The Annals of Probability
1986, Vol. 14, No. 2, 663–676

A SEMIGROUP APPROACH TO POISSON APPROXIMATION

BY P. DEHEUVELS AND D. PFEIFER

Université Paris VI and Technical University Aachen

The aim of this paper is twofold: first, to show that Poisson approximation problems for independent summands can in a natural way be treated in a suitable operator semigroup framework, allowing at the same time for an asymptotically precise evaluation of the leading term with respect to the total variation distance; second, to determine asymptotically those Poisson distributions which minimize this distance for given Bernoulli summands. Besides semigroup methods, coupling techniques as well as direct computations are used.

1. Introduction. Let $X_1, \dots, X_n$ be independent Bernoulli random variables with $p_i = P(X_i = 1) = 1 - P(X_i = 0)$, $0 < p_i < 1$, $i = 1, 2, \dots, n$, and $Y_1, \dots, Y_n$ be independent Poisson random variables with expectations μ_i, $i = 1, 2, \dots$. Let further $S_n = \sum_{i=1}^n X_i$ and $T_n = \sum_{i=1}^n Y_i$. We are interested in the approximation of the distribution of S_n by the distribution of T_n with respect to the total variation distance

$$d(S_n, T_n) = \sup_{M \subset \mathbb{Z}^+} |P(S_n \in M) - P(T_n \in M)|$$

$$= \tfrac{1}{2} \sum_{k=0}^{\infty} |P(S_n = k) - P(T_n = k)|.$$

Estimations of this distance have been given by different authors, for instance Le Cam (1960), Kerstan (1964), Chen (1974, 1975), Serfling (1975, 1978), and most recently by Barbour and Hall (1984), however with a special emphasis on the case $\mu_i = p_i$ in most of these papers. Besides this choice, also $\mu_i = \lambda_i = -\log(1 - p_i)$ is of importance since for $n = 1$, this minimizes $d(S_1, T_1)$ with respect to μ_1 as was shown by Serfling (1975, 1978) using coupling arguments. If we especially assume X_i and Y_i to be maximally coupled, then by Doeblin's inequality,

$$d(S_n, T_n) \le P(S_n \ne T_n) \le 1 - \prod_{i=1}^n (1 - P(X_i \ne Y_i)) = 1 - \prod_{i=1}^n \left((1 + \lambda_i) e^{-\lambda_i}\right)$$

$$\le \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) e^{-\lambda_i} \le \tfrac{1}{2} \sum_{i=1}^n \lambda_i^2,$$

which is a better estimation than those known for the case $\mu_i = p_i$ if $\sum_{i=1}^n \lambda_i$ is small in some sense. Moreover, $\mu_i = \lambda_i$ is, in an asymptotic sense, also the best possible choice as can be seen from the following result.

Received July 1984; revised March 1985.
AMS 1980 *subject classifications.* Primary 60F05, 62E20; secondary 47D05, 20M30.
Key words and phrases. Poisson approximation, operator semigroups, coupling techniques.

THEOREM 1.1. *If* $0 < \sum_{i=1}^n \lambda_i \le 1$, *then for all choices of* μ_i, $i = 1, 2, \dots, n$, *we have*

$$d(S_n, T_n) \ge \left\{ \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) \right\} \exp\left(- \sum_{i=1}^n \lambda_i \right)$$
$$\ge \tfrac{1}{2} \left\{ \sum_{i=1}^n \lambda_i^2 \right\} \exp\left(- \sum_{i=1}^n \lambda_i \right).$$

Consequently, if $p_1, \dots, p_n$ depend on n such that $\sum_{i=1}^n p_i \to 0$ for $n \to \infty$, then uniformly in n, we have

$$\inf_\mu d(S_n, T_n(\mu)) \sim d(S_n, T_n(\lambda)) \sim \tfrac{1}{2} \sum_{i=1}^n p_i^2$$

while

$$d(S_n, T_n(p)) \sim \sum_{i=1}^n p_i^2 \quad \text{only},$$

where for $\mu = (\mu_1, \dots, \mu_n)$, $T_n(\mu)$ denotes a Poisson random variable with expectation $\sum_{i=1}^n \mu_i$, and the inf is taken over all admissible values of μ. The last result also shows that the estimation of Theorem 1 in Barbour and Hall (1984) which is

$$d(S_n, T_n(p)) \le \left\{ \sum_{i=1}^n p_i \right\}^{-1} \left(1 - \exp\left(- \sum_{i=1}^n p_i \right) \right) \sum_{i=1}^n p_i^2,$$

is sharp in this case. For $\sum_{i=1}^n p_i$ being large, however, it can be shown that $d(S_n, T_n(p)) < d(S_n, T_n(\lambda))$ which follows from a general evaluation of the leading term in $d(S_n, T_n(\mu))$ by means of an appropriate semigroup approach. If we assume that $\sum_{i=1}^n p_i$ tends to infinity in a certain way for $n \to \infty$, then it can be shown that the choice $\mu = p$ is indeed asymptotically optimal.

THEOREM 1.2. *If* $\sum_{i=1}^n p_i \to \infty$ *and* $\max(p_1, \dots, p_n) \to 0$ *for* $n \to \infty$, *then*

$$d(S_n, T_n(p)) \sim (2\pi e)^{-1/2} \left\{ \sum_{i=1}^n p_i^2 \right\} \Big/ \left\{ \sum_{i=1}^n p_i \right\}.$$

If additionally $\{\sum_{i=1}^n p_i\} \max(p_1, \dots, p_n) \to 0$, *then also*

$$\inf_\mu d(S_n, T_n(\mu)) \sim d(S_n, T_n(p)).$$

Note that the first relation above corresponds to an evaluation of Kerstan's (1964) leading term, improving at the same time asymptotically the inequality (2.7) in Barbour and Hall (1984) for this case. It is also possible to derive results for intermediate cases, for instance if $\sum_{i=1}^n p_i \to a$ with $n \to \infty$ for some $0 < a < \infty$.

THEOREM 1.3. *Suppose that* $\sum_{i=1}^{n} p_i \to a$ *with* $0 < a < \infty$, *and* $\max(p_1, \dots, p_n) \to 0$ *for* $n \to \infty$. *Then*

$$d(S_n, T_n(\lambda)) \sim \tfrac{1}{2}\left\{\sum_{i=1}^{n} p_i^2\right\}\left\{\frac{a^{[a]}}{[a]!}e^{-a}\right\},$$

and

$$\inf_{\mu} d(S_n, T_n(\mu)) \sim d(S_n, T_n(\lambda)) \quad \textit{iff } a \le \sqrt{2}.$$

Here $[a]$ *denotes the integer part of* a.

Further, for $a > \sqrt{2}$, *there exists* $0 \le \xi_a \le \frac{1}{2}$ *such that*

$$\inf_{\mu} d(S_n, T_n(\mu)) \sim d\big(S_n, T_n(p + \xi_a p^2)\big),$$

where $p^2 = (p_1^2, \dots, p_n^2)$.

Precise evaluations for the last expression as well as for ξ_a will be given in the sequel.

2. The semigroup approach. Consider the Banach space l^1 of all absolutely summable sequences, and let $\mathscr{M}$ denote the set of all probability measures with support contained in the nonnegative integers $\mathbb{Z}^+$. For $m \in \mathscr{M}$, identify m with the element $(m(\{0\}), m(\{1\}), \dots) \in l^1$. Let further $f * g$ denote the convolution of $f, g \in l^1$, i.e.,

$$f * g(n) = \sum_{k=0}^{n} f(k)g(n-k), \qquad n \ge 0. \tag{2.1}$$

Then $f * g \in l^1$, and $\|f * g\| \le \|f\|\,\|g\|$. Define a contraction B on l^1 by $Bf = \varepsilon_1 * f$ where ε_k denotes the unit mass at point $k \in \mathbb{Z}^+$. Then any measure $m \in \mathscr{M}$ can be interpreted as operator on l^1 via

$$mf = m * f = \sum_{k=0}^{\infty} m(\{k\})B^k f, \qquad f \in l^1. \tag{2.2}$$

Further, if I stands for the identity mapping from l^1 to l^1, and if $A = B - I$, then A is the infinitesimal generator of the Poisson convolution semigroup, given by

$$\begin{aligned} e^{tA}f &= \sum_{k=0}^{\infty} \frac{t^k}{k!}A^k f \\ &= \sum_{k=0}^{\infty} e^{-t}\frac{t^k}{k!}\varepsilon_k * f \\ &= \mathrm{Po}(t) * f, \qquad t \ge 0,\ f \in l^1, \end{aligned} \tag{2.3}$$

where $\mathrm{Po}(t)$ denotes the Poisson distribution with mean t.

Since for measures $m_1, m_2 \in \mathscr{M}$, we have

$$d(m_1, m_2) = \tfrac{1}{2} \sum_{k=0}^{\infty} |m_1(\{k\}) - m_2(\{k\})| = \tfrac{1}{2}\|(m_1 - m_2)\varepsilon_0\|, \tag{2.4}$$

where again d is the total variation distance, we can easily formulate Poisson approximation problems for independent summands in this Banach space setting. Namely, if $W_1, \dots, W_n$ are independent random variables with distributions $m_1, \dots, m_n \in \mathscr{M}$, and if $U_n = \sum_{i=1}^n W_i$, we can write

$$d(U_n, T_n(\mu)) = \tfrac{1}{2}\left\|\left(\exp\left(\sum_{i=1}^n \mu_i\right)A - \prod_{i=1}^n m_i\right)\varepsilon_0\right\|. \tag{2.5}$$

Since all operators involved are contractions, a simple estimation for (2.5) is

$$d(U_n, T_n(\mu)) \le \tfrac{1}{2} \sum_{i=1}^n \|\exp(\mu_i A) - m_i\|. \tag{2.6}$$

For instance, if m_i is the binomial distribution on $\{0, 1\}$ with $m_i(\{1\}) = p_i$, then also $m_i = I + p_i A$, hence

$$d(U_n, T_n(p)) \le \sum_{i=1}^n p_i^2, \tag{2.7}$$

which follows for instance by Proposition 1.1.6 in Butzer and Berens (1967) [in fact, m_i represents the two first terms in the Taylor expansion of the Poisson semigroup at $t = p_i$; cf. (2.3)]. Likewise, if m_i is the geometric distribution over $\mathbb{Z}^+$ with $m_i(\{0\}) = q_i = 1 - p_i$, then also $m_i = (q_i/p_i)R(q_i/p_i)$ where for $s > 0$, $R(s) = (sI - A)^{-1}$ denotes the resolvent of the semigroup, hence

$$d(U_n, T_n(\mu)) \le \sum_{i=1}^n (p_i/q_i)^2, \tag{2.8}$$

where $\mu_i = p_i/q_i$ [see e.g., Pfeifer (1985a), Theorem 7.5 and Pfeifer (1985b)]. This generalizes results of Vervaat (1969) to the case of non-i.i.d. summands. While in the introduction it was shown that the estimate (2.7) is sharp for $\sum_{i=1}^n p_i$ small and n large, it can be proved by methods developed in Pfeifer (1985a) that the same is true for relation (2.8).

As a third example, take $m_i = (1 - p_i + p_i^2/2)\varepsilon_0 + p_i(1 - p_i)\varepsilon_1 + p_i^2/2\varepsilon_2 = I + p_i A + p_i^2/2A^2$; then

$$d(U_n, T_n(p)) \le \tfrac{2}{3} \sum_{i=1}^n p_i^3, \tag{2.9}$$

which can be proved similarly to (2.7); just note that in general, we have $\|A^k\| = 2^k$ for all $k \in \mathbb{Z}^+$.

Although the semigroup approach resembles the operator technique used by Le Cam (1960) and Chen (1975), it has the advantage of covering different problems such as the one above, and allowing for an immediate translation of the results obtained to the situation when other distance measures than the total variation distance are considered, for instance the cumulative distribution distance d_0 given by

$$d_0(S_n, T_n) = \sup_{k \geq 0} |P(S_n \leq k) - P(T_n \leq k)|. \tag{2.10}$$

This is possible by a simple change from l^1 to l^∞ [see Pfeifer (1985b)].

To shorten matters, we shall for the remainder of this paper restrict ourselves to the discussion of Poisson approximation for Bernoulli summands with respect to the distance d. However, most of the results given can also easily be formulated for the more general setting outlined above.

THEOREM 2.1. *For $p_1, \dots, p_n$ arbitrary, we have*

$$d(S_n, T_n(p)) = \tfrac{1}{4}\left\{\sum_{i=1}^n p_i^2\right\}\left\|\exp\left\{\left(\sum_{i=1}^n p_i\right)A\right\}A^2\varepsilon_0\right\| + r_n(p) \tag{2.11}$$

with

$$r_n(p) \leq 2.6\frac{\sum_{i=1}^n p_i^3}{\sum_{i=1}^n p_i}$$

$$\leq 2.6\frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n p_i}\max(p_1, \dots, p_n) \quad \text{if } \max\{p_i\} \leq \tfrac{1}{4}.$$

In general, we have

$$\begin{aligned} d(S_n, T_n(\mu)) = \tfrac{1}{2}\Bigg\| &\sum_{i=1}^n (\mu_i - p_i)\exp\left\{\left(\sum_{i=1}^n p_i\right)A\right\}A\varepsilon_0 + \cdots \\ &+ \tfrac{1}{2}\sum_{i=1}^n p_i^2\exp\left\{\left(\sum_{i=1}^n p_i\right)A\right\}A^2\varepsilon_0\Bigg\| + r_n^*(p) + s_n(p, \mu) \end{aligned} \tag{2.12a}$$

with

$$r_n^*(p) \leq 3\sum_{i=1}^n p_i^3 + 2\left\{\sum_{i=1}^n p_i^2\right\}\max(p_1, \dots, p_n)$$

and

$$s_n(p, \mu)$$

$$\leq \tfrac{1}{4}\left\{\sum_{i=1}^n (\mu_i - p_i)\right\}^2 \max\left\{\left\|\exp\left\{\left(\sum_{i=1}^n \mu_i\right)A\right\}A^2\varepsilon_0\right\|, \left\|\exp\left\{\left(\sum_{i=1}^n p_i\right)A\right\}A^2\varepsilon_0\right\|\right\}$$

$$\leq \max\left\{\min\left(1, \frac{1}{\sum_{i=1}^n \mu_i}\right); \min\left(1, \frac{1}{\sum_{i=1}^n p_i}\right)\right\}\left\{\sum_{i=1}^n (\mu_i - p_i)\right\}^2.$$

It should be pointed out that the error estimate for $r_n(p)$ is basically the one given in Kerstan (1964), relation (5). For a discussion of this estimate, cf. also Barbour and Hall (1984).

The following result gives a precise evaluation of the norm terms in Theorem 2.1.

THEOREM 2.2. *For $t > 0$, $\gamma \in \mathbb{R}$ we have*

$$\|e^{tA}A\varepsilon_0\| = 2e^{-t}\frac{t^{[t]}}{[t]!} \sim \frac{2}{\sqrt{2\pi t}} \qquad (t \to \infty), \tag{2.12b}$$

$$\begin{aligned}\|e^{tA}A^2\varepsilon_0\| &= 2\left\{\frac{t^{\alpha-1}(\alpha - t)}{\alpha!} - \frac{t^{\beta-1}(\beta - t)}{\beta!}\right\}e^{-t}\\ &\sim \frac{4}{t\sqrt{2\pi e}} \qquad (t \to \infty),\end{aligned} \tag{2.13}$$

where

$$\alpha = \left[t + \tfrac{1}{2} + \left(t + \tfrac{1}{4}\right)^{1/2}\right] \quad \text{and} \quad \beta = \left[t + \tfrac{1}{2} - \left(t + \tfrac{1}{4}\right)^{1/2}\right], \tag{2.14}$$

$$\begin{aligned}\left\|\gamma t^{-1/2}e^{tA}A\varepsilon_0 + e^{tA}A^2\varepsilon_0\right\| &= 2\left\{\frac{t^{\delta-1}(\delta - t + \gamma\sqrt{t})}{\delta!} - \frac{t^{\eta-1}(\eta - t + \gamma\sqrt{t})}{\eta!}\right\}e^{-t}\\ &\sim \frac{2}{t\sqrt{2\pi}}\left\{\zeta\exp\left(-\tfrac{1}{2}\zeta^{-2}\right) + \frac{1}{\zeta}\exp\left(-\tfrac{1}{2}\zeta^{2}\right)\right\}\\ &\geq \frac{4}{t\sqrt{2\pi e}} \qquad (t \to \infty),\end{aligned} \tag{2.15}$$

where $\delta = [t - \rho + (t + \rho^2)^{1/2}]$, $\eta = [t - \rho - (t + \rho^2)^{1/2}]$ with $\rho = \frac{1}{2}(\gamma\sqrt{t} - 1)$ and $\zeta = \gamma/2 + (1 + \gamma^2/4)^{1/2}$ ($\eta! = \infty$ for $\eta < 0$).

Note that relation (2.13) is just an evaluation of Kerstan's (1964) leading term (†) as can be seen from the proof of Theorem 2.2, giving a simple proof of the right-hand side in Theorem 1.2 via Theorem 2.1, with $t = \sum_{i=1}^n p_i$. On the other hand, if $t = \sum_{i=1}^n p_i \to 0$ for $n \to \infty$, we have $\|e^{tA}A^2\varepsilon_0\| \to \|A^2\varepsilon_0\| = 4$, hence $d(S_n, T_n(p)) \sim \sum_{i=1}^n p_i^2$ as was stated in the introduction. Similarly, the first part of Theorem 1.3 is readily obtained from (2.15) and Theorem 2.1; we only have to observe that we may choose $\gamma^2 = t = \sum_{i=1}^n p_i$ giving $\delta = [t + 1]$ and $\eta = 0$, and that $\lambda_i - p_i = \frac{1}{2}p_i^2 + O(p_i^3)$.

A comparison between (2.12b) and (2.13) in the light of Theorem 2.1 shows that in general, the choice $\mu = \lambda$ is (in an asymptotic way) better than $\mu = p$, as long as

$$\frac{a^{[a]}}{[a]!} < \frac{a^{\alpha-1}(\alpha - a)}{\alpha!} - \frac{a^{\beta-1}(\beta - a)}{\beta!},$$

where α and β are as in (2.14). Straightforward numerical computations show that this occurs as long as $a < a_0$, where $1.59 < a_0 < 1.60$. In this range, we have $[a] = 1$, $\alpha = 3$, and $\beta = 0$, which implies that $x = a_0 - 1$ is the root of the equation

$$x^3 + 3x - 2 = 0$$

which is $x = (\sqrt{2} + 1)^{1/3} - (\sqrt{2} - 1)^{1/3}$, giving $a_0 \doteq 1.596071\ldots$. This proves:

COROLLARY 2.1. *Let* $\sum_{i=1}^n p_i \to a$ *with* $0 \le a < \infty$, *and* $\max(p_1, \ldots, p_n) \to 0$ *for* $n \to \infty$. *Then asymptotically*

$$d(S_n, T_n(\lambda)) < d(S_n, T_n(p))$$

whenever $a < a_0$, *while the converse is true for* $a > a_0$.

3. Asymptotic optimality. First we may observe that for any p, $\inf_\mu d(S_n, T_n(\mu))$ is actually attained by the continuity of $d(m, \mathrm{Po}(t))$ in $t \ge 0$, for any measure $m \in \mathscr{M}$. Also, by Theorem 1.1, Serfling's approach was proven to be asymptotically optimal for $\sum_{i=1}^n p_i \to 0$ $(n \to \infty)$. Suppose now that $\sum_{i=1}^n p_i \to a$ with $0 < a < \infty$, and that $\max(p_1, \ldots, p_n) \to 0$. Then for any optimal choice of μ, we have

$$\begin{aligned} d(T_n(\mu), T_n(p)) &\le d(S_n, T_n(\mu)) + d(S_n, T_n(p)) \\ &\le 2d(S_n, T_n(p)) \to 0 \end{aligned}$$

by Theorem 2.1, and since

$$\begin{aligned} 2d(T_n(\mu), T_n(p)) &\ge |P(T_n(\mu) = 0) - P(T_n(p) = 0)| \\ &= \left| \exp\left(- \sum_{i=1}^n \mu_i \right) - \exp\left(- \sum_{i=1}^n p_i \right) \right|, \end{aligned}$$

we must have $\sum_{i=1}^n \mu_i \sim \sum_{i=1}^n p_i$. But then we can conclude from Theorem 2.1 again that for any such μ there exists some real ξ_a with

$$\sum_{i=1}^n \mu_i = \sum_{i=1}^n p_i + \xi_a \sum_{i=1}^n p_i^2 + o\left(\sum_{i=1}^n p_i^2 \right), \tag{3.1}$$

implying that for any optimal choice of μ,

$$\mu \sim p + \xi_a p^2 \quad \text{with } p^2 = (p_1^2, \ldots, p_n^2), \tag{3.2}$$

giving

$$\begin{aligned} &d(S_n, T_n(\mu)) \\ &\quad \sim \tfrac{1}{4} \sum_{i=1}^n p_i^2 \| 2\xi_a e^{aA} A \varepsilon_0 + e^{aA} A^2 \varepsilon_0 \| \\ &\quad = \tfrac{1}{2} \sum_{i=1}^n p_i^2 \left\{ \frac{a^{\delta-1}(\delta - (1 - 2\xi_a)a)}{\delta!} - \frac{a^{\eta-1}(\eta - (1 - 2\xi_a)a)}{\eta!} \right\} e^{-a} \end{aligned} \tag{3.3}$$

by Theorem 2.2 and the continuity of the left-hand side of (2.15) in $t > 0$. Here

$$\begin{aligned} \delta &= \left[(1-\xi_a)a + \tfrac{1}{2} + \left\{\xi_a^2 a^2 + (1-\xi_a)a + \tfrac{1}{4}\right\}^{1/2}\right], \\ \eta &= \left[(1-\xi_a)a + \tfrac{1}{2} - \left\{\xi_a^2 a^2 + (1-\xi_a)a + \tfrac{1}{4}\right\}^{1/2}\right], \end{aligned} \tag{3.4}$$

which shows that for optimality, we must have $0 \le \xi_a \le \frac{1}{2}$, and

$$\xi_a = \inf\left\{\xi \in \left[0, \tfrac{1}{2}\right] \,\middle|\, a^{\delta(\xi)-\eta(\xi)} \frac{\eta(\xi)!}{\delta(\xi)!} > 1\right\} \quad \text{for } a > \sqrt{2}, \tag{3.5}$$

where $\delta(\xi)$, $\eta(\xi)$ are as in (3.4) with ξ_a being replaced by ξ. Similarly, we can see that for $a \le \sqrt{2}$, $\xi_a = \frac{1}{2}$ is optimal as long as $a^{[a+1]}/[a+1]! \le 1$, which is equivalent to $a \le \sqrt{2}$. This proves Theorem 1.3 completely. Similar arguments show that under the situation of Theorem 1.2, we must have

$$\mu \sim p + \frac{\gamma}{2}\left\{\sum_{i=1}^{n} p_i\right\}^{-1} p^2 \quad \text{for some } \gamma \ge 0; \tag{3.6}$$

but then the right-hand side of relation (2.15) indicates that for an optimal choice, we must have $\zeta = 1$ which corresponds to $\gamma = 0$. This proves Theorem 1.2 completely.

We shall conclude with a discussion of relation (3.5) which allows for a precise evaluation of the second-order term in the minimizing μ. Let $D^{\pm}(\rho) = a - \rho \pm (a + \rho^2)^{1/2}$ for $\rho = a\xi_a - \frac{1}{2}$. Since $D^{+}(\rho)$ is monotonically decreasing in ρ and $-\frac{1}{2} \le \rho \le (a-1)/2$ (since $0 \le \xi_a \le \frac{1}{2}$), and $D^{\pm}(\rho)$ must be an integer by (3.3), we thus have

$$\begin{aligned} a + 1 \le D^{+}(\rho) &\le \left[a + \tfrac{1}{2} + \left(a + \tfrac{1}{4}\right)^{1/2}\right], \\ 0 \le D^{-}(\rho) &\le \left[a + \tfrac{1}{2} - \left(a + \tfrac{1}{4}\right)^{1/2}\right]. \end{aligned} \tag{3.7}$$

This proves the following result.

THEOREM 3.1. *Let* $\sum_{i=1}^{n} p_i \to a$ *with* $\sqrt{2} < a < \infty$ *and* $\max(p_1, \dots, p_n) \to 0$ *for* $n \to \infty$. *Then*

$$\begin{aligned} \xi_a \in \Bigg\{ \frac{1}{2}\left(1 - \frac{N}{a}\frac{N-(a+1)}{N-a}\right) \,\Bigg|\, &]a+1[\;\le N \le \left[a + \tfrac{1}{2} + \left(a + \tfrac{1}{4}\right)^{1/2}\right] \\ &\text{or } \; 0 \le N \le \left[a + \tfrac{1}{2} - \left(a + \tfrac{1}{4}\right)\right] \Bigg\} \cup \{0\}, \end{aligned} \tag{3.8}$$

where $]x[$ *denotes the smallest integer not less than* x.

For example, if $\sqrt{2} < a < \sqrt[3]{6}$, then the only possible value for N in Theorem 3.1 is $N = 3$, from which we obtain

$$\xi_a = \frac{1}{2}\left(1 - \frac{3}{a}\frac{2-a}{3-a}\right). \tag{3.9}$$

Of course, by Theorem 2.2, we have $\xi_a \to 0$ for $a \to \infty$.

4. Proofs of theorems. It remains to prove Theorem 1.1, Theorem 2.1, and Theorem 2.2 only.

PROOF OF THEOREM 1.1. Let m denote the distribution of S_n and let $t = \sum_{i=1}^n \lambda_i$. Then

$$\begin{aligned} 2d(m, \mathrm{Po}(t)) \geq & |m(\{0\}) - \mathrm{Po}(t)(\{0\})| + |m(\{1\}) - \mathrm{Po}(t)(\{1\})| \\ & + |m(\{2,3,\dots\}) - \mathrm{Po}(t)(\{2,3,\dots\})|. \end{aligned}$$

Put $\Lambda = \sum_{i=1}^n \lambda_i$ and $h = t - \Lambda$. It follows that

$$\begin{aligned} 2d(m, \mathrm{Po}(t)) \geq e^{-\Lambda} \Bigg\{ & |1 - e^{-h}| + \left| \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) + \Lambda - e^{-h}(\Lambda + h) \right| \\ & + \left| 1 + \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) + \Lambda - e^{-h}(\Lambda + 1 + h) \right| \Bigg\} \\ = A(h) e^{-\Lambda}. & \end{aligned}$$

For $0 < \Lambda \leq 1$ and $h \to 0$, it can be seen that

$$\begin{aligned} A(h) & \sim 2 \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) + 2h\Lambda - h + |h| \\ & \geq \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) = A(0). \end{aligned}$$

The result now follows from the fact that

$$\begin{aligned} \tfrac{1}{2}A(h) & \geq 1 - e^{-h} + \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) + \Lambda - e^{-h}(\Lambda + h) \\ & \geq \tfrac{1}{2}A(0) \quad \text{for } h \geq 0, \end{aligned}$$

and

$$\begin{aligned} \tfrac{1}{2}A(h) & \geq \sum_{i=1}^n \left(e^{\lambda_i} - 1 - \lambda_i\right) + \Lambda - e^{-h}(\Lambda + h) \\ & \geq \tfrac{1}{2}A(0) \quad \text{for } h \leq 0. \end{aligned}$$

The proof of Theorem 2.1 relies on the following auxiliary result.

LEMMA 4.1. *For $0 \leq s, t < \infty$ we have*

$$e^{sA}\varepsilon_0 - e^{tA}\varepsilon_0 = (s - t)e^{tA}A\varepsilon_0 + R(s,t)$$

with

$$\|R(s,t)\| \le 2\max\left\{\min\left(1,\frac{1}{s}\right);\min\left(1,\frac{1}{t}\right)\right\}(s-t)^2.$$

PROOF. Let $f \in l^1$ be arbitrary. Then from the general Taylor expansion for the semigroup [cf. Pfeifer (1985a), Lemma 4.1] we have

$$e^{sA}f - e^{tA}f = (s-t)e^{tA}Af + \int_t^s (s-u)e^{uA}A^2 f\,du$$

with

$$\left\|\int_t^s (s-u)e^{uA}A^2 f\,du\right\| \le \left|\int_t^s |s-u| \max\{\|e^{sA}A^2 f\|, \|e^{tA}A^2 f\|\}\,du\right|$$
$$= \tfrac{1}{2}(s-t)^2 \max\{\|e^{sA}A^2 f\|, \|e^{tA}A^2 f\|\}.$$

Now everything follows from the observation that by (2.13), we have $\|e^{tA}A^2\varepsilon_0\| \le 4\min(1, 1/t)$.

PROOF OF THEOREM 2.1. For abbreviation, let $t = \sum_{i=1}^n p_i$, $s = \sum_{i=1}^n \mu_i$, $v = \sum_{i=1}^n p_i^2$, and $t_i = \sum_{j \ne i} p_j$. We have

$$A\varepsilon_0 = (-1,1,0,0,\dots) \quad\text{and}\quad A^2\varepsilon_0 = (1,-2,1,0,0,\dots), \tag{4.1}$$

hence

$$\|e^{tA}A\varepsilon_0\| = e^{-t}\sum_{k=0}^{\infty} \frac{t^{k-1}}{k!}|t-k|, \tag{4.2}$$

$$\|e^{tA}A^2\varepsilon_0\| = e^{-t}\sum_{k=0}^{\infty} \frac{t^{k-2}}{k!}\left|t^2 - 2kt + k(k-1)\right|. \tag{4.3}$$

This proves (2.11) using relation (5) in Kerstan (1964), and the Schwarz inequality. In general, since the semigroup and the infinitesimal generator commute, using the decomposition in the proof of Theorem 1 of Le Cam (1960), we obtain for any $f \in l^1$

$$e^{tA}f - \prod_{i=1}^n (I + p_i A)f - \tfrac{1}{2}\sum_{i=1}^n p_i^2 e^{t_i A}A^2 f$$
$$= \sum_{k=1}^n \exp\left(\sum_{i=k+1}^n p_i\right)A\left\{\prod_{i=1}^{k-1}(I+p_iA)\left(e^{p_kA} - (I+p_kA)\right)f \right.$$
$$\left. - \frac{p_k^2}{2}\exp\left(\left(\sum_{i=1}^{k-1} p_i\right)A\right)A^2 f\right\}$$
$$= \sum_{k=1}^n \exp\left(\sum_{i=k+1}^n p_i\right)A\left\{\prod_{i=1}^{k-1}(I+p_iA)\left(e^{p_kA} - \left(I + p_kA + \frac{p_k^2}{2}A^2\right)\right)f \right.$$
$$\left. + \frac{p_k^2}{2}A^2\left[\prod_{i=1}^{k-1}(I+p_iA) - \exp\left(\sum_{i=1}^{k-1} p_i\right)A\right]f\right\},$$

which implies that

$$(4.4)\quad \left\| e^{tA}\varepsilon_0 - \prod_{i=1}^{n}(1+p_iA)\varepsilon_0 - \tfrac{1}{2}ve^{tA}A^2\varepsilon_0 \right\| \le \tfrac{16}{3}\sum_{i=1}^{n} p_i^3 + 4\sum_{k=2}^{n} p_k^2 \frac{\sum_{i=1}^{k-1}p_i^2}{\sum_{i=1}^{k-1}p_i},$$

using the contraction property, Proposition 1.1.6 in Butzer and Berens (1967), relation (2.5), the estimation (2.2) in Barbour and Hall (1984) and the fact that

$$\|e^{tA}A^2f - e^{t_iA}A^2f\| \le p_i\|e^{t_iA}A^3f\| \le 8p_i\|f\|,$$

which is to be proved similarly to Lemma 4.1.

The result now follows from the observation that

$$\begin{aligned} 2d(S_n, T_n(\mu)) &= \left\| e^{sA}\varepsilon_0 - \prod_{i=1}^{n}(I+p_i)A\varepsilon_0 \right\| \\ &= \left\| (e^{sA} - e^{tA})\varepsilon_0 + e^{tA}\varepsilon_0 - \prod_{i=1}^{n}(I+p_iA)\varepsilon_0 \right\| \\ &= \left\| (s-t)e^{tA}A\varepsilon_0 + \frac{v}{2}e^{tA}A^2\varepsilon_0 \right\| + 2r_n^*(p) + 2s_n(p,\mu), \end{aligned}$$

where

$$r_n^*(p) \le \tfrac{8}{3}\sum_{i=1}^{n} p_i^3 + 2\sum_{k=1}^{n} p_k^2 \max(p_1,\dots,p_n)$$

by (4.4), and the estimation for $s_n(p,\mu)$ is due to Lemma 4.1.

Proof of Theorem 2.2. It suffices to prove relations (2.12b) and (2.15) since (2.13) and (2.14) are obtained from (2.15) for $\gamma = 0$. Similar to (4.2) and (4.3), we have

$$\begin{aligned} &\|\gamma t^{-1/2}e^{tA}A\varepsilon_0 + e^{tA}A^2\varepsilon_0\| \\ &\quad = e^{-t}\sum_{k=0} \frac{t^{k-2}}{k!}\left| k^2 - 2k\left(t + \frac{1}{2} - \frac{\gamma\sqrt{t}}{2}\right) + t(t-\gamma\sqrt{t})\right|, \end{aligned}$$

where

$$k^2 - 2k\left(t + \frac{1}{2} - \frac{\gamma\sqrt{t}}{2}\right) + t(t-\gamma\sqrt{t}) > 0$$

if and only if

$$k < t - \rho - (\rho^2+t)^{1/2} \quad \text{or} \quad k > t - \rho + (\rho^2+t)^{1/2}.$$

This gives the left-hand side of relation (2.15).

Since by Stirling's formula, we have

$$(4.5)\qquad \frac{t^w}{w!} \sim \exp\left(t - \frac{(w-t)^2}{2t}\right)\{2\pi t\}^{-1/2}$$

whenever w is integer such that $w - t = O(\sqrt{t})$, the left-hand side of (2.15) is asymptotically equal to

$$\frac{2}{\sqrt{t}}e^{-t}\left(\zeta\frac{t^{\delta}}{\delta!}+\frac{1}{\zeta}\frac{t^{\eta}}{\eta!}\right)\sim\frac{2}{t\sqrt{2\pi}}\left(\zeta\exp\left(-\frac{1}{2\zeta^{2}}\right)+\frac{1}{\zeta}\exp\left(-\frac{\zeta^{2}}{2}\right)\right),$$

as requested. To complete the proof, note that

$$g(x)=x\exp\left(-\frac{1}{2x^{2}}\right)+\frac{1}{x}\exp\left(-\frac{x^{2}}{2}\right)\qquad(x>0)$$

is minimal for $x = 1$ with $g(1) = 2/\sqrt{e}$. The proof of (2.12b) is similar.

Note added in proof. Under the assumption that $\max\{p_1,\dots,p_n\} < \frac{1}{2}$ the remainder term estimations in (2.11) and (2.12a) can be sharpened as follows. Let A be a bounded operator on a Banach space $\mathscr{X}$ with $\|A\| < 1$. Then

$$\log(I+A) = -\sum_{k=1}^{\infty}\frac{(-1)^k}{k}A^k \tag{4.6}$$

exists as a bounded operator on $\mathscr{X}$ with

$$\|\log(I+A)\| \le -\log(1-\|A\|), \tag{4.7}$$

and

$$\exp\{\log(I+A)\} = I + A. \tag{4.8}$$

Under the situation of Theorem 2.1 (cf. also the corresponding proof) we thus have, letting again $t = \sum_{i=1}^{n} p_i$,

$$\begin{aligned}&\left\|e^{tA}f-\prod_{i=1}^{n}(I+p_iA)f-\tfrac{1}{2}\left\{\sum_{i=1}^{n}p_i^2\right\}e^{tA}A^2f\right\|\\&\qquad\le\|e^{tA}Cf\|+\|e^{tA}D^2f\|\sum_{k=0}^{\infty}\frac{\|D\|^k}{(k+2)!},\end{aligned} \tag{4.9}$$

where

$$C=\sum_{k=3}^{\infty}\frac{1}{k}\left\{\sum_{i=1}^{n}p_i^k\right\}(-A)^k,\qquad D=C+\tfrac{1}{2}\left\{\sum_{i=1}^{n}p_i^2\right\}A^2. \tag{4.10}$$

From here it follows that with $M_n = \max\{p_1,\dots,p_n\}$, $L_n = -\log(1-2M_n)$, we have

$$\|e^{tA}Cf\| \le \tfrac{1}{3}\left\{\sum_{i=1}^{n}p_i^3\right\}\|e^{tA}A^3f\|\exp(L_n), \tag{4.11}$$

$$\|e^{tA}D^2f\| \le \tfrac{1}{4}\left\{\sum_{i=1}^{n}p_i^2\right\}^2\|e^{tA}A^4f\|(1+4M_n(1+L_n))^2. \tag{4.12}$$

Generalizing (4.2) and (4.3), it is easily seen that

$$\|e^{tA}Af\| = \|e^{tA}A\|, \qquad \|e^{tA}A^2f\| = \|e^{tA}A^2\| \tag{4.13}$$

for all f with $\|f\| = 1$ (especially $f = \varepsilon_0$), such that the r.h.s. of (4.9) can now be estimated by

$$\begin{aligned}&\tfrac{1}{3}\left\{\sum_{i=1}^n p_i^3\right\}\left\|\exp\left(\frac{t}{2}A\right)A\right\|\left\|\exp\left(\frac{t}{2}A\right)A^2\right\|\exp(L_n)\\ &+\tfrac{1}{8}\left\{\sum_{i=1}^n p_i^2\right\}^2\left\|\exp\left(\frac{tA}{2}\right)A^2\right\|^2\exp\left((1+2L_n)\sum_{i=1}^n p_i^2\right)(1+4M_n(1+L_n))^2,\end{aligned} \tag{4.14}$$

where the norm terms are now again given in (4.2) and (4.3). Especially, (4.14) is an upper bound for the remainder terms $2r_n(p)$ and $2r_n^*(p)$ in Theorem 2.1, improving also the bounds in Kerstan (1964) and Barbour and Hall (1984) (Corollary to Theorem 3) for large values of $\sum_{i=1}^n p_i$ since by Theorem 2.2, if $\sum_{i=1}^n p_i \to \infty$ and $\max(p_1, \ldots, p_n) \to 0$ $(n \to \infty)$, (4.14) is asymptotically

$$\begin{aligned}&\frac{16}{3\pi\sqrt{2e}}\left\{\sum_{i=1}^n p_i^3\right\}\Big/\left\{\sum_{i=1}^n p_i\right\}^{3/2}\\ &\quad+\frac{4}{\pi e}\exp\left(\sum_{i=1}^n p_i^2\right)\left\{\sum_{i=1}^n p_i^2\right\}^2\Big/\left\{\sum_{i=1}^n p_i\right\}^2\\ &\approx 0.728\left\{\sum_{i=1}^n p_i^3\right\}\Big/\left\{\sum_{i=1}^n p_i\right\}^{3/2}\\ &\quad+0.468\exp\left(\sum_{i=1}^n p_i^2\right)\left\{\sum_{i=1}^n p_i^2\right\}^2\Big/\left\{\sum_{i=1}^n p_i\right\}^2.\end{aligned} \tag{4.15}$$

The foregoing remarks show that the additional condition in Theorem 1.2 can be weakened to

$$\sum_{i=1}^n p_i^2 = O(1) \quad (n \to \infty).$$

Acknowledgments. The authors wish to thank the referee for his constructive suggestions which led to a more general presentation of the results, and Robert Serfling for partial financial support of this research under Office of Naval Research Contract N00014-76-C-0608.

REFERENCES

BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

BUTZER, P. L. and BERENS, H. (1967). *Semi-Groups of Operators and Approximation.* Springer, New York.

CHEN, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions. *Ann. Probab.* **2** 178–180.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

KERSTAN, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrsch. verw. Gebiete* **2** 173–179.

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

PFEIFER, D. (1985a). Approximation-theoretic aspects of probabilistic representations for operator semigroups. *J. Approx. Theory* **43** 271–296.

PFEIFER, D. (1985b). A semigroup setting for distance measures in connexion with Poisson approximation. *Semigroup Forum* **31** 199–203.

SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.* **3** 726–731.

SERFLING, R. J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials. *SIAM Rev.* **20** 567–579.

VERVAAT, W. (1969). Upper bounds for the distance in total variation between the binomial and negative binomial and the Poisson distribution. *Statist. Neerlandica* **23** 79–86.

7 AVENUE DU CHÂTEAU
92340 BOURG-LA-REINE
FRANCE

RWTH AACHEN
DEPARTMENT OF STATISTICS
WÜLLNERSTRASSE 3
D-5100 AACHEN
WEST GERMANY