Signed Poisson approximation: a possible alternative to normal and Poisson laws

VYDAS ČEKANAVIČIUS and JULIUS KRUOPIS

Department of Mathematics, Vilnius University, Naugarduko 24, 2006 Vilnius, Lithuania

Signed Poisson approximation is a signed measure, has the structure of the Poisson distribution and can be regarded as a special sort of asymptotic expansion when the expansion is in the exponent. For certain lattice distributions signed Poisson approximation combines advantages of both the normal and Poisson approximations. For the generalized binomial distribution estimates with respect to the total variation and Wasserstein distances are obtained. The results are exemplified by Bernoulli decomposable variables.

Keywords: generalized binomial distribution; signed Poisson measure; total variation norm; Wasserstein distance

1. Introduction

Usually the normal and Poisson approximations are regarded as opposites: the former is associated with the common and typical, the latter with the rare and exceptional. Though the normal approach still dominates in limit theorems, the last decade has seen rapidly growing interest in Poisson laws; see Aldous (1989), Barbour *et al.* (1992) and references therein. However, both the normal and Poisson approaches have their faults. For discrete variables it is meaningless to use the normal approximation if the accuracy is measured in total variation because the answer is always 1. Besides, even for the lattice variables and uniform distance cumbersome summands must be added to the Edgeworth asymptotic expansion. On the other hand, the Poisson approximation ensures the matching of one moment only.

For some distributions the advantages of both approaches can be combined by using the signed Poisson or signed compound Poisson measures. Though in general such approximations are not distributions, they have the same structures as Poisson or compound Poisson distributions. Signed Poisson measures as approximations were introduced by Presman (1983) and Kornya (1983). They were applied in actuarial mathematics (see Hipp 1986; De Pril and Dhaene 1992), probabilistic number theory (Šiaulys and Čekanavičius 1989) and even in a such general setting as the first uniform Kolmogorov theorem (see Čekanavičius 1996). For lattice distributions, the signed Poisson approximations were explored by Kruopis (1986a; 1986b). For Markov chains, signed Poisson approximations were introduced by Borovkov and Pfeifer (1996). General aspects of the signed Poisson approach were discussed in Čekanavičius (1997).

Our paper deals with one special signed Poisson approximation introduced in Kruopis (1986a). We introduce a new unified approach for the total variation and Wasserstein

distances, considering the generalized binomial distribution. We use a discrete Poisson structured measure which for some values of the parameters behaves as the Poisson law, while for other values it is quite comparable with the normal law. Moreover, if the limiting distribution is Poisson, then the analogue of the Simons–Johnson theorem holds. Wasserstein distance is the main distance considered in this paper. We exemplify our results by the so-called Bernoulli decomposable distributions. Though it is more natural to use signed Poisson approximations for the lattice distributions we demonstrate that, even in the general case, such approximations are comparable with the Berry–Esseen theorem. Moreover, the signed Poisson approximation of this paper satisfies simple recursive formulae and can be expressed in terms of Bessel functions.

We use the following notation. Denote by $\mathbb{R}$ the set of real numbers, and by $\mathbb{N}$ the set of natural numbers. Let E_a denote the distribution concentrated at a point a, $E \equiv E_0$. C_m is used to denote positive absolute constants. Products and powers of measures are understood in the convolution sense: $FG = F * G$, $F^n = F^{*n}$, $F^0 = E$. For any finite variation measure V we denote by $\hat{V}(t)$ its Fourier–Stieltjes transform, and by $\exp\{V\} = \sum_{k=0}^{\infty} V^k/k!$ its exponential measure. We also denote by $|V| = \sup_x |V\{(-\infty, x)\}|$ the analogue of the uniform Kolmogorov distance, by $V(x) = V\{(-\infty, x)\}$ the analogue of distribution function, and by $\|V\|$ the total variation norm of V. Note that $\widehat{\exp}\{V\}(t) = \exp(\hat{V}(t))$ and, for V concentrated on $\mathbb{N}$, $\|V\| = \sum_{k=-\infty}^{\infty} |V\{k\}|$. The total variation norm is equivalent to the total variation distance. Moreover, if two measures of finite variation V_1, V_2 satisfy the relation $V_1\{\mathbb{R}\} = V_2\{\mathbb{R}\}$, then

$$\|V_1 - V_2\| = 2d_{TV}(V_1, V_2),$$

where d_{TV} denotes the total variation distance. For V concentrated on integers Wasserstein distance (also known as Kantorovich, Dudley or Fortet–Mourier distance) is defined by

$$\|V\|_W = \sum_{k=-\infty}^{\infty} |V(k)| = \sum_{-\infty}^{\infty} |V\{(-\infty, k)\}|.$$

For further discussion on probability distances we refer readers to Barbour *et al.* (1992, pp. 253–255).

For a distribution F and $h \geqslant 0$ we denote Levy's concentration function by

$$Q(F, h) = \sup_x F\{[x, x + h]\}.$$

Now we introduce the signed Poisson approximation which is used in this paper. Let $\lambda_1 > 0$, $\lambda_2 \in \mathbb{R}$. Set

$$G(\lambda_1, \lambda_2) = \exp\{\lambda_1(E_1 - E) - \lambda_2(E_{-1} - E)\}. \tag{1.1}$$

By definition

$$\hat{G}(\lambda_1, \lambda_2)(t) = \exp\{\lambda_1(e^{it} - 1) - \lambda_2(e^{it} - 1)\}.$$

For any integer k we have the inversion formula

$$G(\lambda_1, \lambda_2)\{k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \hat{G}(\lambda_1, \lambda_2)(t) \, dt. \tag{1.2}$$

Furthermore, for $\lambda_2 > 0$, we have

$$G(\lambda_1, \lambda_2)\{k\} = \exp\{-\lambda_1 + \lambda_2\}\left(\frac{\lambda_1}{\lambda_2}\right)^{k/2} J_k(2\sqrt{\lambda_1\lambda_2}), \tag{1.3}$$

and, for $\lambda_2 < 0$,

$$G(\lambda_1, \lambda_2)\{k\} = \exp\{-\lambda_1 + \lambda_2\}\left(\frac{\lambda_1}{|\lambda_2|}\right)^{k/2} I_k(2\sqrt{\lambda_1|\lambda_2|}). \tag{1.4}$$

Here $J_k(\cdot)$, $I_k(\cdot)$ are the Bessel functions of the first and second kind, respectively. Noting that both of them satisfy quite simple recursive formulae we obtain, for $\lambda_2 \neq 0$,

$$\lambda_2 G(\lambda_1, \lambda_2)\{k+1\} = k G(\lambda_1, \lambda_2)\{k\} - \lambda_1 G(\lambda_1, \lambda_2)\{k-1\}. \tag{1.5}$$

For practical calculations it is possible to use (1.2) and the fast Fourier transforms, recursive formula (1.5), or the asymptotic formulae for Bessel functions. That technical problems of calculation can be succesfully solved was clearly demonstrated by Hipp (1986), Kuon *et al.* (1987) and Dhaene and De Pril (1994), who considered much more general signed Poisson measures.

2. Main results

Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent Bernoulli variables, $P(\xi_j = 1) = 1 - P(\xi_j = 0) = p_j$, $S_n = \xi_1 + \ldots + \xi_n$, and let H_j be the distribution of ξ_j, i.e. $H_j = (1 - p_j)E + p_j E_1$. Obviously, $\prod H_j$ is the distribution of S_n. We shall call $\prod H_j$ a generalized binomial distribution. In the literature the term 'Poisson binomial distribution' is also used (Le Cam 1960).

Approximation of $\prod H_j$ by the Poisson distribution is one of the most popular subjects in discrete limit theorems; see Le Cam (1960), Serfling (1975), Barbour and Hall (1984), Barbour *et al.* (1995), Witte (1990), Wang (1993) and references therein. Signed Poisson approximations were applied by Kruopis (1986a) and Borovkov (1988). Here we shall recall some of their results. Set

$$\lambda = \sum_{j=2}^{n} p_j, \qquad \theta = \max\left(1, \sum_{j=1}^{n} p_j(1-p_j)\right), \tag{2.1}$$

$$a = \lambda - \sum_{j=1}^{n} p_j^2/2, \qquad b = \sum_{j=1}^{n} p_j^2/2. \tag{2.2}$$

We denote the Poisson distribution with parameter λ by $\exp\{\lambda(E_1 - E)\}$. $G(a, b)$ is defined as in Section 1, i.e.

$$G(a, b) = \exp\left\{\sum_{j=1}^{n} p_j\left(1 - \frac{p_j}{2}\right)(E_1 - E) - \sum_{j=1}^{n} \frac{p_j^2}{2}(E_{-1} - E)\right\}. \tag{2.3}$$

The following estimates hold:

$$\left\| \prod_{j=1}^{n} H_j - \exp\{\lambda(E_1 - E)\} \right\| \le 2(1 - e^{-\lambda}) \sum_{j=1}^{n} p_j^2 / \lambda; \qquad (2.4)$$

$$\left\| \prod_{j=1}^{n} H_j - \exp\{\lambda(E_1 - E)\} \right\|_W \le -\frac{1}{2}\frac{\sqrt{e\lambda}}{\sqrt{2\pi}} \ln\left(1 - 2\exp\{2\max p_j\}\frac{2b}{\lambda}\right), \qquad (2.5)$$

if $\dfrac{2b}{\lambda} < \dfrac{1}{2} \exp\{-2\max p_j\}$; and

$$\left| \prod_{j=1}^{n} H_j - G(a, b) \right| \le \varepsilon, \qquad (2.6)$$

where

$$\varepsilon = \min\left(0.2\left(3\sum_{j=1}^{n} p_j^2 - 2\sum_{j=1}^{n} p_j^3\right)\theta^{-3/2} + 2\sum_{j=1}^{n} p_j^2\theta^{-2} + 4.2\left(\sum_{j=1}^{n} p_j^2\right)^2\theta^{-3},\right.$$

$$\left. 0.4\left(3\sum_{j=1}^{n} p_j^2 - 2\sum_{j=1}^{n} p_j^3\right) + 2.7\sum_{j=1}^{n} p_j^2 + 2.3\left(\sum_{j=1}^{n} p_j^2\right)^2\right).$$

Expression (2.4) is due to Barbour and Hall (1984), (2.5) to Witte (1990) and (2.6) to Kruopis (1986a).

Remark 2.1. Note that we use the total variation norm instead of the total variation distance. Therefore the estimate from Barbour and Hall (1984) was multiplied by 2.

Remark 2.2. If $p_j \le C_1 < 1$, $j = 1, \ldots, n$, then

$$C_2 \min\left(\sum_{j=1}^{n} p_j^2, \sum_{j=1}^{n} p_j^2\lambda^{-3/2}\right) \le \varepsilon \le C_3 \min\left(\sum_{j=1}^{n} p_j^2, \sum_{j=1}^{n} p_j^2\lambda^{-3/2}\right). \qquad (2.7)$$

We shall show that, for the total variation norm, the estimate analogous to (2.6) holds and that, for the Wasserstein distance, the rate of convergence for $G(a, b)$ is much better than the rate in (2.5). Moreover, we shall introduce a new special functional which depends on the interval length and combines both distances in one. Set

$$P = \exp\{\theta(E_1 - E)/4 + \theta(E_{-1} - E)/4\}. \qquad (2.8)$$

Note that $\hat{P}(t) = \exp\{-\theta \sin^2(t/2)\}$.

Theorem 2.1. *Let $h \geqslant 0$. Then the following inequality holds:*

$$\sum_{k=-\infty}^{\infty} \left| \prod_{j=1}^{n} H_j\{[k, \, k + h]\} - G(a, \, b)\{[k, \, k + h]\} \right| \leqslant C_4 \theta^{-1} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta} \right)^2 \right) Q(P, \, h).$$

(2.9)

The proof of the theorem is carried out in Section 3.

By the properties of Levy's concentration function and the inversion formula,

$$Q(P, \, 0) \leqslant \sup_k |P\{k\}| \leqslant C_5 \int_{-\pi}^{\pi} \hat{P}(t) \, dt \leqslant C_6 \theta^{-1/2}.$$

(2.10)

Therefore, putting $h = 0$ in (2.9), we can deduce the estimate in variation.

Theorem 2.2. *The following inequality holds:*

$$\left\| \prod_{j=1}^{n} H_j - G(a, \, b) \right\| \leqslant C_7 \theta^{-3/2} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta} \right)^2 \right).$$

(2.11)

In Section 3 we shall prove that (2.9) holds also as $h \to \infty$. Note that $\prod H_j\{\mathbb{R}\} = G(a, \, b)\{\mathbb{R}\}$. Consequently,

$$\sum_{k=-\infty}^{\infty} \left| \prod_{j=1}^{n} H_j\{[k, \, \infty)\} - G(a, \, b)\{[k, \, \infty)\} \right| = \sum_{k=-\infty}^{\infty} \left| \prod_{j=1}^{n} H_j(k) - G(a, \, b)(k) \right|$$

$$= \left\| \prod_{j=1}^{n} H_j - G(a, \, b) \right\|_W.$$

(2.12)

Obviously, $Q(P, \, h) \to 1$ as $h \to \infty$. Therefore, taking into account previous remarks, we can state the following result for the Wasserstein distance.

Theorem 2.3. *The following inequality holds:*

$$\left\| \prod_{j=1}^{n} H_j - G(a, \, b) \right\|_W \leqslant C_8 \theta^{-1} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta} \right)^2 \right).$$

(2.13)

As follows from (2.7), under very mild restrictions, the estimate (2.11) is of the same order as (2.6). In the sense of order, (2.11) and (2.13) can significantly improve (2.4) and (2.5). We shall demonstrate this by considering the case $p_i \equiv p \to 0$, as $n \to \infty$, i.e. the case of some binomial distribution. Then the estimate in (2.4) is of order $\min(np^2, \, p)$, while the estimate in (2.11) is of order $\min(np^2, \, \sqrt{p}/\sqrt{n})$. Notably this estimate is not just of the same or better order than that provided by the Poisson approximation, but is also of better order than that provided by the normal approximation. The advantages of the signed Poisson approximation are even more evident for Wasserstein distance. Considering the above-mentioned binomial distribution, we see that the Poisson approximation in (2.5) is of

order $O(p\sqrt{np})$, i.e. for convergence to zero it should necessarily be $p = o(n^{-1/3})$. Meanwhile the estimate (2.13) is of order p (i.e. is vanishing under $p = o(1)$). Note that (2.9) also provides estimates for other values of h, not only for $h = 0$ or $h = \infty$.

The above-formulated results can be applied to the Bernoulli-decomposable variables. A random variable η is called Bernoulli decomposable if, for some n,

$$\eta = \eta_1 + \eta_2 + \ldots + \eta_n, \tag{2.14}$$

where $\eta_1, \eta_2, \ldots, \eta_n$ are independent Bernoulli variables. If the generating function has only real roots then the corresponding random variable satisfies (2.14). Decomposition (2.14) may not have any physical interpretation. It seems that Harper (1967) was the first to use (2.14) in combinatorial setting. Probably the most recent paper on the subject (including other discrete decompositions) is by Quine (1994). We are unaware of any result for Bernoulli-decomposable variables obtained with respect to the Wasserstein distance though other distances have been applied; see Vatutin and Mikhailov (1982) and Kruopis (1986a). In this section we compare the accuracy of Poisson and signed Poisson approximation when the Wasserstein distance is applied. We do not consider the degenerate case $\mathbf{E}\eta \to 0$.

Usualy it suffices to establish only the existence of decomposition (2.14) (but not the exact distributions of η_j) because the estimates depend on the moments of η, rather than on the individual probabilities of η_j. Thus we can rewrite Theorem 2.3 in the following manner.

Theorem 2.4. *Let the distribution H correspond to the Bernoulli-decomposable random variable η. Then*

$$\left\| H - G\left(\frac{\mathbf{E}\eta + \mathbf{D}\eta}{2}, \frac{\mathbf{E}\eta - \mathbf{D}\eta}{2}\right) \right\|_W \leqslant C_9 \frac{\mathbf{E}\eta - \mathbf{D}\eta}{\max(1, \mathbf{D}\eta)}\left(1 + \left(\frac{\mathbf{E}\eta}{\mathbf{D}\eta}\right)^2\right). \tag{2.15}$$

Remark 2.3. As can be seen from (2.5) the Poisson approximation with parameter $\mathbf{E}\eta$ can provide (at best) an estimate of order

$$O\left(\frac{\mathbf{E}\eta - \mathbf{D}\eta}{\mathbf{E}\eta}\sqrt{\mathbf{E}\eta}\right). \tag{2.16}$$

Inspired by the papers of Vatutin and Mikhailov (1982) and Quine (1994) we shall consider some classical cases of Bernoulli decomposable variables.

Example 1 *Hypergeometric distribution*. Let $2M \leqslant N$, $2n \leqslant N$ and let Y have the hypergeometric distribution, i.e., for $k = 1, 2, \ldots, \min(n, M)$,

$$P(Y = k) = \binom{M}{k}\binom{N - M}{n - k}\bigg/\binom{N}{n}, \qquad \mathbf{E}Y = \frac{nM}{N}, \qquad \mathbf{D}Y = \mathbf{E}Y\frac{(N - M)(N - n)}{N(N - 1)}.$$

As was shown in Vatutin and Mikhailov (1982), Y satisfies (2.14). Assuming that n, $N, M \to \infty$, $M/N \to 0$, $n/N \to 0$, we obtain that the estimate in (2.15) is of order

$O(M/N + n/N)$. Meanwhile the Poisson approximation is of order $O((M/N + n/N)\sqrt{nM/N})$, which can be much slower.

Example 2 *Number of permutations cycles.* Any permutation of $(1, 2, \ldots, n)$ can be expressed in terms of cycles, where a cycle is a group of elements permuted amongst themselves. It was proved in Feller (1968, p. 258) that the number of cycles Y_c is Bernoulli decomposable. Moreover, in decomposition (2.14),

$$P(\eta_j = 1) = \frac{1}{n - j + 1}, \qquad j = 1, 2, \ldots, n.$$

Consequently the estimate in (2.15) is of order $O(1/\ln n)$. Meanwhile the Poisson approximation is of order $O(1/\sqrt{\ln n})$.

3. Non-uniform estimates

In this section we obtain non-uniform estimates for measures defined in Section 2. We employ the same approach as in Čekanavičius (1993), considering estimates that are non-uniform with respect to the interval's length and its beginning point. We use the same notation as in the previous section. Set

$$d = \min(|m - \lambda|, |m - \lambda + h + 1|), \qquad \Delta = \prod_{j=1}^{n} H_j - G(a, b). \tag{3.1}$$

Theorem 3.1. *Let* $m \in \mathbb{N}$, $h > 0$. *Then*

$$|\Delta\{[m, m + h]\}| \left(1 + \frac{d^2}{\theta}\right) \leq C_{10}\theta^{-3/2} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta}\right)^2\right) Q(P, h). \tag{3.2}$$

Proof. For the sake of brevity, set

$$G_j = \exp\left\{\left(p_j - \frac{p_j^2}{2}\right)(E_1 - E) - \frac{p_j^2}{2}(E_{-1} - E)\right\}, \tag{3.3}$$

$$U(t) = \sum_{j=1}^{n} \prod_{l=1}^{j-1} (\hat{H}_l(t)e^{-itp_l}) \prod_{l=j+1}^{n} (\hat{G}_l(t)e^{-itp_j}) (\hat{H}_j(t) - \hat{G}_j(t))(1 - e^{-it})^{-1}. \tag{3.4}$$

Note that

$(\hat{H}_j(t) - \hat{G}_j(t))(1 - e^{-it})^{-1}$

$$= -\exp\left\{p_j(e^{it} - 1) - \frac{p_j^2}{2}(e^{it} - 1)^2\right\} \frac{p_j^2}{2}(e^{it} - 1)^2 \int_0^1 \exp\{\tau p_j^2(e^{it} - 1)^2(1 - e^{-it})/2\}\, d\tau$$

$$+ \frac{p_j^3}{2}e^{it}(e^{it} - 1)^2 - \frac{p_j^4}{8}e^{it}(e^{it} - 1)^3 - \frac{p_j^3}{2}e^{it}(e^{it} - 1)^2\left(1 - \frac{p_j}{2}(e^{it} - 1)\right)^3 \int_0^1 (1 - \tau)^2$$

$$\times\ \exp\left\{\tau(p_j(e^{it} - 1) - \frac{p_j^2}{2}(e^{it} - 1)^2\right\} d\tau. \tag{3.5}$$

Using (3.5) and the trivial inequality $\exp\{2p_j(1 - p_j)\sin^2(t/2)\} \leqslant \exp\{2\}$, after quite standard calculation we obtain

$$|U(t)| \leqslant C_{11} \sum_{j=1}^{n} p_j^2 \sin^2\left(\frac{t}{2}\right)\hat{P}(t) \leqslant C_{12} \sum_{j=1}^{n} p_j^2 \theta^{-1}, \tag{3.6}$$

$$|U'(t)| \leqslant C_{13} \sum_{j=1}^{n} p_j^2 \left|\sin\left(\frac{t}{2}\right)\right|\hat{P}^2(t)\left(1 + \lambda\sin^2\frac{t}{2}\right) \leqslant C_{14} \sum_{j=1}^{n} p_j^2 \theta^{-1/2}\left(1 + \left(\frac{\lambda}{\theta}\right)^2\right)\hat{P}(t), \tag{3.7}$$

$$|U''(t)| \leqslant C_{15} \sum_{j=1}^{n} p_j^2 \left(1 + \lambda^2\sin^4\left(\frac{t}{2}\right)\right)\hat{P}^2(t) \leqslant C_{16} \sum_{j=1}^{n} p_j^2\left(1 + \left(\frac{\lambda}{\theta}\right)^2\right)\hat{P}(t). \tag{3.8}$$

Let $h \in \mathbb{N}$. Summing the inversion formula (1.2), we obtain

$$\Delta\{[m, m + h]\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} U(t)(e^{-it(m-\lambda)} - e^{-it(m-\lambda+h+1)})\, dt. \tag{3.9}$$

Evidently

$$\left|e^{-it(m-\lambda)} - e^{-it(m-\lambda+h+1)}\right| \leqslant \min\left(2, 2(h + 1)\left|\sin\left(\frac{t}{2}\right)\right|\right).$$

Therefore, by (3.6),

$$|\Delta\{[m, m + h]\}| \leqslant C_{17}\theta^{-1} \sum_{j=1}^{n} p_j^2 \int_{-\pi}^{\pi} \min\left(1, \frac{h + 1}{\sqrt{\theta}}\right)\hat{P}(t)\, dt. \tag{3.10}$$

Note that $\exp\{i\pi\} = \exp\{-i\pi\}$. Taking into account (3.5) and integrating by parts for $d > 0$, we obtain

$$\left|\int_{-\pi}^{\pi} U(t)e^{-it(m-\lambda)}\, dt\right| = \left|-\frac{1}{(m-\lambda)^2}\int_{-\pi}^{\pi} U''(t)e^{-it(m-\lambda)}\, dt\right| \leqslant \frac{1}{d^2}\int_{-\pi}^{\pi} |U''(t)|\, dt. \tag{3.11}$$

An analogous estimate holds if to replace $m - \lambda$ by $m - \lambda + h + 1$. Therefore, from (3.8) and (3.9) we deduce, for all d,

$$|\Delta\{[m, m + h]\}| d^2 \leqslant C_{18} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta}\right)^2\right) \int_{-\pi}^{\pi} \hat{P}(t)\, dt. \tag{3.12}$$

Again let $d > 0$. Then, integrating by parts in (3.9) and using (3.6)–(3.8), we obtain

$$2\pi|\Delta\{[m, m + h]\}| \leqslant \left|\int_{-\pi}^{\pi} U(t)(1 - e^{-it(h+1)})e^{-it(m-\lambda)}\, dt\right|$$

$$\leqslant \frac{1}{|m - \lambda|} \left|\int_{-\pi}^{\pi} e^{-it(m-\lambda)} U'(t)(1 - e^{it(h+1)})\, dt\right| + \frac{h + 1}{|m - \lambda|} \left|\int_{-\pi}^{\pi} U(t)e^{-it(m-\lambda+h+1)}\, dt\right|$$

$$\leqslant \frac{1}{|(m - \lambda)^2|} \int_{-\pi}^{\pi} |(U'(t)(1 - e^{-it(h+1)}))'|\, dt + \frac{h + 1}{(m - \lambda)(m - \lambda + h + 1)} \int_{-\pi}^{\pi} |U'(t)|\, dt$$

$$\leqslant \frac{C_{19}}{d^2} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta}\right)^2\right) \int_{-\pi}^{\pi} \frac{h + 1}{\sqrt{\theta}} \hat{P}(t)\, dt. \tag{3.13}$$

Combining (3.10), (3.12) and (3.13), we obtain, for $d \geqslant 0$,

$$|\Delta\{[m, m + h]\}| \left(1 + \frac{d^2}{\theta}\right) \leqslant C_{27}\theta^{-1} \sum_{j=1}^{n} p_j^2 \left(1 + \left(\frac{\lambda}{\theta}\right)^2\right) \int_{-\pi}^{\pi} \min\left(1, \frac{h + 1}{\sqrt{\theta}}\right) \hat{P}(t)\, dt. \tag{3.14}$$

So far we have considered integer values of h. But by Čekanavičius (1993) it was proved that in this case

$$\int_{-\pi}^{\pi} \min\left(1, \frac{h + 1}{\sqrt{\theta}}\right) \hat{P}(t)\, dt \leqslant \frac{C_{21}}{\sqrt{\theta}} Q(P, h). \tag{3.15}$$

From (3.14) and (3.15) we obtain (3.2). Now let $h > 0$ be not necessarily be an integer. Denote the integer part of h by $[h]$. Then $\Delta\{[m, m + h]\} = \Delta\{[m, m + [h]]\}$ and

$$|\Delta\{[m, m + h]\}| \left(1 + \frac{d^2}{\theta}\right) \leqslant |\Delta\{[m, m + [h]]\}| 2\left(1 + \frac{(\min(|\{m - \lambda\}|, |m - \lambda + [h] + 1|))^2}{\theta}\right). \tag{3.16}$$

Now we can apply (3.2) which has already been proved for $[h]$, and notice that $Q(P, [h]) \leqslant Q(P, h)$. $\qquad\square$

Remark 3.1. It is impossible to deduce (3.2) from the results of Čekanavičius (1993).

Proof of Theorem 2.1. Let $h > 0$. We have

$$\sum_{m=-\infty}^{\infty} \left(1 + \frac{d^2}{\theta}\right)^{-1} \leqslant \sum_{m=-\infty}^{\infty} \left(\left(1 + \frac{(m - \lambda)^2}{\theta}\right)^{-1} + \left(1 + \frac{(m - \lambda + h + 1)^2}{\theta}\right)^{-1}\right)$$

$$\leqslant 2(1 + \pi\sqrt{\theta}) \leqslant C_{29}\sqrt{\theta}. \tag{3.17}$$

Now (2.9) follows from (3.17) and (3.2). If $h = 0$ then we note that, for integer k, $P\{[k, k + \frac{1}{2}]\} = P\{k\}$ and similar relations hold for $\prod H_j$ and $G(a, b)$. Therefore, this case coincides with the case $h = \frac{1}{2}$, for which (3.2) is already proved.   □

Remark 3.2. From (3.3), taking $h = \frac{1}{2}$ or $h \to \infty$, we can obtain non-uniform estimates for $\Delta\{m\}$ and $\Delta(m)$. From these estimates the proof of (2.13) imediately follows.

4. Remarks on the accuracy of approximation

One question still unanswered is how bad the accuracy of approximation G can become if the assumptions of Section 2 are violated. Roughly speaking, we shall show in this section that in comparison with the normal and Poisson laws the signed Poisson approximation is always on the safe side. By this we mean that in the sense of order one cannot get much worse by using the signed Poisson approximation instead of the normal or Poisson laws even under very general assumptions. It was mentioned above that the total variation is meaningless for the Gaussian approximation of the lattice distributions. Therefore we shall consider the uniform distance.

Let X_j, $j = 1, 2, \ldots, n$, be independent random variables with

$$\mathbf{E}X_j = \mu_j, \qquad \mathbf{D}X_j = \sigma_j^2, \qquad \mathbf{E}|X_j - \mu_j|^3 = \beta_{3j} < \infty. \tag{4.1}$$

Set $Z_n = X_1 + X_2 + \ldots + X_n$, $F_j(x) = P(X_j < x)$, $\mu(Z_n) = \sum_1^n \mu_j$, $\sigma^2(Z_n) = \sum_1^n \sigma_j^2$, $\beta_3(Z_n) = \sum_1^n \beta_{3j}$.

Let Φ_{μ, σ^2} be the normal distribution with mean $\mu(Z_n)$ and variance $\sigma^2(Z_n)$. The well-known Berry–Esseen theorem states that

$$\left| \prod_{j=1}^n F_j - \Phi_{\mu, \sigma^2} \right| \leqslant C_{23} \frac{\beta_3(Z_n)}{\sigma^3(Z_n)}. \tag{4.2}$$

Theorem 4.1. *The following inequality holds:*

$$\left| \prod_{j=1}^n F_j - G\left(\frac{\mu(Z_n) + \sigma^2(Z_n)}{2}, \frac{\mu(Z_n) - \sigma^2(Z_n)}{2} \right) \right| \leqslant C_{24} \left(\frac{\beta_3(Z_n)}{\sigma^3(Z_n)} + \frac{\max(|\mu(Z_n)|, \sigma^2(Z_n))}{\sigma^3(Z_n)} \right). \tag{4.3}$$

The proof of the theorem is deferred until the end of the section.

Of course, there are many distributions that have the same rate of accuracy as the normal – see, for example, the so-called scaled Poisson distributions in Rachev and Rüschendorf (1990). The main advantage of the signed Poisson approximation is its latticeness, i.e., the fact that it is concentrated on integers. It seems reasonable to approximate lattice distributions by measures having the same support. Comparing (4.3) with (4.2), we see that there is an additional summand in (4.3). It appeared because we have not used any centring. For the lattice distributions we can get rid of it. Indeed, let us assume that all X_i are

concentrated on integers. *Integer centring* involves centring Z_n while retaining the lattice structure and ensuring that

$$|\mu_j| = \min_a |\mathbf{E}X_j - a|. \tag{4.4}$$

Here the minimum is taken over all integers a. From (4.4) we see that $|\mu_j| \leq \frac{1}{2}$. From the fact that all F_j are concentrated on integers and from (4.4) we obtain

$$\sum_{k=-\infty}^{\infty} |k| F_j\{k\} \leq \sum_{k=-\infty}^{\infty} k^2 F_j\{k\} \leq \sum_{k=-\infty}^{\infty} |k|^3 F_j\{k\},$$

$$|k|^3 \leq 8(|k - \mu_j|^3 + |\mu_j|^3) \leq 8(|k - \mu_j|^3 + 2^{-3}) \leq 16|k - \mu_j|^3. \tag{4.5}$$

Therefore

$$\max(|\mu(Z_n)|, \sigma^2(Z_n)) \leq \sum_{j=1}^{n} \mathbf{E}|X_j|^3 \leq 16\beta_3(Z_n). \tag{4.6}$$

Thus we obtain the following corollary.

Corollary 4.1. *Let X_j, $j = 1, \ldots, n$, be concentrated on integers and let the assumptions (4.1) and (4.4) hold. Then (4.3) holds with the estimate $C_{25}\beta_3(Z_n)/\sigma^3(Z_n)$, i.e. the rate of accuracy is as in (4.2).*

It should be noted that Corollary 4.1 is the generalization of the result of Kruopis (1986b) where identical distributed lattice variables were considered.

There is no such general result as (4.2) for the Poisson approximation. However, we can compare the signed Poisson approximation with the Poisson law. Let $\mu(Z_n) > 0$. Then under (4.1) we have

$$\left| \exp\{\mu(Z_n)(E_1 - E)\} - G\left(\frac{\mu(Z_n) + \sigma^2(Z_n)}{2}, \frac{\mu(Z_n) - \sigma^2(Z_n)}{2}\right) \right|$$

$$\leq C_{26}|\sigma^2(Z_n) - \mu(Z_n)|\min(1, 1/\min(\sigma^2(Z_n), \mu(Z_n))). \tag{4.7}$$

Note that in (4.7) we do not assume that X_j are lattice variables. From (4.7) we see that G is always close to the corresponding Poisson law whenever $\mu(Z_n)$ and $\sigma^2(Z_n)$ are close. But those are exactly the conditions that ensure applicability of the Poisson approximation. Thus we can expect that G will be as good as the Poisson approximation. If $\mu(Z_n) < 0$ then we should compare G with $\exp\{|\mu(Z_n)|(E_{-1} - E)\}$.

The proof of (4.7) is based on the following estimates. Let $\sigma^2(Z_n) - \mu(Z_n) > 0$. Then

$$|\exp\{\mu(Z_n)(e^{it} - 1)\} - \exp\{(\sigma^2(Z_n) + \mu(Z_n))(e^{it} - 1)/2 + (\sigma^2(Z_n) - \mu(Z_n))(e^{-it} - 1)/2\}|$$

$$\leq |\exp\{\mu(Z_n)(e^{it} - 1)\}| \left|1 - \exp\left\{-2(\sigma^2(Z_n) - \mu(Z_n))\sin^2\left(\frac{t}{2}\right)\right\}\right|$$

$$\leq \exp\left\{-2\mu(Z_n)\sin^2\left(\frac{t}{2}\right)\right\} C_{27}|\sigma^2(Z_n) - \mu(Z_n)|\sin^2\left(\frac{t}{2}\right). \tag{4.8}$$

Analogously, for $\sigma^2(Z_n) - \mu(Z_n) \leq 0$, we obtain that the left-hand side of (4.8) is less than or equal to $\exp\{-2\sigma^2(Z_n)\sin^2(t/2)\}C|\sigma^2(Z_n) - \mu(Z_n)|\sin^2(t/2)$. By the Tsaregradskii (1958) inequality (see also Kruopis 1986a), for any lattice measure V:

$$|V| \leq \frac{1}{4}\int_{-\pi}^{\pi} \frac{|\hat{V}(t)|}{|t|}\, dt.$$

Therefore, (4.7) follows from (4.8).

Proof of Theorem 4.1. For the sake of brevity we shall write G omitting its parameters. By the triangle inequality and (4.2) it is sufficient to estimate

$$J = \Phi_{\mu,\sigma^2} - G.$$

(It is impossible to use (4.2) twice because, in general, G is not a distribution.) By a variant of Esseen's smoothing inequality (see, for example, Petrov 1975, Chapter 5), we have

$$|J| \leq C_{28}\int_0^{\pi} \frac{|\hat{J}(t)|}{t}\, dt + C_{29}\sup_x \int_{|y| \leq C_5} |G(x + y) - G(x)|\, dy. \tag{4.9}$$

It is easy to check that, for $|t| \leq \pi$,

$$\left|\sin\left(\frac{t}{2}\right)\right| \geq \left|\frac{t}{\pi}\right|. \tag{4.10}$$

By the inversion formula (1.2), we obtain

$$\sup_x |G(x + y) - G(x)| \leq (|y| + 1)\sup_x |G(x)|$$

$$\leq (|y| + 1)\int_{-\pi}^{\pi} \exp\left\{-2\sigma^2(Z_n)\sin^2\frac{t}{2}\right\} dt \leq \frac{C_{30}}{\sigma(Z_n)}(|y| + 1). \tag{4.11}$$

Let

$$A(t) = it\mu(Z_n) - \frac{\sigma^2(Z_n)t^2}{2} - \frac{\mu(Z - n) + \sigma^2(Z_n)}{2}(e^{it} - 1) + \frac{\mu(Z_n) - \sigma^2(Z_n)}{2}(e^{-it} - 1).$$

It is easy to check that the real part of $A(t)$ is always non-positive and

$$|A(t)| \leq C_{31}(|\mu(Z_n)| + \sigma^2(Z_n))|t|^3. \tag{4.12}$$

Hence, from (4.11) and (4.12), we obtain

$$|\hat{J}(t)| = |\hat{G}(t)| \, |\exp\{A(t)\} - 1| = |\hat{G}(t)| \left| \int_0^1 \exp\{A(t)\tau\} A(t) \, d\tau \right|$$

$$\leqslant |\hat{G}(t)| \, |A(t)| \int_0^1 |\exp\{A(t)\tau\}| \, d\tau \leqslant |\hat{G}(t)| \, |A(t)|$$

$$\leqslant C_{32}(|\mu(Z_n)| + \sigma^2(Z_n)) |t|^3 \exp\{-2\sigma^2(Z_n)t^2/\pi\}. \tag{4.13}$$

Combining (4.11) and (4.13) with (4.9), we obtain the statement of the theorem. $\qquad\square$

5. Concluding remarks

The theory of signed Poisson approximations is far from complete. Especially serious problems arise in the case of signed *compound* approximations. However, for situations as considered in this paper the signed Poisson approximation is a vigorous competitor to the Poisson and normal laws. We have emphasized the rate of accuracy, choosing the form of the estimates that was most convenient for comparisons. As a consequence our absolute constants are quite large.

Simons and Johnson (1971) proved that for the binomial distribution convergence to Poisson law is stronger than convergence in total variation. Their result was extended to the generalized binomial, negative binomial and compound Poisson cases (and sometimes to more general spaces) by Chen (1974; 1975), Dasgupta (1992) and Wang (1991; 1993). Many further generalizations can be found in the recent paper by Barbour *et al.* (1995).

However, for our purposes the generalized binomial law will be sufficient. Let S_n be defined as in Section 2. Assume that

$$\sum_{j=1}^{n} p_j = \lambda \tag{5.1}$$

where λ does not depend on n, and

$$\max_{1\leqslant j\leqslant n} p_j \to 0, \qquad \text{as } n \to \infty. \tag{5.2}$$

Then

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} v(k) \left| P(S_n = k) - \frac{\lambda^k}{k!} e^{-\lambda} \right| = 0 \tag{5.3}$$

for all non-negative functions $v(k)$, with

$$\sum_{k=0}^{\infty} v(k) \frac{\lambda^k}{k!} e^{-\lambda} < \infty. \tag{5.4}$$

(see Chen 1974). One of the most striking examples is

$$\lim_{n \to \infty} \sum_{k=0}^{\infty} \mathrm{e}^{vk} \left| P(S_n = k) - \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \right| = 0, \tag{5.5}$$

valid for any constant $v > 0$.

We shall prove that if (5.2) holds then the signed Poisson approximation considered in this paper is so close to Poisson distribution that the analogue of (5.3) holds. Let a, b be defined as in Section 3, i.e.,

$$a = \lambda - \sum_{j=1}^{n} p_j^2 / 2, \qquad b = \sum_{j=1}^{n} p_j^2 / 2. \tag{5.6}$$

Proposition. *Let (5.1) and (5.2) be satisfied, $v(x)$ be a non-negative even function satisfying (5.4). Then*

$$\lim_{n \to \infty} \sum_{k=-\infty}^{\infty} v(k) |P(S_n = k) - G(a, b)\{k\}| = 0. \tag{5.7}$$

Proof. From (1.5), we obtain

$$G(a, b)\{k\} = \mathrm{e}^{-a+b} \sum_{m=0}^{\infty} \frac{a^{k+m}(-b)^m}{m!(m+k)!} \qquad \text{if } k \geq 0, \tag{5.8}$$

$$G(a, b)\{k\} = \mathrm{e}^{-a+b} \sum_{m=0}^{\infty} \frac{a^m(-b)^{m+|k|}}{m!(m+|k|)!} \qquad \text{if } k < 0. \tag{5.9}$$

From (5.6), we see that

$$\sum_{k=0}^{\infty} v(k) \frac{a^k}{k!} \leq \sum_{k=0}^{\infty} v(k) \frac{\lambda^k}{k!} < \infty, \tag{5.10}$$

$$\sum_{k=1}^{\infty} v(k) \frac{b^k}{k!} \leq \frac{1}{2} \max p_j \sum_{k=1}^{\infty} v(k) \frac{\lambda^k}{k!} \to 0, \qquad \text{as } n \to \infty. \tag{5.11}$$

Hence

$$\sum_{k=-\infty}^{-1} v(k) |P(S_n = k) - G(a, b)\{k\}| \leq \sum_{k=1}^{\infty} v(k) \mathrm{e}^{-a+b+ab} \frac{b^k}{k!} = o(1). \tag{5.12}$$

On the other hand, for $k \geq 0$,

$$\left| G(a, b)\{k\} - \mathrm{e}^{-a+b} \frac{a^k}{k!} \right| \leq \mathrm{e}^{-a+b} \sum_{m=1}^{\infty} \frac{a^{k+m}b^m}{m!(m+k)!} \leq \mathrm{e}^{-a+b} \frac{a^k}{k!}(\mathrm{e}^{ab} - 1), \tag{5.13}$$

$$\left| \mathrm{e}^{-a+b} \frac{a^k}{k!} - \mathrm{e}^{-\lambda} \frac{a^k}{k!} \right| = \mathrm{e}^{-\lambda} \frac{a^k}{k!}(\mathrm{e}^{2b} - 1). \tag{5.14}$$

Note that $a = \lambda - b/2 \to \lambda$, $b \to 0$ as $n \to \infty$. From (5.13) and (5.14), we obtain

$$\sum_{k=0}^{\infty} v(k) |P(S_n = k) - G(a, b)\{k\}| = \sum_{k=1}^{\infty} v(k) \frac{(\lambda^k - a^k)}{k!} \mathrm{e}^{-\lambda} + o(1). \qquad (5.15)$$

But the first summand on the right-hand side of (5.15) vanishes due to the dominant convergence theorem. From (5.12) and (5.15), we obtain (5.7). □

Example. Let $v > 0$ be any fixed constant and let (5.6) be satisfied. Then

$$\lim_{n \to \infty} \sum_{k=-\infty}^{\infty} \mathrm{e}^{vk} |P(S_n = k) - G(a, b)\{k\}| = 0.$$

Acknowledgement

References

Aldous, D.J. (1989) *Probability Approximations via the Poisson Clumping Heuristic.* New York: Springer-Verlag.

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

Barbour, A.D., Holst, L. and Janson, S. (1992) *Poisson Approximations.* Oxford: Oxford University Press.

Barbour, A.D., Chen, L.H.Y. and Choi, K.P. (1995) Poisson approximation for unbounded functions, I: Independent summands. *Statist. Sinica*, **5**, 749–766.

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

Borovkov, K.A. and Pfeifer, D. (1996) Pseudo-Poisson approximation for Markov chains. *Stochastic Process. Appl.*, **61**, 163–180.

Čekanavičius, V. (1993) Non-uniform theorems for discrete measures. *Lithuanian Math. J.*, **33**, 114–125.

Čekanavičius, V. (1996) On multivariate Le Cam theorem and compound Poisson measures. *Statist. Probab. Lett.*, **28**, 33–39.

Čekanavičius, V. (1997) Asymptotic expansions in the exponent: a compound Poisson approach. *Adv. Appl. Probab.*, **29**, 374–387.

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

Chen, L.H.Y. (1975) An approximation theorem for convolutions of probability measures. *Ann. Probab.*, **3**, 992–999.

Dasgupta, R. (1992) Nonuniform rates of convergence to the Poisson distribution. *Sankhyā Ser. A.*, **54**, 460–463.

De Pril, N. and Dhaene, J. (1992) Error bounds for compound Poisson approximations of the individual risk model. *ASTIN Bull.*, **22**, 135–148.

Dhaene, J. and De Pril, N. (1994) On a class of approximative computation methods in the individual risk model. *Insurance Math. Econom.*, **14**, 181–196.

Feller, W. (1968) *Introduction to Probability Theory and its Applications*, 3rd edition. New York: Wiley.

Harper, L.H. (1967) Stirling behavior is asymptotically normal. *Ann. Math. Statist.*, **38**, 410–414.

Hipp, C. (1986) Improved approximations for the aggregate claims distribution in the individual model. *ASTIN Bull.*, **16**, 89–100.

Kornya, P. (1983) Distribution of aggregate claims in the individual risk theory model. *Trans. Soc. Actuaries*, **35**, 823–858.

Kruopis, J. (1986a) Precision of approximations of the generalized binomial distribution by convolutions of Poisson measures. *Lithuanian Math. J.*, **26**, 37–49.

Kruopis, J. (1986b) Approximations for distributions of sums of lattice random variables I. *Lithuanian Math. J.*, **26**, 234–244.

Kuon, S., Reich, A. and Reimers, L. (1987) Panjer vs. Kornya vs. De Pril: a comparison from a practical point of view. *ASTIN Bull.*, **17**, 183–191.

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

Petrov, V.V. (1975) *Sums of Independent Random Variables*. Berlin: Springer-Verlag.

Presman, E.L. (1983) Approximation of binomial distributions by infinitely divisible ones. *Theory Probab. Appl.*, **28**, 393–403.

Quine, M.P. (1994) Probability approximations for divisible discrete distributions. *Austral. J. Statist.*, **36**, 339–349.

Rachev, S.T. and Rüschendorf, L. (1990) Approximation of sums by compound Poisson distributions with respect to stop-loss distances. *Adv. Appl. Probab.*, **22**, 350–374.

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

Šiaulys, J. and Čekanavičius, V. (1989) Approximation of distributions of integer-valued additive functions by discrete charges, II. *Lithuanian Math. J.*, **29**, 80–95.

Simons, G. and Johnson, N.L. (1971) On the convergence of binomial to Poisson distribution. *Ann. Math. Statist.*, **42**, 1735–1736.

Tsaregradskii, I.P. (1958) On uniform approximation of a binomial distribution by infinitely divisible laws. *Teor. Veroyatnost. i Primenen.* **3**, 470–474 (in Russian).

Vatutin, V.A. and Mikhailov, V.G. (1982) Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory Probab. Appl.*, **27**, 734–743.

Wang, Y.H. (1991) A compound Poisson convergence theorem. *Ann. Probab.*, **19**, 452–455.

Wang, Y.H. (1993) On the number of successes in independent trials. *Statist. Sinica*, **3**, 295–312.

Witte, H.J. (1990) A unification of some approaches to Poisson approximation. *J. Appl. Probab.*, **27**, 611–621.