Bernoulli **11**(6), 2005, 1115–1128

Approximation of sums of conditionally independent variables by the translated Poisson distribution

ADRIAN RÖLLIN

Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. E-mail: adrian.roellin@math.unizh.ch

It is shown that the sum of a Poisson and an independent approximately normally distributed integer-valued random variable can be well approximated in total variation by a translated Poisson distribution, and further that a mixed translated Poisson distribution is close to a mixed translated Poisson distribution with the same random shift but fixed variance. Using these two results, a general approach is then presented for the approximation of sums of integer-valued random variables, having some conditional independence structure, by a translated Poisson distribution. We illustrate the method by means of two examples. The proofs are mainly based on Stein's method for distributional approximation.

Keywords: translated Poisson distribution; Stein's method; total variation metric

1. Introduction

The Berry–Esseen theorem provides a uniform bound for the accuracy of the central limit theorem when approximating the probabilities of sets A of the form $(-\infty, a)$, $a \in \mathbb{R}$. If more complicated sets A are to be considered, some additional 'smoothness' condition is typically required. McDonald (1979) and Burgess and McDonald (1995) assumed a so-called 'Bernoulli part' to deduce a local limit theorem from a central limit theorem. Čekanavičius and Vaĭtkus (2001) used the smoothing property of a sum of independent Bernoulli random variables to approximate this sum with a translated Poisson distribution in total variation. Barbour and Čekanavičius (2002) incorporate a measure of the smoothness of the distribution of the individual independent integer-valued summands as a component of their estimate of the distance between the distribution of their sum and a translated Poisson distribution; see the discusion in the next section.

This paper combines ideas from the above papers to show that the distribution of many sums of dependent integer-valued random variables can be approximated in total variation by the translated Poisson distribution with the same order of accuracy as that of the Berry–Esseen theorem. Previous attempts are limited to simple examples (Barbour and Xia, 1999; Čekanavičius and Vaĭtkus, 2001). Analogous results hold also for local limit approximations.

Much in the spirit of McDonald (1979), we begin by considering the sum of an integer-valued random variable Φ, which is close in distribution to the normal, and an independent Poisson random variable, which acts as the smoothing component. We show that this sum can be well approximated in total variation by a translated Poisson distribution with the same first two moments (Theorem 1) and that a similar approximation follows for a local limit metric. The translated Poisson distribution, being concentrated on the integers, is a more natural approximation than the normal in the context of integer-valued variables, and the stronger results are a reflection of this.

We then show that Theorem 1 can be used as the basis of a rather general method which yields good results in a number of dependent settings; see Theorem 3. We illustrate the method with two examples. For the proofs, we use Stein's method for distributional approximation, introduced by Stein (1972), adapted to the Poisson setting; see Barbour *et al*. (1992).

1.1. Notation

We say that an integer-valued random variable Y has a *translated Poisson distribution* with parameters μ and σ^2 and write

$$\mathcal{L}(Y) = \mathrm{TP}(\mu, \sigma^2)$$

if $\mathcal{L}(Y - \mu + \sigma^2 + \gamma) = \mathrm{Po}(\sigma^2 + \gamma)$, where $\gamma = \langle \mu - \sigma^2 \rangle$ and $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that $\mathbb{E}Y = \mu$ and that $\sigma^2 \leqslant \operatorname{var} Y = \sigma^2 + \gamma \leqslant \sigma^2 + 1$. Note also that $\mathrm{Po}(\sigma^2) = \mathrm{TP}(\sigma^2, \sigma^2)$.

We say that an integer-valued random variable Y has an *F-mixed translated Poisson distribution* and write

$$\mathcal{L}(Y) = \mathrm{TP}[F]$$

if F is a probability measure on $\mathbb{R} \times \mathbb{R}^+$ and, for all $j \in \mathbb{Z}$,

$$\mathbb{P}[Y = j] = \int_{\mathbb{R}\times\mathbb{R}^+} \mathrm{TP}(x, y)\{j\} F(\mathrm{d}x, \mathrm{d}y).$$

Thus a mixed Poisson distribution $\mathrm{Po}[G]$ with mixing distribution G is $\mathrm{TP}[F]$, where F is concentrated on the diagonal and has marginals G.

In this paper, the measure F will often be generated by two random variables Φ and Λ on a common probability space, that is, $F := \mathcal{L}(\Phi, \Lambda)$. We treat Φ as the 'random shift' and Λ as the 'random variance' of Y. Note that, due to our definition of $\mathrm{TP}(\mu, \sigma^2)$, Φ need not be integer-valued.

Throughout the paper, we shall be concerned with two metrics for probability distributions on the integers, the total variation metric d_{TV} and the local limit metric d_{loc}, where for two probability distributions P and Q,

$$d_{\mathrm{TV}}(P, Q) := \sup_{A\subset\mathbb{Z}} |P(A) - Q(A)|, \qquad d_{\mathrm{loc}}(P, Q) := \sup_{k\in\mathbb{Z}} |P(\{k\}) - Q(\{k\})|.$$

Let further δ_x denote the unit mass at $x \in \mathbb{R}$ and $*$ the convolution of measures.

2. Main results

2.1. Poisson smoothing

In this paper, we assume the random translation Φ to be approximately Gaussian. In terms of Stein's method of distributional approximation, this is expressed as follows. Denote by $\|\cdot\|$ the essential supremum norm and define the function space

$$\mathcal{F} = \{f \in C^1(\mathbb{R}) | f' \text{ absolutely continuous, } \|f\| + \|f'\| + \|f''\| < \infty\}.$$

Then, we shall assume that, for some $\varepsilon \geqslant 0$,

$$|\mathbb{E}\{f'(\Phi_c) - \Phi_c f(\Phi_c)\}| \leqslant \varepsilon \|f''\|, \qquad \text{for all } f \in \mathcal{F}, \tag{2.1}$$

where $\Phi_c := (\Phi - \mu)/\tau$, and μ and τ^2 are the mean and variance of Φ.

Theorem 1. *Let Φ be a random variable with mean μ and variance τ^2 such that estimate* (2.1) *holds for some $\varepsilon \geqslant 0$. Then, for any $\lambda > 0$,*

$$d_{\mathrm{TV}}(\mathrm{TP}[\mathcal{L}(\Phi) \times \delta_\lambda], \mathrm{TP}(\mu, \tau^2 + \lambda)) \leqslant \frac{c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 3\sqrt{\lambda}}{(\tau^2 + \lambda)\sqrt{\lambda}}, \tag{2.2}$$

$$d_{\mathrm{loc}}(\mathrm{TP}|\mathcal{L}(\Phi) \times \delta_\lambda|, \mathrm{TP}(\mu, \tau^2 + \lambda)) \leqslant \frac{4c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 12\sqrt{\lambda}}{(\tau^2 + \lambda)\lambda}, \tag{2.3}$$

where $c_0 = 1 + \sqrt{2}$.

So, suppose that $(\Phi^{(n)})_{n\geqslant 1}$ is a sequence obeying a central limit theorem, in the sense that $\Phi_c^{(n)}$ converges to the standard normal and that the corresponding sequence $(\varepsilon_n)_{n\geqslant 1}$ from (2.1) tends to zero. Suppose also that $\tau_n^2 := \operatorname{var}\Phi^{(n)}$ and λ_n tend to infinity at the same rate as $n \to \infty$. Then the estimate (2.2) is of order $\mathrm{O}(\varepsilon_n + \lambda_n^{-1/2})$ and (2.3) is of order $\mathrm{O}(\lambda_n^{-1/2}\varepsilon_n + \lambda_n^{-1})$. In typical situations, say $\tau_n^2 \asymp n$ in a central limit theorem for sums of locally dependent variables, we recover the expected order $\mathrm{O}(n^{-1/2})$ for (2.2) and $\mathrm{O}(n^{-1})$ for (2.3) if $\lambda_n \asymp n$; compare these with the second example in the next section.

2.2. Translated Poisson approximation

Let W be an integer-valued random variable with mean μ and variance σ^2 and X a random element of a Polish space on the same probability space. Assume that we want to approximate $\mathcal{L}(W)$ by a translated Poisson distribution with parameters μ and σ^2. Put $\mu_X = \mathbb{E}(W|X)$, $\sigma_X^2 = \operatorname{var}(W|X)$ and $\lambda = \mathbb{E}(\sigma_X^2)$ and consider the following, simple application of the triangle inequality for a metric d:

$$
\begin{aligned}
d(\mathcal{L}(W),\, \mathrm{TP}(\mu,\, \sigma^2)) \leqslant\; & d(\mathcal{L}(W),\, \mathrm{TP}[\mathcal{L}(\mu_X,\, \sigma_X^2)]) \\
& + d(\mathrm{TP}[\mathcal{L}(\mu_X,\, \sigma_X^2)],\, \mathrm{TP}[\mathcal{L}(\mu_X) \times \delta_\lambda]) \qquad (2.4) \\
& + d(\mathrm{TP}[\mathcal{L}(\mu_X) \times \delta_\lambda],\, \mathrm{TP}(\mu,\, \sigma^2)),
\end{aligned}
$$

where in this paper we shall take either d_{TV} or d_{loc}.

The second term on the right in (2.4) can be bounded using Stein's method, as in the next theorem.

Theorem 2. *Let Φ be a real-valued random variable and let Λ be a non-negative random variable with expectation $\lambda > 0$ and variance ν^2. Then*

$$d_{\mathrm{TV}}(\mathrm{TP}[\mathcal{L}(\Phi,\, \Lambda)],\, \mathrm{TP}[\mathcal{L}(\Phi) \times \delta_\lambda]) \leqslant \frac{2+\nu}{\lambda} + \frac{1}{\lambda^2}, \qquad (2.5)$$

$$d_{\mathrm{loc}}(\mathrm{TP}[\mathcal{L}(\Phi,\, \Lambda)],\, \mathrm{TP}[\mathcal{L}(\Phi) \times \delta_\lambda]) \leqslant \frac{2\sqrt{2}(1+\nu) + \sqrt{\lambda}}{\lambda^{3/2}} + \frac{1+4\nu^2}{\lambda^2}. \qquad (2.6)$$

The bounds (2.2)–(2.3) and (2.5)–(2.6) will be used for large λ, and typically with τ^2 and ν^2 large as well. It is, however, interesting to note that they do not tend to 0 if τ^2 and ν^2 tend to 0, as might have been expected. The reason is that the distributions $\mathrm{TP}(\mu,\, \sigma^2)$, although indexed by two continuous parameters, all belong to the set $\delta_m * \mathrm{Po}(\lambda)$ for $(m,\, \lambda) \in \mathbb{Z} \times \mathbb{R}_+$. This is reflected by the fact that the distributions $\mathrm{TP}(\mu,\, \sigma^2)$ do not change continuously with respect to either μ or σ^2 when $\mu - \sigma^2 \in \mathbb{Z}$. For example, $\mathrm{TP}(2-\varepsilon,\, 1) = \mathrm{Po}(2-\varepsilon)$, but $\mathrm{TP}(2,\, 1) = \delta_1 * \mathrm{Po}(1)$. Because of this fundamental discontinuity, $\tau^2 \to 0$ and $\nu^2 \to 0$ cannot imply that the bounds (2.2)–(2.3) and (2.5)–(2.6) tend to zero.

Barbour *et al.* (1992, Theorem 1.C) gave a bound for the distance $d_{\mathrm{TV}}(\mathrm{Po}[\mathcal{L}(\Lambda)],\, \mathrm{Po}(\lambda))$ of order $\mathrm{O}(\nu^2/\lambda)$. However, $\mathcal{L}(\Lambda)$ influences both the mean and variance of the distribution $\mathrm{Po}[\mathcal{L}(\Lambda)]$, whereas in Theorem 2 it only mixes the variance of $\mathrm{TP}[\mathcal{L}(\Phi,\, \Lambda)]$, leading to qualitatively different bounds.

To bound the third term on the right in (2.4) we can apply Theorem 1, provided that μ_X satisfies inequality (2.1) for some small ε. Combining all the above facts, we have the following theorem.

Theorem 3. *Let W be an integer-valued random variable with expectation μ and variance σ^2 and let X be a random element of a Polish space on the same probability space. Define μ_X, σ_X^2 and λ as at the beginning of this section and let $\tau^2 = \mathrm{var}(\mu_X)$, $\nu^2 = \mathrm{var}(\sigma_X^2)$. Assume that there exists $\varepsilon \geqslant 0$ such that $(\mu_X - \mu)/\tau$ satisfies* (2.1). *Then*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \mathbb{E}D_{\mathrm{TV}}(X) + \frac{2+\nu}{\lambda} + \frac{1}{\lambda^2} + \frac{c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 3\sqrt{\lambda}}{\sigma^2\sqrt{\lambda}},$$

$$d_{\mathrm{loc}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \mathbb{E}D_{\mathrm{loc}}(X) + \frac{2\sqrt{2}(1+\nu) + \sqrt{\lambda}}{\lambda^{3/2}} + \frac{1+4\nu^2}{\lambda^2}$$
$$+ \frac{4c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 12\sqrt{\lambda}}{\sigma^2\lambda},$$

where $D_{\mathrm{TV}}(X) := d_{\mathrm{TV}}(\mathcal{L}(W|X), \mathrm{TP}(\mu_X, \sigma_X^2))$, $D_{\mathrm{loc}}(X) := d_{\mathrm{loc}}(\mathcal{L}(W|X), \mathrm{TP}(\mu_X, \sigma_X^2))$ *and* $c_0 = 1 + \sqrt{2}$.

Now we are already able to bound $\mathbb{E}D_{\mathrm{TV}}(X)$ and $\mathbb{E}D_{\mathrm{loc}}(X)$ if the conditional distribution $\mathcal{L}(W|X)$ can be represented as a sum of independent integer random variables, since, as in Barbour and Čekanavičius (2002) or Čekanavičius and Vaĭtkus (2001), we can then approximate $\mathcal{L}(W|X)$ by the corresponding translated Poisson distribution.

To see this in more detail, recall Theorem 3.1 of Barbour and Čekanavičius (2002). Let $\tilde{W} = \sum_{i=1}^n Z_i$ be a sum of independent integer-valued random variables, such that $\mathbb{E}Z_i = \mu_i$, $\operatorname{var} Z_i = \sigma_i^2$ and $\mathbb{E}|Z_i^3| < \infty$. Put

$$\tilde{W}_i := \tilde{W} - Z_i, \qquad d := \max_{1\leqslant i\leqslant n} d_{\mathrm{TV}}(\mathcal{L}(\tilde{W}_i), \mathcal{L}(\tilde{W}_i + 1)), \tag{2.7}$$

$$\psi_i := \sigma_i^2\mathbb{E}\{Z_i(Z_i-1)\} + |\mu_i - \sigma_i^2|\mathbb{E}\{(Z_i-1)(Z_i-2)\} + \mathbb{E}|Z_i(Z_i-1)(Z_i-2)|. \tag{2.8}$$

Then, with $\tilde{\mu} = \sum\mu_i$, $\tilde{\sigma}^2 = \sum\sigma_i^2$, and $\psi = \sum\psi_i$,

$$d_{\mathrm{TV}}(\mathcal{L}(\tilde{W}), \mathrm{TP}(\tilde{\mu}, \tilde{\sigma}^2)) \leqslant \frac{2+d\psi}{\tilde{\sigma}^2}. \tag{2.9}$$

The factor d may be expressed in terms of the smoothness of the individual Z_i. With

$$v_i := \min\{\tfrac{1}{2}, 1 - d_{\mathrm{TV}}(\mathcal{L}(Z_i), \mathcal{L}(Z_i+1))\} \tag{2.10}$$

we have the simpler bound

$$d \leqslant \left(\sum_{i=1}^n v_i - \max_{1\leqslant i\leqslant n} v_i\right)^{-1/2}. \tag{2.11}$$

For analogous bounds in the d_{loc} case, we need some further notation. Proceeding as Barbour and Čekanavičius (2002, Section 4), define

$$d' := \tfrac{1}{2}\max_{1\leqslant i\leqslant n} \|\mathcal{L}(W_i) * (\delta_1 - \delta_0)^{*2}\|. \tag{2.12}$$

Using (4.4) and (4.8), just slight adaptations to the proof of Theorem 3.1 in (Barbour and Čekanavičius 2002) are needed to show that

$$d_{\mathrm{loc}}(\mathcal{L}(\tilde{W}), \mathrm{TP}(\tilde{\mu}, \tilde{\sigma}^2)) \leqslant \frac{2+d'\psi}{\tilde{\sigma}^2} \tag{2.13}$$

From equation (4.9) in Barbour and Čekanavičius (2002) we obtain the bound

$$d' \leqslant 4\left(\sum_{i=1}^{n} v_i - 4\max_{1\leqslant i\leqslant n} v_i\right)^{-1}. \tag{2.14}$$

3. Applications

3.1. Random sum of independent and identically distributed random variables

Theorem 4. *Let N be a non-negative, integer-valued random variable with expectation $a > 8$ and variance b^2 such that (2.1) holds for $N_c := (N - a)/b$ and some $\varepsilon \geqslant 0$, and let $Z_1, Z_2, \ldots$ be independent and identically distributed integer-valued random variables with expectation r and variance s^2, independent also of N; put $W = \sum_{i=1}^{N} Z_i$. Let ψ_1 and v_1 be as in (2.8) and (2.10) for Z_1, and assume that $v_1 > 0$. Then, with $\mu = \mathbb{E}W = ar$ and $\sigma^2 = \mathrm{var}W = as^2 + b^2r^2$,*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \frac{1.5\psi_1}{s^2\sqrt{v_1(a-2)}} + \frac{5\varepsilon b^3r^3 + 5b^2r^2 + 2.5br}{(as^2 + b^2r^2)sa^{1/2}} + \frac{1 + 9as^2 + a\mathrm{bs}^4 + 4b^2s^4}{a^2s^4}$$

$$d_{\mathrm{loc}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \frac{8\psi_1}{s^2v_1(a-8)} + \frac{15 + 3bs^2}{a^{3/2}s^3} + \frac{1 + 5as^2 + 8b^2s^4}{a^2s^4} + \frac{10(2\varepsilon b^3r^2 + 2b^2r^2 + br)}{(as^2 + b^2r^2)as^2}.$$

A random variable W of the form considered in this example arises in the study of the Reed–Frost epidemic process treated by Barbour and Utev (2004). In their Theorem 3.1, a local limit theorem is proved using Fourier arguments under the assumption that the Laplace transform of N is close to that of the normal distribution. Our result is formulated in very much simpler terms, and in addition gives an explicit approximation error. If we assume that $a \asymp n$ and $b^2 \asymp n$ and that $\varepsilon = \mathrm{O}(n^{-1/2})$, the total variation bound above is of order $\mathrm{O}(n^{-1/2})$.

Barbour and Utev (2004, Theorem 3.2) also prove a stronger local limit approximation, but at the cost of very much more restrictive conditions than ours.

Proof. We apply Theorem 3. In accordance with the notation of the previous section, let

$$\mu_N := \mathbb{E}(W|N) = Nr, \qquad \tau^2 := \mathrm{var}(\mu_N) = b^2r^2;$$

$$\sigma_N^2 := \mathrm{var}(W|N) = Ns^2, \qquad \lambda := \mathbb{E}(\sigma_N^2) = as^2, \qquad \nu^2 := \mathrm{var}(\sigma_N^2) = b^2s^4.$$

Then, given $N = k$, we can apply Theorem 3.1 from Barbour and Čekanavičius (2002) to W in order to bound $D_{\mathrm{TV}}(k)$ and $D_{\mathrm{loc}}(k)$. To this end, define $d(k)$ as in (2.7) and $d'(k)$ as in (2.12) with $n = k$. From (2.11), we obtain the estimate $d(k) \leqslant (kv_1 - v_1)^{-1/2}$, and from (2.14) $d'(k) \leqslant 4(kv_1 - 4v_1)^{-1}$ and hence, applying (2.9) and (2.13),

$$D_{\mathrm{TV}}(k) \leqslant \begin{cases} \dfrac{4}{as^2} + \dfrac{\psi_1\sqrt{2}}{s^2\sqrt{v_1(a-2)}}, & \text{if } k \geqslant a/2, \\ 1, & \text{if } k < a/2, \end{cases}$$

$$D_{\mathrm{loc}}(k) \leqslant \begin{cases} \dfrac{4}{as^2} + \dfrac{8\psi_1}{s^2 v_1(a-8)}, & \text{if } k \geqslant a/2, \\ 1, & \text{if } k < a/2. \end{cases}$$

Using Chebyshev's inequality to bound $\mathbb{P}[N < a/2]$, we therefore obtain

$$\mathbb{E}D_{\mathrm{TV}}(N) \leqslant \frac{4b^2}{a^2} + \frac{4}{as^2} + \frac{\psi_1\sqrt{2}}{s^2\sqrt{v_1(a-2)}}, \qquad \mathbb{E}D_{\mathrm{loc}}(N) \leqslant \frac{4b^2}{a^2} + \frac{4}{as^2} + \frac{8\psi_1}{s^2 v_1(a-8)}.$$

The remaining elements in Theorem 3 are immediate; we use $\sqrt{2} \leqslant 1.5$ and hence $c_0 \leqslant 2.5$. □

3.2. k-runs

Theorem 5. *Let $\xi_0, \ldots, \xi_{n-1}$ be independent and identically distributed random variables with $\mathbb{P}[\xi_0 = 1] = 1 - \mathbb{P}[\xi_0 = 0] = p$ for some $p \in (0, 1)$, where $n = m(2k-1)$ for some integers $k, m \geqslant 2$. To avoid edge effects, put $\xi_{n+i} := \xi_i$ for $i = 0, \ldots, 2k-2$. Define $U_j := \prod_{i=j}^{j+k-1} \xi_i$ and put $W = \sum_{j=0}^{n-1} U_j$. Then, with $\mu = \mathbb{E}W = np^k$ and $\sigma^2 = \operatorname{var} W = np^k\{1 + p - p^k(2 + (2k-1)(1-p))\}/(1-p)$,*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \frac{K_1}{\sqrt{n}}, \qquad d_{\mathrm{loc}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \frac{K_2}{n}$$

for some constants $K_i = K_i(k, p)$, $i = 1, 2$, which are independent of n.

The formulas for $K_i(k, p)$ that we establish here are rather crude and complicated but explicit. For $k = 2$, a bound of the same order was given by Barbour and Xia (1999), but their method of proof was extremely involved. Here, we can apply Theorem 3 and (2.9) directly, to obtain a result for arbitrary k. Some numerical comparisons with the bound of Barbour and Xia (1999) for $k = 2$ are given in Table 1, deduced by a more careful examination of the error terms.

Proof. Once again we apply Theorem 3. Split the indices $N_n := \{0, \ldots, n-1\}$ into m blocks $J_b = J_b^1 \cup J_b^2$, $b \in N_m$, of size $s := 2k-1$ with $J_b^1 = \{bs, \ldots, bs+k-2\}$ and $J_b^2 = \{bs+k-1, \ldots, (b+1)s-1\}$, and set

$$X = \left\{ \xi_i : i \in \bigcup_{b \in N_m} J_b^1 \right\}.$$

Let L_b (R_b) be the number of consecutive 1s of the ξ_i at the beginning (end) of block J_b^1.

Table 1. Numerical comparison for the 2-runs example: total variation distance estimate using the method in (a) this paper and (b) Barbour and Xia (1999). Missing values are due to parameter restrictions

(a)

	p				
n	0.1	0.25	0.5	0.75	0.9
10^6	0.4463	0.2334	0.1747	0.5528	>1
10^8	0.0445	0.0233	0.0175	0.0553	0.2554
10^{10}	0.0045	0.0023	0.0017	0.0055	0.0255

(b)

	p				
n	0.1	0.25	0.5	0.75	0.9
10^6	0.0304	–	0.1251	0.6014	–
10^8	0.0030	–	0.0125	0.0601	–
10^{10}	0.0003	–	0.0013	0.0060	–

With $W = \sum_{b\in N_m} W_b$, where $W_b = \sum_{j\in J_b} U_j$, the W_b are conditionally independent given X. Note that $\mathcal{L}(W_b|X) = \mathcal{L}(W_1|R_1, L_2)$. We have

$$\mathbb{E}\{W_1|R_1 = r, L_2 = l\} = \frac{p^{k-r} + p^{k-l} - 2p^k}{1-p} + p^k$$

and so

$$\mu_X = mp^k + \sum_{b\in N_m} \frac{p^{k-R_b} + p^{k-L_{b+1}} - 2p^k}{1-p} = mp^k + \sum_{b\in N_m} V_b,$$

where the $V_b := (p^{k-R_b} + p^{k-L_b} - 2p^k)/(1-p)$ are independent and identically distributed with $\mathbb{E}V_1 = 2(k-1)p^k$. Some simple calculations give

$$\mathbb{E}\{p^{-2L_1}\} = \frac{1+p-p^k}{p^{k-1}}, \qquad \mathbb{E}\{p^{-L_1}\} = k-(k-1)p,$$

$$\mathbb{E}\{p^{-L_1-R_1}\} = p^{-k+1} + (k-1)(1-p) + \tfrac{1}{2}(k-1)(k-2)(1-p)^2,$$

hence

$$\tau_1^2(k, p) := \operatorname{var} V_1 = \frac{p^{k+1}}{(1-p)^2}(4 + 2p - (3k^2+k)p^{k-1} + (6k^2-4k-4)p^k - (3k^2-5k+2)p^{k+1})$$

and

$$\tau^2 := \operatorname{var} \mu_X = m\tau_1^2 \leqslant \frac{mp^{k+1}(4+2p)}{(1-p)^2}. \tag{3.1}$$

As $|V_b - \mathbb{E}V_b| \leqslant 2p/(1-p)$ almost surely, we have $\mathbb{E}|V_b - \mathbb{E}V_b|^3 \leqslant 2p\tau_1^2/(1-p)$. Now, an inequality of the form (2.1) is easily derived (see, for example, Reinert 1998, Theorem 2.1). For a sum of independent random variables $\sum Z_i$ with zero expectation and variances σ_i^2 such that $\sum \sigma_i^2 = 1$, inequality (2.1) holds with $\varepsilon = \sum(\sigma_i^3 + \frac{1}{2}\mathbb{E}|Z_i^3|)$, and we may therefore take

$$\varepsilon = \frac{1}{\sqrt{m}}\left\{1 + \frac{2p}{(1-p)\tau_1}\right\} =: \frac{1}{\sqrt{m}}\varepsilon_1(k, p). \tag{3.2}$$

For (2.9), we have the following rather crude bounds. First, note that

$$\mathbb{P}[W_b = 0 | R_b, L_{b+1}] \geqslant (1-p)^2, \qquad \mathbb{P}[W_b = 1 | R_b, L_{b+1}] \geqslant p^k(1-p)^2 \tag{3.3}$$

almost surely. Hence, from (3.3),

$$d_{\mathrm{TV}}(\mathcal{L}(W_b | R_b, L_{b+1}), \mathcal{L}(W_b + 1 | R_b, L_{b+1})) \leqslant 1 - p^k(1-p)^2$$

and, with (2.11) and (2.14),

$$d \leqslant p^{-k/2}(1-p)^{-1}(m-1)^{-1/2}, \qquad d' \leqslant 4p^{-k}(1-p)^{-2}(m-4)^{-1} \tag{3.4}$$

Furthermore, it follows from (3.3) that

$$\operatorname{var}(W_b | R_b, L_{b+1}) \geqslant p^k(1-p)^2, \tag{3.5}$$

and, noting that $0 \leqslant W_b \leqslant sI[W_b \geqslant 1]$,

$$\operatorname{var}(W_b | R_b, L_{b+1}) \leqslant \mathbb{E}(\mathrm{W}_b^2 | R_b, L_{b+1}) \leqslant s^2 P_b,$$

where $P_b := \mathbb{P}[W_b \geqslant 1 | R_b, L_{b+1}]$. Thus $\psi_b(R_b, L_{b+1}) \leqslant s^3 P_b^2(1+2s) + s^3 P_b$ and, since

$$\mathbb{E}P_b^2 \leqslant \mathbb{E}P_b = \mathbb{P}[W_b \geqslant 1] \leqslant \mathbb{E}W_b = sp^k, \tag{3.6}$$

it follows that

$$\mathbb{E}\psi_b(R_b, L_{b+1}) \leqslant 2p^k s^4(1+s). \tag{3.7}$$

Thus, from (2.9), (3.4), (3.5) and (3.7),

$$\mathbb{E}D_{\mathrm{TV}}(X) \leqslant \frac{2}{mp^k(1-p)^2} + \frac{4k(2k-1)^4}{(1-p)^3\sqrt{(m-1)p^k}}, \tag{3.8}$$

$$\mathbb{E}D_{\mathrm{loc}}(X) \leqslant \frac{2}{mp^k(1-p)^2} + \frac{16k(2k-1)^4}{p^k(1-p)^4(m-4)}. \tag{3.9}$$

To complete the bound in Theorem 3, we still need a lower bound for λ and an upper bound for ν^2, both of which are properties of the distribution of

$$\sigma_X^2 = \operatorname{var}\left(\sum_{b=0}^{m-1} W_b \middle| X\right) = \sum_{b=0}^{m-1} \operatorname{var}(W_b | R_b, L_{b+1}) =: \sum_{b=0}^{m-1} Y_b.$$

It is immediate from (3.3) that

$$\lambda = \mathbb{E}(\sigma_X^2) = m\mathbb{E}Y_1 \geqslant mp^k(1-p)^2, \tag{3.10}$$

and, since the Y_b are 1-dependent,

$$\nu^2 = \operatorname{var}(\sigma_X^2) = \sum_{j=0}^{m-1} \operatorname{var} Y_b + 2\sum_{j=0}^{m-1} \operatorname{cov}(Y_b, Y_{b+1}) \leqslant 3m \operatorname{var} Y_1.$$

Now $\operatorname{var} Y_1 \leqslant \mathbb{E}Y_1^2$ and

$$Y_1 = \operatorname{var}(W_1 | R_1, L_2) \leqslant \mathbb{E}(W_1^2 | R_1, L_2) \leqslant s^2 P_1$$

almost surely, so that, with (3.6), $\operatorname{var} Y_1 \leqslant s^4\mathbb{E}(P_1^2) \leqslant p^k s^5$; hence

$$\nu^2 \leqslant 3mp^k(2k-1)^5. \tag{3.11}$$

Combining (3.1), (3.2), (3.8), (3.9), (3.10) and (3.11) with the bounds in Theorem 3, it follows that d_{TV} is of order $\mathrm{O}(m^{-1/2})$ and d_{loc} of order $\mathrm{O}(m^{-1})$, and recalling that $m = n/(2k-1)$ completes the proof. □

4. Proofs

4.1. Stein approach for the translated Poisson distribution

To use Stein's method for approximation in the d_{TV} and d_{loc} metrics we start with the Poisson case; for details, see Barbour *et al*. (1992).

Let W be an integer-valued random variable with expectation μ and variance $\sigma^2 > 0$, and let $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \langle \mu - \sigma^2 \rangle$, where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that, if $Y \sim \mathrm{TP}(\mu, \sigma^2)$, $Y - s \sim \mathrm{Po}(\sigma^2 + \gamma)$. Let $\mathcal{A}g(j) = (\sigma^2+\gamma)g(j+1) - jg(j)$ be the usual Stein operator for the Poisson distribution with mean $\sigma^2+\gamma$, and for $A \subset \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$ let $g_A : \mathbb{Z} \to \mathbb{R}$ be the (bounded) solution of

(i) $g(j) = 0$ for all $j \leqslant 0$,
(ii) $\mathcal{A}g(j) = I[j \in A] - \mathrm{Po}(\sigma^2+\gamma)\{A\}$ for all $j > 0$.

We can thus bound the total variation distance with

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) &= d_{\mathrm{TV}}(\mathcal{L}(W-s), \mathrm{Po}(\sigma^2+\gamma)) \\ &= \sup_{B\subset\mathbb{Z}} |\mathbb{E}I[W-s \in B] - \mathrm{Po}(\sigma^2+\gamma)\{B\}| \\ &\leqslant \sup_{A\subset\mathbb{Z}_+} |\mathbb{E}\mathcal{A}g_A(W-s)| + \mathbb{P}[W-s<0], \end{aligned} \tag{4.1}$$

and analogously

$$d_{\mathrm{loc}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)) \leqslant \sup_{k\in\mathbb{Z}_+} |\mathbb{E}\mathcal{A}g_{\{k\}}(W-s)| + \mathbb{P}[W-s<0]. \tag{4.2}$$

The last terms in (4.1) and (4.2) are usually bounded using Chebyshev's inequality.

From Barbour *et al*. (1992) we obtain the well-known bounds on the supremum norm of g_A,

$$\|g_A\| \leqslant (\sigma^2+\gamma)^{-1/2} \leqslant \sigma^{-1}, \qquad \|\Delta g_A\| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}, \tag{4.3}$$

where $\Delta g_A(j) := g_A(j+1) - g_A(j)$. If $A = \{k\}$ for some $k \in \mathbb{Z}$, we have the better estimate

$$\|g_{\{k\}}\| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}. \tag{4.4}$$

With $\tilde{g}_A(j) := g_A(j-s)$ we can rewrite the Stein operator, obtaining

$$\begin{aligned}\mathcal{A}g_A(W-s) &= (\sigma^2+\gamma)g_A(W-s+1) - (W-s)g_A(W-s)\\ &= \sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W) + \gamma\Delta\tilde{g}_A(W).\end{aligned} \tag{4.5}$$

The bounds on $\tilde{g}_A$ are of course the same as on g_A in (4.3) and (4.4). Thus, the last term is easily bounded by

$$|\mathbb{E}\{\gamma\Delta\tilde{g}_A(W)\}| \leqslant \gamma\sigma^{-2} \leqslant \sigma^{-2}. \tag{4.6}$$

To obtain better estimates than in Poisson approximation, we proceed as Barbour and Čekanavičius (2002). To this end, let U and V be independent integer-valued random variables. Then it is easy to see that, for any bounded function F,

$$|\mathbb{E}\Delta F(U)| \leqslant 2\|F\| d_{\mathrm{TV}}(\mathcal{L}(U), \mathcal{L}(U+1)), \tag{4.7}$$

$$|\mathbb{E}\Delta^2 F(U+V)| \leqslant 4\|F\| d_{\mathrm{TV}}(\mathcal{L}(U), \mathcal{L}(U+1)) d_{\mathrm{TV}}(\mathcal{L}(V), \mathcal{L}(V+1)). \tag{4.8}$$

4.2. Proofs of the theorems

Lemma 1. *Let* Φ *be a random variable with* $\mathbb{E}\Phi = \mu$ *and* $\operatorname{var}\Phi = \tau^2$*, such that* $\Phi_c = (\Phi-\mu)/\tau$ *satisfies* (2.1) *for some* $\varepsilon \geqslant 0$*. Then, for any random variable Z obeying* $\mathbb{E}(Z|\Phi) = 0$ *and* $\mathbb{E}(Z^2|\Phi) \leqslant 1$*,*

$$|\mathbb{E}\{\tau^2 f'(\Phi+Z) - (\Phi-\mu)f(\Phi+Z)\}| \leqslant (\varepsilon\tau^3 + \tau^2 + \tfrac{1}{2}\tau)\|f''\|, \qquad \textit{for all } f \in \mathcal{F}. \tag{4.9}$$

Proof. We write (2.1) in the form

$$|\mathbb{E}\{\tau^2 f'(\Phi) - (\Phi-\mu)f(\Phi)\}| \leqslant \varepsilon\tau^3\|f''\|, \qquad \text{for all } f \in \mathcal{F}. \tag{4.10}$$

By Taylor expansion of f around Φ we obtain

$$\mathbb{E}\{\tau^2 f'(\Phi+Z) - (\Phi-\mu)f(\Phi+Z)\}$$

$$= \mathbb{E}\left\{\tau^2\left[f'(\Phi) + Z\int_0^1 f''(\Phi+sZ)\mathrm{d}s\right] - (\Phi-\mu)\left[f(\Phi) + Zf'(\Phi) + Z^2\int_0^1(1-s)f''(\Phi+sZ)\mathrm{d}s\right]\right\}.$$

With $\mathbb{E}\{(\Phi-\mu)Zf'(\Phi)\} = 0$ the estimate is easily obtained. □

Proof of Theorem 1. First, we prove inequality (2.2). Let Z' be a random variable with

$\mathcal{L}(Z'|\Phi) = \mathrm{Po}(\gamma'_\Phi)$, where $\gamma'_\Phi = \langle \Phi - \lambda \rangle$, and let $Y \sim \mathrm{Po}(\lambda)$ be independent of (Φ, Z'). Set $Z = Z' - \gamma'_\Phi$ and $W = \Phi + Z + (Y - \lambda)$. Then, $W \sim \mathrm{TP}[\mathcal{L}(\Phi) \times \delta_\lambda]$, and, with $s = \lfloor \mu - \tau^2 - \lambda \rfloor$ and $\gamma = \langle \mu - \tau^2 - \lambda \rangle$, taking $\sigma^2 = \tau^2 + \lambda$ in (4.5), we have

$$\begin{aligned}\mathbb{E}\mathcal{A}g_A(W - s) &= \mathbb{E}\{(\tau^2 + \lambda)\Delta \tilde{g}_A(W) - (W - \mu)\tilde{g}_A(W) + \gamma \Delta \tilde{g}_A(W)\} \\ &= \mathbb{E}\{\tau^2 \Delta \tilde{g}_A(W) - (\Phi - \mu)\tilde{g}_A(W)\} + \mathbb{E}\{(\gamma - \gamma'_\Phi)\Delta \tilde{g}_A(W)\} \\ &= \mathbb{E}\{\tau^2 \Delta h_A(\Phi + Z - \lambda) - (\Phi - \mu) h_A(\Phi + Z - \lambda)\} \\ &\quad + \mathbb{E}\{(\gamma - \gamma'_\Phi)\Delta \tilde{g}_A(W)\}\end{aligned} \tag{4.11}$$

where for the second equality we use the fact that

$$\mathbb{E}\{Y g(Y)\} = \mathbb{E}\{\lambda g(Y + 1)\} \tag{4.12}$$

(see Barbour *et al*. 1992, p. 5) and for the third equality we put

$$h_A(j) := \mathbb{E}\{\tilde{g}_A(W) | \Phi + Z - \lambda = j\} = \mathbb{E}\{\tilde{g}_A(j + Y)\}$$

and use the independence of Y.

The second term in (4.11) is simply estimated with (4.3). To estimate the main term we use (4.9) for an appropriate interpolation function h_A.

Hence, we construct a function $f_A \in \mathcal{F}$, satisfying the conditions $f_A(j) = h_A(j)$ and $f'_A(j) = \Delta h_A(j)$ for all $j \in \mathbb{Z}$. For $j \in \mathbb{Z}$ and $x \in [0, 1)$, define the function

$$f_A(j + x) := h_A(j) + \Delta h_A(j) x + \Delta^2 h_A(j) \cdot \begin{cases} -c_0 x^2/2 & \text{if } x \leqslant c_0^{-1} 2^{-1/2}, \\ c_0(1 - x)(3 - 2\sqrt{2} - x)/2 & \text{if } x > c_0^{-1} 2^{-1/2}, \end{cases}$$

where $c_0 = 1 + \sqrt{2}$. Clearly, f satisfies the desired conditions, and we can then use calculus to show that

$$\| f''_A \| \leqslant c_0 \| \Delta^2 h_A \|.$$

The interpolation of h with the function f is optimal in the sense that the factor c_0 cannot be improved in the above inequality.

Using (4.7) for $F := \Delta \tilde{g}_A$ and invoking the bounds (4.3), we have

$$|\Delta^2 h_A(j)| = |\mathbb{E}\{\Delta^2 \tilde{g}_A(j + Y)\}| \leqslant 2 \|\Delta \tilde{g}_A\| d_{\mathrm{TV}}(\mathcal{L}(Y), \mathcal{L}(Y + 1)) \leqslant \frac{2}{(\tau^2 + \lambda)\sqrt{\lambda}}, \tag{4.13}$$

where we have used the fact that $d_{\mathrm{TV}}(\mathcal{L}(Y), \mathcal{L}(Y + 1)) \leqslant 1/\sqrt{\lambda}$, which can easily be proved with Stein's method for the Poisson case using (4.12).

Applying Lemma 1 to (4.11) with $f(x) = f_A(x - \lambda)$, we obtain the final bound

$$\begin{aligned}|\mathbb{E}\mathcal{A}g_A(W - s)| &\leqslant (\varepsilon\tau^3 + \tau^2 + \tfrac{1}{2}\tau)\|f''_A\| + \|\Delta \tilde{g}_A\| \\ &\leqslant \frac{c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau)}{(\tau^2 + \lambda)\sqrt{\lambda}} + \frac{1}{\tau^2 + \lambda}.\end{aligned}$$

As $\operatorname{var} W \leqslant \tau^2 + \lambda + 1$, it follows from Chebyshev's inequality that

$$\mathbb{P}[W - s < 0] \leqslant \left\{ \frac{\operatorname{var} W}{(\tau^2 + \lambda)^2} \wedge 1 \right\} \leqslant \left\{ \left(\frac{1}{\tau^2 + \lambda} + \frac{1}{(\tau^2 + \lambda)^2} \right) \wedge 1 \right\} \leqslant \frac{2}{\tau^2 + \lambda},$$

and hence, from (4.1), inequality (2.2) is proved.

For inequality (2.3), write $Y = Y_1 + Y_2$, where Y_1, Y_2 are independent, $\mathrm{Po}(\lambda/2)$ distributed random variables. Using (4.8) for $F := \tilde{g}_{\{k\}}$ and invoking (4.4), we replace the estimate (4.13) by

$$|\Delta^2 h_{\{k\}}(j)| \leqslant 4 \| \tilde{g}_{\{k\}} \| d_{\mathrm{TV}}(\mathcal{L}(Y_1), \mathcal{L}(Y_1 + 1))^2 \leqslant \frac{8}{(\tau^2 + \lambda)\lambda}. \qquad \square$$

Proof of Theorem 2. We first prove (2.5). Write $X = (\Phi, \Lambda)$. Given X fixed, let $Y \sim \mathrm{Po}(\Lambda)$ and $Z' \sim \mathrm{Po}(\gamma')$ be independent, where $\gamma' = \langle \Phi - \Lambda \rangle$, and set $W = \Phi + (Z' - \gamma') + (Y - \Lambda)$. Then $\mathcal{L}(W|X) = \mathrm{TP}(\Phi, \Lambda)$; we now use (4.1) with the conditional distribution $\mathbb{P}^X$ of W given X with $\mu = \Phi$ and $\sigma^2 = \lambda = \mathbb{E}\Lambda$ to obtain our estimate. From (4.5), with $s = \lfloor \Phi - \lambda \rfloor$ and $\gamma = \langle \Phi - \lambda \rangle$, it follows that

$$\begin{aligned}
\mathbb{E}^X \mathcal{A} g_A(W - s) &= \mathbb{E}^X \{ \lambda \Delta \tilde{g}_A(W) - (W - \Phi) \tilde{g}_A(W) \} + \gamma \mathbb{E}^X \Delta \tilde{g}_A(W) \\
&= \mathbb{E}^X \{ (\lambda - \Lambda) \Delta \tilde{g}_A(W) \} + \mathbb{E}^X \{ (\gamma - \gamma') \Delta \tilde{g}_A(W) \},
\end{aligned}$$

where we have used (4.12) for $Y + Z' \sim \mathrm{Po}(\Lambda + \gamma')$, and hence, using (4.3),

$$|\mathbb{E}^X \mathcal{A} g_A(W - s)| \leqslant \lambda^{-1}(|\lambda - \Lambda| + 1).$$

Moreover, by Chebyshev's inequality,

$$\mathbb{P}^X[W - s < 0] \leqslant \frac{\lambda + \gamma}{\lambda^2}. \tag{4.14}$$

Hence, we can bound (4.1) to give

$$d_{\mathrm{TV}}(\mathcal{L}(W|X), \mathrm{TP}(\Phi, \Lambda)) \leqslant \frac{|\lambda - \Lambda|}{\lambda} + \frac{1}{\lambda} + \frac{\lambda + \gamma}{\lambda^2}.$$

Taking expectation over X, the claim follows.

To prove inequality (2.6), use (4.7) for $F := \tilde{g}_{\{k\}}$ and the bound (4.4) to obtain

$$|\mathbb{E}^X \{ \mathcal{A} g_{\{k\}}(W - s) \}| \leqslant 2(|\lambda - \Lambda| + 1) \| \tilde{g}_{\{k\}} \| d_{\mathrm{TV}}(\mathcal{L}(Y), \mathcal{L}(Y + 1)) \leqslant \frac{2(|\lambda - \Lambda| + 1)}{\lambda \sqrt{\Lambda}}.$$

By Chebyshev's inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left\{ 1 \wedge \frac{2(|\lambda - \Lambda| + 1)}{\lambda \sqrt{\Lambda}} \right\} &\leqslant \mathbb{E}\left\{ I[\Lambda \leqslant \lambda/2] + I[\Lambda > \lambda/2] \frac{2\sqrt{2}(|\lambda - \Lambda| + 1)}{\lambda^{3/2}} \right\} \\
&\leqslant \frac{4\nu^2}{\lambda^2} + \frac{2\sqrt{2}(\nu + 1)}{\lambda^{3/2}}
\end{aligned}$$

and hence, with (4.14) and (4.1), the claim. $\square$

Acknowledgements

I thank Andrew Barbour and Małgorzata Roos for helpful discussions and the referees for their comments which improved the results and the presentation.

References

Barbour, A.D. and Čekanavičius, V. (2002) Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.*, **30**, 509–545.

Barbour, A.D. and Utev, S. (2004) Approximating the Reed–Frost epidemic process. *Stochastic Process. Appl.*, **113**, 173–197.

Barbour, A.D. and Xia, A. (1999) Poisson perturbations. *ESAIM Probab. Statist.*, **3**, 131–159.

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

Burgess, D. and McDonald, D. (1995) An elementary proof of the local central limit theorem. *J. Theoret. Probab.*, **8**, 693–701.

Čekanavičius, V. and Vaĭtkus, P. (2001) Centered Poisson approximation by the Stein method. *Lithuanian Math. J.*, **41**, 319–329.

McDonald, D. (1979) On local limit theorem for integer valued random variables. *Theory Probab. Appl.*, **24**, 613–619.

Reinert, G. (1998) Coupling for normal approximations with Stein's method. In D. Aldous and J. Propp (eds), *Microsurveys in Discrete Probability*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 41, pp. 193–207. Providence, RI: American Mathematical Society.

Stein, C. (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In L. LeCam, J. Neyman and E.L. Scott (eds), *Proceedings of the Sixth Berkeley Symposium on Mathematical statistics and Probability*, Vol. 2, pp. 583–602. Berkeley: University of California Press.

Received May 2004 and revised April 2005