Bernoulli **12**(6), 2006, 943–954

On Stein's factors for Poisson approximation in Wasserstein distance

A.D. BARBOUR[1] and AIHUA XIA[2]

[1]*Angewandte Mathematik, Universität Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland. E-mail: a.d.barbour@math.unizh.ch*
[2]*Department of Mathematics and Statistics, University of Melbourne, Parkville, Vic. 3010, Australia. E-mail: xia@ms.unimelb.edu.au*

We provide a probabilistic proof of various Stein's factors for Poisson approximation in terms of the Wasserstein distance.

Keywords: Poisson approximation; Stein factors; Stein's method; Wasserstein distance

1. Introduction and main results

Stein's method for Poisson approximation was adapted by Chen (1975) from Stein's (1972) method for normal approximation, and it has proved to be an extremely powerful tool for establishing Poisson approximations to sums of dependent integer-valued random variables (see Barbour *et al.* 1992: Chapter 1). The method is based on the observation that a random variable Y follows the Poisson distribution with mean λ, denoted as $\mathrm{Pn}(\lambda)$, if and only if $\mathrm{E}[\lambda g(Y+1) - Yg(Y)] = 0$ for all functions $g : \mathbb{N} \to \mathbb{R}$ such that $\mathrm{E}|Yg(Y)| < \infty$. This can be converted into approximation theorems with respect to any of a general class of distances $d_{\mathcal{F}}$ on probability measures on $\mathbb{Z}_+$, defined by

$$d_{\mathcal{F}}(P, Q) := \sup_{f \in \mathcal{F}} |P\{f\} - Q\{f\}|,$$

where $\mathcal{F}$ is any suitably rich set of test functions $f : \mathbb{Z}_+ \to \mathbb{R}$. To do so, take any $f \in \mathcal{F}$, and recursively solve for the function g_f which satisfies the equations

$$\lambda g_f(i+1) - i g_f(i) = f(i) - \mathrm{Pn}(\lambda)\{f\}, \qquad i \in \mathbb{Z}_+, \tag{1.1}$$

where $\mathrm{Pn}(\lambda)\{f\} := \mathrm{E}f(Y)$ with $Y \sim \mathrm{Pn}(\lambda)$. Then, for any random variable W on $\mathbb{Z}_+$, we have

$$\mathrm{E}f(W) - \mathrm{Pn}(\lambda)\{f\} = \mathrm{E}[\lambda g_f(W+1) - W g_f(W)], \tag{1.2}$$

so long as the expectations all exist. Let

$$M_l(g) := \sup_{w \in \mathbb{Z}_+} |\Delta^l g(w)|, \qquad l \in \mathbb{Z}_+,$$

where $\Delta g(w) := g(w+1) - g(w)$. If it can be shown that

$$|\mathrm{E}\{\lambda g(W+1) - Wg(W)\}| \leqslant \varepsilon_0 M_0(g) + \varepsilon_1 M_1(g) + \varepsilon_2 M_2(g),$$

1350–7265

for all functions g for which $M_l(g) < \infty$, $l = 0, 1, 2$, then it follows from (1.2) that

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathrm{Pn}(\lambda)) = \sup_{f\in\mathcal{F}} |\mathrm{E}f(W) - \mathrm{Pn}(\lambda)\{f\}|$$

$$\leqslant \varepsilon_0 \sup_{f\in\mathcal{F}} M_0(g_f) + \varepsilon_1 \sup_{f\in\mathcal{F}} M_1(g_f) + \varepsilon_2 \sup_{f\in\mathcal{F}} M_2(g_f).$$

The set of test functions considered in this note is the set of Lipschitz functions on $\mathbb{Z}_+$, that is,

$$\mathcal{F} = \mathcal{F}_W = \{f : \mathbb{Z}_+ \to \mathbb{R} : |f(x) - f(y)| \leqslant |x - y|\},$$

and the corresponding metric is the Wasserstein metric. The estimates of Stein's constants $\sup_{f\in\mathcal{F}} M_l(g_f)$ for the Wasserstein metric are summarized in the following theorem. Here, since (1.1) does not involve the value of $g_f(0)$, it is convenient to define it by $g_f(0) := g_f(1)$.

Theorem 1.1. *Defining* $g_f(0) = g_f(1)$, *with* $c_1 \wedge c_2 := \min\{c_1, c_2\}$, *we have*

$$\sup_{f\in\mathcal{F}_W} M_0(g_f) = 1, \tag{1.3}$$

$$\sup_{f\in\mathcal{F}_W} M_1(g_f) \leqslant 1 \wedge \frac{8}{3\sqrt{2\mathrm{e}\lambda}} \leqslant 1 \wedge \frac{1.1437}{\sqrt{\lambda}}, \tag{1.4}$$

$$\sup_{f\in\mathcal{F}_W} M_2(g_f) \leqslant \frac{4}{3} \wedge \frac{2}{\lambda}. \tag{1.5}$$

Remark 1.1. The bounds (1.3) and (1.4) are stated in Barbour *et al.* (1992: Remark 1.1.6), without detailed proof.

Remark 1.2. The bound of (1.5) is tight. From (2.24) below, we have

$$\sup_{f\in\mathcal{F}_W} \Delta^2 g_f(1) = \frac{2}{\lambda} - \frac{4(1 - \mathrm{e}^{-\lambda} - \mathrm{e}^{-\lambda}\lambda)}{\lambda^3} \sim \begin{cases} \dfrac{2}{\lambda}, & \text{as } \lambda \to \infty, \\ \dfrac{4}{3}, & \text{as } \lambda \to 0. \end{cases}$$

In Poisson approximation, the usual distance of choice is the total variation distance d_{TV}, for which $\mathcal{F} := \{\mathbf{1}_A, A \subset \mathbb{Z}_+\}$; the analogues of (1.3) and (1.4) for d_{TV}, given in Barbour *et al.* (1992: Remark 1.1.2), are

$$\sup_{f\in\mathcal{F}_W} M_0(g_f) \leqslant 1 \wedge \sqrt{\frac{2}{\mathrm{e}\lambda}}, \qquad \sup_{f\in\mathcal{F}_W} M_1(g_f) \leqslant 1 \wedge \frac{1}{\lambda}. \tag{1.6}$$

The Wasserstein distance between probability measures on $\mathbb{Z}_+$ takes into account not only the amounts by which their probabilities differ, as in the total variation distance, but also where the differences occur. In particular, when approximating by $\mathrm{Pn}(\lambda)$, differences in probabilities 'typically' occur at places separated by distances of order $\sqrt{\lambda}$, the standard deviation of $\mathrm{Pn}(\lambda)$, and are thus magnified by a factor of order $\sqrt{\lambda}$; hence d_W can be expected typically to be of order $\sqrt{\lambda}\, d_{\mathrm{TV}}$. This is reflected in the comparison between the Stein factors for d_W, given in (1.3) and (1.4), and those for d_{TV}, given in (1.6); the former are of larger order by a factor of $\sqrt{\lambda}$. Correspondingly, in the very simplest example of a sum of independent random variables $W = \sum_{i=1}^n X_i$, where $X_i \sim \mathrm{Be}(p_i)$, we have

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{Pn}(\lambda)) \leqslant \lambda^{-1}\lambda_2, \tag{1.7}$$

from Barbour *et al.* (1992: (1.23)), where $\lambda := \sum_{i=1}^n p_i$ and $\lambda_2 := \sum_{i=1}^n p_i^2$. Exactly the same Stein argument, but using (1.4) to bound $M_1(g)$, gives the bound

$$d_W(\mathcal{L}(W), \mathrm{Pn}(\lambda)) \leqslant 1.1437\lambda^{-1/2}\lambda_2, \tag{1.8}$$

larger by a factor of order $\sqrt{\lambda}$. Nonetheless, by considering the Lipschitz function $f(j) = |j - \lambda|$, it is easy to see that the bound in (1.8) is of the correct order in λ.

In order to get a d_W-bound of the same order as in (1.7), it is necessary to approximate not by $\mathrm{Pn}(\lambda)$, but by a distribution which matches the variance as well as the mean; for example, by

$$P' := \mathrm{Pn}(\lambda - b) * \delta_b,$$

where $*$ denotes convolution, δ_b the point mass at b, and $b = \lfloor \lambda_2 \rfloor$. If, in fact, λ_2 is an integer, then Stein's method, together with (1.5), can be used to derive the following neat bound.

Proposition 1.2. *Let $W = \sum_{i=1}^n X_i$, where the $X_i \sim \mathrm{Be}(p_i)$ are independent, and define $\lambda = \sum_{i=1}^n \mathrm{p}_i$, $\lambda_2 := \sum_{i=1}^n p_i^2$. If λ_2 is an integer, set $b = \lambda_2$, $a = \lambda - \lambda_2$; then we have*

$$d_W(\mathcal{L}(W), P') \leqslant 2a^{-1}\sum_{i=1}^n p_i^2(1 - p_i) + 2\mathrm{E}[(b - W)\mathbf{1}_{\{W \leqslant b-1\}}] \leqslant 4a^{-1}\lambda_2. \tag{1.9}$$

Of course, there are also bounds of better order than $\lambda^{-1}\lambda_2$ for d_{TV}-approximation of $\mathcal{L}(W)$ by distributions matching the first two moments; see, for example, Barbour and Hall (1984: Theorem 3), and Čekanavičius and Vaitkus (2001). However, it is important now to note that the arguments needed are more complicated, in part because there is no universal analogue of (1.5), with a bound of order $\lambda^{-3/2}$, for d_{TV}. In particular, it is for this reason not so easy to improve on the Poisson approximation bounds in d_{TV}, when W is a general sum of weakly dependent random variables, just by fitting the second moment as well, though various techniques have been used in particular contexts: see Barbour and Eagleson (1987) and Röllin (2005), for example. In contrast, for proving d_W-approximation to distributions which match both mean and variance, this problem does not arise, because of the *uniform* bound given in (1.5).

2. The proofs

Splitting $g_f(i) = h_f(i) - h_f(i-1)$, $i \geqslant 1$, we can reformulate (1.1) as

$$\lambda(h_f(i+1) - h_f(i)) + i(h_f(i-1) - h_f(i)) = f(i) - \mathrm{Pn}(\lambda)\{f\}, \tag{2.1}$$

where the left-hand side of (2.1) is the generator of the immigration–death process with constant immigration rate λ and unit per-capita death rate applied to h_f. If $\{Z_i(t),\, t \geqslant 0\}$ denotes the immigration–death process with this generator and initial value $Z_i(0) = i$, then the solution to Stein equation (2.1) can be written as

$$h_f(i) = -\int_0^\infty \{\mathrm{E}[f(Z_i(t))] - \mathrm{Pn}(\lambda)\{f\}\}\,\mathrm{d}t; \tag{2.2}$$

see Barbour (1988) or Brown and Xia (2001) for details.

***Proof of* (1.3).** Since g_f is the first difference of h_f, we have for $i \geqslant 1$ that

$$g_f(i) = h_f(i) - h_f(i-1) = -\int_0^\infty \mathrm{E}[f(Z_i(t)) - f(Z_{i-1}(t))]\,\mathrm{d}t. \tag{2.3}$$

Now let S be a negative exponential random variable with mean 1 and independent of $\{Z_{i-1}(t),\, t \geqslant 0\}$. Construct

$$Z_i(t) = Z_{i-1}(t) + \mathbf{1}_{\{S>t\}}. \tag{2.4}$$

We obtain from (2.3) that

$$g_f(i) = -\int_0^\infty \mathrm{e}^{-t}\mathrm{E}[f(Z_{i-1}(t)+1) - f(Z_{i-1}(t))]\,\mathrm{d}t = -\int_0^\infty \mathrm{e}^{-t}\mathrm{E}\Delta f(Z_{i-1}(t))\,\mathrm{d}t. \tag{2.5}$$

Hence, for $f \in \mathcal{F}_W$,

$$|g_f(i)| \leqslant \int_0^\infty \mathrm{e}^{-t}\,\mathrm{d}t = 1,$$

with equality when $f(k) = k$, $k \in \mathbb{Z}_+$. □

We shall mostly use the notation of Brown and Xia (2001) for the proof of the remaining results. For each $i \geqslant 1$, write

$$\tau_i^+ = \inf\{t : Z_i(t) = i+1\}, \qquad \tau_i^- = \inf\{t : Z_i(t) = i-1\},$$

$$\pi_i = \mathrm{Pn}(\lambda)\{i\}, \qquad e_i^+ = \mathrm{E}(\tau_i^+), \qquad e_i^- = \mathrm{E}(\tau_i^-),$$

and for convenience, set

$$\tau_0^- = \infty, \qquad e_0^- = \infty.$$

Applying Lemma 2.2 of Brown and Xia (2001) with immigration rate $\alpha_k = \lambda$ and death rate $\beta_k = k$ gives

$$e_j^+ = \frac{F(j)}{\lambda \pi_j}, \qquad e_j^- = \frac{\overline{F}(j)}{j\pi_j}, \tag{2.6}$$

where

$$F(j) = \sum_{k=0}^{j} \pi_k, \qquad \overline{F}(j) = \sum_{k=j}^{\infty} \pi_k. \tag{2.7}$$

Let g_j be the solution of the Stein equation (1.1) for the function $f = 1_{\{j\}}$. Then Brown and Xia (2001: (2.9) and (2.10)) state that, for $i \geqslant 1$ and $j \geqslant 0$,

$$g_j(i) = \begin{cases} -\pi_j e_{i-1}^+, & i \leqslant j, \\ \pi_j e_i^-, & i \geqslant j+1. \end{cases}$$

Since it is clear that

$$g_f(i) = \sum_{j=0}^{\infty} f(j) g_j(i),$$

we obtain

$$g_f(i) = -e_{i-1}^+ \sum_{j \geqslant i} \pi_j f(j) + e_i^- \sum_{j \leqslant i-1} \pi_j f(j), \qquad i \geqslant 1. \tag{2.8}$$

Proof of **(1.4).** By the definition of $g_f(0) = g_f(1)$, we have $\Delta g_f(0) = 0$, so it remains to consider $\Delta g_f(i)$ for $i \geqslant 1$. It follows from (2.8) that

$$\Delta g_f(i) = -(e_i^+ - e_{i-1}^+) \sum_{j \geqslant i+1} \pi_j f(j) + (e_{i+1}^- - e_i^-) \sum_{j \leqslant i-1} \pi_j f(j) + \pi_i f(i)(e_{i-1}^+ + e_{i+1}^-). \tag{2.9}$$

Replacing f by $f - f(i)$ if necessary, we may assume that $f(i) = 0$, so (2.9) becomes

$$\Delta g_f(i) = -(e_i^+ - e_{i-1}^+) \sum_{j \geqslant i+1} \pi_j f(j) + (e_{i+1}^- - e_i^-) \sum_{j \leqslant i-1} \pi_j f(j). \tag{2.10}$$

Lemma 2.4 of Brown and Xia (2001) shows that e_k^+ is increasing in k and e_k^- is decreasing in k, so we obtain from (2.10) that, for each $f \in \mathcal{F}_W$,

$$|\Delta g_f(i)| \leqslant (e_i^+ - e_{i-1}^+) \sum_{j \geqslant i+1} \pi_j (j - i) + (e_i^- - e_{i+1}^-) \sum_{j \leqslant i-1} \pi_j (i - j), \tag{2.11}$$

the maximum in (2.10) being achieved by $f_{i1}(j) = -|j - i|$, $j \in \mathbb{Z}_+$. This identifies the extremal function f for evaluating $\Delta g_f(i)$.

On the other hand, (2.5) implies that

$$\Delta g_f(i) = -\int_0^\infty \mathrm{e}^{-t} \mathrm{E}[\Delta f(Z_i(t)) - \Delta f(Z_{i-1}(t))] \, \mathrm{d}t. \tag{2.12}$$

We now use the coupling (2.4) to obtain from (2.12) that

$$\Delta g_f(i) = -\int_0^\infty \mathrm{e}^{-2t}\mathrm{E}[\Delta^2 f(Z_{i-1}(t))]\,\mathrm{d}t, \tag{2.13}$$

which implies that

$$\begin{aligned}\sup_{f\in\mathcal{F}_W}|\Delta g_f(i)| &= \Delta g_{f_{i1}}(i) = 2\int_0^\infty \mathrm{e}^{-2t}P(Z_{i-1}(t)=i-1)\,\mathrm{d}t \qquad (2.14)\\ &\leqslant 2\int_0^\infty \mathrm{e}^{-2t}\max_j P(Z_0(t)=j)\,\mathrm{d}t \leqslant 2\int_0^\infty \mathrm{e}^{-2t}\left(1\wedge\frac{1}{\sqrt{2\mathrm{e}\lambda(1-\mathrm{e}^{-t})}}\right)\mathrm{d}t\\ &= \begin{cases} 1, & \text{if } 2\mathrm{e}\lambda\leqslant 1,\\ \dfrac{8}{3\sqrt{2\mathrm{e}\lambda}}-\dfrac{1}{\mathrm{e}\lambda}+\dfrac{1}{12\mathrm{e}^2\lambda^2}\leqslant\dfrac{8}{3\sqrt{2\mathrm{e}\lambda}}, & \text{if } 2\mathrm{e}\lambda>1,\end{cases}\end{aligned}$$

where the second inequality comes from the fact that $Z_0(t)\sim \mathrm{Pn}(\lambda(1-\mathrm{e}^{-t}))$ and from Barbour *et al.* (1992: 262). □

To prove (1.5), we first need a technical lemma.

Lemma 2.1. *For* $i\geqslant 1$,

$$e^+_{i+1}-2e^+_i+e^+_{i-1}\geqslant 0, \tag{2.15}$$

$$e^-_{i+2}-2e^-_{i+1}+e^-_i\geqslant 0. \tag{2.16}$$

Proof. By (2.6) and (2.7), we have

$$\begin{aligned}e^+_{i+1}-2e^+_i+e^+_{i-1} &= \frac{F(i+1)}{\lambda\pi_{i+1}}-2\frac{F(i)}{\lambda\pi_i}+\frac{F(i-1)}{\lambda\pi_{i-1}}\\ &= \frac{1}{\lambda\pi_{i+1}}\sum_{j=0}^{i+1}\pi_j-\frac{2}{\lambda\pi_i}\sum_{j=1}^{i+1}\pi_{j-1}+\frac{1}{\lambda\pi_{i-1}}\sum_{j=2}^{i+1}\pi_{j-2}\\ &= \lambda^{-i-2}(i+1)!+\lambda^{-i-1}[(i+1)!-2(i!)]\\ &\quad+\sum_{j=2}^{i+1}\lambda^{-(i+2-j)}\left(\frac{(i+1)!}{j!}-2\frac{i!}{(j-1)!}+\frac{(i-1)!}{(j-2)!}\right).\end{aligned} \tag{2.17}$$

It is straightforward to check that all of these coefficients are non-negative for $i\geqslant 1$, and hence (2.15) follows.

Likewise, we obtain from (2.6) and (2.7) that

$$e^-_{i+2} - 2e^-_{i+1} + e^-_i = \frac{\overline{F}(i+2)}{(i+2)\pi_{i+2}} - 2\frac{\overline{F}(i+1)}{(i+1)\pi_{i+1}} + \frac{\overline{F}(i)}{i\pi_i}$$
$$= \sum_{j=i}^{\infty} \lambda^{j-i}\left(\frac{(i+1)!}{(j+2)!} - 2\frac{i!}{(j+1)!} + \frac{(i-1)!}{j!}\right),$$

and (2.16) follows from the fact that all of the coefficients are positive, for $i \geqslant 1$. □

***Proof of* (1.5).** Replacing f by $-f$ if necessary, it suffices to show that $\Delta^2 g_f(i) \leqslant 4/3 \wedge 2/\lambda$ for all $f \in \mathcal{F}_W$. For $i = 0$, it follows from (2.14) and because $g_f(0) = g_f(1)$ that

$$\Delta^2 g_f(0) = \Delta g_f(1) \leqslant \int_0^\infty 2\mathrm{e}^{-2t} P(Z_0(t) = 0)\,\mathrm{d}t$$
$$= \int_0^\infty 2\mathrm{e}^{-2t-\lambda(1-e^{-t})}\,\mathrm{d}t = \frac{2}{\lambda} - \frac{2}{\lambda^2}(1 - \mathrm{e}^{-\lambda}) \leqslant 1 \wedge \frac{2}{\lambda},$$

where, again, $Z_0(t) \sim \mathrm{Pn}(\lambda(1 - \mathrm{e}^{-t}))$, so that we may take $i \geqslant 1$ for the rest of the proof.

Using (2.9), we obtain

$$\Delta^2 g_f(i) = \Delta g_f(i+1) - \Delta g_f(i)$$
$$= -\sum_{j\geqslant i+2} \pi_j f(j)(e^+_{i+1} - 2e^+_i + e^+_{i-1}) + \sum_{j\leqslant i-1} \pi_j f(j)(e^-_{i+2} - 2e^-_{i+1} + e^-_i)$$
$$+ \pi_{i+1} f(i+1)(2e^+_i - e^+_{i-1} + e^-_{i+2}) + \pi_i f(i)(e^-_{i+2} - 2e^-_{i+1} - e^+_{i-1}). \qquad (2.18)$$

Replacing f by $\tilde{f} = f - f(i)$, we may assume that $f(i) = 0$, so it follows from (2.18) and Lemma 2.1 that

$$\Delta^2 g_f(i) = -(e^+_{i+1} - 2e^+_i + e^+_{i-1}) \sum_{j\geqslant i+2} (f(j) - f(i+1))\pi_j + (e^-_{i+2} - 2e^-_{i+1} + e^-_i) \sum_{j\leqslant i-1} \pi_j f(j)$$
$$+ f(i+1)\left[\pi_{i+1}(2e^+_i - e^+_{i-1} + e^-_{i+2}) - (e^+_{i+1} - 2e^+_i + e^+_{i-1}) \sum_{j\geqslant i+2} \pi_j\right]$$
$$\leqslant (e^+_{i+1} - 2e^+_i + e^+_{i-1}) \sum_{j\geqslant i+2} (j - (i+1))\pi_j + (e^-_{i+2} - 2e^-_{i+1} + e^-_i) \sum_{j\leqslant i-1} \pi_j (i-j)$$
$$+ f(i+1)\left[\pi_{i+1}(2e^+_i - e^+_{i-1} + e^-_{i+2}) - (e^+_{i+1} - 2e^+_i + e^+_{i-1}) \sum_{j\geqslant i+2} \pi_j\right], \qquad (2.19)$$

with equality if

$$f(j) = \begin{cases} f(i+1) + (i+1) - j, & \text{for } j \geqslant i+2, \\ i - j, & \text{for } j \leqslant i. \end{cases}$$

If

$$\pi_{i+1}(2e_i^+ - e_{i-1}^+ + e_{i+2}^-) - (e_{i+1}^+ - 2e_i^+ + e_{i-1}^+) \sum_{j \geqslant i+2} \pi_j < 0, \tag{2.20}$$

we may take $f(i+1) = -1$, in which case the corresponding f which achieves the maximum of $\Delta^2 g_f(i)$ is $f_{i2}(j) = i - j$, $j \in \mathbb{Z}_+$. We shall show that this is impossible. In fact, we can use (2.13) to deduce that

$$\Delta^2 g_f(i) = -\int_0^\infty \mathrm{e}^{-2t} \mathrm{E}[\Delta^2 f(Z_i(t)) - \Delta^2 f(Z_{i-1}(t))] \, \mathrm{d}t,$$

which, together with the coupling (2.4), ensures that

$$\Delta^2 g_f(i) = -\int_0^\infty \mathrm{e}^{-3t} \mathrm{E}[\Delta^3 f(Z_{i-1}(t))] \, \mathrm{d}t. \tag{2.21}$$

Hence, $\Delta^2 g_{f_{i2}}(i) = 0$, and consequently

$$\pi_{i+1}(2e_i^+ - e_{i-1}^+ + e_{i+2}^-) - (e_{i+1}^+ - 2e_i^+ + e_{i-1}^+) \sum_{j \geqslant i+2} \pi_j \geqslant 0,$$

contradicting to (2.20). Hence the function f which maximizes $\Delta^2 g_f(i)$ over $f \in \mathcal{F}_W$ has $f(i+1) = +1$, and is given by

$$f_{i3}(j) = \begin{cases} i - j, & \text{if } j \leqslant i, \\ i + 2 - j, & \text{if } j \geqslant i + 1. \end{cases}$$

Defining

$$f_{i4}(j) = f_{i3}(j) - f_{i2}(j) = \begin{cases} 0, & \text{for } j \leqslant i, \\ 2, & \text{for } j \geqslant i + 1, \end{cases}$$

it follows from (2.18) and because $\Delta^2 g_{f_{i2}}(i) = 0$ that

$$\begin{aligned} \sup_{f \in \mathcal{F}_W} \Delta^2 g_f(i) &= \Delta^2 g_{f_{i4}}(i) \\ &= -2(e_{i+1}^+ - 2e_i^+ + e_{i-1}^+) \sum_{j \geqslant i+2} \pi_j + 2\pi_{i+1}(2e_i^+ - e_{i-1}^+ + e_{i+2}^-) \\ &= -2(e_{i+1}^+ - 2e_i^+ + e_{i-1}^+) \sum_{j \geqslant i+1} \pi_j + 2\pi_{i+1}(e_{i+1}^+ + e_{i+2}^-) \qquad (2.22) \\ &\leqslant 2\pi_{i+1}(e_{i+1}^+ + e_{i+2}^-), \end{aligned}$$

where the inequality is due to (2.15); and then, by (2.6),

$$\pi_{i+1}(e_{i+1}^+ + e_{i+2}^-) = \pi_{i+1}\left(\frac{F(i+1)}{\lambda \pi_{i+1}} + \frac{\overline{F}(i+2)}{\lambda \pi_{i+1}}\right) = \frac{1}{\lambda}. \tag{2.23}$$

Finally, it follows from (2.21) that

$$\Delta^2 g_{f_{i4}}(i) = -\int_0^\infty e^{-3t}[2P(Z_{i-1}(t) = i-2) - 4P(Z_{i-1}(t) = i-1) + 2P(Z_{i-1}(t) = i)]\,dt$$
$$\leqslant 4\int_0^\infty e^{-3t}\,dt = \frac{4}{3},$$

completing the proof. □

In fact, it follows from (2.22) and (2.23) that

$$\sup_{f\in\mathcal{F}_W} \Delta^2 g_f(1) = \frac{2}{\lambda} - 2(e_2^+ - 2e_1^+ + e_0^+)\sum_{j\geqslant 2}\pi_j;$$

hence, since $e_2^+ - 2e_1^+ + e_0^+ = 2/\lambda^3$ from (2.17), we have

$$\sup_{f\in\mathcal{F}_W} \Delta^2 g_f(1) = \frac{2}{\lambda} - \frac{4(1 - e^{-\lambda} - e^{-\lambda}\lambda)}{\lambda^3}, \tag{2.24}$$

which is enough to show that the bound (1.5) is asymptotically sharp.

Proof of Proposition 1.4. It is straightforward to check that (1.9) is equivalent to

$$d_W(\mathcal{L}(W-b), \mathrm{Pn}(a)) \leqslant 2a^{-1}\sum_{i=1}^n p_i^2(1-p_i) + 2\mathrm{E}[(b-W)\mathbf{1}_{\{W\leqslant b-1\}}] \leqslant 4a^{-1}\lambda_2. \tag{2.25}$$

To see (2.25), first observe that

$$\mathrm{E}\{Wg(W)\} = \sum_{i=1}^n p_i \mathrm{E}g(W_i+1)$$

for all functions $g:\mathbb{Z}\to\mathbb{R}$ such that $\mathrm{E}[W|g(W)|] < \infty$, where $W_i := W - X_i$. Hence

$$\sum_{i=1}^n p_i\mathrm{E}g(W+1) - \mathrm{E}\{Wg(W)\} = \sum_{i=1}^n p_i^2 \mathrm{E}\Delta g(W_i+1),$$

and also

$$\mathrm{E}\{\Delta g(W_i+1) - \Delta g(W)\} = (1-p_i)\mathrm{E}\Delta^2 g(W_i);$$

thus we have

$$(\lambda-\lambda_2)\mathrm{E}g(W+1) - \mathrm{E}\{(W-\lambda_2)g(W)\} = \sum_{i=1}^n p_i^2(1-p_i)\mathrm{E}\Delta^2 g(W_i). \tag{2.26}$$

Now, for $f:\mathbb{Z}\to\mathbb{R}$, let $g_f:\mathbb{N}\to\mathbb{R}$ be the solution to the equation

$$ag_f(j+1) - jg_f(j) = f(j) - \mathrm{Pn}(a)\{f\}, \qquad j\geqslant 0,$$

and set $g_f(0) = g_f(1)$ and $g_f(j) = 0$ for $j<0$; then define $\tilde{g}_f(j) := g_f(j-b)$. Directly from (2.26), it follows that

$$
\begin{aligned}
a\mathrm{E}\tilde{g}_f(W+1) - \mathrm{E}\{(W-b)\tilde{g}_f(W)\} &= \sum_{i=1}^{n} p_i^2(1-p_i)\mathrm{E}\Delta^2 g_f(W_i - b) \\
&= \sum_{i=1}^{n} p_i^2(1-p_i)\mathrm{E}\big\{[\Delta^2 g_f(W_i - b)]\mathbf{1}_{\{W_i \geqslant b-2\}}\big\}.
\end{aligned}
$$

On the other hand, from the definition of $\tilde{g}$ and because $g_f(0) = g_f(1)$,

$$
\begin{aligned}
a\tilde{g}_f(j+1) - (j-b)\tilde{g}_f(j) &= ag_f(j-b+1) - (j-b)g_f(j-b) \\
&= \begin{cases} f(j-b) - \mathrm{Pn}(a)\{f\}, & \text{if } j \geqslant b, \\ ag_f(0) = ag_f(1), & \text{if } j = b-1, \\ 0, & \text{if } j \leqslant b-2, \end{cases}
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\mathrm{E}\{(f(W-b) - \mathrm{Pn}(a)\{f\})\mathbf{1}_{[b,\infty)}(W)\} \\
&\quad = \sum_{i=1}^{n} p_i^2(1-p_i)\mathrm{E}\big\{[\Delta^2 g_f(W_i - b)]\mathbf{1}_{\{W_i \geqslant b-2\}}\big\} - ag_f(1)P(W = b-1) \\
&\quad = \sum_{i=1}^{n} p_i^2(1-p_i)\mathrm{E}\big\{[\Delta^2 g_f(W_i - b)]\mathbf{1}_{\{W_i \geqslant b\}}\big\} \qquad (2.27) \\
&\qquad + g_f(1)\left\{\sum_{i=1}^{n} p_i^2(1-p_i)[P(W_i = b-2) - P(W_i = b-1)] - aP(W = b-1)\right\}.
\end{aligned}
$$

Now, arguing carefully, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} p_i^2(1-p_i)[P(W_i = b-2) - P(W_i = b-1)] - aP(W = b-1) \\
&\quad = \sum_{i=1}^{n} p_i^2[P(W_i = b-2) - P(W = b-1)] - aP(W = b-1),
\end{aligned}
$$

this last because

$$
P(W = j) = (1-p_i)P(W_i = j) + p_iP(W_i = j-1), \qquad j \geqslant -2; \qquad (2.28)
$$

hence we deduce that

$$\sum_{i=1}^{n} p_i^2(1-p_i)[P(W_i = b-2) - P(W_i = b-1)] - aP(W = b-1)$$

$$= \sum_{i=1}^{n} p_i^2 P(W_i = b-2) - \lambda P(W = b-1)$$

$$= \sum_{i=1}^{n} p_i^2[P(W_i \leqslant b-2) - P(W_i \leqslant b-3)] - \lambda P(W = b-1)$$

$$= \sum_{i=1}^{n} p_i[P(W_i \leqslant b-2) - P(W \leqslant b-2)] - \lambda P(W = b-1)$$

$$= \mathrm{E}[(W-\lambda)\mathbf{1}_{[0,b)}(W)], \tag{2.29}$$

where the penultimate equality again follows from (2.28).

Without real loss of generality, we may take $f(0) = 0$, so that then $g_f(1) = -\mathrm{Pn}(a)\{f\}/a$. Thus we have from (2.27), (2.29) and (1.5) that

$$|\mathrm{E}(f(W-b) - \mathrm{Pn}(a)\{f\})|$$

$$\leqslant \frac{2}{a}\sum_{i=1}^{n} p_i^2(1-p_i) + |\mathrm{E}[(f(W-b) - \mathrm{Pn}(a)\{f\})\mathbf{1}_{[0,b)}(W)] + g_f(1)\mathrm{E}[(W-\lambda)\mathbf{1}_{[0,b)}(W)]|$$

$$= \frac{2}{a}\sum_{i=1}^{n} p_i^2(1-p_i)$$

$$+ |\mathrm{E}\{[f(W-b) - a^{-1}\mathrm{Pn}(a)\{f\}(W-b) - a^{-1}\mathrm{Pn}(a)\{f\}(a+b-\lambda)]\mathbf{1}_{[0,b)}(W)\}|$$

$$\leqslant \frac{2}{a}\sum_{i=1}^{n} p_i^2(1-p_i) + 2\mathrm{E}[(b-W)\mathbf{1}_{[0,b)}(W)] \leqslant 2a^{-1}\lambda_2 + 2\lambda_2 P(W \leqslant b-1),$$

since $|\mathrm{Pn}(a)\{f\}| \leqslant a$, $|f(W-b)| \leqslant |W-b|$, $b = \lambda_2$ and $a + b = \lambda$. Finally, by Chebyshev's inequality,

$$P(W \leqslant b-1) \leqslant P(|W-\lambda| \geqslant \lambda + 1 - b) \leqslant \frac{\mathrm{E}\{(W-\lambda)^2\}}{(\lambda+1-b)^2} \leqslant \frac{1}{a},$$

completing the proof of (2.25). Here, the Chernoff lower bound could be used instead, normally resulting in a bound of much smaller order for this contribution. □

If λ_2 is not an integer, there is a correction due to the fact that we cannot take $b = \lambda_2$ and have the random variable $W - b$ on the integers. However, if δ is such that $b - \delta$ is an integer, then a can be replaced by $a + \delta$ and b by $b - \delta$ in (2.25), and the error bound then has to be increased by an amount $\delta\|\Delta \mathrm{g}\| \leqslant 1.1437\delta(\mathrm{a}+\delta)^{-1/2}$.

Acknowledgements

The research for this paper was partially supported by Schweizerischer Nationalfondsprojekt Nr. 20-107935/1 (ADB) and by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (ADB, AX), whose contributions are gratefully acknowledged. We are also grateful to the referees for a number of helpful suggestions.

References

Barbour, A.D. (1988) Stein's method and Poisson process convergence. *J. Appl. Probab.*, **25A**, 175–184.

Barbour, A.D. and Eagleson, G.K. (1987) An improved Poisson limit theorem for sums of dissociated random variables. *J. Appl. Probab.*, **24**, 586–599.

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 Approximation*. Oxford: Oxford University Press.

Brown, T.C. and Xia, A. (2001) Stein's method and birth–death processes. *Ann. Probab.*, **29**, 1373–1403.

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

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

Röllin, A. (2005) Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli*, **11**, 1115–1128.

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

Received November 2005 and revised February 2006