \sup\{\|T(t)f-T(s)f\| ; \; 0 \leq s,t \leq b, \; |s-t| < \delta\} , \quad f \in \mathfrak{X}, \; \delta > 0 , \tag{1.6}$$

he proved that

$$\|\exp(A_h t)f-T(t)f\| \leq \omega^b(n^{1/3},f)+K\|f\|h^{1/3}, \quad f \in \mathfrak{X} , \; h>0 \tag{1.7}$$

(such that $t+h^{1/3}<b$) where K is independent of f, h and t (but possibly dependent of b, M and ω).

He also raised the question whether the exponent of h in (1.7) could be simultaneously increased for both summands (cf. also Butzer and Berens [5]). A positive answer for this problem was given by Ditzian [10] who proved that $h^{1/3}$ could be replaced by h^{χ} with $0<\chi<1/2$, and that χ could not be extended to values larger than 1/2. He also gave a complete treatise for the case $x=1/2$ [11]

Another important operator in semigroup theory is the resolvent $R(\lambda)$ defined as the Laplace transform

$$R(\lambda)f = \int_0^\infty e^{-\lambda s}T(s)f\,ds \tag{1.8}$$

which is a bounded linear operator for $\lambda>\omega$; in this case also

$$R(\lambda) = (\lambda I - A)^{-1} \tag{1.9}$$

which shows its connexion with the infinitesimal generator A (see also Butzer and Berens [5] for a more general discussion). Since $\{\lambda R(\lambda);\ \lambda>\omega\}$ forms a strong approximation process on $\mathfrak{X}$, two further semigroup representations - in terms of the resolvent - are of special interest, due to Widder and Phillips (see [16]):

$$T(t)f = \lim_{n\to\infty} \{\tfrac{n}{t}R(\tfrac{n}{t})\}^n f, \qquad f\in\mathfrak{X} \tag{1.10}$$

$$T(t)f = \lim_{\lambda\to\infty} \exp(-t\lambda I + t\lambda^2 R(\lambda))f, \qquad f\in\mathfrak{X}. \tag{1.11}$$

For these representations, Ditzian [12] showed that

$$\|\{\tfrac{n}{t}R(\tfrac{n}{t})\}^{n+1}f - T(t)f\| \leq K\omega^b(n^{-1/2},f), \quad f\in\mathfrak{X}, \tag{1.12}$$

where $0\leq t<b-\delta$ ($0<\delta<b$ being fixed), and $n>\omega b+1$, and

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \leq L\omega^b(\lambda^{-1/2},f), \quad f\in\mathfrak{X} \tag{1.13}$$

where again $0\leq t<b-\delta$, and $\lambda>\omega$. Here, K and L are independent of n, λ and t. Better estimations are obtained for $f\in D(A)$; in this case,

$$\|\{\tfrac{n}{t}R(\tfrac{n}{t})\}^{n+1}f - T(t)f\| \leq K^*n^{-1/2}\omega^b(n^{-1/2},Af), \quad n>\omega b+1 \tag{1.14}$$

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \leq L^*\lambda^{-1/2}\omega^b(\lambda^{-1/2},Af), \qquad \lambda>\omega, \tag{1.15}$$

where again K^* and L^* are independent of n, λ and t. The rate results in (1.14) and (1.15) are best possible as was also shown by Ditzian (loc. cit.), i.e. even for $f\in\bigcap_{r=1}^{\infty} D(A^r)$, the results cannot be improved except for the constants involved.

Surprisingly enough, it is possible to give very elementary probabilistic proofs for the estimations (1.12) to (1.15), which at the same time allow for extensions of these to arbitrary semigroup representations (27). This will be treated in more detail in chapter 3.

Also, it can be seen that the representation theorems mentioned above have a certain probabilistic form originating from a special version of the famous law of large numbers for a random number of summands. This aspect will be worked out in more detail in the following chapter.

2. The probabilistic setting of semigroup representations

We begin with a simple intuitive approach via the Widder inversion formula (1.10). Writing

$$\lambda R(\lambda)f = \int_0^\infty \lambda e^{-\lambda s} T(s)\, f\, ds \;, \quad f \in \mathscr{X} \;, \; \lambda > \omega \;, \tag{2.1}$$

we see that this expression could also be thought of as some "expectation" $E[T(X)]f$ with an exponentially distributed random variable X with mean $1/\lambda$. Since for $\lambda \to \infty$, this distribution tends to the Dirac measure ε_0 concentrated in 0, and applying the strong continuity of the semigroup, we have

$$\lambda R(\lambda)f = E[T(X)]f \to T(0)f = f \;, \quad \text{for} \quad \lambda \to \infty \;, \tag{2.2}$$

by weak convergence in the probabilistic sense (which here is at the same time strong convergence in $\mathscr{X}$), showing in a probabilistic way that $\{\lambda R(\lambda);\ \lambda > \omega\}$ is a strong approximation process on $\mathscr{X}$. Similarly, if $\lambda = 1/t$ $(t > 0)$ and $\{X_k;\ k \in \mathbb{N}\}$ are independent copies of X, we have

$$\begin{aligned}\{\tfrac{n}{t} R(\tfrac{n}{t})\}^n f &= \prod_{k=1}^n E[\,T(\tfrac{X_k}{n})]\, f = E[\,\prod_{k=1}^n T(\tfrac{X_k}{n})]f \\ &= E[\,T(\tfrac{1}{n}\sum_{k=1}^n X_k)]f \to T(t)f \;, \quad f \in \mathscr{X} \quad (n \to \infty),\end{aligned} \tag{2.3}$$

by the independence of $\{X_k\colon k \in \mathbb{N}\}$ (giving the second equality) and the (simple) weak law of large numbers (giving the limit relation), i.e. $\frac{1}{n}\sum_{k=1}^n X_k \to E(X) = t$ in distribution (in fact, under our conditions, convergence holds even almost surely). Relation (2.3) is basically also given in Chung's paper [8], but was not proved to hold generally. Of course, this relation needs some clarification. In this context, three main problems have to be considered:

(A) The measurability of the mapping $t \mapsto T(t)$, and the precise definition of an "expectation" $E[T(X)]$ as some element of $\mathcal{E}(\mathcal{X})$.

(B) The possibility of interchanging product and expectation under (stochastic) independence (as is true in the case of merely real - valued random variables).

(C) The extension of weak convergence of measures to the case of operator-valued random variables.

Concerning the first part of A), there is a negative result in general.

Theorem 2.1 If $\liminf_{t \downarrow t_0} \|T(t)-T(t_0)\| > 0$ for some $t_0>0$ (i.e. the semigroup is not uniformly continuous from the right in some point $t_0>0$), then the mapping $t \mapsto T(t)$ is neither strongly nor Borel-measurable (i.e. measurable with respect to the σ-field generated by the operator topology), nor separably valued.

Proof see [25].

A simple example of a non-measurable semigroup is the semigroup of left translations on the space $\mathcal{X} = USC(\mathbb{R})$ of all uniformly continuous and bounded functions on $\mathbb{R}$, given by

$$T(t)f(x) = f(x+t), \quad x \in \mathbb{R}, \quad t \geq 0, \quad f \in \mathcal{X}. \tag{2.4}$$

Here always $\|T(t)-T(s)\| = 2$ for all $s,t \geq 0$, $s \neq t$. The question whether a strongly continuous operator semigroup is generally weakly measurable remains undecided since the dual space $\mathcal{E}(\mathcal{X})^*$ is not explicitly known in most cases. We thus do not know whehter $E[T(X)]$ possibly exists as a Pettis integral in $\mathcal{E}(\mathcal{X})$. However, a slight modification of Pettis' integral as introduced in [25] gives a solution of this problem.

Definition 2.1 Let $(\mathcal{M}, \mathcal{F}, \mu)$ be a measure space and $S: \mathcal{M} \to \mathcal{E}(\mathcal{X})$ a mapping such that $f^*(S(\cdot)f)$ is measurable (in the ordinary sense) for all $f \in \mathcal{X}$ and $f^* \in \mathcal{X}^*$. S is called μ-integrable if there exists an element $J \in \mathcal{E}(\mathcal{X})$ such that

$$f^*(J(f)) = \int_{\mathcal{M}} f^*(S(\cdot)f)d\mu \quad \text{for all } f \in \mathfrak{X}, \ f^* \in \mathfrak{X}^*. \tag{2.5}$$

J is then called the μ-integral of S: $J = \int_{\mathcal{M}} S\, d\mu$. If μ is a probability measure, then J will also be called expectation of S: $J = E(S)$.

The point here is that the integral can (by the Hahn-Banach Theorem) already be uniquely defined by less linear functionals than the whole dual space $\mathcal{E}(\mathfrak{X})^*$. A sufficient condition for the existence of the (extended) integral is the following [25].

Lemma 2.1 If $S(\cdot)f$ is Borel-measurable and separably valued for every $f \in \mathfrak{X}$ such that $\|S(\cdot)\|$ is dominated by some μ-integrable function $g \geq 0$ (in the ordinary sense), then $S(\cdot)$ is extended Pettis-integrable, and

$$\left\| \int_{\mathcal{M}} S\, d\mu \right\| \leq \int_{\mathcal{M}} g\, d\mu \tag{2.6}$$

holds.

In our situation, $S(\cdot) = T(X)$ with $\|T(X)\| \leq Me^{\omega X}$, and $T(X)f$ is measurable by the strong continuity of $T(\cdot)f$, hence $E[T(X)]$ exists uniquely as an element of $\mathcal{E}(\mathfrak{X})$ whenever $E(e^{\omega X}) = \psi_X^*(\omega)$, the moment-generating function (or Laplace transform) of X, exists at ω. Relation (2.6) then translates into

$$\|E[T(X)]\| \leq M\, \psi_X^*(\omega) . \tag{2.7}$$

The concept of extended Pettis integration also answers problem B). In fact, the following result holds [25].

Theorem 2.2 If X and Y are non-negative, independent real random variables such that for the moment-generating functions, $\psi_X^*(\omega) < \infty$, $\psi_Y^*(\omega) < \infty$, then

$$E[T(X)T(Y)] = E[T(X)]\, E[T(Y)] . \tag{2.8}$$

Note that by independence, $\psi_{X+Y}^*(\omega) = \psi_X^*(\omega)\psi_Y^*(\omega)$, hence by the semigroup property, $E[T(X)T(Y)]$ exists as an element of $\mathcal{E}(\mathfrak{X})$.

Similarly, problem C) can be solved, at least for the law of large numbers [25].

Theorem 2.3 If $\{X_k\colon k\in\mathbb{N}\}$ are independent copies of a non-negative real random variable X for which $\psi_X^*(\delta) < \infty$ for some positive δ, and $E(X)=t$, then for $n>\omega/\delta$, $E[T(\frac{1}{n}\sum_{k=1}^{n} X_k)] \in \mathscr{E}(\mathscr{X})$, and the limit relation

$$\|E[T(\frac{1}{n}\sum_{k=1}^{n} X_k)]\| \le M\{\psi_X^*(\frac{\omega}{n})\}^n \to M e^{\omega t} \qquad (n\to\infty) \tag{2.9}$$

as well as weak convergence in the probabilistic sense holds under operators, i.e.

$$E[T(\frac{1}{n}\sum_{k=1}^{n} X_k)]\, f \to T(t)f\ , \qquad f\in\mathscr{X} \quad (n\to\infty)\ . \tag{2.10}$$

(in the strong sense in $\mathscr{X}$)

In fact, it can be shown that all known representation formulas in product form are of such probabilistic type [25], and that under some positivity conditions, only such probabilistic representations are possible [24]. This again emphasizes the importance of probabilistic methods in the analysis of such approximation problems.

Before we are going to specialize on different distributions in (2.10) in order to reobtain the known representation theorems, we shall develop a further (seemingly more general, but in fact equivalent) representation theorem. For this purpose, the probability generating function ψ_N of a non - negative, integer - valued random variable N will be needed, given by $\psi_N(s) = \psi_N^*(\log s)$, $s>0$. Equivalently, $\psi_N(s) = \sum_{k=0}^{n} P(N=k)\ s^k$, $s\ge 0$, which explains for the name. A further useful concept in the probabilistic approach to semigroup theory is that of a random sum of random variables as was already pointed out by Chung [8], i.e. we consider $X = \sum_{k=1}^{N} Y_k$, where $\{Y_k\colon k\in\mathbb{N}\}$ are independent copies of some non-negative random variable Y which is independent of N. Then $\psi_X^*(\cdot) = \psi_N(\psi_Y^*(\cdot))$, and it can be proved that if $\psi_N(\psi_Y^*(\omega)) < \infty$, then

$$E[T(X)] = \psi_N(E[T(Y)]) = \sum_{k=0}^{\infty} P(N=k)\ \{E[T(Y)]\}^k \tag{2.11}$$

holds (see [25]). In this setting, Theorem 2.3 can be formulated as a product representation formula.

Theorem 2.4 Let N be a non-negative integer-valued random variable and Y be a non-negative real random variable such that $\psi_N(\delta_1) < \infty$ for some $\delta_1 > 1$ and $\psi_Y^*(\delta_2) < \infty$ for some $\delta_2 > 0$. Then the expectations $E(N) = \zeta$ and $E(Y) = \gamma$ (say) exist, and for sufficiently large n, $S_n = \psi_N(E[T(\frac{Y}{n})]) \in \mathcal{E}(\mathfrak{X})$ with

$$\|S_n\| \leq M\ \psi_N(\psi_Y^*(\tfrac{\omega}{n})) \ . \tag{2.12}$$

Further, a strong semigroup representation in product form

$$\{\psi_N(E[T(\tfrac{Y}{n})])\}^n f \to T(t)f, \qquad f \in \mathfrak{X} \quad (n \to \infty) \tag{2.13}$$

holds with $t = \zeta\gamma$.

It is also possible to establish a continuous analogue of relation (2.13) by means of stochastic processes instead of sequences of random variables [25].

Theorem 2.5 Let $\{N(\tau);\ \tau \geq 0\}$ be a stochastic process ranging through the non-negative integers and Y be a non-negative real random variable, fulfilling the following conditions:

$$\psi_{N(\tau)}(\delta_1) < \infty \quad \text{for some } \delta_1 > 1 \text{ and all } \tau; \tag{2.14}$$

$$\psi_Y^*(\delta_2) < \infty \quad \text{for some } \delta_2 > 0; \tag{2.15}$$

$$\limsup_{\tau \to \infty} \psi_{N(\tau)}(\psi_Y^*(\tfrac{r\omega}{\tau})) < \infty \quad \text{for some } r > 1; \tag{2.16}$$

$$\frac{1}{\tau} N(\tau) \to \zeta \in \mathbb{R} \quad (\tau \to \infty) \text{ in probability.} \tag{2.17}$$

Then for sufficiently large τ, $S_\tau = \psi_{N(\tau)}(E[T(\frac{Y}{\tau})]) \in \mathcal{E}(\mathfrak{X})$ with

$$\|S_\tau\| \leq M\ \psi_{N(\tau)}(\psi_Y^*(\tfrac{\omega}{\tau})) \ . \tag{2.18}$$

Further, a strong semigroup representation of the form

$$\psi_{N(\tau)}(E[T(\tfrac{Y}{\tau})]f \to T(t)f \ , \qquad f \in \mathfrak{X} \quad (\tau \to \infty) \tag{2.19}$$

holds with $t = \zeta\gamma$, where $\gamma = E(Y)$.

The following example shows that relation (2.19) also covers the discrete version (2.13): simply take independent copies

$\{N_k;\ k\in\mathbb{N}\}$ of N and let $N(\tau)=\sum_{1\leq k\leq\tau} N_k$, $\tau\geq 0$. Then by the law of large numbers, $\frac{1}{\tau}N(\tau)\to\xi=E(N)$ $(\tau\to\infty)$ even almost surely, and $\psi_{N(\tau)}=\psi_N^{int(\tau)}$. For $\tau\in\mathbb{N}$, (2.13) now follows from (2.19) (note that in this example, conditions (2.14) to (2.16) are fulfilled under the assumptions of Theorem 2.4).

It should be pointed out that the proof of Theorem 2.5 depends on a version of the law of large numbers for random sums, i.e.

$$\frac{1}{\tau}\sum_{k=1}^{N(\tau)} Y_k \to \zeta\gamma \quad \text{in probability} \quad (\tau\to\infty) \tag{2.20}$$

(see [25], Lemma 2) where $\{Y_k:\ k\in\mathbb{N}\}$ are independent copies of Y, indenendent of the process $\{\dot{N}(\tau);\ \tau\geq 0\}$.

The two most important subcases of Theorems 2.4 and 2.5 are given by $Y\equiv\gamma$ being a constant, leading to first main theorems, and Y being exponentially distributed with mean γ, leading to second main theorems (involving the resolvent). In these cases, relation (2.19) translates into

$$\psi_{N(\tau)}(T(\tfrac{t}{\zeta\tau}))f \to T(t)f\ , \qquad f\in\mathcal{X} \qquad (\tau\to\infty) \tag{2.21}$$

$$\psi_{N(\tau)}(\tfrac{\zeta\tau}{t}R(\tfrac{\zeta\tau}{t}))f \to T(t)f, \qquad f\in\mathcal{X} \qquad (\tau\to\infty). \tag{2.22}$$

For instance, if $\{N(\tau);\ \tau\geq 0\}$ is a Poisson process with parameter t (i.e. $E(N(\tau))=\tau t$), we have $\psi_{N(\tau)}(s)=\exp(\tau t(s-1))$, $s\geq 0$, giving Hille's first exponential formula (1.5) and Phillips' exponential formula (1.11) (with $\zeta=t$ here) while Widder's inversion formula (1.10) is obtained from (2.13) with $N\equiv 1$. A similar distinction as above can also be made for Theorem 2.4. We then have

$$\{\psi_N(T(\tfrac{\zeta}{n}))\}^n f \to T(t)f, \qquad f\in\mathcal{X} \qquad (n\to\infty) \tag{2.23}$$

$$\{\psi_N(\tfrac{n}{\zeta}R(\tfrac{n}{\zeta}))\}^n f \to T(t)f, \quad f\in\mathcal{X} \qquad (n\to\infty) \tag{2.24}$$

where again $t=\zeta\gamma$.

Considering especially binomial and geometric distributions over $\{0,1\}$ and $\{0,1,2,\cdots\}$, respectively, the probability generating functions are

$$\psi_N(s) = (1-\zeta) + \zeta s \ , \qquad s \geq 0 \tag{2,25}$$

and

$$\psi_N(s) = \frac{1}{1+\zeta-\zeta s} \ , \qquad 0 \leq s < 1+1/\zeta \ . \tag{2.26}$$

respectively. This gives rise to the following representation theorems.

$$\{(1-t)I + tT(\tfrac{1}{n})\}^n f \to T(t)f \ , \quad f \in \mathscr{X} \ (n\to\infty)(\text{Kendall [19]}) \tag{2.27}$$

$$\{(1+t)I - tT(\tfrac{1}{n})\}^{-n} f \to T(t)f \ , \quad f \in \mathscr{X} \ (n\to\infty)(\text{Shaw [30]}) \tag{2.28}$$

$$\{2I - T(\tfrac{t}{n})\}^{-n} f \to T(t)f, \qquad f \in \mathscr{X} \ (n\to\infty)(\text{Shaw [30]}) \tag{2.29}$$

$$\{(1-t)I + t\, nR(n)\}^n f \to T(t)f, \quad f \in \mathscr{X} \ (n\to\infty)(\text{Chung [8]}) \tag{2.30}$$

$$\{(1+t)I - t\, nR(n)\}^{-n} f \to T(t)f, \quad f \in \mathscr{X} \ (n\to\infty)(\text{Chung [8]}) \tag{2.31}$$

$$\{2I - \tfrac{n}{t} R(\tfrac{n}{t})\}^{-n} f \to T(t)f, \quad f \in \mathscr{X} \ (n\to\infty)(\text{Pfeifer [26]}) \tag{2.32}$$

Notice that in Chung's formula (2.31) it is not necessary to assume $\omega=0$ as in [8], and that relations (2.27) and (2.30) are only valid for $0 \leq t \leq 1$ in the general case (for uniformly continuous semigroups, see [24], Theorem 2).

Of course, a lot of further representation theorems of probabilistic type are immediately available, among them product formulas as (2.13) even with unequal factors [25], or formulas in which the semigroup $T(t)$ is replaced by a truncated Taylor series if the infinitesimal generator A is bounded (see [23-26]).

Conversely, the following theorem holds [24].

Theorem 2.6 Let ψ_t be a real analytic function is some interval $[0,\delta]$, $\delta > 1$, with non-negative coefficients. Then if

$$\{\psi_t(T(\tfrac{1}{n}))\}^n f \to T(t)f, \quad f \in \mathscr{X} \quad (n\to\infty) \tag{2.33}$$

or

$$\{\psi_t(nR(n))\}^n f \to T(t)f, \quad f \in \mathscr{X} \quad (n\to\infty) \tag{2.34}$$

holds for an arbitrary strongly continuous non-periodic operator semigroup with $\|T(t)\| > 0$, then ψ_t is necessarily the probability generating function of a non-negative integer-valued random variable N with E(N)=t, i.e. the representations (2.33) and (2.34) are probabilistic.

For a slightly more general version of this theorem, see [26]. In fact, we are not aware of a representation theorem for strongly continuous operator semigroups in product form which is not probabilistic (i.e. for which ψ_t has at least one negative coefficient; see also [24], Theorem 2). However, for uniformly continuous semigroups, extensions from probabilistic to non-probabilistic representations are possible (loc. cit.).

3. Implications for approximation theory

a. Direct theorems

In the preceding chapters it has become apparent that generating functions of random variables in connexion with the law of large numbers are the main tool for the derivation of strong representation theorems in semigroup theory. We shall show here that with respect to approximation-theoretic questions, the variance of the underlying random variables together with the corresponding generating functions will play the central role in this area. For this purpose, a more general Taylor expansion of the semigroup as developed in [26] and [29]) will be needed.

Theorem 3.1 Let $r \geq 1$ and $f \in D(A^r)$. Then for arbitrary s, $t \geq 0$,

$$T(t)f-T(s)f = \sum_{k=1}^{r-1} \frac{(t-s)^k}{k!} T(s)A^k f + \int_s^t \frac{(t-u)^{r-1}}{(r-1)!} T(U)A^r f\, du . \tag{3.1}$$

Since the remainder part in formula (3.1) is a strongly measurable function of the variables s and t by the strong continuity of the semigroup, we can easily derive a probabilistic estimation for the rate of convergence for the different representation theorems worked out in chapter 2. For instance, if X is a non-negative real random variable with expectation E(X)=t such that for the moment-generating function, $\psi_X^*(3\omega) < \infty$, we have

$$E[T(X)]f-T(t)f = \frac{\sigma^2}{2} T(t)A^2 f + E\Big[\int_t^X \frac{(x-u)^2}{2} T(u)A^3 f\, du\Big] \tag{3.2}$$

for $f \in D(A^3)$, where $\sigma^2 = \sigma^2(X)$ denotes the variance $E((X-t)^2)$ of X.

Moreover, the remainder expectation can be estimated by

$$\|A^3 f\| \frac{M}{6} \{e^{\omega t} E(|X-t|^3) + \omega E((X-t)^4 e^{\omega X})\} \tag{3.3}$$

which itself is dominated by (say)

$$\|A^3 f\| \frac{M}{6} \{ e^{\omega t} E(|X-t|^3) + \omega \{E((X-t)^6)\}^{2/3} \cdot \sqrt[3]{\psi_X^*(3\omega)}\} \tag{3.4}$$

which can be concluded from Hölder's inequality. Since by Theorem 2.3, in semigroup representation theorems typically random variables of the form $\frac{1}{n}\sum_{k=1}^{n} X_k$ are considered where $\{X_k : k \in \mathbb{N}\}$ are independent copies of X, and in the case of independence, variances are additive, we obtain from (3.2) to (3.4):

$$E[T(\frac{1}{n}\sum_{k=1}^{n} X_k)]f - T(t)f = \frac{\sigma^2}{2n} T(t)A^2 f + O(n^{-3/2}) \quad (n \to \infty) \tag{3.5}$$

for $f \in D(A^3)$ which means that the best rate of approximation in any of the discrete semigroup representations covered by Theorems 2.3 and 2.4 is $O(n^{-1})$ for $n \to \infty$ (and $O(\tau^{-1})$ for Theorem 2.5 for $\tau \to \infty$, respectively, if for the variance $\sigma^2(N(\tau)) = o(\tau^2)$; (see [29]). Using an extension of the famous central limit theorem in probability theory based on uniform integrability conditions (which here are fulfilled by the existence of the moment-generating functions) it is even possible to give exact general estimations for the rate of convergence for all probabilistic representation theorems in terms of the underlying generating functions. For the sake of simplicity, we shall present some of these results under the situation of Theorem 2.4 only (for further details, see [29]).

Theorem 3.2 Under the assumptions and with the notations of Theorem 2.4, we have

$$\begin{aligned} &\| \{\psi_N(E[T(\tfrac{Y}{n})])\}^n f - T(t)f \| \\ &\quad \leq M e^{\omega t} \|Af\| \Big\{ \frac{1}{\sqrt{n}} \sqrt{\zeta\sigma^2(Y) + \gamma^2\sigma^2(N)} + \cdots \\ &\quad + \frac{4\omega}{\eta^2 n} \psi_N(\psi_Y^*(\eta)) \exp\{\frac{\omega^2 \psi_N(\psi_Y^*(\eta))}{n(\eta - 2\omega/n)^2}\} \Big\} \end{aligned} \tag{3.6}$$

for $n > \max(\frac{2\omega}{\eta}, \frac{60}{\psi_N(\psi_Y^*(\eta))})$ and $0 < \eta \leq \delta$, if $f \in D(A)$;

$$\|\{\psi_N(E[T(\frac{Y}{n})])\}^n f - T(t)f\| \leq \frac{M}{2}\, e^{\omega t}\|A^2 f\|\Big\{\frac{1}{n}(\zeta\sigma^2(Y)+\gamma^2\sigma^2(N))+\cdots$$

$$+\ \frac{4\omega}{\eta^3 n\sqrt{n}}\ \{\psi_N(\psi_Y^*(\eta))\}^{3/2}\exp\{\frac{2\omega^2\psi_N(\psi_Y^*(\eta))}{n(\eta-4\omega/n)^2}\}\Big\} \tag{3.7}$$

for $n > \max(\frac{4\omega}{\eta}, \frac{60}{\psi_N(\psi_Y^*(\eta))})$ and $0<\eta\leq\delta$, if $f\in D(A^2)$;

$$\{\psi_N(E[T(\frac{Y}{n})])\}^n f - T(t)f$$

$$= \frac{1}{2n}(\zeta\sigma^2(Y)+\gamma^2\sigma^2(N))\ T(t)A^2 f + O(n^{-3/2}) \tag{3.8}$$

for $n\to\infty$, if $f\in D(A^3)$, where the remainder can be estimated by

$$\frac{M}{6}\ e^{\omega t}\ \|A^3 f\|\Big\{\frac{3}{n\ \sqrt{n}}\ \frac{1}{\eta^3}\{\psi_N(\psi_Y^*(\eta))\}^{3/2} + \frac{8\omega}{\eta^4 n^2}\{\psi_N(\psi_Y^*(\eta))\}^2\times\cdots$$

$$\times\exp\{\frac{\omega^2\psi_N(\psi_Y^*(\eta))}{n(\eta-3\omega/n)^2}\}\Big\} \tag{3.9}$$

for $n > \max(\frac{3\omega}{\eta}, \frac{45}{\psi_N(\psi_Y^*(\eta))})$ and $0<\eta\leq\delta$.

It should be pointed out that in the above relations (3.6) to (3.9), the expression $\zeta\sigma^2(Y)+\gamma^2\sigma^2(N)$ is just the variance of the composed random variable $X=\sum_{k=1}^{N} Y_k$ (cf. p.11, and Lemma 3 in Chung [8]).

Of course, refined estimations for the remainder terms above are possible for the individual representation theorems using characteristic properties of the underlying distributions; likewise for Theorem 2.5. For instance, in Hille's and Phillips' exponential formulas (1.5) and (1.11), we can provide the following (probabilistic) estimations.

Theorem 3.3 Let $t\geq 0$. Then

$$\|\exp(A_h t)f-T(t)f\| \leq M\, e^{\omega t}\|Af\|\{\sqrt{ht}+2\omega ht\ \exp(\omega^2 hte^{2\omega h})\} \tag{3.10}$$

for $h\leq t$, if $f\in D(A)$;

$$\|\exp(A_h t)f-T(t)f\| \leq \frac{M}{2}e^{\omega t}\|A^2 f\|\{ht+2\sqrt{2h^3t^3}\ \exp(2\omega^2 hte^{4\omega h})\} \tag{3.11}$$

for $h\leq t$, if $f\in D(A^2)$;

$$\exp(A_h t)f - T(t)f = \frac{ht}{2} T(t)A^2 f + O(h^{3/2}) \quad (h \to 0) \tag{3.12}$$

for $f \in D(A^3)$, where the remainder term can be estimated by

$$\frac{M}{6} e^{\omega t} \|A^3 f\| \{3h^{3/2}\exp(\frac{t}{2} e^{\sqrt{h}}) + 38\omega h^2 \exp(\frac{t}{3}e^{\sqrt{h}} + \frac{3}{2}\omega^2 hte^{3\omega h})\} . \tag{3.13}$$

Also,

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \le Me^{\omega t}\|Af\| \left\{\sqrt{\frac{2t}{\lambda}} + 4 \frac{e^t}{\lambda} \exp(\frac{2\omega^2 t}{\lambda - 2\omega})\right\} \tag{3.14}$$

for $\lambda > \max(4; 2\omega)$, if $f \in D(A)$;

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \le \frac{M}{2}e^{\omega t}\|A^2 f\|\{ \frac{2t}{\lambda} + 6 \frac{e^t}{\sqrt{\lambda^3}} \exp(\frac{4\omega^2 t}{\lambda - 4\omega})\} \tag{3.15}$$

for $\lambda > \max(16; 4\omega)$, if $f \in D(A^2)$;

$$\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f = \frac{t}{\lambda} T(t)A^2 f + O(\lambda^{-3/2}) \quad (\lambda \to \infty) \tag{3.16}$$

for $f \in D(A^3)$, where the remainder term can be estimated by

$$\frac{M}{6} e^{\omega t}\|A^3 f\| \left\{ \frac{3}{\sqrt{\lambda^3}} e^{2t/3} + \frac{38\omega}{\lambda^2} e^t \exp(\frac{3\omega^2 t}{\lambda - 3\omega})\right\} \text{ for } \lambda > \max(9;3\omega). \tag{3.17}$$

The proof of this (which is given in [29]) depends on the fact that for a Poisson process $\{N(\tau);\ \tau > 0\}$ with parameter t we have

$$\psi^*_{N(\tau)}(s) = \exp(\tau t(e^s - 1)) , \quad s \ge 0 , \tag{3.18}$$

and that for an exponentially distributed random variable Y with mean 1,

$$\psi^*_Y(s) = \frac{1}{1-s} , \quad 0 \le s < 1 . \tag{3.19}$$

Note that corresponding direct theorems for Widder's inversion formula (1.10) are covered by Theorem 3.2 with $\sigma^2(Y) = t^2$, $\zeta = 1$

and $\sigma^2(N) = 0$ (for details, see [26] or [29], relation (4.37)). Finally, it should be pointed out that also improvements of the rate of approximation in the representation theorems are possible using Bernstein's (2) and Voronovskaja's (31) approach of variance elimination. For instance, for Theorems 2.4 and 3.2, we have the following result [29].

Theorem 3.4 Under the conditions of Theorem 2.4, we have

$$\|\{\psi_N(E[T(\tfrac{Y}{n})])\}^n f - T(t)f - \frac{1}{2n}(\zeta\sigma^2(Y)+\gamma^2\sigma^2(N))\cdot$$
$$\cdot\{\psi_N(E[T(\tfrac{Y}{n})])\}^n A^2 f\| = O(n^{-2}) \quad (3.20)$$

for $n \to \infty$, if $f \in D(A^4)$, which in general is the best possible rate.

It is possible to improve the order of approximation to any order $O(n^{-k})$ for $n \to \infty$ $(k>2)$ by successive application of the above method (and for f being smooth enough); however, since moments higher than the third are involved in this case, it seems hardly possible to give a nice simple formula here. If, however, binomial distributions are considered (i.e. Kendall's [19] representation, see (2.27)), the corresponding moments still are easily to compute. In this case, the following result holds true [26].

Theorem 3.5 For $0 \leq t \leq 1$ and $f \in D(A^6)$ we have

$$\|T(t)f - \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^k \{T(\tfrac{k}{n})f - \frac{1}{2n}\, t(1-t)T(\tfrac{k}{n})A^2 f - \cdots$$
$$- \frac{1}{6n^2}\, t(1-t)(1-2t)T(\tfrac{k}{n})A^3 f + \frac{1}{8n^2} t^2(1-t)^2 T(\tfrac{k}{n})A^4 f\}\|$$
$$= O(n^{-3}) \quad (3.21)$$

for $n \to \infty$.

The proof of this theorem depends on the fact that if $X_1,\cdots,X_n$ are independent binomially distributed random variables over $\{0,1\}$ with mean t, then

$$E((\bar{X}_n-t)^3) = \frac{1}{n^2}\, t(1-t)(1-2t)\,,$$
$$E((\bar{X}_n-t)^4) = \frac{3}{n^2}\, t^2(1-t)^2 + \frac{1}{n^3}\, t(1-t)(6t^2-6t+1) \quad (3.22)$$

where $\bar{X}_n = \sum_{k=1}^{n} X_k$ denotes the arithmetic mean of $X_1,\cdots,X_n$.

b. Moduli estimations

In this section, we shall come back to questions treated in chapter 1 concerning estimations of semigroup representations in terms of various kinds of moduli of continuity. Besides the rectified modulus of continuity introduced in (1.6) we shall also deal with the local modulus of continuity ω^* given by

$$\omega^*(\delta, t, f) = \sup\{\|T(t)f - T(s)f\| \ ; \quad s \geq 0, \quad |s-t| < \delta\} \tag{3.23}$$

for $\delta > 0$, $t \geq 0$, $f \in \mathfrak{X}$, and the second modulus of continuity

$$\omega_2(t,f) = \sup\{\|(T(s) - I)^2 f\| \quad ; \quad 0 \leq s \leq t\} \tag{3.24}$$

for $t \geq 0$, $f \in \mathfrak{X}$. Since for $b > 0$, we have

$$\omega^b(\delta, f) \geq \sup\{\omega^*(\delta,t,f); \quad 0 \leq t \leq b-\delta\} \tag{3.25}$$

for $\delta < b$, we only need to consider the local modulus ω^*; relation (3.25) then allows for an immediate translation of the results obtained to those involving the rectified modulus ω^b. Following an idea of Chung [8], writing

$$\|E[T(X)]f - T(t)f\|$$

$$\leq \left[\int_{|X-t| \leq \varepsilon} + \int_{|X-t| > \varepsilon}\right] \|T(X)f - T(t)f\| \, dP \tag{3.26}$$

for a suitable non-negative random variable X with mean $E(X) = t$ (cf. (3.2)), where P is the underlying probability measure, and using exponential tail-probability estimations for $P(|X-t|>\varepsilon)$ via Markov's inequality and the moment-generating function, the following fundamental estimation can be obtained [26].

Theorem 3.6 If for the moment generating function, $\psi_X^*(\delta) < \infty$ for some $\delta > 2\omega$, then for arbitrary $\varepsilon > 0$, $\eta \leq \delta$, $f \in \mathfrak{X}$ we have

$$\|E[T(X)]f - T(t)f\| \leq \omega^*(\varepsilon,t,f) + Me^{-\varepsilon\eta}\|f\|(e^{-2\eta t}\psi_X^*(2\eta) + \cdots$$

$$+ e^{2\eta t}\psi_X^*(-2\eta))^{\frac{1}{2}}\left(\sqrt{\psi_X^*(2\omega)} + e^{\omega t}\right) . \tag{3.27}$$

An immediate consequence of Theorem 3.6 is the following result.

Theorem 3.7 Under the assumptions and with the notations of Theorem 2.4, we have

$$\|\{\psi_N(E[T(\frac{Y}{n})])\}^n f - T(t)f\| \leq \omega^*(n^{-\chi},t,f) + Me^{\omega t}\|f\| e^{-n^{(\frac{1}{2}-\chi)}} \times \cdots$$

$$\times \sqrt{2}\, \exp\left(\frac{\psi_N(\psi_Y^*(\eta))}{(\eta-2/\sqrt{n})^2}\right) \left\{1+\exp\left(\frac{\omega^2\psi_N(\psi_Y^*(\eta))}{n(\eta-2\omega/\sqrt{n})^2}\right)\right\} \qquad (3.28)$$

for $n > \max(\frac{4}{\eta^2}, \frac{2\omega}{\eta})$ and $0<\eta<\delta$, where $0<\chi<\frac{1}{2}$.

Of course, similar estimations are also possible in the setting of Theorem 2.5. For the sake of simplicity, we shall only state the corresponding results for Hille's and Phillips' formulas (1.7) and (1.11).

Theorem 3.8 For $f \in \mathfrak{X}$ and $0<\chi<\frac{1}{2}$, we have

$$\|\exp(A_h t)f - T(t)f\| \leq \omega^*(h^{\chi},t,f) + Me^{\omega t}\|f\| \; e^{-h^{(\chi-\frac{1}{2})}} \times \cdots$$

$$\times \sqrt{2}\, \exp(t\; e^{2\sqrt{h}})\; \{1+\exp(\omega^2 ht\; e^{2\omega h})\}, \quad h>0 \qquad (3.29)$$

and

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \leq \omega^*(\lambda^{-\chi},t,f) + Me^{\omega t}\|f\| e^{-\lambda^{(\frac{1}{2}-\chi)}} \times \cdots$$

$$\times \sqrt{2}\, \exp\left(\frac{2t\sqrt{\lambda}}{\sqrt{\lambda}-2}\right) \left\{1+\exp\left(\frac{2\omega^2 t}{\lambda-2\omega}\right)\right\}, \quad \lambda > \max(4;2\omega). \qquad (3.30)$$

Of course, Theorem 3.8 immediately leads to (1.7) for $\chi = \frac{1}{3}$, and to Ditzian's [12] results. Note that corresponding estimations for Widder's inversion formula (1.10) are covered by Theorem 3.7. From a probabilistic point of view, it is no surprise that in theorems 3.7 and 3.8, no extension to the case $\chi = \frac{1}{2}$ or more is possible in general. This is again due to the central limit theorem which says that under our conditions, if again $\bar{X}_n$ denotes the arithmetic means of $X_1,\cdots,X_n$ (being distributed as X), we have $\sqrt{n}(\bar{X}_n - t) \xrightarrow{\mathcal{L}} Z$ $(n\to\infty)$ where Z is a normally distributed random variable with zero mean and variance $\sigma^2=\sigma^2(X)$. Thus for the tail probabilities $P(|\bar{X}_n - t| > \varepsilon/\sqrt{n})$ we have a strictly positive limit for $n\to\infty$ given by $P(|Z|>\varepsilon)$ for every $\varepsilon>0$. In fact, it is possible to show that the counter example of Ditzian [10, Example 3.2] using the translation semigroup (cf. (2.4)) not only works for Hille's exponential formula (1.5), but also generally for probabilistic representation theorems.

Finally, we want to show that also for estimations involving the second modulus of continuity ω_2 the variance of the underlying random variables plays the central role as was claimed in Butzer and Hahn [6] only for groups of isometric operators.

Theorem 3.9 Let us assume that $\{T(t);\ t \geq 0\}$ is a contraction semigroup, i.e. $M = 1$ and $\omega=0$. Then, udner the conditions of Theorem 2.5, we have

$$\| \psi_{N(\tau)}(E[T(\tfrac{Y}{\tau})])f - T(t)f\| \leq K\omega_2(\frac{1}{\sqrt{2\tau}}\ \{\zeta\sigma^2(Y) + \frac{\gamma^2}{\tau}\ \sigma^2(N(\tau))\}^{\frac{1}{2}}, f) \tag{3.31}$$

for $f \in \mathscr{X}$ and $\tau > 0$.
Alternatively, with the notations of Theorme 2.4, we have

$$\| \{\psi_N(E[T(\tfrac{Y}{n})])\}^n f - T(t)f\| \leq K\omega_2(\frac{1}{\sqrt{2n}}\ \{\zeta\sigma^2(Y) + \gamma^2\sigma^2(N)\}^{\frac{1}{2}}, f) \tag{3.32}$$

for $f \in \mathscr{X}$ and $n \in \mathbb{N}$.
In both cases K denotes a generic positive constant. Note that due to $\omega=0$, the conditions on the existence of the generating functions in Theorem 2.4 need not to be imposed here (see [29]). To give some examples, the corresponding estimations for the representations (2.27) to (2.32) are listed below (cf. [26]).

Corollary 3.1 For $n \in \mathbb{N}$ and appropriate choices of t, we have for $f \in \mathscr{X}$

$$\| \{(1-t)I + tT(\tfrac{1}{n})\}^n f - T(t)f\| \leq K\omega_2(\sqrt{\frac{t(1-t)}{2n}}, f), \tag{3.33}$$

$$\| \{(1+t)I - tT(\tfrac{1}{n})\}^{-n} f - T(t)f\| \leq K\omega_2(\sqrt{\frac{t(1+t)}{2n}}, f), \tag{3.34}$$

$$\| \{2I - T(\tfrac{t}{n})\}^{-n} f - T(t)f\| \leq K\omega_2(\frac{t}{\sqrt{n}}, f), \tag{3.35}$$

$$\| \{(1-t)I + nR(n)\}^n f - T(t)f\| \leq K\omega_2(\sqrt{\frac{t(2-t)}{2n}}, f), \tag{3.36}$$

$$\| \{(1+t)I - tnR(n)\}^{-n} f - T(t)f\| \leq K\omega_2(\sqrt{\frac{t(2+t)}{2n}}, f), \tag{3.37}$$

$$\| \{2I - \tfrac{n}{t}R(\tfrac{n}{t})\}^{-n} f - T(t)f\| \leq K\omega_2(\sqrt{\frac{3t^2}{2n}}, f). \tag{3.38}$$

Correspondingly, for Hille's and Phillips' exponential formulas (1.5) and (1.11), the following estimations are valid.

Corollary 3.2 For h, $\lambda>0$ and $f \in \mathscr{X}$ we have

$$\|\exp(A_h t)f - T(t)f\| \leq K\omega_2\left(\sqrt{\frac{ht}{2}}\,,\ f\right) \quad , \tag{3.39}$$

$$\|\exp(-t\lambda I + t\lambda^2 R(\lambda))f - T(t)f\| \leq K\omega_2\left(\sqrt{\frac{t}{\lambda}}\,,\ f\right) . \tag{3.40}$$

For the remainder of this chapter we shall discuss in short some probabilistic aspects of Ditzian's [12] estimations in relations (1.12) to (1.15) which were worked out in [27]. In fact, some of Ditzian's (loc. cit.) ideas can be used to establish the following fundamental probabilistic estimations.

Theorem 3.10 Let $0 \leq t \leq b$ and assume that X is a random variable which is concentrated on the interval $[0,b]$ with mean $E(X) = t$. Then ψ_X^* exists everywhere, and $\sigma^2 = \sigma^2(X)$ is finite. In this case, for all $\varepsilon > 0$, we have

$$\|E[T(X)]f - T(t)f\| \leq (1 + \frac{\sigma}{\varepsilon})\omega^b\ (\varepsilon\,,\ f) \quad \text{for} \quad f \in \mathscr{X} \ , \tag{3.41}$$

$$\|E[T(X)]f - T(t)f\| \leq \sigma(1 + \frac{\sigma}{\varepsilon})\ \omega^b(\varepsilon\,,\ Af) \ \text{for} \ f \in D(A). \tag{3.42}$$

Especially, if X is replaced by the arithmetic mean $\bar{X}_n$ of independent copies of X as in Theorem 2.3, we have $\sigma^2(\bar{X}_n) = \sigma^2/n$, hence with the choice $\varepsilon = n^{-\frac{1}{2}}$, the factors of ω^b in (3.41) and (3.42) become independent of n. By a splitting technique as in (3.26) now Ditzian's [11,12] estimations are similarly reobtained, and even proved to hold basically for all probabilistic representation theorems. Moreover, the factors for ω^b are now given a probabilistic meaning since they can be expressed by the variances of the underlying random variables as well as the generating functions being involved (cf. Theorems 3.6 and 3.7). For instance, Theorem 2.4 can be reformulated in this setting as follows.

Theorem 3.11 Under the conditions and with the notations of Theorem 2.4, if $0 \leq t < b - \delta$ ($0 < \delta < b$ being fixed), there exist constants K and L such that for sufficiently large n, we have

$$\|\{\psi_N(E[T(\tfrac{Y}{n})])\}^n f - T(t)f\| \leq K\omega^b(n^{-\frac{1}{2}},\ f\,) \quad \text{for} \quad f \in \mathscr{X} \ , \tag{3.43}$$

$$\|\{\psi_N(E[T(\tfrac{Y}{n})])\}^n f - T(t)f\| \leq L\ n^{-\frac{1}{2}}\ \omega^b(n^{-\frac{1}{2}}, Af) \quad \text{for } f \in D(A). \tag{3.44}$$

Of course, similar estimations are available also for Theorem

2.5, leading to the estimations in (1.13) and (1.15), whereas the estimations in (1.12) and (1.14) are covered by Theorem 3.11.

4. Applications to other fields of approximation theory

As has been shown by Hahn [13,14] probability theory can successfully be applied not only within the framework of operator semigroups but also in other fields connected with approximation theory (see also Lindvall [20]). However, a closer look shows that the basic ideas there are almost the same as developed in this article; they are all in the spirit of Bernstein's [1] early paper of 1912. Moreover, it is possible to reobtain some of these results by specializing on the semigroup of translations introduced in (2.4); in fact, many convergence theorems involving for example Bernstein polynomials, Szász - Mirakjan and Baskakov operators (which are all exponential operators and hence share many properties of probabilistic operators, see Ismail and May [18]) can be derived from semigroup representation theorems, even with estimations of the rate of convergence. For the three examples above, simply take Kendall's (2.27), Hille's (1.5) and Shaw's (2.28) representations. To give an example, we shall list some results for the Bernstein polynomials [29], which are derived from Theorems 3.2, 3.4, 3.5, 3.9, and 3.10.

Example 4.1 For $g \in C[0,1]$ and $0 \leqq t \leqq 1$ we have with $B_n(g;t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}$:

$$\|B_n(g;\cdot) - g(\cdot)\| \to 0 \qquad (n \to \infty) , \tag{4.1}$$

$$|B_n(g;t) - g(t)| \leqq \|g'\| \sqrt{\frac{t(1-t)}{n}} \leqq \frac{\|g'\|}{2\sqrt{n}} , \quad n \in \mathbb{N}, \quad \text{if } g' \in C[0,1] , \tag{4.2}$$

$$|B_n(g;t) - g(t)| \leqq \|g''\| \frac{t(1-t)}{n} \leqq \frac{\|g''\|}{8n} , \quad n \in \mathbb{N} , \quad \text{if } g'' \in C[0,1] , \tag{4.3}$$

$$B_n(g;t) - g(t) = g''(t) \frac{t(1-t)}{2n} + O(n^{-3/2}) , \quad \text{if } g''' \in C[0,1] \quad (n \to \infty) \tag{4.4}$$

$$|B_n(g;t) - \frac{t(1-t)}{2n} B_n(g'';t) - g(t)| \leqq \frac{\|g'''\|}{6n^2} |t(1-t)(1-2t)| + \cdots$$
$$+ \frac{3\|g^{(4)}\|}{8n^2} t^2(1-t)^2 + \frac{\|g^{(4)}\|}{24n^3} t(1-t)(6t^2-6t+1)$$

$$\leq \frac{\|g'''\|}{60n^2} + \frac{3\|g^{(4)}\|}{128n^2} + \frac{\|g^{(4)}\|}{480n^3} \quad , \quad n \in \mathbb{N}, \quad \text{if} \quad g^{(4)} \in C[0,1] \, , \tag{4.5}$$

$$\|B_n(g;t) - \frac{t(1-t)}{2n} B_n(g'';t) - \frac{t(1-t)(1-2t)}{6n^2} B_n(g''';t) + \cdots + \frac{t^2(1-t)^2}{8n^2} B_n(g^{(4)};t) - g(t)\| = O(n^{-3}) \, , \quad \text{if} \quad g^{(6)} \in C[0,1] \; (n \to \infty) \tag{4.6}$$

$$|B_n(g;t) - g(t)| \leq K\omega_2\left(\sqrt{\frac{t(1-t)}{2n}}\, , g\right) \; (K \text{ a positive constant}), \tag{4.7}$$

$$|B_n(g;t) - g(t)| \leq (1 + \sqrt{t(1-t)})\omega^1(n^{-\frac{1}{2}}, g) \quad , \tag{4.8}$$

$$|B_n(g;t) - g(t)| \leq \sqrt{\frac{t(1-t)}{n}}\,(1 + \sqrt{t(1-t)})\omega^1(n^{-\frac{1}{2}}, g') \quad \text{for} \quad g' \in C[0,1] \tag{4.9}$$

where ω_2 and ω^1 now are the corresponding moduli of continuity for $C[0,1]$.

Of course, similar estimations can be given for the other two examples mentioned above (and many more); see [29].

For the remainder of this paper, we shall turn to some approximation problems in probability theory in connexion with the Poisson convolution semigroup. A first operator-theoretic approach to the limit theorems in this area is due to LeCam [7], while a rigorous treatment by means of semigroup methods was only recently given by the author [22,28] and Deheuvels and Pfeifer [9]. In fact, the Poisson convolution semigroup can be considered as the discrete analogue to the semigroup of translations, involving now a bounded generator A. To be more precise, consider the Banach space ℓ^1 of all summable sequences, consisting of elements $f = (f(0), f(1), \cdots)$. Let further ε_k denote the Dirac measure concentrated in $k \in \mathbb{Z}^+$, and $*$ denote the convolution operation, i.e.

$$f * g(n) = \sum_{k=0}^{n} f(k)g(n-k) \, , \quad n \geq 0 \tag{4.10}$$

for $f, g \in \ell^1$. Then again $f * g \in \ell^1$, and $\|f * g\| \leq \|f\|\,\|g\|$. Also, with respect to convolution, probability measures P over $\mathbb{Z}^+$ will be identified with the element $(P(\{0\}), P(\{1\}), \cdots) \in \ell^1$. Then

$$Bf = \varepsilon_1 * f , \qquad f \in \ell^1 \tag{4.11}$$

defines a contraction on ℓ^1 , and $A = B - I$ is the (bounded) generator of the Poisson convolution semigroup, i.e.

$$T(t)f = e^{At} f = \sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} \varepsilon_k * f \ = P(t) * f , \qquad f \in \ell^1 \tag{4.12}$$

where $P(t)$ denotes the Poisson distribution over $\mathbb{Z}^+$ with mean t. In this setting,

$$(I + \frac{t}{n} A)^n f = B(n, \frac{t}{n}) * f , \qquad f \in \ell^1 , \quad n \geq t \tag{4.13}$$

where now $B(n,p)$ denotes the binomial distribution over $\{0,1,\cdots, n\}$ with success parameter $p \in [0,1]$. Similarly,

$$\{ \frac{n}{t} R (\frac{n}{t}) \}^n f = \bar{B}(n, \frac{n}{n+t}) * f , \qquad f \in \ell^1 \tag{4.14}$$

with $\bar{B}(n,p)$ denoting the negative binomial distribution over $\mathbb{Z}^+$ with parameters n and p (see [26] and [29]).

Since in probability theory, many convergence theorems are stated only in terms of convergence in distribution $\overset{\mathcal{D}}{\to}$, it is necessary to look for an equivalent metric in order to achieve results on the degree of approximation. One such here is the total variation distance d defined for probability measures P, Q over $\mathbb{Z}^+$ by

$$d(P,Q) = \sup\{ |P(A)-Q(A)| ; \ A \subseteq \mathbb{Z}^+ \} = \frac{1}{2} \sum_{k=0}^{\infty} |P(\{k\})-Q(\{k\})| . \tag{4.15}$$

A classical result of Poisson then says that

$$d(B(n, \frac{t}{n}); \ P(t)) \to 0 \qquad (n \to \infty), \tag{4.16}$$

$$d(\bar{B}(n, \frac{n}{n+t}); \ P(t)) \to 0 \qquad (n \to \infty). \tag{4.17}$$

But it is easy to see that with $f = (1,0,0,\cdots) \in \ell^1$ the distances in (4.16) and (4.17) can also be expressed as

$$d(B(n, \frac{t}{n}); \ P(t)) = \frac{1}{2} \|(I + \frac{t}{n} A)^n f - T(t)f\| , \quad n \geq t , \tag{4.18}$$

$$d(B(n, \frac{n}{n+t}); P(t)) = \frac{1}{2} \|\{ \frac{n}{t} R (\frac{n}{t})\}^n f - T(t)f \| , \quad n \geq 1. \qquad (4.19)$$

While relation (4.18) is connected with an exponential formula for bounded generators (see [16] or [22,28]), relation (4.19) is just an application of Widder's inversion formula (1.10). The estimations worked out in chapter 3 now can be used to give very precise results on the degree of approximation for (4.19) (and basically, also for (4.18), see [26]). For instance, in both cases, the right hand side in (4.18) and (4.19) can be estimated by

$$\frac{t^2}{4n} \|T(t)A^2 f\| + O(n^{-2}) \qquad (n \to \infty) \quad , \qquad (4.20)$$

where for large t, we have

$$\|T(t)A^2 f\| = \frac{4(1+o(1))}{t\sqrt{2\pi e}} \qquad (t \to \infty) \qquad (4.21)$$

(see Deheuvels and Pfeifer [9]). Of course, also other Poisson convergence theorems can be handled this way [26], and it is even possible to choose different metrics [28]. More research work in this area, using semigroup theory and its approximation-theoretic aspects, is in progress.

Acknowledgements I am very much indebted to Professor L.C. Hsu and the Organizing Committee of the Dalian Conference on Approximation Theory for their friendly invitation and hospitality, and also for giving me the opportunity of many fruitful discussions with several Chinese mathematicians on questions concerning various aspects of approximation theory.

References

[1] N.S. Bernstein, *Démonstration du théorème de Weierstrass, fondée sur le calcul des probabilités*. Soob. Kharkov Mat. Obs. 13(1912), 1-2.

[2] N.S. Bernstein, *Complément à l'article de E. Voronovskaja*. C.R. Acad. Sci. URSS, 86-92 (1932).

[3] P. Billingsley, *Probability and Measure*, Wiley, New York, 1979.

[4] P.L. Butzer, *Halbgruppen von linearen Operatoren und eine Anwendung in der Approximationstheorie*, J. reine und angew. Math., 197(1957), 112-120.

[5] P.L. Butzer, and H. Berens, *Semi-Groups of Operators and Approximation*, Springer, Berlin, 1967.

[6] P.L. Butzer and L. Hahn, *A probabilistic approach to representation formulae for semigroups of operators with rates of convergence*, Semigroup Forum, 21(1980), 257-272.

[7] L. LeCam, *An approximation theorem for the Poisson binomial distribution*. Pacific J. Math., 10(1960), 1181-1197.

[8] K.L. Chung, *On the exponential formulas of semi-group theory*. Math., Scand, 10(1962), 153-162.

[9] P. Deheuvels, and D. Pfeifer, *A semi-group approach to Poisson approximation*, To appear in Ann. Prob., 1984.

[10] Z. Ditzian, *On Hille's first exponential formula*, Proc. Amer. Math. Soc., 22(1969), 351-355.

[11] Z. Ditzian, *Note on Hille's exponential formula*, Proc. Amer. Math. Soc., 24(1970), 351-352.

[12] Z. Ditzian, *Exponential formulae for semi-groups of operators in terms of the resolvent*, Israel J. Math., 9(1971), 541-553.

[13] L. Hahn, *Approximation by operators of probabilistic type*, J. Approx. Theory, 30(1980), 1-10.

[14] L. Hahn, *A note on stochastic methods in connexion with approximation theorems for positive linear operators*, Pacific J. Math., 101(1982), 307-319.

[15] E. Hille, *Functional Analysis and Semi-Groups*, Amer. Math. Soc. Colloq. Publ.,1948, Providence, R.I.

[16] E. Hille and R.S. Phillips, *Functional Analysis and Semi-Groups*. Amer. Math. Soc. Colloq. Publ.,31(1957), Providence, R.I.

[17] L.C. Hsu, *An estimation for the first exponential formula in the theory of semi-groups of linear operators*, Czechoslovak Math. J., 10 (85)(1960) 323-328.

[18] M.E.H. Ismail, and C.P. May, *On a family of approximation operators*. J. Math. Anal. Appl., 63(1978), 446-462.

[19] D.G. Kendall, *Bernstein polynomials and semigroups of operators*. Math. Scand., 2(1954), 185-186.

[20] T. Lindvall, *Bernstein polynomials and the law of large numbers*. Math. Sci., 7(1982), 127-139.

[21] D. Pfeifer, *On a general probabilistic representation formula for semigroups of operators*. J. Math. Res. Exponsition, 2 (4)(1982), 93-98.

[22] D. Pfeifer, *A semi-group theoretic proof of Poisson's limit law*. Semigroup Forum 26(1983), 379-382.

[23] D. Pfeifer, *On a probabilistic representation theorem of operator semigroups with bounded generator*. J. Math. Res. Exposition, 4 (1)(1984), 79-81.

[24] D. Pfeifer, *A note on probabilistic representations of operator semigroups*, Semigroup Forum, 28(1984), 335-340.

[25] D. Pfeifer, *Probabilistic representations of operator semigroups - a unifying approach*, Semigroup Forum, 30(1984), 17-34.

[26] D. Pfeifer, *Stochastic Methods in the Theory of Semigroups of Linear Operators (in German)*, Habilitationsschrift, RWTH Aachen., 1984.

[27] D. Pfeifer, *A probabilistic approach to Ditzian's estimations for operator semigroup representations*, Technical report, RWTH Aachen., 1984.

[28] D. Pfeifer, *A semigroup setting for distance measures in connexion with Poisson approximation*. To appear in Semigroup Forum, 1984.

[29] D. Pfeifer, *Approximation-theoretic aspects of probabilistic representations for operator semigroups*, J. Approx. Theory, 43(1985), 271-296.

[30] S.Y. Shaw, *Approximation of unbounded functions and applications to representations of semigroups*, J. Approx. Theory, 28(1980), 238-259.

[31] E.V. Voronovskaja, *The asymptotic properties of the approximation of functions with Bernstein polynomials (in Russian)*, Dokl. Akad. Nauk. SSSR(A), 79-85, 1932.

[32] K. Weierstrass, *Über die analytische Darstellbarkeit sogenannter Willkürlicher Funktionen einer reellen Veränderlichen*, Sitz. -Ber. Akad. Wiss. Berlin, 633-639, 789-805, 1885.