PSEUDO–POISSON APPROXIMATION FOR MARKOV CHAINS

K. A. Borovkov

Statistics Department, University of Melbourne, Australia

D. Pfeifer

Fachbereich Mathematik, Universität Hamburg, Germany

We consider the problem of approximating the distribution of a Markov chain with 'rare' transitions in an arbitrary state space by that of the corresponding pseudo–Poisson process. Sharp estimates for both first– and second–order approximations are obtained. The remarkable fact is that the convergence rate in this setup can be better than that in the ordinary Poisson theorem: the ergodicity of the embedded 'routing' Markov chain improves essentially the degree of approximation. This is of particular importance if the accumulated transition intensity of the chain is of a moderate size so that neither the usual estimates from the Poisson theorem nor the existence of a stationary distribution alone provide good approximation results. On the other hand, the estimates also improve the known results in the ordinary Poisson theorem.

AMS 1991 Subject Classifications: 60F05, 60J10, 60J25.

Poisson approximation ∗ Markov chain ∗ pseudo–Poisson process

1. Introduction and main results

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with success probabilities

$$\boldsymbol{P}(X_k = 1) = 1 - \boldsymbol{P}(X_k = 0) = p_k, \quad 0 \le p_k \le 1, \quad k = 1, \ldots, n. \tag{1.1}$$

The problem of approximating the distribution of the sum $S_n = X_1 + \ldots + X_n$ by a Poisson distribution has already a long history since Poisson (1837) proved the convergence of the binomial distributions $\mathcal{B}(n,p)$ (as $n \to \infty$, $p = \lambda/n$) to the law which was named later after him, and since von Bortkewitsch (1898) demonstrated the role of this convergence in the statistical analysis of rare events. A lot of

Correspondence to: D. Pfeifer, Institut für Mathematische Stochastik, Universität Hamburg, Bundesstr. 55, D–20146 Hamburg, Germany.

* Supported by the Alexander von Humboldt Foundation.

publications has been devoted to the estimation of convergence rates in the Poisson limit theorem, and some of the related results will be discussed briefly below.

In the present paper, we consider a more general problem of pseudo–Poisson approximation of Markov chains. Although in some cases this can be reduced to ordinary Poisson approximation, we should stress that even then our results given here are new and improve the known estimates for the latter. Our approach here is a development of the semigroup method used in Deheuvels and Pfeifer (1986, 1988).

Let S_k, $k \geq 0$, be a Markov chain in an arbitrary state space $(\mathcal{X}, \mathcal{S})$ with initial distribution m_0 and transition function

$$P_k(x, B) = \boldsymbol{P}(S_k \in B |\, S_{k-1} = x), \quad x \in \mathcal{X},\ B \in \mathcal{S},\ k \geq 1.$$

Suppose that

$$P_k = (1 - p_k)I + p_k R_k, \quad 0 \leq p_k \leq 1,\ k \geq 1, \tag{1.2}$$

where $I(x, B) = \mathbf{1}(x \in B)$, and the R_k are some stochastic kernels. Thus, on the k–th step, no changes occur with probability $1 - p_k$, and, with probability p_k, there occurs a transition governed by the kernel R_k. Clearly, the sums S_k of independent indicators considered in the first paragraph form the simplest chain of the sort with $\mathcal{X} = \mathbf{Z}$ and transition function (1.2) with $R_k = P$, where $P(i, \{j\})$ is 1 if $i = j$ and 0 otherwise. Thus our notations for this classical case and for more general Markov chain case under consideration are consistent and should cause no confusion.

The total variation distance between two measures m_1 and m_2 on $(\mathcal{X}, \mathcal{S})$ will correspondingly be denoted by

$$d(m_1, m_2) = \sup_{B \in \mathcal{S}} |m_1(B) - m_2(B)|.$$

The pseudo–Poisson process Y_t is defined as a homogeneous Markov process with transition function

$$Q_t(x, B) = \boldsymbol{P}(Y_{t+s} \in B |\, Y_s = x) = \exp(t(P - I)) = \mathrm{e}^{-t} \sum_{k=0}^{\infty} \frac{1}{k!} t^k P^k$$

(cf. e.g. Feller (1971), X,1). Here $P^k = P^{k-1}P$ and, for two kernels P and R,

$$PR(x, B) = \int_{\mathcal{X}} R(x, dy)\, P(y, B).$$

We shall also use the notation $Rm(B) = \int_{\mathcal{X}} m(dx)\, R(x, B)$ for a kernel R and a measure m. In this notation, $\mathcal{L}(S_n) = P_n \ldots P_2 P_1 m_0$.

Pseudo–Poisson processes form an important class of Markov processes, which are used for modeling throughout many areas of science and engineering. They can be also used to approximate homogeneous Markov processes, an analytical counterpart of this situation being closely related to the theory of semigroups (see e.g. Chapter X in Feller (1971)). On the other hand, the assumption of exponential distributions for interarrival times in queueing systems, lifetimes in reliability theory etc., is closely related to the lack of memory (aging) property. Sometimes this is only an approximation to the real situation when time is discrete – as is typically the case in computing systems, so that during each quantum of time, we have independent Bernoulli trials with small success probability. Hence each success 'switches on' a transition of the system, which can often be described by a Markov transition function of the type (1.2). When these time quanta are sufficiently small, the approximation of these models by pseudo–Poisson models is justified by the usual Poisson theorem. However, if one has to deal with more delicate 'intermediate' situations, when the size of the quanta is relatively large, the question of whether such an approximation is appropriate becomes non–trivial. Our paper is also devoted to solving this problem.

We now give some relevant results on the behaviour of the total variation distance between the laws $\mathcal{L}(S_n)$ of the sum S_n in (1.1) and the approximating Poisson distribution

$$\Pi_\lambda(\{k\}) = \mathrm{e}^{-\lambda}\lambda^k/k!,\ \ k \in \mathbf{Z}_+$$

with mean λ. Denote by $\pi(\lambda)$ a random variable having the law Π_λ and recall that the total variation distance

$$\begin{aligned} d(\mathcal{L}(S_n),\, \Pi_\lambda) &= \sup_{M\subset\mathbf{Z}} |\boldsymbol{P}(S_n\in M) - \boldsymbol{P}(\pi(\lambda)\in M)| = \\ &= \frac{1}{2}\sum_{k=0}^{\infty} |\boldsymbol{P}(S_n = k) - \boldsymbol{P}(\pi(\lambda) = k)|. \end{aligned}$$

Let

$$\lambda_k = \sum_{j=1}^{n} p_j^k, \quad \lambda = \lambda_1, \quad \theta = \frac{\lambda_2}{\lambda}, \quad p_0 = \max_{1\le j\le n} p_j.$$

Some of the first estimates for the non–i.i.d. case were obtained by Le Cam (1960):

$$d(\mathcal{L}(S_n),\, \Pi_\lambda) \le \lambda_2, \quad \text{and} \quad d(\mathcal{L}(S_n),\, \Pi_\lambda) \le 8\,\theta \ \text{ if } \ p_0 \le 1/4,$$

as well as by Kerstan (1964):

$$d(\mathcal{L}(S_n),\, \Pi_\lambda) \le 1.05\,\theta \ \text{ if } \ p_0 \le 1/4.$$

For small p_j, the choice of $q = \sum_{j=1}^{n} q_j$, $\quad q_j = -\log(1-p_j)$, for the mean of the approximating Poisson law is preferable; A. Borovkov (1976) showed that

$$d(\mathcal{L}(S_n),\, \Pi_q) \le \frac{1}{2}\lambda_2^+, \quad \lambda_2^+ = \sum_{j=1}^{n} p_j^2(1+p_j) \le (1+p_0)\lambda_2 \tag{1.3}$$

(Serfling (1978) obtained independently almost the same result). Later, Deheuvels and Pfeifer (1986) proved that this choice is, in a sense, almost optimal.

By a different approach, taking account of what happens when λ becomes large, Barbour and Hall (1984) obtained the estimate

$$d(\mathcal{L}(S_n),\, \Pi_\lambda) \le (1-\mathrm{e}^{-\lambda})\theta. \tag{1.4}$$

The following results, giving the asymptotic behaviour of the distance between the distributions of S_n and $\pi(\lambda)$, were obtained in Deheuvels and Pfeifer (1986, 1988). Let

$$G(t) = \frac{\mathrm{e}^{-t}}{2}\left(\frac{t^{r^+-1}(r^+-t)}{r^+!} - \frac{t^{r^--1}(r^--t)}{r^-!}\right),$$

with $r^{\pm} = \left\lfloor t + 1/2 \pm \sqrt{(t+1/4)}\right\rfloor$, where $\lfloor\cdot\rfloor$ denotes the integer part; note that $G(t) \sim 1/(t\sqrt{2\pi\mathrm{e}})$ as $t \to \infty$. Then

$$d(\mathcal{L}(S_n),\, \Pi_\lambda) = \begin{cases} G(\lambda)\lambda_2 + 2.6\varepsilon\dfrac{\lambda_3}{\lambda} \le \left(\dfrac{1}{\mathrm{e}} + 2.6\, p_0\right)\theta & \text{for } p_0 \le 1/4, \\ G(\lambda)\lambda_2 + \dfrac{\varepsilon\eta^3}{2(1-\eta)} \le \left(\dfrac{1}{\mathrm{e}} + \dfrac{\eta}{1-\eta}\right)\theta & \text{for } \theta < 1/2, \end{cases} \tag{1.5}$$

where $\eta = (2\theta)^{1/2}$ and $|\varepsilon| < 1$ in both cases.

Thus, if $\theta \to 0$ as $n \to \infty$, then

$$d(\mathcal{L}(S_n),\, \Pi_\lambda) \sim G(\lambda)\lambda_2, \tag{1.6}$$

and if $\lambda \to \infty$, the latter is $\sim \theta/\sqrt{2\pi\mathrm{e}}$. Note that in the i.i.d. case (i.e. when $\mathcal{L}(S_n) = \mathcal{B}(n,p)$ with $\theta = p$, $\lambda = np$), Prokhorov (1953) obtained an estimate with the same main term.

We mention here also the papers by Franken (1966), Presman (1983), Witte (1990, 1993) concerning the case of independent X's, and by Sevastyanov (1972), Banis (1975), Grigelionis (1966), Brown (1983), Serfozo (1986), Wang (1981) in which various schemes of dependent variables have been treated; unfortunately, it is impossible to mention here all the papers devoted to the problem. Recently a special monograph by Barbour et al. (1992) has been published devoted mainly to the use of the Stein – Chen method for approximating the distributions of the sums of both independent and dependent random indicators.

These results show the limitations of Poisson approximations; in many situations, when the number of summands is sufficiently large to make direct computations rather cumbersome, the values of θ (or λ_2) are not so small that the desired accuracy of approximation can be ensured. Thus, a refinement of Poisson approximation is both of practical and theoretical interest. Here we mention the work by Shorgin (1977), Barbour and Hall (1984), Kruopis (1986), K. Borovkov (1988), and Barbour et al. (1992). In particular, note the following results on the 'second–order' approximation. Denote by $\Pi_{t,s}$ the (generally speaking, signed) measure on $\mathbf{Z}$ with generating function

$$\sum_{k=0}^{\infty} z^k \Pi_{t,s}(\{k\}) = \exp\bigl(t(z-1) - s(z^2-1)\bigr). \tag{1.7}$$

Note that the coefficients $r_k = \Pi_{t,s}\bigl(\{k\}\bigr)$ admit a straightforward calculation. A simple recursive formula for them can e.g. be found in Johnson and Kotz (1969):

$$r_{-1} := 0, \quad r_0 := \mathrm{e}^{-t+s}, \quad r_{k+1} = \frac{1}{k+1}\bigl(tr_k - 2sr_{k-1}\bigr), \quad k \geq 0. \tag{1.8}$$

Kruopis (1986) showed that, for $\Lambda = \sum_{j=1}^{n} p_j(1-p_j)$,

$$d(\mathcal{L}(S_n), \Pi_{\lambda+\lambda_2,\, \lambda_2/2}) \leq \\ \leq 5\mathrm{e}^{2p_0}\lambda_3 \min\Bigl(1.2\,\Lambda^{-3/2} + 4.2\,\lambda_2\Lambda^{-3},\, 2 + \Lambda + 3.4\,\lambda_2\Bigr), \tag{1.9}$$

Further, it was proved by K. Borovkov (1988) that, for another choice of parameters of the approximation, one can get a better coefficient in the main term λ_3 of the bound. Put

$$\mu_j = \frac{p_j}{1-p_j}, \quad \nu_1 = \sum_{j=1}^{n} \mu_j, \quad \nu_2 = \sum_{j=1}^{n} (\mu_j - \log(1+\mu_j)), \quad \nu_4 = \sum_{j=1}^{n} \mu_j^4.$$

Then

$$d(\mathcal{L}(S_n),\, \Pi_{\nu_1,\,\nu_2}) \leq \frac{1}{6}(\mathrm{e}^{\nu_2}+1)(\lambda_3 + 3\nu_4). \tag{1.10}$$

The estimate (1.9) is preferable for large values of λ, while (1.10) is better for small and moderate values of λ. Moreover, the latter estimate remains valid — as well as the estimate (1.3) — also for the total variation distance between the corresponding approximant and the law of the whole sequence $(S_1, \ldots, S_n)$. This is due to the coupling method employed in the proofs of both estimates. Note also that if, instead of $\Pi_{\nu_1,\,\nu_2}$ in (1.10), we use the law $\Pi_{\nu_1} * \mathrm{e}^{\nu_2}(I_0 - \nu_2 I_2)$, where I_k is the degenerate probability measure at point k, then we should add to the right hand side of (1.10) the term $\frac{1}{2}(\mathrm{e}^{\nu_2} - 1)^2$ (which is $\sim \frac{1}{2}\lambda_2^2$ for small λ_2).

A somewhat different approximation was proposed in Barbour and Hall (1984). Let $\mathbf{1}(B)$ be the indicator random variable of the event B and

$$\begin{aligned}\Pi^*_{\lambda,\lambda_2}(A) =& \Pi_\lambda(A) - \frac{\lambda_2}{2\lambda^2}\boldsymbol{E}\left[\mathbf{1}(\pi(\lambda)\in A)\big(\pi(\lambda)^2 - (2\lambda+1)\pi(\lambda) + \lambda^2\big)\right] = \\ =& \sum_{k\in A} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\left(1 - \frac{\lambda_2}{2\lambda^2}\big(k^2 - (2\lambda+1)k + \lambda^2\big)\right).\end{aligned}$$

Then

$$\begin{aligned}d(\mathcal{L}(S_n),\, \Pi^*_{\lambda,\,\lambda_2}) \le& 2\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\min(1,\, 1.4\,\lambda^{-1/2})\lambda_3 + 2(1-\mathrm{e}^{-\lambda})^2\theta^2 \le \\ &\le 4(1-\mathrm{e}^{-\lambda})(1 - \frac{1}{2}\mathrm{e}^{-\lambda})\lambda_3\lambda^{-1}. \qquad (1.11)\end{aligned}$$

Note that in the relation (1.11), as in the above–mentioned modification of (1.10), there is a term containing λ_2^2 (or θ^2), which is due to the structure of the approximant and can be estimated only by a term of the order $\lambda_3\lambda^{-1}$. For the $\Pi_{t,s}$–approximant, as one can see from (1.9) and from our Theorem 2 below, it is possible to get an estimate of the order $\lambda_3\lambda^{-3/2}$.

Note also that, for both (1.10) and (1.11), similar higher order expansions are available from K. Borovkov (1988) and Barbour et al. (1992) respectively, but we restrict ourselves here only to 'second order' approximations.

Shur (1984) was apparently the first to consider the problem of estimating the rates of approximation in the general Markov chain setup with kernels (1.2). He put

$$U = \exp(P_n - I)\dots\exp(P_2 - I)\exp(P_1 - I)$$

and used Um_0 to approximate the law $\mathcal{L}(S_n)$ (in the case when $\mathcal{X} = \mathbf{Z}$). Note that in general operators with kernels P_k do not commute, so that Um_0 cannot be expressed in a 'more computable' form of the distribution of the pseudo–Poisson process $Q_\lambda m_0$, where the kernel

$$Q_\lambda = \exp\left(\sum_{k=1}^{n}(P_k - I)\right) = \exp(\lambda(P - I)),$$

has 'spectral mixture' $P = \lambda^{-1}\sum_{k=1}^{n} p_k R_k$. He applied a straightforward argument using the properties of the exponential in a Banach algebra (cf. Section 2 below) to get the estimate

$$d(\mathcal{L}(S_n),\, Um_0) \le \lambda_2 \qquad (1.12)$$

for any initial distribution m_0.

However, it is easy to see that the coupling argument (see e.g. Serfozo (1986) and K. Borovkov (1988) for the use of the latter in a similar situation for compound Poisson approximation) applies equally in this case (to the indicators of non–trivial transitions on particular steps, which occur with probabilities p_k) and yields the same estimate. Moreover, the choice of parameters q_k for the exponentials:

$$\widetilde{U} = \exp(r_n(P_n - I)) \dots \exp(r_2(P_2 - I)) \exp(r_1(P_1 - I)), \quad r_k = q_k/p_k,$$

leads to the following estimate (parallel to (1.3)): for any initial distribution m_0,

$$d(\mathcal{L}(S_n), \widetilde{U} m_0) \le \frac{1}{2} \lambda_2^+. \tag{1.13}$$

In what follows we consider a narrower class of Markov chains, which could be called 'semi–homogeneous', and which occur quite often in applications (for an example see Section 4 below). Namely, we suppose that all R_k in (1.9) coincide: for a common stochastic kernel P,

$$P_k = (1 - p_k) I + p_k P, \quad 0 \le p_k \le 1, \quad k \ge 1. \tag{1.14}$$

Clearly, in this case all P_k commute, so that (1.12) and (1.13) now take the form

$$d(\mathcal{L}(S_n), Q_\lambda m_0) \le \lambda_2, \quad d(\mathcal{L}(S_n), Q_q m_0) \le \frac{1}{2} \lambda_2^+$$

(obviously, $\widetilde{P} = P$ and $\widetilde{Q}_q = Q_q$ for (1.14)). Moreover, it is not difficult to see that the estimate (1.4) and the *inequalities* in (1.5) continue to hold in this setup, too (this follows from the corresponding estimates for indicators of non–trivial transitions, governed by the kernel P, which are independent Bernoulli random variables with success probabilities p_k).

However, the *equalities* in (1.5) and hence the asymptotics of the form (1.6) need not hold in the case of a general semi–homogeneous Markov chain, since the existence of an invariant distribution can essentially improve the rate of convergence when λ becomes large. Our main results here provide rather sharp estimates for these cases. On the other hand, these estimates improve also the known results for the ordinary scheme of Bernoulli summands, as this scheme is just a special case of our setup. Thus, the estimates (1.15) and (1.24) of our Theorems 1 and 2 below are valid for this scheme, too. In 'regular cases' (when λ_2 is small), these are several times better than the estimates (1.4) and (1.11), respectively. For example, in the i.i.d. case, when $p_i \equiv \lambda/n$, we have, say, the following values of the ratios of the (appropriate parts of the) right hand sides of these inequalities: for $n = 100$,

$$\lambda = 2: \qquad \frac{\text{r.h.s. of (1.4)}}{\text{r.h.s. of (1.15)}} = 2.942, \qquad \frac{\text{r.h.s. of (1.11)}}{\text{r.h.s. of (1.24)}} = 3.792,$$

$$\lambda = 3: \qquad \frac{\text{r.h.s. of (1.4)}}{\text{r.h.s. of (1.15)}} = 2.206, \qquad \frac{\text{r.h.s. of (1.11)}}{\text{r.h.s. of (1.24)}} = 7.573.$$

Note that for $n = 1000$, the last ratio is 8.901.

Now we formulate our main results. To take account of the ergodicity of the kernel P, it will be convenient to introduce the following functions. For a function $\delta_\cdot : \mathbf{R}_+ \mapsto \mathbf{R}_+$, characterizing the rate of convergence of the iterations P^k, we put

$$l(t) = l_\delta(t) = \boldsymbol{E}\delta_{\pi(t)}, \qquad L(t) = L_\delta(t) = \boldsymbol{E}\left|\frac{\pi(t)}{t} - 1\right|\delta_{\pi(t)}.$$

Set

$$\tau = \lambda_2 + \frac{8}{3}\lambda_3 \mathrm{e}^{2p_0} + 2p_0.$$

Theorem 1. *For any initial distribution* m_0*,*

$$d(\mathcal{L}(S_n), Q_\lambda m_0) \le \begin{cases} \mathrm{e}^{\tau-\lambda}\lambda_2 & \text{for } \lambda \le 2, \\ \mathrm{e}^{\tau-1}\theta & \text{for all } \lambda > 0. \end{cases} \tag{1.15}$$

If, for some distribution m_∞ *on* $(\mathcal{X}, \mathcal{S})$*,*

$$d(P^k m_0, m_\infty) \le \delta_k, \quad k \ge 0, \quad P m_\infty = m_\infty, \tag{1.16}$$

then

$$d(\mathcal{L}(S_n), Q_\lambda m_0) \le 2\mathrm{e}^\tau l(\lambda)\lambda_2. \tag{1.17}$$

If (1.16) *holds uniformly for all probability measures* m_0 *on* $(\mathcal{X}, \mathcal{S})$*, then, for any initial distribution* m_0*, one has, in addition to* (1.17)*, the estimate*

$$d(\mathcal{L}(S_n), Q_\lambda m_0) \le \mathrm{e}^\tau L^2(\lambda/2)\lambda_2. \tag{1.18}$$

In particular, if P is uniformly ergodic (see Nummelin (1984)), then there exists a stationary distribution m_∞, and the function δ_k can be chosen identical for all m_0 and decreasing exponentially fast:

$$d(P^k m_0, m_\infty) \le \delta_k = C\rho^k, \quad 0 < C < \infty,\ 0 < \rho < 1. \tag{1.19}$$

To specify the form of the estimates (1.17) and (1.18) in this special case, we shall need the function

$$M_\rho(t) = \rho\boldsymbol{E}\left|\frac{\pi(\rho t)}{\rho t} - 1\right| = 2\rho\mathrm{e}^{-\rho t}\frac{(\rho t)^{\lfloor\rho t\rfloor}}{\lfloor\rho t\rfloor!}, \tag{1.20}$$

where again $\lfloor t\rfloor$ is the integer part of t. Since

$$2\mathrm{e}^{-t}\frac{t^{\lfloor t\rfloor}}{\lfloor t\rfloor!} = 2\mathrm{e}^{-t} \text{ for } 0 \le t \le 1 \quad \text{and} \quad \le \sqrt{2/\mathrm{e}t} \text{ for all } t > 0 \tag{1.21}$$

(note also that left hand side of (1.21) is $\sim \sqrt{2/\pi t}$ as $t \to \infty$, see e.g. Deheuvels and Pfeifer (1988)), we have the estimates

$$M_\rho(\lambda) \leq \begin{cases} 2\rho \mathrm{e}^{-\rho\lambda} & \text{for } \lambda \leq \rho^{-1}, \\ \sqrt{2\rho/\mathrm{e}\lambda} & \text{for all } \lambda > 0. \end{cases} \tag{1.22}$$

Corollary 1. *If* (1.19) *holds for all* m_0, *then*

$$d(\mathcal{L}(S_n),\, Q_\lambda m_0) \leq C \min\Big\{2,\, C\big(M_\rho(\lambda/2) + 1 - \rho\big)^2\Big\}\, \mathrm{e}^{\tau-(1-\rho)\lambda}\lambda_2. \tag{1.23}$$

To formulate the next result, we introduce the two–parameter signed kernel

$$Q_{t,s} = \exp\big(t(P-I) - s(P^2 - I)\big) = \sum_{k=0}^{\infty} P^k\, \Pi_{t,s}(\{k\}),$$

recall (1.7) and (1.8). Put $\tau' = \tau + \lambda_2$.

Theorem 2. *For any initial distribution* m_0,

$$d(\mathcal{L}(S_n),\, Q_{\lambda+\lambda_2,\,\lambda_2/2} m_0) \leq \begin{cases} \mathrm{e}^{\tau'-\lambda}\left(\frac{4}{3}\lambda_3 + \lambda_4\right) & \text{for } \lambda \leq 3, \\ \mathrm{e}^{\tau'}(c_1\lambda_3\lambda^{-3/2} + c_2\lambda_4\lambda^{-2}) & \text{for all } \lambda > 0, \end{cases} \tag{1.24}$$

where $c_1 = \sqrt{6}\mathrm{e}^{-3/2} \leq 0.547$, $c_2 = 4\mathrm{e}^{-2} \leq 0.542$. *If* (1.16) *holds for some distribution* m_∞ *on* $(\mathcal{X},\, \mathcal{S})$, *then*

$$d(\mathcal{L}(S_n),\, Q_{\lambda+\lambda_2,\,\lambda_2/2} m_0) \leq 2\mathrm{e}^{\tau'} l(\lambda)\left(\frac{4}{3}\lambda_3 + \lambda_4\right). \tag{1.25}$$

If (1.16) *holds uniformly for all probability measures* m_0 *on* $(\mathcal{X},\, \mathcal{S})$, *then, for any initial distribution* m_0, *one has, in addition to* (1.25), *the estimate*

$$d(\mathcal{L}(S_n),\, Q_{\lambda+\lambda_2,\,\lambda_2/2} m_0) \leq \mathrm{e}^{\tau'}\left(\frac{4}{3}L^3(\lambda/3)\lambda_3 + L^4(\lambda/4)\lambda_4\right). \tag{1.26}$$

Corollary 2. *If* (1.19) *holds for all* m_0, *then*

$$\begin{aligned} d(\mathcal{L}(S_n),\, Q_{\lambda+\lambda_2,\,\lambda_2/2} m_0) &\leq \\ &\leq C\mathrm{e}^{\tau'-(1-\rho)\lambda}\Big(\frac{4}{3}\min\{2,\, C^2\big(M_\rho(\lambda/3) + 1 - \rho\big)^3\}\lambda_3 + \\ &\qquad + \min\{2,\, C^3\big(M_\rho(\lambda/4) + 1 - \rho\big)^4\}\lambda_4\Big). \end{aligned} \tag{1.27}$$

Remark 1. By the presence of the factor e^{τ} (or $e^{\tau'}$) in the above estimates, a significant improvement is evidently only achieved if the term λ_2 is sufficiently small. It should, however, be stressed that in the general case, even for large values of λ_2, this factor is often compensated by the presence of the term $l(\lambda)$ which can for instance result in the leading factor $e^{-(1-\rho)\lambda}$.

2. Auxiliary results for Banach algebras

Since the transition rule in a Markov chain reduces just to the multiplication of integral operators, which form a Banach algebra, our problem can be re–formulated as an approximation problem for products in a Banach algebra. In this section we prove two lemmas in this more abstract setup, which are also of independent interest (see e.g. Kato (1980), Ch.IX, 2, on approximation of a continuous semigroup by discrete ones). But first we carry out the above–mentioned re–formulation of the problem.

Let $\mathcal{M}$ be the space of finite signed measures m on $(\mathcal{X}, \mathcal{S})$ endowed with the total variation norm $\|m\| = |m(\mathcal{X})|$, where $|m|$ is the total variation of m. Note that, if $m(\mathcal{X}) = 0$, then

$$\sup_{C\in\mathcal{S}} |m(C)| = \frac{1}{2}\|m\|. \tag{2.1}$$

Denote by $\mathcal{A}$ the set of integral operators on $\mathcal{M}$:

$$(Am)(\,\cdot\,) = \int A(x,\,\cdot\,)\, m(dx),$$

with kernels $A(x, C)$, $x \in \mathcal{X}$, $C \in \mathcal{S}$, having finite norm

$$\|A\| = \sup_{x\in\mathcal{X}} \|A(x,\,\cdot\,)\|.$$

Multiplication is defined in $\mathcal{A}$ in a natural way by

$$(AB)(x,\,\cdot\,) = \int A(y,\,\cdot\,)\, B(x, dy).$$

It is not hard to verify that $\mathcal{A}$ forms a Banach algebra with unity I having the kernel $I(x, C) = \mathbf{1}(x \in C)$.

Further, it is clear that if m_0 is the initial distribution of the Markov chain $\{S_k\}$ with transition function P_k, then S_n has distribution $P_n \ldots P_2 P_1 m_0$, and hence in the case (1.14) one has

$$P_n \ldots P_2 P_1 m_0 = \prod_{i=1}^{n} (I + p_i A) m_0, \quad A = P - I. \tag{2.2}$$

The pseudo–Poisson process Y_t with transition function P of the embedded chain then follows the distribution

$$\exp(tA)\, m_0, \tag{2.3}$$

and we come to the problem of approximating the product (2.2) by the expression (2.3).

In the rest of the section, $\mathcal{A}$ is an abstract Banach algebra with norm $\|\cdot\|$ and unit element I. Denote, for an element $A \in \mathcal{A}$ and a set of reals p_i, $i = 1, \dots, n$,

$$P = A + I, \quad \alpha = \alpha(A) = \|P\| + 1 = \|A + I\| + 1,$$

$$\Delta_1 = \exp(\lambda A) - \prod_{i=1}^{n}(I + p_i A),$$

$$\Delta_2 = \exp(\lambda A - \frac{1}{2}\lambda_2 A^2) - \prod_{i=1}^{n}(I + p_i A),$$

$$\gamma = \gamma(\alpha) = \Big(\frac{\alpha^2}{2} - 1\Big)\lambda_2 + \frac{\alpha^3}{3}\mathrm{e}^{\alpha p_0}\lambda_3 + \alpha p_0, \quad \beta = \max\{0,\, \alpha^2/2 - 1\},$$

recall that $\lambda_k = \sum\limits_{j=1}^{n} p_j^k$, $\lambda = \lambda_1$, $p_0 = \max\limits_{1\le j\le n} p_j$.

Remark 2. Note that, in the special context of the Banach algebra of integral operators, when P is a probability transition function, one has $\|P\| = 1$. Hence one has $\alpha = 2$, $\beta = 1$ in this case, and therefore the quantity $\gamma = \gamma(2)$ is equal now to τ introduced before formulating Theorem 1.

Lemma 1. *For any element $A \in \mathcal{A}$ and any $p_i \ge 0$, $i = 1, \dots, n$, with $p_0 = \max_{1\le j\le n} p_j \le \min\{2^{1/2},\, \alpha/\beta\}$,*

$$\|\Delta_1\| \le \mathrm{e}^{\gamma}\|A^2\mathrm{e}^{\lambda A}\|\frac{\lambda_2}{2}, \tag{2.4}$$

$$\|\Delta_2\| \le \mathrm{e}^{\gamma+\beta\lambda_2}\left(\|A^3\mathrm{e}^{\lambda A}\|\frac{\lambda_3}{3} + \|A^4\mathrm{e}^{\lambda A}\|\frac{\lambda_4}{8}\right). \tag{2.5}$$

Proof. Since all the elements under consideration commute (they are either polynomials of A or exponentials of such polynomials), we have, putting

$$F(t) = \mathrm{e}^{-tA}(I + tA) \quad \text{and} \quad \varepsilon_i = \mathrm{e}^{\lambda A}\big(I - F(p_i)\big),$$

by the 'telescoping argument', that

$$\begin{aligned}\Delta_1 &= \prod_{j=1}^{n}\mathrm{e}^{p_j A} - \prod_{j=1}^{n}(I + p_j A) = \\ &= \sum_{i=1}^{n}\Big(\prod_{j=1}^{i-1}(I + p_j A)\Big)\Big(\mathrm{e}^{p_i A} - I - p_i A\Big)\exp\Big\{\sum_{k=i+1}^{n} p_k A\Big\} = \\ &= \sum_{i=1}^{n}\Big(\prod_{j=1}^{i-1}F(p_j)\Big)\varepsilon_i.\end{aligned} \tag{2.6}$$

First note that

$$
\begin{aligned}
F(t) =& I + \int\limits_0^t F'(u)\,du = I - A^2 \int\limits_0^t u\mathrm{e}^{-uA} du = \\
=& I - \frac{1}{2}t^2A^2 + A^2 \int\limits_0^t u(I - \mathrm{e}^{-uA})\,du = \\
=& I - \frac{1}{2}t^2A^2 + A^3 \int\limits_0^t u\,du \int\limits_0^u \mathrm{e}^{-vA} dv,
\end{aligned}
\tag{2.7}
$$

where we have made use of the relation

$$
I - \mathrm{e}^{-uA} = A \int\limits_0^u \mathrm{e}^{-vA} dv. \tag{2.8}
$$

Further, since $P = A + I$, we see that

$$
I - \frac{1}{2}t^2A^2 = \left(1 - \frac{t^2}{2}\right) I + \frac{t^2}{2}(2P - P^2),
$$

and therefore, for $t^2 \leq 2$,

$$
\|I - \frac{1}{2}t^2A^2\| \leq 1 - \frac{t^2}{2} + \frac{t^2}{2}\big(2(\alpha-1) + (\alpha-1)^2\big) = 1 + \left(\frac{\alpha^2}{2} - 1\right) t^2 \leq 1 + \beta t^2.
$$

To estimate the last term in (2.7), note that $\|A\| \leq \alpha$ and, for $v \geq 0$,

$$
\|\mathrm{e}^{-vA}\| \leq \mathrm{e}^{v\|A\|} \leq \mathrm{e}^{\alpha v}. \tag{2.9}
$$

Hence it follows from (2.7) that

$$
\begin{aligned}
\|F(t)\| \leq& 1 + \beta t^2 + \alpha^3\,\mathrm{e}^{\alpha t} \int\limits_0^t u\,du \int\limits_0^u dv = \\
=& 1 + \beta t^2 + \frac{\alpha^3}{3} t^3 \mathrm{e}^{\alpha t} \leq \exp\left\{\beta t^2 + \frac{\alpha^3}{3} t^3 \mathrm{e}^{\alpha t}\right\}, \quad 0 \leq t \leq 2^{1/2},
\end{aligned}
\tag{2.10}
$$

and thus

$$
\begin{aligned}
\Big\| \prod_{j=1}^{i-1} F(p_j) \Big\| &\leq \prod_{j=1}^{i-1} \|F(p_j)\| \leq \\
&\leq \exp\Big\{\beta \sum_{j=1}^{i-1} p_j^2 + \frac{\alpha^3}{3}\mathrm{e}^{\alpha p_0} \sum_{j=1}^{i-1} p_j^3\Big\} \leq \exp\Big\{\beta\lambda_2 + \frac{\alpha^3}{3}\mathrm{e}^{\alpha p_0}\lambda_3\Big\}.
\end{aligned}
\tag{2.11}
$$

To complete the proof of (2.4), it remains to estimate $\|\varepsilon_i\|$. We have, similar to (2.8),

$$I - F(t) = A^2 \int\limits_0^t u\mathrm{e}^{-uA} du, \tag{2.12}$$

so that it follows from (2.9) that

$$\|\varepsilon_i\| = \big\|\mathrm{e}^{\lambda A}\big(I - F(p_i)\big)\big\| \le \big\|A^2\mathrm{e}^{\lambda A}\big\|\, \mathrm{e}^{\alpha p_0} \int\limits_0^{p_i} u\, du = \|A^2\mathrm{e}^{\lambda A}\|\, \frac{1}{2} p_i^2 \mathrm{e}^{\alpha p_0}. \tag{2.13}$$

The first estimate of Lemma 1 follows from (2.6), (2.11), and (2.13).

Now we shall estimate the norm of Δ_2. Similar to (2.6), we have, for

$$H(t) = \exp\left\{-\frac{1}{2}t^2A^2\right\} - \mathrm{e}^{-tA}(I + tA) \quad \text{and} \quad \varepsilon_i' = \mathrm{e}^{\lambda A} H(p_i),$$

the representation

$$\Delta_2 = \sum_{i=1}^n \Big(\prod_{j=1}^{i-1} F(p_j)\Big) \exp\Big\{-\frac{1}{2}A^2 \sum_{k=i+1}^n p_k^2\Big\}\varepsilon_i'. \tag{2.14}$$

Since

$$-\frac{1}{2}t^2A^2 = -\frac{1}{2}t^2(P^2 - I) + t^2(P - I), \quad \|\mathrm{e}^{t^2(P-I)}\| = \mathrm{e}^{-t^2}\|\mathrm{e}^{t^2P}\| \le \exp\{t^2(\|P\| - 1)\},$$

and $\|P\| = \alpha - 1$, one has

$$\begin{aligned} \Big\|\exp\left\{-\frac{1}{2}t^2A^2\right\}\Big\| &\le \exp\left\{\frac{t^2}{2}\big(\|P\|^2 + 1\big) + t^2(\|P\| - 1)\right\} = \\ &= \exp\left\{\frac{t^2}{2}\big((\alpha - 1)^2 + 1\big) + t^2(\alpha - 2)\right\} = \exp\left\{\left(\frac{\alpha^2}{2} - 1\right)t^2\right\} \le \mathrm{e}^{\beta t^2}. \end{aligned} \tag{2.15}$$

Now it follows from the estimates (2.11) and (2.15) that the norm of the ith summand on the right hand side of (2.14) does not exceed

$$\Big(\prod_{j=1}^{i-1} \|F(p_j)\|\Big)\mathrm{e}^{\beta\lambda_2}\|\varepsilon_i'\| \le \exp\left\{2\beta\lambda_2 + \frac{\alpha^3}{3}\mathrm{e}^{\alpha p_0}\lambda_3\right\}\|\varepsilon_i'\|,$$

so that

$$\|\Delta_2\| \le \exp\left\{2\beta\lambda_2 + \frac{\alpha^3}{3}\mathrm{e}^{\alpha p_0}\lambda_3\right\}\sum_{i=1}^n \|\varepsilon_i'\|. \tag{2.16}$$

It remains to estimate the last sum. Since

$$H'(t) = A^2 t\left[\mathrm{e}^{-tA} - \exp\Big\{-\frac{1}{2}t^2A^2\Big\}\right],$$

we have

$$H(t) = \int_0^t H'(u)\,du = A^2 \int_0^t u \left[\left(\mathrm{e}^{-uA} - I\right) + \left(I - \exp\left\{ -\frac{1}{2} u^2 A^2 \right\} \right) \right] du =$$
$$= -A^3 \int_0^t u \int_0^u \mathrm{e}^{-vA}\,dv\,du + A^4 \int_0^t u \int_0^{u^2/2} \mathrm{e}^{-vA^2}\,dv\,du,$$

and hence

$$\begin{aligned} \varepsilon_i' =& \mathrm{e}^{\lambda A} H(p_i) = \\ =& -A^3 \mathrm{e}^{\lambda A} \int_0^{p_i} u \int_0^u \mathrm{e}^{-vA}\,dv\,du + A^4 \mathrm{e}^{\lambda A} \int_0^{p_i} u \int_0^{u^2/2} \mathrm{e}^{-vA^2}\,dv\,du. \end{aligned} \tag{2.17}$$

Note that (2.15) yields, for $0 \le v \le p_i^2/2$,

$$\left\| \mathrm{e}^{-vA^2} \right\| \le \mathrm{e}^{\beta p_i^2} \le \mathrm{e}^{\beta p_0^2}, \qquad i = 1, \ldots, n,$$

and hence (2.9) and (2.17) imply the estimate

$$\begin{aligned} \|\varepsilon_i'\| \le& \|A^3 \mathrm{e}^{\lambda A}\| \mathrm{e}^{\alpha p_0} \int_0^{p_i} u\,du + \|A^4 \mathrm{e}^{\lambda A}\| \mathrm{e}^{\beta p_0^2} \int_0^{p_i} u \int_0^{\frac{1}{2}} u^3\,du = \\ =& \|A^3 \mathrm{e}^{\lambda A}\| \mathrm{e}^{\alpha p_0} \frac{p_i^3}{3} + \|A^4 \mathrm{e}^{\lambda A}\| \mathrm{e}^{\beta p_0^2} \frac{p_i^4}{8}. \end{aligned} \tag{2.18}$$

Since $\beta p_0^2 \le \alpha p_0$ from the assumptions of Lemma 1, the estimate (2.5) follows now from (2.16) and (2.18). ■

The right hand sides of (2.4) and (2.5) contain terms of the form $\|A^k \mathrm{e}^{\lambda A}\|$ with $k = 2, 3, 4$. The next lemma provides estimates for these quantities. Note that the second of the estimates in the case of $\alpha = 2$ follows from (3.8) and (3.15) in Deheuvels and Pfeifer (1988).

Lemma 2. *Let $A \in \mathcal{A}$ and $\alpha = \|A + I\| + 1$. Then, for all $k \ge 1$,*

$$\|A^k \mathrm{e}^{tA}\| \le \begin{cases} \mathrm{e}^{(\alpha-2)t} \left(2(\alpha-1)\mathrm{e}^{-(\alpha-1)t/k} + |\alpha - 2| \right)^k & \text{for } 0 < t \le k/(\alpha-1), \\ \mathrm{e}^{(\alpha-2)t} \left(\left(\dfrac{2k(\alpha-1)}{\mathrm{e}t} \right)^{1/2} + |\alpha-2| \right)^k & \text{for all } t > 0. \end{cases}$$

Proof. It is easy to see that

$$A\mathrm{e}^{tA} = \mathrm{e}^{-t}(P - I) \sum_{k=0}^{\infty} \frac{1}{k!} t^k P^k = \mathrm{e}^{-t} t^{-1} \sum_{k=0}^{\infty} \frac{k-t}{k!} t^k P^k, \tag{2.19}$$

and therefore

$$\begin{aligned}\|A\mathrm{e}^{tA}\| \le & \mathrm{e}^{-t}t^{-1}\sum_{k=0}^{\infty}\frac{|k-t|}{k!}t^k(\alpha-1)^k = \\ = & \mathrm{e}^{(\alpha-2)t}\frac{\alpha-1}{(\alpha-1)t}\mathrm{e}^{-(\alpha-1)t}\sum_{k=0}^{\infty}\frac{\big(t(\alpha-1)\big)^k}{k!}|k-(\alpha-1)t+(\alpha-2)t| \le \\ \le & \mathrm{e}^{(\alpha-2)t}\big((\alpha-1)M_1((\alpha-1)t)+|\alpha-2|\big), \end{aligned} \tag{2.20}$$

where $M_\rho(s)$ is defined in (1.20). From the first relation in (1.22) we have, for $t \le (\alpha-1)^{-1}$,

$$\|A\mathrm{e}^{tA}\| \le \mathrm{e}^{(\alpha-2)t}\left(2(\alpha-1)\mathrm{e}^{-(\alpha-1)t}+|\alpha-2|\right),$$

and since

$$\|A^k\mathrm{e}^{tA}\| = \|\big(A\mathrm{e}^{At/k}\big)^k\| \le \|A\mathrm{e}^{At/k}\|^k, \tag{2.21}$$

we have, for $0 < t \le k/(\alpha-1)$,

$$\|A^k\mathrm{e}^{tA}\| \le \mathrm{e}^{(\alpha-2)t}\left(2(\alpha-1)\mathrm{e}^{-(\alpha-1)t/k}+|\alpha-2|\right)^k.$$

Now the second relation in (1.22) implies the estimate

$$\|A\mathrm{e}^{tA}\| \le \mathrm{e}^{(\alpha-2)t}\left(\left(\frac{2(\alpha-1)}{\mathrm{e}t}\right)^{1/2}+|\alpha-2|\right),$$

which completes, together with (2.21), the proof of Lemma 2. ∎

3. Proof of theorems

Proof of Theorem 1. To prove (1.15), it suffices, in view of the re–formulation of our problem in terms of approximation of (2.2) by (2.3), to make use of (2.1), (2.4), and Lemma 2 (recall that, in this case, $\alpha = 2,\ \gamma = \tau$).

To prove (1.17), we should use the following representation based on (2.6) and (2.12):

$$\Delta_1 = BA^2\mathrm{e}^{\lambda A}, \qquad B = \sum_{i=1}^{n}\Big(\prod_{j=1}^{i-1}F(p_i)\Big)\int_0^{p_i} u\,\mathrm{e}^{-uA}du, \tag{3.1}$$

where

$$\|B\| \le \frac{1}{2}\mathrm{e}^{\tau}\lambda_2 \tag{3.2}$$

(the latter follows from the estimates (2.11) and (2.9)). Now $Am_\infty = 0$ by (1.16), and hence, putting

$$\zeta_j(m) = P^j m - m(\mathcal{X})m_\infty, \qquad m \in \mathcal{M},$$

we have

$$A^k \mathrm{e}^{\lambda A} m_0 = \mathrm{e}^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} A^k P^j m_0 = \mathrm{e}^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} A^k \zeta_j(m_0), \quad k \geq 1.$$

Since $\|\zeta_j(m_0)\| \leq 2\delta_j$ by (1.16) and (2.1), we have $\|A^k \zeta_j\| \leq \|A\|^k \cdot 2\delta_j = 2^{k+1}\delta_j$, and therefore

$$\|A^k \mathrm{e}^{\lambda A} m_0\| \leq 2^{k+1} \boldsymbol{E} \delta_{\pi(\lambda)} = 2^{k+1} l(\lambda), \tag{3.3}$$

Thus, as soon as (1.16) holds, we derive from (2.1), (2.4), and (3.3) (for $k = 2$) the bound (1.17).

Now let (1.16) hold uniformly for all initial distributions m_0. We shall show that, in this case, the norm $\|A^k \mathrm{e}^{\lambda A}\|$ admits another bound, different from those given in Lemma 2. For any $m \in \mathcal{M}$, let m_+ and m_- be positive and negative parts of the Jordan decomposition $m = m_+ - m_-$ of the measure m respectively. Since both $m_\pm^{-1}(\mathcal{X})\, m_\pm$ are probability measures, it follows form (1.16) and (2.1) that the norms

$$\|P^k m_\pm - m_\pm(\mathcal{X}) m_\infty\| \leq m_\pm(\mathcal{X}) \cdot 2\delta_k,$$

and hence

$$\|\zeta_k(m)\| \leq \|P^k m_+ - m_+(\mathcal{X}) m_\infty\| + \|P^k m_- - m_-(\mathcal{X}) m_\infty\| \leq 2\delta_k \|m\|,$$

for $\|m\| = m_+(\mathcal{X}) + m_-(\mathcal{X})$. On the other hand, since

$$\sum_{k=0}^{\infty} \frac{t^k}{k!}(k - t) = 0,$$

we have from (2.19) that

$$A\mathrm{e}^{tA} m = \mathrm{e}^{-t} t^{-1} \sum_{k=0}^{\infty} \frac{t^k}{k!}(k - t)\zeta_k(m),$$

and hence

$$\|A\mathrm{e}^{tA} m\| \leq 2\|m\| e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \left| \frac{k}{t} - 1 \right| \delta_k = 2L(t)\|m\|.$$

Thus $\|A\mathrm{e}^{tA}\| \leq 2L(t)$ and (2.21) implies that

$$\|A^k \mathrm{e}^{tA}\| \leq 2^k L^k(t/k). \tag{3.4}$$

To prove (1.18), it remains now to make use of (2.4). Theorem 1 is proved. ∎

Proof of Corollary 1. It suffices to note that, for the function $\delta_\cdot$ from (1.19), $l(t) = C\boldsymbol{E}\rho^{\pi(t)} = Ce^{-(1-\rho)t}$ and, similar to (2.20),

$$\begin{aligned} L(t) =& Ce^{-t}t^{-1}\sum_{k=0}^{\infty}\frac{t^k}{k!}|k-t|\rho^k \le \\ \le& Ce^{-(1-\rho)t}\frac{e^{-\rho t}}{t}\sum_{k=0}^{\infty}\frac{(\rho t)^k}{k!}|k-\rho t+(\rho-1)t| \le \\ \le& Ce^{-(1-\rho)t}\big(M_\rho(t)+1-\rho\big). \end{aligned} \tag{3.5}$$

Hence in this case

$$L^2(\lambda/2) = C^2e^{-(1-\rho)\lambda}\big(M_\rho(\lambda/2)+1-\rho\big)^2$$

and (1.23) follows immediately from (1.17) and (1.18). ∎

Proof of Theorem 2. To prove (1.24), we note first that

$$Q_{\lambda+\lambda_2,\lambda_2/2} = \exp\Big(\lambda A - \frac{1}{2}\lambda_2 A^2\Big).$$

Making use of (2.1) and (2.5), we get

$$d(\mathcal{L}(S_n),\, Q_{\lambda+\lambda_2,\,\lambda_2/2}m_0) \le \frac{1}{2}e^{\tau'}\left(\|A^3e^{\lambda A}\|\frac{\lambda_3}{3} + \|A^4e^{\lambda A}\|\frac{\lambda_4}{8}\right) \tag{3.6}$$

(recall again that, in this context, $\alpha = 2$, $\beta = 1$, and hence $\gamma + \beta\lambda_2 = \tau + \lambda_2 = \tau'$). Now Lemma 2 gives

$$\|A^3e^{\lambda A}\| \le \begin{cases} 8e^{-\lambda} & \text{for } 0 < \lambda \le 3, \\ \left(\dfrac{6}{e\lambda}\right)^{3/2} & \text{for all } \lambda > 0, \end{cases} \qquad \|A^4e^{\lambda A}\| \le \begin{cases} 16e^{-\lambda} & \text{for } 0 < \lambda \le 4, \\ \left(\dfrac{8}{e\lambda}\right)^{2} & \text{for all } \lambda > 0, \end{cases}$$

which yields, together with (3.6), the estimate (1.24).

As for (1.25) and (1.26), one has, similar to (3.1) and (3.2), that

$$\Delta_2 = B_1A^3e^{\lambda A} + B_2A^4e^{\lambda A}, \qquad \|B_1\| \le \frac{1}{3}e^{\tau}\lambda_3, \quad \|B_2\| \le \frac{1}{8}e^{\tau}\lambda_4. \tag{3.7}$$

To complete the proof of Theorem 2, it remains to use the estimates (3.3) and (3.4) for $k = 3, 4$. ∎

Proof of Corollary 2. This follows immediately from Theorem 2 and (3.5). ∎

4. Reflected random walk.

In this section we give an example illustrating the use of our results in a typical for queueing situation. It deals in fact with a 'quantitative justification' of the use of continuous–time models when treating discrete–time systems with small 'time quanta'.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. real random variables, independent of S_0, and let

$$S_k = \max(0,\ S_{k-1} + \xi_k), \quad k = 1, 2, \ldots, n. \tag{4.1}$$

Such Markov chains arise in a natural way in queueing theory and, for many models, the study of the behaviour of a queue reduces just to the study of the solution of (4.1). Now suppose that the 'driver' $\{\xi_k\}$ is of low 'intensity', that is the probability

$$\boldsymbol{P}(\xi_k \neq 0) =: \lambda n^{-1} \tag{4.2}$$

is small (thus we are in the 'triangular array' setup). The simplest example of this sort is a Geometric/Geometric/1 queue in discrete time with 'flabby' arrivals and services. In this case, S_k is the length of the queue at time k, and

$$\boldsymbol{P}(\xi_k = 1) = p(1-r), \qquad \boldsymbol{P}(\xi_k = -1) = (1-p)r,$$

$$\boldsymbol{P}(\xi_k = 0) = 1 - p(1-r) - (1-p)r,$$

where p and r are the arrival and service rates resp. (see e.g. Section 9.2 in Hunter (1983)). With respect to the distribution of S_n, it is well known that equation (4.1) has the solution

$$S_n = Z_n - \min\left(-S_0,\ Z_1,\ Z_2, \ldots, Z_n\right), \quad Z_k = \xi_1 + \ldots \xi_k. \tag{4.3}$$

Thus, if one tries to apply the (compound) Poisson approximation here (what is quite natural in view of (4.2)), the error will be of the order $\lambda^2 n^{-1}$ (cf. the remark after (1.10); in fact, this is the best possible estimate in the 'functional' setup, when approximating the whole sequence $Z_1, \ldots, Z_n$).

Now let $\eta_1, \eta_2, \ldots$ be i.i.d. real random variables following the conditional distribution of ξ_1 given $\xi_1 \neq 0$. Consider the new Markov chain

$$U_k = \max(0,\ U_{k-1} + \eta_k), \quad k \geq 1, \quad U_0 = S_0,$$

and denote its transition kernel by P (this is clearly consistent with our notation in Section 1).

If $\boldsymbol{E}\eta_1 < 0$ then $\{U_k\}$ is ergodic (with the stationary distribution m_∞ coinciding with the law of $\sup_{k\geq 0} \sum_{i=0}^{k} \eta_i$) and, moreover, if $\boldsymbol{E}\mathrm{e}^{c\eta_1} < \infty$ for some $c > 0$, then the chain $\{U_k\}$ is geometrically ergodic (see e.g. Example 5.5(d) and Theorem 6.14 in Nummelin (1984)), so that there exists a constant $\rho < 1$ and a measurable function $\mu(x)$, $x \in [0, \infty)$ (integrable w.r.t. the stationary distribution m_∞ of $\{U_k\}$), such that, for any $x \in [0, \infty)$,

$$d\big(P^n(x, \cdot), m_\infty(\cdot)\big) \le \mu(x)\rho^n, \quad n \ge 1$$

(and due to the monotonicity of the mapping in (4.1) and the role of 0 as a recurrent state, $\mu(x)$ can be chosen to be nondecreasing).

Therefore, for any initial distribution m_0 with $S_0 \le x_0$ a.s., we have (1.16) with $\delta_k = \mu(x_0)\rho^k$, $k \ge 0$, and hence (1.17) implies that in this case

$$d(\mathcal{L}(S_n), Q_\lambda m_0) \le 2\mu(x_0)\mathrm{e}^{\tau-(1-\rho)\lambda}\lambda^2 n^{-1}, \tag{4.4}$$

which contains the exponential factor $\mathrm{e}^{-(1-\rho)\lambda}$ (cf. the comment after (4.3) and also Remark 1). The transition kernel Q_λ corresponds to the reflected compound Poisson process (just the simple birth–and–death process with constant birth and death rates p and r respectively in the case of the Geometric/Geometric/1 queue, which corresponds to the continuous time M/M/1 queue). Note that, by the presence of the exponential term $\mathrm{e}^{-(1-\rho)\lambda}$, the right hand side of (4.4) decreases when λ is large, hence the estimate in (4.4) is clearly superior to what is known otherwise in the literature, giving estimates of order $\lambda^2 n^{-1}$ only, which are in general worse.

Acknowledgements

This joint work was enabled by a grant of the Alexander von Humboldt Foundation while the first author was visiting the Carl von Ossietzky Universität Oldenburg, which is gratefully appreciated.

We are also grateful to the referees whose useful comments helped to improve the exposition of the paper and remove a small error.

REFERENCES

R. Banis, Limit Poisson processes in the scheme of summation of dependent integer–valued processes, Lit. Math. Sb. 4 (1975) 5 – 15. [In Russian.]

A. Barbour and P. Hall, On the rate of Poisson convergence, Math. Proc. Camb. Phil. Soc. 95 (1984) 473 – 480.

A. Barbour, L. Holst and S. Janson, Poisson Approximation (Clarendon Press, Oxford, 1992).

A. A. Borovkov, Wahrscheinlichkeitstheorie (Birkhäuser, Basel – Stuttgart, 1976).

K. A. Borovkov, Refinement of Poisson approximation. Theory Probab. Appl. 33 (1988) 343 – 347.

L. von Bortkewitsch, Das Gesetz der kleinen Zahlen (Teubner, Leipzig, 1898).

T. C. Brown, Some Poisson approximation using compensators, Ann. Probab. 11 (1983) 726 – 744

P. Deheuvels and D. Pfeifer, A semigroup approach to Poisson approximation, Ann. Probab. 14 (1986) 665 – 678.

P. Deheuvels and D. Pfeifer, On a relationship between Uspensy's theorem and Poisson approximation, Ann. Inst. Statist. Math. 40 (1988) 671 – 681.

W. Feller, An Introduction to Probability Theory and Its Applications, Vol.2 (Wiley, New York, 1971, 2nd ed.).

P. Franken, Approximation der Verteilungen von Summen unabhängiger nichtnegativer ganzzahliger Zufallsgrößen durch Poissonische Verteilung, Mathematische Nachrichten, 27 (1966) 303 – 340.

B. Grigelionis, On convergence of sums of step processes to Poisson process, Lit. Math. Sb. 6 (1966) 241 – 244. [In Russian.]

J. J. Hunter, Mathematical Techniques of Applied Probability, Vol. II (Academic Press, New York, 1983).

N. L. Johnson and S. Kotz, Discrete Distributions (Wiley, New York, 1969).

T. Kato, Perturbation Theory for Linear Operators (Springer, New York, 1980, 2nd ed.).

J. Kruopis, Precision of approximation of the generalized binomial distribution by convolution of Poisson measures, Lithuanian Math. J. 26 (1986) 37 – 49.

L. Le Cam, An approximation theorem for the Poisson binomial distribution, Pacific J. Math. 10 (1960) 1181 – 1197.

J. Kerstan, Verallgemeinerung eines Satzes von Prochorov und Le Cam, Z. für Wahrscheinlichkeitstheorie verw. Gebiete 2 (1964), 173 – 179.

E. Nummelin, General Irreducible Markov Chains and Non-negative Operators (Cambridge Univ. Press, Cambridge, 1984).

S. D. Poisson, Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités

(Bachelier, Paris, 1837).
E. L. Presman, Approximation of binomial distributions by infinitely divisible ones, Theory Probab. Appl. 28 (1983) 393 – 403.
Yu. V. Prokhorov, Asymptotic behaviour of the binomial distribution, Uspekhi Matem. Nauk 8, no. 3(55) (1953) 135 – 142. [In Russian.]
R. J. Serfling, Some elementary results on Poisson approximation in a sequence of Bernoulli trials, SIAM Rev. 20 (1978) 567 – 579.
R. F. Serfozo, Compoud Poisson approximation for sums of random variables, Ann. Probab. 14 (1986) 1391 – 1398.
R. F. Serfozo, Compound Poisson approximation for sums of random variables. Correction, Ann. Probab. 16 (1988) 429.
B. A. Sevastyanov, Poisson limit law for a scheme of sums of dependent random variables, Theory Probab. Appl. 17 (1972) 695 – 698.
S. Ya. Shorgin, Approximation of a generalized binomial distribution, Theory Probab. Appl. 22 (1977) 846 – 850.
M. G. Shur, The Poisson theorem and Markov chains, Theory Probab. Appl. 29 (1984) 124 – 126.
Y. H. Wang, On the limit of the Markov binomial distribution, J. Appl. Probab. 18 (1981) 937 – 942.
H.–J. Witte, A unification of some approaches to Poisson approximation, J. Appl. Probab. 27 (1990) 611 – 621.
H.–J. Witte, On the optimality of multivariate Poisson approximation, Stoch. Proc. Appl. 44 (1993), 75 – 88.