The Annals of Probability
2002, Vol. 30, No. 2, 509–545

TOTAL VARIATION ASYMPTOTICS FOR SUMS OF INDEPENDENT INTEGER RANDOM VARIABLES

BY A. D. BARBOUR[1] AND V. ČEKANAVIČIUS

Universität Zürich and Vilnius University

Let $W_n := \sum_{j=1}^n Z_j$ be a sum of independent integer-valued random variables. In this paper, we derive an asymptotic expansion for the probability $\mathbb{P}[W_n \in A]$ of an arbitrary subset $A \in \mathbb{Z}$. Our approximation improves upon the classical expansions by including an explicit, uniform error estimate, involving only easily computable properties of the distributions of the Z_j: an appropriate number of moments and the total variation distance $d_{\mathrm{TV}}(\mathcal{L}(Z_j), \mathcal{L}(Z_j + 1))$. The proofs are based on Stein's method for signed compound Poisson approximation.

1. Introduction. The asymptotic theory of sums of independent random variables has been extensively studied and is very well understood; see, for example, Petrov (1975, 1995). Suppose that $(Z_j, j \geq 1)$ is a sequence of independent random variables having $(r+1)$th moments, and set $W_n := s_n^{-1}(S_n - \mathbb{E}S_n)$, where $S_n := \sum_{j=1}^n Z_j$ and $s_n^2 = \operatorname{Var} S_n$. Then, under certain conditions on the characteristic functions of the Z_j, the central limit theorem can be refined to an Edgeworth expansion of the form

$$|\mathbb{P}[W_n \leq x] - \Pi_{n,r}(x)| = o\big(n^{-(r-1)/2}\big), \tag{1.1}$$

which holds uniformly in $x \in \mathbb{R}$. Here,

$$\Pi_{n,r}(x) := \Phi(x)\left\{1 + \sum_{l=1}^{r-1} Q_{ln}(x) n^{-l/2}\right\}, \tag{1.2}$$

the polynomials $Q_{ln}(x)$ have coefficients specified in terms of the moments of the Z_j, and Φ denotes the distribution function of the standard normal distribution [Petrov (1975), Theorem 7, page 175].

If the Z_j are integer random variables, the conditions on the characteristic functions are not satisfied, and the Edgeworth expansion of W_n is much more complicated. In the case of identically distributed summands, again under some extra conditions, it takes the form

$$\begin{aligned}\left|\mathbb{P}[W_n \leq x] - \Pi_{n,r}(x) + \sum_{l=1}^{r-1} (-1)^{\lfloor (l-1)/2) \rfloor} (\sigma^2 n)^{-l/2} S_l(\sigma x \sqrt{n}) \frac{d^l}{dx^l} \Pi_{n,r}(x)\right| \\ = o(n^{-(r-1)/2}),\end{aligned} \tag{1.3}$$

Received April 2000; revised June 2001.

[1]Supported in part by Schweizer Nationalfonds Grants 20-50686.97 and 20-61753.00.

AMS 2000 subject classifications. 60G50, 60F05, 62E20.

Key words and phrases. Compound Poisson, Stein's method, total variation distance, Kolmogorov's problem.

holding uniformly in $x \in \mathbb{R}$. Here,

$$S_{2j}(x) = 2\sum_{k=1}^{\infty}(2\pi k)^{-2j}\cos(2\pi kx), \qquad S_{2j+1}(x) = 2\sum_{k=1}^{\infty}(2\pi k)^{-2j-1}\sum_{i=1}^{n}(2\pi kx),$$

and the remaining notation is as before; see Ibragimov and Linnik (1971), Theorem 3.3.4. This expansion is not as useful as for random variables with absolutely continuous component, because of the difference between the supports of $\mathcal{L}(W_n)$ and Φ; as a consequence, the estimate (1.3) contains cumbersome summands $S(\cdot)$, and the approximating measure is inappropriate for use in conjunction with total variation distance. The situation is even worse if the integer-valued summands are not identically distributed, since then still more complicated formulas are required [Pipiras (1970)]; this expansion does not even seem to be widely known, let alone used.

Nonetheless, there are expansions for the individual point probabilities under similar moment conditions, but with different additional conditions, of the form

$$\left| s_n\mathbb{P}[S_n = j] - \Phi'(x_j)\left\{1 + \sum_{l=1}^{r-1} q_{ln}(x_j)n^{-l/2}\right\}\right| = o(n^{-r/2}), \tag{1.4}$$

now uniformly for $j \in \mathbb{Z}$, where $x_j = s_n^{-1}(j - \mathbb{E}S_n)$ [Petrov (1975), Theorem 12, page 204]. By adding these local estimates of the individual probabilities, one can, in principle, obtain an approximation in total variation, but it is rather unwieldy. A further drawback to both (1.3) and (1.4) is that the error terms are far from explicit, and neither expansion is applicable to triangular arrays.

In view of these problems, there has been much research into more adequate discrete approximations of sums of integer-valued random variables. One important area is that in which the summands are Bernoulli $\mathrm{Be}(p)$ random variables with small p, in which case Poisson approximation can be very useful; see Barbour, Holst and Janson (1992). In other circumstances, compound Poisson approximation may be better [Barbour, Chen and Loh (1992) and Barbour and Utev (1999)]. However, under the conditions typically used to show that the distribution of W_n is close to a normal limit, a discrete analogue of the normal law, with two moments to be fitted, would seem to be a more appropriate starting point. This was the motivation behind the signed compound Poisson (SCP) approach. SCP measures are obtained as a generalization of compound Poisson distributions by allowing negative parameters λ_l in the compound Poisson generating function (2.1), and gain in flexibility as a result; however, the measures are, in general, signed measures, rather than probability distributions.

The SCP approach has proved to be quite useful. In many cases, the corresponding approximation is sharper than both normal and Poisson approximations. For example, Presman (1983) proved that the accuracy of the SCP approximation to the binomial law $\mathrm{Bi}(n, p)$ with $p \le 1/2$ is of the order $O(\min(np^3,\ p/\sqrt{np}))$, which, in general, is better than the normal approximation [$O(1/\sqrt{np})$, and only for uniform distance] and the Poisson approximation [$O(\min(np^2, p))$]. SCP approximations have been explored in a number of papers, including Kruopis (1986), Hipp (1986), Borovkov and Pfeifer (1996), Čekanavičius (1997, 1998) and Čekanavičius and Mikalauskas (1999). These authors all employed the method of characteristic functions, and the error terms in their results rarely contained (small) explicit constants.

In Barbour and Xia (1999), Stein's method was adapted for proving SCP approximations with explicit error bounds. In particular, they derived an analogue of the Berry–Esseen theorem in total variation for sums of independent integer-valued random variables, in which the approximating distribution is a translate of a compound Poisson distribution, itself almost a Poisson distribution, and in which the error estimate is explicitly expressed in terms of the first three moments of the Z_j, together with the total variation distances $d_{\mathrm{TV}}(\mathcal{L}(Z_j), \mathcal{L}(Z_j+1))$. Here, we develop their approach to treat asymptotic expansions.

The main result is Theorem 5.1. Let $W = \sum_{i=1}^n Z_i$ be a sum of independent integer-valued random variables Z_i having finite $(r+1)$th moments, which has been (integrally) translated so that $|\mathbb{E}W - \operatorname{Var} W| \le 1$; in other words, prepared for a Poisson approximation, instead of being centered at the mean for a normal approximation. Under the rather mild conditions (5.2), an explicit bound is given for the accuracy in total variation of the approximation of $\mathcal{L}(W)$ by the signed measure

$$\tilde{\nu}_r\{j\} := \mathrm{Po}(\mathbb{E}W)\{j\}\left\{1 + \sum_{u=1}^{S}(-1)^u b_u C_u(j;\mu)\right\},$$

where $C_u(\cdot\,;\cdot)$ denotes the uth Charlier polynomial, $S = \max\{1, 3(r-1)\}$ and the b_u are defined in terms of the first $(r+1)$ moments of W; see (5.8). For instance, if the Z_i's are identically distributed with strongly aperiodic distribution and have finite $(r+1+\delta)$th moment for some $0 < \delta \le 1$, then the error bound is of order $O(n^{-(r-1+\delta)/2})$. However, there is no need to demand identical distributions; a similar order of error is also valid under rather simple uniformity conditions. The measure $\tilde{\nu}_r$, although a signed measure, is completely explicit, and is only a rather small perturbation of the Poisson distribution $\mathrm{Po}(\mathbb{E}W)$ when $\mathbb{E}W$ is large; note that, when comparing with more traditional asymptotics, $\mathbb{E}W$ plays the role of the variance $\operatorname{Var} W$, as is clear from the choice of location.

The proof of this approximation is far from routine. A major problem is that, although the class of SCP measures for which the solutions of the Stein equation have good properties is extended in Corollary 2.2 beyond that of Barbour and Xia (1999), it is still, in general, not large enough to include the SCP measures required in the proof of Theorem 5.1. Here, a novel technique is introduced, in which the solutions to the Stein equation for a different distribution, chosen from the "good" class, are used as surrogates. As in the Berry–Esseen theorem of

Barbour and Xia (1999), the error estimates are explicitly expressed in terms of the moments of the Z_j and the total variation distances $d_{\mathrm{TV}}(\mathcal{L}(Z_j), \mathcal{L}(Z_j+1))$, quantities which are relatively simple to work with. As a result, the expansions are also directly applicable to triangular arrays.

In the early 1950s, Kolmogorov formulated a question about the accuracy of the best possible infinitely divisible approximation to the sums of *arbitrarily* chosen independent and identically distributed random variables, not necessarily having finite moments. Kolmogorov, Prokhorov, Le Cam, Meshalkin, Arak and many others contributed to this problem, which became known as the first uniform Kolmogorov theorem. The search for a solution inspired the development of new methods, such as the Tsaregradskii inequality and the triangle function method; led to new results, such as the Kolmogorov–Rogozin inequality for concentration functions; and gave rise to new approximations, such as Presman's SCP approximation. Kolmogorov's problem in uniform distance was finally solved by Arak (1981), who proved that any sum of independent and identically distributed random variables can be approximated within the class $\mathcal{D}$ of all infinitely divisible laws with accuracy $Cn^{-2/3}$, C being an (implicit) absolute constant. For a comprehensive history of the problem, see Arak and Zaitsev (1988).

In general, as proved by Zaitsev (1991), Kolmogorov's problem is insoluble in total variation, without additional assumptions on the random variables. However, improving earlier results of Tsaregradskii (1958) and Meshalkin (1961), Presman (1983) solved Kolmogorov's problem in total variation for the binomial distribution. For triangular arrays, Kolmogorov's theorem in total variation has so far only been explored for Bernoulli variables. Here, as an application of short expansions ($r = 2$) in terms of compound Poisson *probability* measures, we obtain estimates in total variation in Kolmogorov's theorem for a large class of triangular arrays of integer-valued random variables (Z_{jn}, $1 \le j \le n$, $n \ge 1$); under certain uniformity conditions, expressed in terms of bounds on their second and fourth moments and on $d_{\mathrm{TV}}(\mathcal{L}(Z_{jn}), \mathcal{L}(Z_{jn}+1))$, the accuracy of approximation is at least of order $O(n^{-1})$ (Corollary 4.5).

The structure of the paper is as follows. In Section 2, we establish properties of the solution of the Stein equation for certain signed measures on the integers; these are basic to the subsequent argument. We then treat the simplest case, that of approximation by a centered Poisson distribution, in Section 3; for a large class of integer-valued random variables, the centered Poisson approximation already extends both the classical Poisson and the normal approximations. In Section 4, we move on to second-order expansions, concentrating on the case when the approximations are probability measures, as is relevant to Kolmogorov's problem. The main asymptotic expansion (Theorem 5.1) is proved in Section 5.

In expansions such as (1.3), $(r+1)$ moments are assumed to exist, the expansion has $(r-1)$ terms refining the limiting approximation and the error is not specified beyond being $o(n^{-(r-1)/2})$. We use the same number of moments to give an expansion with $(r-1)$ refining terms, together with an explicit error bound. If we assume the existence of the $(r+1+\delta)$th moment, for any $0<\delta\le 1$, this takes the form of an explicit "Liapounov"-style error bound of order $O(n^{-(r-1+\delta)/2})$; in general, the error bound is of "Lindeberg" style.

2. Solving the Stein equation. Given $\gamma \in \mathbb{Z}$, $t \in \mathbb{N}$ and

$$\boldsymbol{\lambda} := (\lambda_{-t}, \lambda_{-t+1}, \dots, \lambda_{-1}, \lambda_1, \lambda_2, \dots, \lambda_t) \in \mathbb{R}^{2t},$$

let $\pi := \pi_{(\boldsymbol{\lambda},\gamma)}$ denote the (possibly signed) measure with generating function

$$\hat{\pi}(z) := \sum_{j\in\mathbb{Z}} \pi(j) z^j = z^\gamma \exp\left\{ \sum_{1\le|l|\le t} \lambda_l (z^l - 1) \right\}, \tag{2.1}$$

for which therefore $\pi(\mathbb{Z}) = 1$. If all the λ_l are nonnegative, this is a compound Poisson distribution on $\mathbb{Z}$, with origin shifted by γ:

$$\pi_{(\boldsymbol{\lambda},\gamma)} = \mathcal{L}\left(\gamma + \sum_{1\le|l|\le t} l N_l \right),$$

where the $N_l \sim \mathrm{Po}(\lambda_l)$ are independent, but we allow the possibility of negative λ_l's. A corresponding Stein operator $\mathcal{A} := \mathcal{A}_{(\boldsymbol{\lambda},\gamma)}$ on functions $g\colon \mathbb{Z}\to\mathbb{R}$ is given by

$$(\mathcal{A}g)(j) := \sum_{1\le|l|\le t} l\lambda_l g(j+l) - (j-\gamma)g(j); \tag{2.2}$$

note that $\pi\{\mathcal{A}g\} = 0$ for all bounded g, as can be seen by differentiating (2.1) with respect to z and equating coefficients.

It is usual, when applying Stein's method, to try to solve the equation $\mathcal{A}g = f - \pi\{f\}$ for the test functions f appropriate, for instance, to the total variation norm. Here, using the perturbation technique of Barbour and Xia (1999), we demonstrate the existence of *approximate* solutions having good properties, under the assumptions

$$\lambda := \sum_{1\le|l|\le t} l\lambda_l > 0 \tag{2.3}$$

and

$$\theta := \lambda^{-1} \sum_{l\in L_t} l(l-1)|\lambda_l| < 1/2, \tag{2.4}$$

where $L_t := \{l \in \mathbb{Z};\ |l| \le t,\ l \ne 0, 1\}$.

THEOREM 2.1. *Under conditions* (2.3) *and* (2.4), *given any bounded* $f\colon\mathbb{Z}\to\mathbb{R}$, *there exists* $g\colon \mathbb{Z}\to\mathbb{R}$ *satisfying*

(1) $g(i) = 0, \qquad i \le 0;$

(2) $\left|(\mathcal{A}_{(\boldsymbol{\lambda},0)}g)(i) - \big(f(i) - \pi_{(\boldsymbol{\lambda},0)}\{f\}\big)\right| \le \dfrac{2}{1-2\theta}\,\eta(\boldsymbol{\lambda})\|f\|, \qquad i \ge 0;$

(3) $\|g\| \le \dfrac{2}{1-2\theta}\{1 \wedge \lambda^{-1/2}\}\|f\|, \qquad \|\Delta g\| \le \dfrac{2}{1-2\theta}\{1 \wedge \lambda^{-1}\}\|f\|,$

where $\eta := \eta(\boldsymbol{\lambda}) := \sum_{j<0} |\pi_{(\boldsymbol{\lambda},0)}\{j\}|$.

Here, for $h\colon \mathbb{Z} \to \mathbb{R}$, we define $\Delta h(j) := h(j+1) - h(j)$, $j \ge 0$, and we use $\|h\|$ to denote the supremum norm. For (signed) measures, $\|\cdot\|$ denotes the total variation norm.

PROOF OF THEOREM 2.1. Let $\mathcal{E}$ denote the set of all bounded functions $g\colon \mathbb{Z} \to \mathbb{R}$, and let $\mathcal{E}'$ denote the Banach space of all functions $\psi \in \mathcal{E}$ with $\mathrm{Po}(\lambda)\{\psi\} = 0$ and with $\psi(j) = 0$ for all $j < 0$, equipped with the norm $\|\cdot\|'$ defined by

$$\|\psi\|' := \tfrac{1}{2}\Big\{\sup_{j\ge 0}\psi(j) - \inf_{j\ge 0}\psi(j)\Big\}.$$

For any $f \in \mathcal{E}$, let Sf denote the solution $g_0 \in \mathcal{E}$ to the equations

$$\lambda g_0(j+1) - jg_0(j) = f(j) - \mathrm{Po}(\lambda)\{f\},\ j \ge 0, \qquad g_0(j) = 0,\ j \le 0. \tag{2.5}$$

Then it follows as in Barbour, Holst and Janson (1992), Lemma 1.1.1, that

$$\begin{aligned} \|Sf\| &\le 2\{1 \wedge \lambda^{-1/2}\}\|f\|' \le 2\{1 \wedge \lambda^{-1/2}\}\|f\|, \\ \|\Delta Sf\| &\le 2\{1 \wedge \lambda^{-1}\}\|f\|' \le 2\{1 \wedge \lambda^{-1}\}\|f\|. \end{aligned} \tag{2.6}$$

Now define $U\colon \mathcal{E} \to \mathcal{E}$ by

$$\begin{aligned} (Ug)(j) &:= (\mathcal{A}_{(\boldsymbol{\lambda},0)}g)(j) - \lambda g(j+1) + jg(j) \\ &= \sum_{l=2}^{t} l\lambda_l \sum_{k=1}^{l-1} \Delta g(j+k) + \sum_{i=1}^{t} i\lambda_{-i} \sum_{k=0}^{i} \Delta g(j-k), \qquad j \in \mathbb{Z}, \end{aligned} \tag{2.7}$$

and $P\colon \mathcal{E} \to \mathcal{E}'$ by

$$(Pg)(j) = g(j) - \mathrm{Po}(\lambda)\{g\},\ j \ge 0, \qquad (Pg)(j) = 0,\ j < 0, \tag{2.8}$$

and consider the operator $Q\colon \mathcal{E}' \to \mathcal{E}'$ defined by

$$Q = P\mathcal{A}_{(\boldsymbol{\lambda},0)}S = I + PUS. \tag{2.9}$$

Direct calculation shows that, for $g \in \mathcal{E}$,

$$\|Ug\| \le \|\Delta g\| \sum_{l \in L_t} l(l-1)|\lambda_l| \le \lambda\theta\|\Delta g\| \tag{2.10}$$

and, since $\|f\|' \le \|f\|$ for any $f \in \mathcal{E}$, it follows from (2.6) that

$$\|PUS\psi\|' = \|US\psi\|' \le \|US\psi\| \le \lambda\theta\|\Delta S\psi\| \le 2\theta\|\psi\|'$$

for all $\psi \in \mathcal{E}'$. Thus $\|PUS\|' \le 2\theta$, and so Q is invertible, with

$$\|Q^{-1}\|' = \|(I - PUS)^{-1}\|' \le (1-2\theta)^{-1}. \tag{2.11}$$

Now, for $f \in \mathcal{E}$, define

$$g_f = SQ^{-1}Pf. \tag{2.12}$$

Then it follows from (2.6) and (2.11) that

$$\begin{aligned} \|g_f\| &\le 2\{1 \wedge \lambda^{-1/2}\}\|Q^{-1}Pf\|' \le \{1 \wedge \lambda^{-1/2}\}\frac{2}{1-2\theta}\|Pf\|' \\ &= \{1 \wedge \lambda^{-1/2}\}\frac{2}{1-2\theta}\|f\|' \le \{1 \wedge \lambda^{-1/2}\}\frac{2}{1-2\theta}\|f\|; \end{aligned} \tag{2.13}$$

similarly,

$$\|\Delta g_f\| \le \{1 \wedge \lambda^{-1}\}\frac{2}{1-2\theta}\|f\|. \tag{2.14}$$

Also, from (2.9), for $j \ge 0$ and $\psi \in \mathcal{E}'$, we have

$$(\mathcal{A}_{(\lambda,0)}S\psi)(j) = (Q\psi)(j) + \mathrm{Po}(\lambda)\{\mathcal{A}_{(\lambda,0)}S\psi\},$$

so that

$$\begin{aligned} (\mathcal{A}_{(\lambda,0)}g_f)(j) &= \big(\mathcal{A}_{(\lambda,0)}S(Q^{-1}Pf)\big)(j) \\ &= (QQ^{-1}Pf)(j) + \mathrm{Po}(\lambda)\{\mathcal{A}_{(\lambda,0)}g_f\} \\ &= f(j) - \mathrm{Po}(\lambda)\{f\} + \mathrm{Po}(\lambda)\{\mathcal{A}_{(\lambda,0)}g_f\} = f(j) - c_f, \end{aligned} \tag{2.15}$$

say. Thus, in view of (2.13)–(2.15), the function g_f of (2.12) has all the properties required for the theorem, if we can estimate $\|\pi_{(\lambda,0)}\{f\} - c_f\|$.

However, writing $\pi\{\cdot\}$ for $\pi_{(\lambda,0)}\{\cdot\}$, we note that $\pi\{\mathcal{A}_{(\lambda,0)}g_f\} = 0$ implies that

$$0 = \sum_{j\ge 0}\pi(j)\big(f(j) - c_f\big) + \sum_{j<0}\pi(j)(Ug_f)(j).$$

Noting that $\pi(\mathbb{Z}) = 1$ and that $\mathrm{Po}(\lambda)(\mathcal{A}_{(\lambda,0)}g_f) = \mathrm{Po}(\lambda)(Ug_f)$, we get

$$\begin{aligned} |\pi\{f\} - c_f| &\le \eta(\|f\| + |c_f| + \|Ug_f\|) \\ &\le \eta(2\|f\| + |\mathrm{Po}(\lambda)\{\mathcal{A}_{(\lambda,0)}g_f\}| + \|Ug_f\|) \\ &\le 2\eta(\|f\| + \|Ug_f\|) \le 2\eta(\|f\| + \lambda\theta\|\Delta g_f\|) \\ &\le \frac{2}{1-2\theta}\eta\|f\|, \end{aligned} \tag{2.16}$$

because of (2.14). This completes the proof. □

Note also that, from (2.15) and (2.16),

$$\|\pi\| \le \frac{2\eta}{1-2\theta} + 1 + \frac{2\theta}{1-2\theta} = \frac{1+2\eta}{1-2\theta}. \tag{2.17}$$

COROLLARY 2.2. *Under conditions* (2.3) *and* (2.4), *given any bounded* $f\colon \mathbb{Z}\to\mathbb{R}$, *there exists* $g\colon \mathbb{Z}\to\mathbb{R}$ *satisfying*

(1) $g(i)=0, \qquad i\le\gamma;$

(2) $\big|(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(i)-\big(f(i)-\pi_{(\boldsymbol{\lambda},\gamma)}\{f\}\big)\big|\le \frac{2}{1-2\theta}\eta(\boldsymbol{\lambda})\|f\|, \qquad i\ge\gamma;$

(3) $\|g\|\le\frac{2}{1-2\theta}\{1\wedge\lambda^{-1/2}\}\|f\|, \qquad \|\Delta g\|\le\frac{2}{1-2\theta}\{1\wedge\lambda^{-1}\}\|f\|,$

where $\eta(\boldsymbol{\lambda}):=\sum_{j<0}|\pi_{(\boldsymbol{\lambda},0)}\{j\}|$ *is as before*.

PROOF. Take $g(i)=\hat g(i-\gamma)$, $i\in\mathbb{Z}$, where $\hat g$ is the function defined in Theorem 2.1 with $\hat f$ for f, where $\hat f(i):=f(i+\gamma)$. □

COROLLARY 2.3. *If* W *is an integer-valued random variable such that*, *for some* $\boldsymbol{\lambda}$ *satisfying* (2.3) *and* (2.4),

$$|\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)|\le\varepsilon_0\|g\|+\varepsilon_1\|\Delta g\| \tag{2.18}$$

for all bounded $g\colon\mathbb{Z}\to\mathbb{R}$, *then*

$$\begin{aligned}&\|\mathcal{L}(W)-\pi_{(\boldsymbol{\lambda},\gamma)}\|\\ &\qquad\le\frac{2}{1-2\theta}\big\{(1\wedge\lambda^{-1/2})\varepsilon_0+(1\wedge\lambda^{-1})\varepsilon_1+\eta(\boldsymbol{\lambda})+\big(1+\eta(\boldsymbol{\lambda})\big)\mathbb{P}[W<\gamma]\big\},\end{aligned}$$

where $\lambda:=\sum_{1\le|l|\le t} l\lambda_l$ *is as before*.

PROOF. Take any bounded $f\colon\mathbb{Z}\to\mathbb{R}$, and let g be as in Corollary 2.2. Then it is immediate that

$$\begin{aligned}&|\mathbb{E}f(W)-\pi_{(\boldsymbol{\lambda},\gamma)}\{f\}|\\ &\qquad\le\big|\mathbb{E}\{f(W)-\pi_{(\boldsymbol{\lambda},\gamma)}\{f\}\}I[W<\gamma]\big|+\big|\mathbb{E}\{f(W)-\pi_{(\boldsymbol{\lambda},\gamma)}\{f\}\}I[W\ge\gamma]\big|\\ &\qquad\le(1+\|\pi\|)\|f\|\mathbb{P}[W<\gamma]+\frac{2}{1-2\theta}\eta\|f\|\\ &\qquad\quad+\big|\mathbb{E}\{(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)\big(1-I[W<\gamma]\big)\}\big|\\ &\qquad\le(1+\|\pi\|)\|f\|\mathbb{P}[W<\gamma]+\frac{2}{1-2\theta}\eta\|f\|\\ &\qquad\quad+\varepsilon_0\|g\|+\varepsilon_1\|\Delta g\|+\|U_{(\boldsymbol{\lambda},\gamma)}g\|\mathbb{P}[W<\gamma],\end{aligned}$$

and the corollary follows from (2.11), (2.17) and Corollary 2.2. □

The result of Corollary 2.3 is clearly simpler if $W\ge\gamma$ a.s., or if $\lambda_l=0$ whenever $l<0$, in which case $\eta(\boldsymbol{\lambda})=0$. Some ways of bounding $\eta(\boldsymbol{\lambda})$ are given in Section 6.

3. Centered Poisson approximation. In what follows, we take W to be a sum $\sum_{i=1}^n Z_i$ of independent integer-valued random variables. Here, we consider the simplest possible approximation by measures of the form $\pi_{(\boldsymbol{\lambda},\gamma)}$, in which $t=1$, $\lambda_1>0$ and $\lambda_{-1}=0$, so that $\pi_{(\boldsymbol{\lambda},\gamma)}\sim\gamma+\mathrm{Po}(\lambda_1)$ is a translate of a Poisson distribution. This simple approximation was considered by Čekanavičius and Vaitkus (1999), a slightly more refined version being employed in Barbour and Xia (1999). For such a choice of $\boldsymbol{\lambda}$, we have $\theta=0$ and $\eta(\boldsymbol{\lambda})=0$, and the Stein operator $\mathcal{A}_{(\boldsymbol{\lambda},0)}$ is the usual Poisson operator.

To apply Corollary 2.3, we need to show that

$$|\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)|\le\varepsilon_0\|g\|+\varepsilon_1\|\Delta g\| \tag{3.1}$$

for all bounded $g\colon\mathbb{Z}\to\mathbb{R}$, when λ_1 and γ are suitably chosen. The strategy, here and subsequently, is to start by choosing coefficients $\lambda_l^{(i)}$ and $\gamma^{(i)}$ for each i, in such a way that the corresponding number of moments of the Z_i are exactly matched. It is then usually necessary to add a rounding correction to $\sum_{i=1}^n\gamma^{(i)}$, to obtain an integral value of γ, best results being obtained when the λ_l are also chosen to be slightly different from $\sum_{i=1}^n\lambda_l^{(i)}$. This procedure makes $|\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)|$ suitably small, as illustrated in the following theorem.

Suppose that $\mathbb{E}Z_i=\mu_i$, $\operatorname{Var}Z_i=\sigma_i^2$ and $\mathbb{E}|Z_i^3|<\infty$. Take $\lambda_1^{(i)}=\sigma_i^2$ and $\gamma^{(i)}=\mu_i-\sigma_i^2$, matching the first two moments of Z_i exactly, and then define

$$\gamma:=\lfloor\mu-\sigma^2\rfloor,\qquad\lambda_1=\sigma^2+\delta, \tag{3.2}$$

where $\mu:=\sum_{i=1}^n\mu_i=\mathbb{E}W$, $\sigma^2:=\sum_{i=1}^n\sigma_i^2=\operatorname{Var}W$ and

$$0\le\delta:=(\mu-\sigma^2)-\lfloor\mu-\sigma^2\rfloor<1.$$

Further, using $d_{\mathrm{TV}}(P,Q)$ for probability measures P and Q to denote $\frac12\|P-Q\|$, set

$$\begin{aligned}
&W_i:=W-Z_i,\qquad d:=\max_{1\le i\le n}\|\mathcal{L}(W_i)-\mathcal{L}(W_i+1)\|,\\
&v_i:=\min\{\tfrac12,1-d_{\mathrm{TV}}(\mathcal{L}(Z_i),\mathcal{L}(Z_i+1))\},\\
&\psi_i:=\sigma_i^2\mathbb{E}\{Z_i(Z_i-1)\}+|\mu_i-\sigma_i^2|\mathbb{E}\{(Z_i-1)(Z_i-2)\}\\
&\qquad+\mathbb{E}|Z_i(Z_i-1)(Z_i-2)|,
\end{aligned} \tag{3.3}$$

observing, from Barbour and Xia [(1999), Proposition 4.6], that if $V:=\sum_{i=1}^n v_i$ and $v^*:=\max_{1\le i\le n}v_i$, then

$$d\le2\{V-v^*\}^{-1/2}. \tag{3.4}$$

THEOREM 3.1. *For γ and λ_1 as in* (3.2),

$$\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| \le (1 \wedge \sigma^{-2})\left\{2\delta + d\sum_{i=1}^{n}\psi_i\right\} + 2\mathbb{P}[W < \gamma]$$
$$\le \sigma^{-2}\left\{4 + d\sum_{i=1}^{n}\psi_i\right\}.$$

PROOF. To obtain a bound of the form (3.1), we write

$$\begin{aligned}\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W) &= \mathbb{E}\{\lambda_1 g(W+1) - (W-\gamma)g(W)\}\\ &= \sigma^2\mathbb{E}g(W+1) + (\mu-\sigma^2)\mathbb{E}g(W)\\ &\quad - \mathbb{E}\{Wg(W)\} + \delta\mathbb{E}\Delta g(W)\\ &= \sum_{i=1}^{n}\{\sigma_i^2\mathbb{E}g(W+1) + (\mu_i - \sigma_i^2)\mathbb{E}g(W) - \mathbb{E}\{Z_i g(W)\}\}\\ &\quad + \delta\mathbb{E}\Delta g(W),\end{aligned} \tag{3.5}$$

the last term being simply bounded using

$$|\mathbb{E}\Delta g(W)| \le \|\Delta g\|. \tag{3.6}$$

Proceeding as in Barbour and Xia [(1999), Theorem 4.3], we write Newton's expansion in the form

$$\begin{aligned} g(w+l) &= g(w+1) + (l-1)\Delta g(w+1)\\ &\quad + \begin{cases} \sum_{s=1}^{l-2}(l-1-s)\Delta^2 g(w+s), & l \ge 3,\\ 0, & 1 \le l \le 2,\\ \sum_{s=0}^{-l}(-l-s+1)\Delta^2 g(w-s), & l \le 0,\end{cases}\end{aligned} \tag{3.7}$$

and we also observe that, for any random variable U, any bounded g and any integer j, we have

$$|\mathbb{E}\Delta^2 g(U+j)| \le \|\Delta g\|\,\|\mathcal{L}(U) - \mathcal{L}(U+1)\|. \tag{3.8}$$

Combining (3.7) and (3.8), we thus obtain the short expansion

$$\begin{aligned}&\left|\mathbb{E}g(W_i+l) - \mathbb{E}g(W_i+1) - (l-1)\mathbb{E}\Delta g(W_i+1)\right|\\ &\qquad \le \tfrac{1}{2}(l-1)(l-2)d\|\Delta g\|.\end{aligned} \tag{3.9}$$

We now use (3.9) to expand the main terms in (3.5), obtaining

$$\begin{aligned}\mathbb{E}g(W+1) &= \sum_{j\in\mathbb{Z}} \mathbb{P}[Z_i=j]\mathbb{E}g(W_i+j+1)\\ &= \mathbb{E}g(W_i+1)+\mu_i\mathbb{E}\Delta g(W+1)+r_{i1},\end{aligned}\tag{3.10}$$

where

$$|r_{i1}| \le \tfrac{1}{2}\mathbb{E}\{Z_i(Z_i-1)\}d\|\Delta g\|,\tag{3.11}$$

and then

$$\mathbb{E}g(W)=\mathbb{E}g(W_i+1)+(\mu_i-1)\mathbb{E}\Delta g(W_i+1)+r_{i0}\tag{3.12}$$

where

$$|r_{i0}| \le \tfrac{1}{2}\mathbb{E}\{(Z_i-1)(Z_i-2)\}d\|\Delta g\|,\tag{3.13}$$

and finally,

$$\begin{aligned}\mathbb{E}\{Z_i g(W)\} &= \sum_{j\in\mathbb{Z}} j\mathbb{P}[Z_i=j]\mathbb{E}g(W_i+j)\\ &= \mu_i\mathbb{E}g(W_i+1)+(\sigma_i^2+\mu_i^2-\mu_i)\mathbb{E}\Delta g(W_i+1)+r_i,\end{aligned}\tag{3.14}$$

where

$$|r_i| \le \tfrac{1}{2}\mathbb{E}|Z_i(Z_i-1)(Z_i-2)|d\|\Delta g\|.\tag{3.15}$$

But now, putting (3.10)–(3.15) into (3.5), we find that

$$\begin{aligned}&|\sigma_i^2\mathbb{E}g(W+1)+(\mu_i-\sigma_i^2)\mathbb{E}g(W)-\mathbb{E}\{Z_i g(W)\}|\\ &\qquad\le \sigma_i^2|r_{i1}|+|\mu_i-\sigma_i^2|\,|r_{i0}|+|r_i| \le \tfrac{1}{2}d\|\Delta g\|\psi_i,\end{aligned}\tag{3.16}$$

since the coefficients of $\mathbb{E}g(W_i+1)$ and $\mathbb{E}\Delta g(W_i+1)$ exactly vanish. Thus, from (3.5), (3.6) and (3.16), we have obtained a bound of the form (3.1), with $\varepsilon_0=0$ and $\varepsilon_1=\delta+\frac{1}{2}d\sum_{i=1}^n\psi_i$; since also $\mathbb{P}[W<\gamma]\le\sigma^{-2}$ by Chebyshev's inequality, and recalling that here $\theta=\eta(\boldsymbol{\lambda})=0$, the theorem follows from Corollary 2.3. □

COROLLARY 3.2. *Suppose that the Z_i satisfy $\sigma_i^2\ge a>0$, $v_i\ge b>0$ and $\sigma_i^{-2}\psi_i\le c<\infty$ for all $1\le i\le n$. Then*

$$\begin{aligned}\|\mathcal{L}(W)-\pi_{(\boldsymbol{\lambda},\gamma)}\| &\le 2c\{nb-1/2\}^{-1/2}+2\delta(na)^{-1}+2\mathbb{P}[W<\gamma]\\ &\le 2c\{nb-1/2\}^{-1/2}+4(na)^{-1}.\end{aligned}$$

In the error bound in Corollary 3.2, the second term is of smaller order. However, in triangular arrays $(Z_{in},\ 1\le i\le n,\ n\ge 1)$, it is natural to impose bounds on the $(Z_{in},\ 1\le i\le n)$ in which $a=a_n$, $b=b_n$ and $c=c_n$, and then the relative orders of magnitude may be different. For example, if $Z_{in} \sim \mathrm{Be}(p_n)$ with $p_n \le 1/2$, in which case $\gamma = \lfloor np_n^2 \rfloor$ and $\lambda_1 = np_n(1-p_n) + (np_n^2 - \gamma)$, then $a_n = p_n(1-p_n)$, $b_n = \min\{p_n, 1-p_n\}$ and $c_n = 2p_n$; if also $np_n \ge 1$, then Corollary 3.2 gives a bound of order

$$O\big(p_n(np_n)^{-1/2} + (np_n)^{-1}\big),$$

with the second term being the larger if $np_n^3 \to 0$. However, for Bernoulli Z_i, the last term can be improved if, for instance, $np_n^2 < 1$, in which case $\gamma = 0$, so that we have the usual Poisson approximation, $\lambda_1 = np_n$, $\delta = np_n^2$ and $\mathbb{P}[W < \gamma] = 0$; then the second term, still the larger, is of order $O(p_n \wedge np_n^2)$, as usual for Poisson approximation. Using Theorem 3.1, similar bounds can be obtained for unequally distributed Bernoulli summands.

Theorem 3.1 generalizes Čekanavičius and Vaitkus [(1999), Theorem 2.1], which only covered Bernoulli random variables, and also Čekanavičius (1998), Theorem 3, which was only for independent and identically distributed sequences and had no *explicit* representation of the constants implied in the error estimates. Because Theorem 3.1 contains very explicit bounds, it can be applied in great generality to triangular arrays.

4. Second-order approximations. In this section, we refine the centered Poisson approximation. First, we take one extra main parameter λ_l in the approximating distribution, either λ_{-1} or λ_2, and establish approximations to $\mathcal{L}(W)$ by *probability* distributions $\pi_{(\boldsymbol{\lambda},\gamma)}$ of accuracy $O(n^{-1})$ under reasonable uniformity conditions, provided that the third cumulant of W is not too far from its variance. This last, unwanted restriction is then removed by considering more general probability distributions $\pi_{(\boldsymbol{\lambda},\gamma)}$ with λ_1, λ_2 and λ_s nonzero, for some $s \notin \{0,1,2\}$; see Corollary 4.5 for the implied contribution to Kolmogorov's problem in total variation.

The first approximation is by the distribution of the *difference* of two independent Poisson random variables, centered appropriately. The extra parameter λ_{-1} enables one more moment to be matched. We suppose that the Z_i are as in the previous section, but now satisfy $\mathbb{E}Z_i^4 < \infty$; we define $c_{3i} := \mathbb{E}\{(Z_i - \mu_i)^3\}$. We also introduce the notation $\kappa_r^{(i)} := \kappa_r(Z_i)$ to denote the rth factorial cumulant of Z_i [Kendall and Stuart (1963), Section 3.17, page 75], so that, in particular,

$$\kappa_1^{(i)} = \mu_i, \qquad \kappa_2^{(i)} = \sigma_i^2 - \mu_i, \qquad \kappa_3^{(i)} = c_{3i} - 3\sigma_i^2 + 2\mu_i. \tag{4.1}$$

The factorial cumulants can be formally fitted using parameters $\lambda_{-1}^{(i)}$, $\lambda_1^{(i)}$ and $\gamma^{(i)}$ from the equations

$$\kappa_r^{(i)} = (-1)^{r-1}(r-1)!\gamma^{(i)} + \binom{1}{r} r!\lambda_1^{(i)} + \binom{-1}{r} r!\lambda_{-1}^{(i)}, \qquad 1 \le r \le 3, \tag{4.2}$$

where, for $m \in \mathbb{Z}$ and $r \in \mathbb{N}$, $\binom{m}{r} := m(m-1)\cdots(m-r+1)/r!$. This gives

$$\gamma^{(i)} = -3\kappa_2^{(i)} - \kappa_3^{(i)}, \qquad \lambda_1^{(i)} = \mu_i + 2\kappa_2^{(i)} + \tfrac{1}{2}\kappa_3^{(i)}, \\ \lambda_{-1}^{(i)} = -\kappa_2^{(i)} - \tfrac{1}{2}\kappa_3^{(i)}, \tag{4.3}$$

leading to the choices

$$\gamma := \left\lfloor \sum_{i=1}^{n} \gamma^{(i)} \right\rfloor = -3\kappa_2 - \kappa_3 - 2\delta = \mu - c_3 - 2\delta, \tag{4.4}$$

with $0 \le \delta < 1/2$, where $\kappa_r := \kappa_r(W)$ and $c_3 := \mathbb{E}(W-\mu)^3$, and

$$\lambda_1 = \mu + 2\kappa_2 + \tfrac{1}{2}\kappa_3 + \delta = \tfrac{1}{2}(c_3 + \sigma^2) + \delta, \\ \lambda_{-1} = -\kappa_2 - \tfrac{1}{2}\kappa_3 - \delta = \tfrac{1}{2}(\sigma^2 - c_3) - \delta. \tag{4.5}$$

In order that $\pi_{(\lambda,\gamma)}$ be a probability distribution, we must therefore have $c_3 \le \sigma^2 - 2\delta$, and, if $\lambda = c_3 + 2\delta > 0$, then

$$\theta = (\sigma^2 - 2\delta - c_3)/(c_3 + 2\delta) < 1/2 \tag{4.6}$$

if $c_3 > (2\sigma^2 - 6\delta)/3$. Thus this approximation is only suitable if the sum W satisfies the restrictive moment condition

$$(2\,\mathrm{Var}\,W - 6\delta)/3 < \mathbb{E}(W-\mu)^3 \le \mathrm{Var}\,W - 2\delta. \tag{4.7}$$

We need some further notation before stating the theorem. We define

$$d' := \max_{1 \le i \le n} \|\mathcal{L}(W_i) * (E_1 - E)^{*2}\|, \tag{4.8}$$

where the measures E and E_1 denote unit mass on 0 and 1, respectively, and $*$ denotes convolution. Since, for probability measures ν_1 and ν_2, we have

$$\|\nu_1 * \nu_2 * (E_1 - E)^{*2}\| \le \|\nu_1 * (E_1 - E)\|\,\|\nu_2 * (E_1 - E)\|,$$

it follows from (3.3) and (3.4) that

$$d' \le 4\left\{1 \wedge \frac{2}{(V - 4v^*)_+}\right\} \le 4\left\{1 \wedge \frac{2}{(V-2)_+}\right\} \le 16V^{-1}. \tag{4.9}$$

Finally, we set

$$\begin{aligned} \psi_i' := {} & \tfrac{1}{2}|c_{3i} - \sigma_i^2|\mathbb{E}|(Z_i - 2)(Z_i - 3)(Z_i - 4)| \\ & + \tfrac{1}{2}|c_{3i} + \sigma_i^2|\mathbb{E}|Z_i(Z_i - 1)(Z_i - 2)| \\ & + |\mu_i - c_{3i}|\mathbb{E}|(Z_i - 1)(Z_i - 2)(Z_i - 3)| \\ & + \mathbb{E}|Z_i(Z_i - 1)(Z_i - 2)(Z_i - 3)|. \end{aligned} \tag{4.10}$$

THEOREM 4.1. *With the above notation, if the independent random variables* Z_i *have* $\mathbb{E}Z_i^4 < \infty$ *and their sum* W *satisfies* (4.7), *then*

$$\begin{aligned}\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| &\\ \leq \frac{2}{1-2\theta}\Biggl\{&(c_3+2\delta)^{-1}\Biggl(d\delta + \frac{1}{6}d'\sum_{i=1}^n \psi_i'\Biggr)\\ &+ (c_3+2\delta)^{-4}\Biggl(3\sigma^4 + \sum_{i=1}^n \mathbb{E}(Z_i-\mu)^4\Biggr) + \exp\Biggl\{-\frac{25\lambda}{128}\Biggr\}\Biggr\},\end{aligned}$$

where θ *is as in* (4.6), d *as in* (3.3) *and* d' *as in* (4.8).

PROOF. Once again, we bound $|\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)|$ and apply Corollary 2.3. In fact,

$$\begin{aligned}\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W) &= \mathbb{E}\{-\lambda_{-1}g(W-1) + \lambda_1 g(W+1) - (W-\gamma)g(W)\}\\ &= \sum_{i=1}^n \mathbb{E}\{\tfrac{1}{2}(c_{3i}-\sigma_i^2)g(W-1) + \tfrac{1}{2}(c_{3i}+\sigma_i^2)g(W+1)\\ &\qquad + (\mu_i - c_{3i})g(W) - Z_i g(W)\} + \delta\mathbb{E}\Delta^2 g(W-1),\end{aligned} \tag{4.11}$$

where the last term is bounded, as in the previous section, by $\delta d\|\Delta g\|$. Much as before, for any random variable U, bounded g and integer j, we have

$$|\mathbb{E}\Delta^3 g(U+j)| \leq \|\mathcal{L}(U) * (E_1 - E)^{*2}\|\,\|\Delta g\|. \tag{4.12}$$

Now write Newton's formula in the form

$$\begin{aligned}g(w+l) &= g(w+1) + (l-1)\Delta g(w+1) + \binom{l-1}{2}\Delta^2 g(w+1)\\ &\quad + \begin{cases}\displaystyle\sum_{s=1}^{l-3}\binom{l-s-1}{2}\Delta^3 g(w+s), & l \geq 4,\\ 0, & 1 \leq l \leq 3,\\ \displaystyle-\sum_{s=0}^{-l}\binom{-l-s+2}{2}\Delta^3 g(w-s), & l \leq 0,\end{cases}\end{aligned} \tag{4.13}$$

noting that

$$\begin{aligned}\sum_{s=1}^{l-3}\binom{l-s-1}{2} &= \frac{1}{6}(l-1)(l-2)(l-3), \qquad l \geq 4,\\ \sum_{s=0}^{-l}\binom{-l-s+2}{2} &= \frac{1}{6}|(l-1)(l-2)(l-3)|, \qquad l \leq 0.\end{aligned}$$

Then, much as in the proof of Theorem 3.1, we have the short expansion

$$\begin{aligned}&\left|\mathbb{E}g(W_i+l)-\mathbb{E}g(W_i+1)-(l-1)\mathbb{E}\Delta g(W_i+1)-\binom{l-1}{2}\mathbb{E}\Delta^2 g(W_i+1)\right|\\ &\quad\le \tfrac{1}{6}|(l-1)(l-2)(l-3)|\|\mathcal{L}(W_i)*(E_1-E)^{*2}\|\,\|\Delta g\|\\ &\quad\le \tfrac{1}{6}|(l-1)(l-2)(l-3)|\,d'\|\Delta g\|.\end{aligned}\tag{4.14}$$

Applying this successively, we obtain

$$\begin{aligned}\mathbb{E}g(W-1)&=\sum_{j\in\mathbb{Z}}\mathbb{P}[Z_i=j]\mathbb{E}g(W_i+j-1)\\ &=\mathbb{E}g(W_i+1)+(\mu_i-2)\mathbb{E}\Delta g(W_i+1)\\ &\quad+\tfrac{1}{2}(\kappa_2^{(i)}-4\mu_i+\mu_i^2+6)\mathbb{E}\Delta^2 g(W_i+1)+r_{i,-1},\end{aligned}\tag{4.15}$$

where

$$|r_{i,-1}|\le \tfrac{1}{6}d'\|\Delta g\|\mathbb{E}|(Z_i-2)(Z_i-3)(Z_i-4)|;\tag{4.16}$$

then

$$\begin{aligned}\mathbb{E}g(W+1)&=\mathbb{E}g(W_i+1)+\mu_i\mathbb{E}\Delta g(W_i+1)\\ &\quad+\tfrac{1}{2}(\kappa_2^{(i)}+\mu_i^2)\mathbb{E}\Delta^2 g(W_i+1)+r_{i1},\end{aligned}\tag{4.17}$$

where

$$|r_{i1}|\le \tfrac{1}{6}d'\|\Delta g\|\mathbb{E}|Z_i(Z_i-1)(Z_i-2)|;\tag{4.18}$$

then

$$\begin{aligned}\mathbb{E}g(W)&=\mathbb{E}g(W_i+1)+(\mu_i-1)\mathbb{E}\Delta g(W_i+1)\\ &\quad+\tfrac{1}{2}(\kappa_2^{(i)}-2\mu_i+\mu_i^2+2)\mathbb{E}\Delta^2 g(W_i+1)+r_{i0},\end{aligned}\tag{4.19}$$

where

$$|r_{i0}|\le \tfrac{1}{6}d'\|\Delta g\|\mathbb{E}|(Z_i-1)(Z_i-2)(Z_i-3)|;\tag{4.20}$$

finally,

$$\begin{aligned}\mathbb{E}\{Z_i g(W)\}&=\mu_i\mathbb{E}g(W_i+1)+(\kappa_2^{(i)}+\mu_i^2)\mathbb{E}\Delta g(W_i+1)\\ &\quad+\tfrac{1}{2}(\kappa_3^{(i)}+3\mu_i\kappa_2^{(i)}+\mu_i^3)\mathbb{E}\Delta^2 g(W_i+1)+r_i,\end{aligned}\tag{4.21}$$

where

$$|r_i|\le \tfrac{1}{6}d'\|\Delta g\|\mathbb{E}|Z_i(Z_i-1)(Z_i-2)(Z_i-3)|.\tag{4.22}$$

Putting (4.15)–(4.22) into (4.11), only the remainders survive, giving a bound as in (2.18) with $\varepsilon_0=0$ and $\varepsilon_1=\{d\delta+\frac{1}{6}d'\sum_{i=1}^n\psi_i'\}$.

To complete the bound in Corollary 2.3, we still need to bound $\eta(\boldsymbol{\lambda})$ and $\mathbb{P}[W < \gamma]$. For the latter, a fourth moment bound gives

$$\mathbb{P}[W < \gamma] \le \frac{\mathbb{E}(W-\mu)^4}{(c_3+2\delta)^4} \le \frac{3\sigma^4 + \sum_{i=1}^n \mathbb{E}(Z_i-\mu)^4}{(c_3+2\delta)^4}. \tag{4.23}$$

For $\eta(\boldsymbol{\lambda})$, it follows from Lemmas 6.4 and 6.5 that

$$\eta(\boldsymbol{\lambda}) \le \exp\{-\tfrac{1}{4}(\lambda - \tfrac{1}{8}\Lambda_2)\}, \tag{4.24}$$

where

$$\lambda = c_3 + 2\delta \quad \text{and} \quad \Lambda_2 = \tfrac{1}{2}(3\sigma^2 - c_3) - \delta. \tag{4.25}$$

Combining (4.7), (4.24) and (4.25), a bound for $\eta(\boldsymbol{\lambda})$ is derived, and the theorem follows from Corollary 2.3. □

REMARK. The estimate of $\mathbb{E}\{Z_i g(W)\}$ derives from (4.14), multiplied by l. If we allow for the bound (3.9) as well, we have

$$\begin{aligned} &\biggl| l\biggl\{\mathbb{E}g(W_i+l) - \mathbb{E}g(W_i+1) - (l-1) \\ &\quad \times \mathbb{E}\Delta g(W_i+1) - \binom{l-1}{2}\mathbb{E}\Delta^2 g(W_i+1)\biggr\}\biggr| \\ &\qquad \le \min\{|l(l-1)(l-2)|\,d, \tfrac{1}{6}|l(l-1)(l-2)(l-3)|\,d'\}\|\Delta g\|, \end{aligned} \tag{4.26}$$

leading to the estimate

$$|r_i| \le \|\Delta g\|\mathbb{E}\{|Z_i(Z_i-1)(Z_i-2)|\min\{d, \tfrac{1}{6}d'|(Z_i-3)|\}\}. \tag{4.27}$$

The bound (4.23) can also be replaced by a third moment Chebyshev estimate, coupled with Rosenthal's inequality. This enables one to bound $\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\|$, assuming the existence of only three moments.

COROLLARY 4.2. *Suppose that the Z_i satisfy the inequalities $\sigma_i^2 \ge a > 0$, $v_i \ge b > 0$ and $\sigma_i^{-2}(\psi_i' + \mathbb{E}(Z_i-\mu_i)^4) \le c < \infty$ for all i, and suppose that their sum W satisfies* (4.7). *Then*

$$\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| \le \frac{K}{1-2\theta}\biggl\{\frac{c}{nb} + \frac{1}{na(nb)^{1/2}} + \frac{1}{(na)^2} + \frac{c}{(na)^3} + e^{-\alpha na}\biggr\},$$

where the constants K and α are uniform in $\sigma^2 \ge 2$.

REMARK. If $\mathbb{E}Z_i^4 = \infty$, replace ψ_i' by the smaller

$$\begin{aligned} \psi_i'' := {}& \tfrac{1}{2}|c_{3i} - \sigma_i^2|\mathbb{E}|(Z_i-2)(Z_i-3)(Z_i-4)| \\ &+ \tfrac{1}{2}|c_{3i} + \sigma_i^2|\mathbb{E}|Z_i(Z_i-1)(Z_i-2)| \\ &+ |\mu_i - c_{3i}|\mathbb{E}|(Z_i-1)(Z_i-2)(Z_i-3)|, \end{aligned} \tag{4.28}$$

and assume that $\sigma_i^{-2}(\psi_i'' + \mathbb{E}|Z_i - \mu_i|^3) \le c < \infty$ for all i. Then $\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\|$ can be bounded by the same expression as in Corollary 4.2, supplemented by the additional term

$$\text{(4.29)} \qquad \frac{1}{na(nb)^{1/2}} \sum_{i=1}^{n} \mathbb{E}\bigl(|Z_i(Z_i-1)(Z_i-2)| \min\{1, n^{-1/2}|Z_i - 3|\}\bigr).$$

This, for instance, gives a bound of order $O(n^{-(1+\delta)/2})$ if also $\mathbb{E}|Z_i|^{3+\delta} \le c' < \infty$ for all i.

An alternative approximation using two Poisson parameters λ_1 and λ_2 and a shift γ can also be derived. Here, the factorial cumulant equations corresponding to (4.2) are

$$\text{(4.30)} \qquad \kappa_r^{(i)} = (-1)^{r-1}(r-1)!\gamma^{(i)} + \binom{1}{r} r!\lambda_1^{(i)} + \binom{2}{r} r!\lambda_2^{(i)}, \qquad 1 \le r \le 3,$$

giving

$$\gamma^{(i)} = \tfrac{1}{2}\kappa_3^{(i)}, \qquad \lambda_1^{(i)} = \mu_i - \kappa_2^{(i)} - \kappa_3^{(i)}, \qquad 2\lambda_2^{(i)} = \kappa_2^{(i)} + \tfrac{1}{2}\kappa_3^{(i)},$$

and leading to the choices

$$\text{(4.31)} \qquad \begin{aligned} \gamma &= \lfloor \tfrac{1}{2}\kappa_3 \rfloor = \tfrac{1}{2}\kappa_3 + \delta, \qquad 0 \le \delta < 1, \\ \lambda_1 &= 2\sigma^2 - c_3 - 2\delta, \qquad 2\lambda_2 = \tfrac{1}{2}(c_3 - \sigma^2) + \delta. \end{aligned}$$

For a probability distribution $\pi_{(\boldsymbol{\lambda},\gamma)}$, we need to have $c_3 \ge \sigma^2 - 2\delta$ and $c_3 \le 2(\sigma^2 - \delta)$; then $\lambda = \frac{1}{2}(3\sigma^2 - c_3) - \delta > 0$ and

$$\text{(4.32)} \qquad \theta = \{c_3 - \sigma^2 + 2\delta\}/\{3\sigma^2 - c_3 - 2\delta\} < 1/2$$

if $c_3 < (5\sigma^2 - 6\delta)/3$. Arguing as for Theorem 4.1, if

$$\text{(4.33)} \qquad \operatorname{Var} W - 2\delta \le \mathbb{E}(W - \mu)^3 < (5 \operatorname{Var} W - 6\delta)/3,$$

then $\mathcal{L}(W)$ is approximated by $\pi_{(\boldsymbol{\lambda},\gamma)}$ with much the same accuracy as that given in Theorem 4.1 and Corollary 4.2; note that here $\eta(\boldsymbol{\lambda}) = 0$ automatically.

In the ranges allowed by (4.7) and (4.33), we have shown that there is an infinitely divisible approximation to $\mathcal{L}(W)$ with accuracy of order $O(n^{-1})$, measured with respect to total variation distance, under uniformity assumptions such as those in Corollary 4.2. For independent and identically distributed summands, this merely requires a finite fourth moment and a nonzero value of $d_{\mathrm{TV}}(\mathcal{L}(Z_1), \mathcal{L}(Z_1 + 1))$, provided that one of (4.7) or (4.33) holds. This is a partial answer to Kolmogorov's problem in total variation, as discussed in the Introduction; furthermore, under the conditions of Corollary 4.2, identical distributions are not required.

We now show that we can circumvent the limitations imposed by (4.7) and (4.33), demonstrating the existence of an infinitely divisible approximation to $\mathcal{L}(W)$ with total variation accuracy of order $O(n^{-1})$, provided only that the order assumptions in Corollary 4.2 hold. The approach is much as above, but the approximating distribution is a little more complicated. We consider γ-shifted compound Poisson distributions with only λ_1, λ_2 and λ_s nonzero, where $s \in \mathbb{Z} \setminus \{0, 1, 2\}$ is to be chosen later. Here, λ_s plays the role that λ_{-1} and λ_2 played in the previous approximations, whereas λ_2 is used so as better to match the terms arising because γ has to be an integer. We can fit the first three moments of W in this way by taking

$$\begin{aligned} \delta &:= \gamma - \{\mu - \sigma^2 + s^{-1}(c_3 - \sigma^2)\}, \\ \lambda_1 &:= \sigma^2 - \frac{s}{s-1}\{s^{-1}(c_3 - \sigma^2) + 2\delta\}, \\ \lambda_2 &:= \frac{s\delta}{2(s-2)}, \qquad \lambda_s := \frac{c_3 - \sigma^2 - 2s\delta(s-2)^{-1}}{s^2(s-1)}. \end{aligned} \tag{4.34}$$

Very much as in the proof of Theorem 4.1, we find that

$$\begin{aligned} &|\mathbb{E}(\mathcal{A}_{(\boldsymbol{\lambda},\gamma)}g)(W)| \\ &\quad \le \frac{d'}{6}\sum_{i=1}^{n}\hat{\psi}_i \|\Delta g\| + \frac{2\delta}{(s-1)(s-2)}\left|\sum_{j=2}^{s-1}\sum_{l=1}^{j-1}\sum_{r=0}^{l-1}\mathbb{E}\Delta^3 g(W+r)\right| \end{aligned}$$

(the sums interpreted appropriately when $s < 0$), where

$$\begin{aligned} \hat{\psi}_i := {} & |\mu_i - \sigma_i^2 + s^{-1}(c_{3i} - \sigma_i^2)|\mathbb{E}|(Z_i - 1)(Z_i - 2)(Z_i - 3)| \\ & + |\sigma_i^2 - (s-1)^{-1}(c_{3i} - \sigma_i^2)|\mathbb{E}|Z_i(Z_i - 1)(Z_i - 2)| \\ & + \left|\frac{c_{3i} - \sigma_i^2}{s(s-1)}\right|\mathbb{E}|(Z_i + s - 1)(Z_i + s - 2)(Z_i + s - 3)| \\ & + \mathbb{E}|Z_i(Z_i - 1)(Z_i - 2)(Z_i - 3)| \end{aligned} \tag{4.35}$$

and $|\mathbb{E}\Delta^3 g(W + r)| \le d'\|\Delta g\|$. Using Corollary 2.3, this gives

$$\begin{aligned} &\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| \\ &\quad \le \frac{2}{1-2\theta}\left\{\frac{d'}{6\lambda}\left(\sum_{i=1}^{n}\hat{\psi}_i + 2|s|\delta\right) + \eta(\boldsymbol{\lambda}) + \big(1 + \eta(\boldsymbol{\lambda})\big)\mathbb{P}[W < \gamma]\right\}, \end{aligned} \tag{4.36}$$

provided that $\theta < 1/2$, where θ and λ are as in (2.3) and (2.4).

We thus need to show that γ and s can be chosen in such a way that $\pi_{(\boldsymbol{\lambda},\gamma)}$ is a probability measure and $\theta < 1/2$. We achieve this by taking

$$\gamma := \lceil \mu - \sigma^2 + s^{-1}(c_3 - \sigma^2) \rceil \quad \text{and} \quad s := -\max\{1, \lceil 8\sigma^{-2}(\sigma^2 - c_3) \rceil\} \tag{4.37}$$

if $c_3 < \sigma^2$; if $\sigma^2 \le c_3 < \sigma^2 + 3$, we take

$$\gamma := \lceil \mu - \sigma^2 + s^{-1}(c_3 - \sigma^2)\rceil + 3 \quad \text{and} \quad s := -2; \tag{4.38}$$

if $c_3 \ge \sigma^2 + 3$, we take

$$\gamma := \lceil \mu - \sigma^2 + s^{-1}(c_3 - \sigma^2)\rceil \quad \text{and} \quad s := \max\{6, \lceil 8\sigma^{-2}(c_3 - \sigma^2)\rceil\}. \tag{4.39}$$

THEOREM 4.3. *If* $\sigma^2 \ge 24$, *then* $\pi_{(\boldsymbol{\lambda},\gamma)}$, *defined using* (4.34) *and* (4.37)–(4.39), *is a probability measure*, *and*

$$\begin{aligned} \|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| \le{}& \frac{2d'}{3\sigma^2}\left\{\sum_{i=1}^n \hat{\psi}_i + 2C\right\} + \frac{10}{3}\exp\left\{-\frac{5\sigma^2}{48C}\right\} \\ &+ 42\sigma^{-4} + 14\sigma^{-8}\sum_{i=1}^n \mathbb{E}(Z_i - \mu_i)^4, \end{aligned} \tag{4.40}$$

where $\hat{\psi}_i$ *is as in* (4.35) *and* $C := \max\{8, \lceil 8\sigma^{-2}|\sigma^2 - c_3|\rceil\}$.

PROOF. Routine calculation shows that, with the choices of γ and s made in (4.37)–(4.39), and with $\sigma^2 \ge 24$, the quantities λ_1, λ_2 and λ_s are all nonnegative; furthermore, $\lambda \ge 5\sigma^2/6$, $\theta \le 1/5$ and $|s\delta| \le C$. Hence $\pi_{(\boldsymbol{\lambda},\gamma)}$ is indeed a probability measure, and it merely remains to examine (4.36) in detail. The first term is easy; and $\mathbb{P}[W < \gamma]$ can be estimated using a fourth moment bound much as in (4.23), since $\mu - \gamma = \lambda \ge 5\sigma^2/6$ and $\eta(\boldsymbol{\lambda}) \le 1$ because $\pi_{(\boldsymbol{\lambda},\gamma)}$ is a probability measure. Finally, $\eta(\boldsymbol{\lambda}) \le \exp\{-\lambda/(8t)\}$, where $t = 2 \vee |s|$, from Lemma 6.6(2). Since also $|s| \le C$ and $\lambda \ge 5\sigma^2/6$, the theorem follows. □

REMARK. If $\mathbb{E}Z_i^4 = \infty$, the bounds (4.36) and (4.40) can be altered in a way similar to that used in (4.28) and (4.29).

COROLLARY 4.4. *Suppose that the* Z_i *satisfy the inequalities* $\sigma_i^2 \ge a > 0$, $v_i \ge b > 0$ *and* $\sigma_i^{-2}(\hat{\psi}_i + \mathbb{E}(Z_i - \mu_i)^4) \le c < \infty$ *for all* $1 \le i \le n$. *Then*

$$\|\mathcal{L}(W) - \pi_{(\boldsymbol{\lambda},\gamma)}\| \le K\left\{\frac{c}{nb} + \frac{1+\sqrt{c}}{n^2ab} + \frac{1}{(na)^2} + \frac{c}{(na)^3} + \exp\left(\frac{-\alpha na}{1+\sqrt{c}}\right)\right\},$$

where the constants K *and* α *are uniform in* $\sigma^2 \ge 24$.

COROLLARY 4.5. *Let* $(Z_i^{(n)},\ 1 \le i \le n,\ n \ge 1)$ *be a triangular array of integer-valued random variables which are independent within rows*, *and suppose that*, *for all* n *and all* $1 \le i \le n$,

$$v_i^{(n)} \ge v_* > 0, \qquad \mathbb{E}\{(Z_i^{(n)} - \mathbb{E}Z_i^{(n)})^4\} \le c(\sigma_i^{(n)})^2.$$

Then, setting $W^{(n)} := \sum_{i=1}^n Z_i^{(n)}$ *and* $(\sigma^{(n)})^2 = \mathrm{Var}\, W^{(n)}$, *it follows that*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W^{(n)}), \mathcal{D}\bigr) = O\bigl(n^{-1} + (\sigma^{(n)})^{-4}\bigr),$$

where $\mathcal{D}$ *denotes the class of all infinitely divisible laws and where the implied constants depend only on* c *and* v_*.

PROOF. It is immediate that $(\sigma_i^{(n)})^2 \le c$ and that $|c_{3i}^{(n)}| \le (\sigma_i^{(n)})^2\sqrt{c}$, and we may assume that $(\sigma^{(n)})^2 \ge 24$. Furthermore, when applying Theorem 4.3 to $W^{(n)}$, we have

$$1 \le |s^{(n)}| \le 9 + 8\sqrt{c} \quad \text{and} \quad |s^{(n)} - 1| \ge 2.$$

Now translate the $Z_i^{(n)}$ integrally so that $|\mu_i^{(n)}| \le 1/2$, implying by Lemma 6.7 that

$$|\mu_i^{(n)}| \le \mathbb{E}|Z_i^{(n)}| \le 2(\sigma_i^{(n)})^2;$$

then it follows easily that

$$\hat{\psi}_i^{(n)} \le K(\sigma_i^{(n)})^2(1 + \sqrt{c})^4$$

for a universal constant K. Applying Theorem 4.3 completes the proof. □

The improvement over Arak's (1981) bound of $O(n^{-2/3})$ is possible because of the extra uniformity conditions imposed.

For Bernoulli random variables, we have

$$\mu = \sum_{i=1}^n p_i, \qquad \sigma^2 = \sum_{i=1}^n p_i(1-p_i), \qquad c_3 = \sum_{i=1}^n p_i(1-p_i)(1-2p_i)$$

and $0 \le \sigma^2 - c_3 \le 2\sigma^2$. This gives a bound of order

$$O\left\{\left(\sum_{i=1}^n p_i(1-p_i)\right)^{-2}\left(1 + \sum_{i=1}^n p_i^2\right)\right\}$$

in Theorem 4.3. In particular, if all the p_i's are equal, we have

$$d_{\mathrm{TV}}\bigl(\mathrm{Bi}(n,p), \mathcal{D}\bigr) = O\bigl\{(np)^{-2} + n^{-1}\bigr\}, \tag{4.41}$$

where we take $p \le 1/2$ without loss of generality. If np is small, one can use the ordinary Poisson approximation to show that

$$d_{\mathrm{TV}}\bigl(\mathrm{Bi}(n,p), \mathcal{D}\bigr) = O\bigl\{\min(p, np^2)\bigr\}. \tag{4.42}$$

The bounds obtained by combining (4.41) and (4.42) are of the correct order of magnitude, as shown by Presman (1983).

5. Higher order expansions. In this section, we look for even more accurate approximations. First, we suppose that the random variables Z_i are (integrally) "centered" in such a way that all the partial sums $S_{rs} := \sum_{i=r}^{s} Z_i$ have second factorial cumulant satisfying

$$|\kappa_2(S_{rs})| = |\operatorname{Var} S_{rs} - \mathbb{E}S_{rs}| \le 1, \qquad 1 \le r, s \le n, \tag{5.1}$$

as is clearly possible. Writing $W := S_n := S_{1n}$, this implies a choice of γ close to that of Theorem 3.1 and (3.2), and, in particular, $\mu := \mathbb{E}W$ and $\sigma^2 := \operatorname{Var} W$ satisfy $|\sigma^2 - \mu| \le 1$, so that also $\sigma^2/\mu \le 2$ if $\sigma^2 \ge 2$, as is implied by (5.2) below. For random variables which are not "centered" in this way, the modification required merely translates W, which makes no essential difference; however, in the formulas for the bounds that we derive, the factorial moments and cumulants appearing are all for the "centered" random variables, and they may well be different from those of the original random variables.

In the spirit of the uniformity conditions of Corollary 4.2, we make the following assumptions:

$$\sigma_i^2 := \operatorname{Var} Z_i \ge 2 \quad \text{and} \quad 1 - d_{\mathrm{TV}}\big(\mathcal{L}(Z_i), \mathcal{L}(Z_i + 1)\big) \ge v_* \tag{5.2}$$

for all $1 \le i \le n$, where $0 < v_* \le 1/2$. If these conditions are not satisfied by the original random variables, they can usually be recovered by forming new random variables $\widetilde{Z}_i$ by adding successive Z_j's, correspondingly reducing n; Theorem 5.2, if applicable, may also give useful bounds. If these procedures do not work, it may well be the case that the approximation would genuinely be poor.

By analogy with the previous expansions, we assume that the random variables Z_i have finite $(r+1)$th moments. We then choose any real numbers $\tau_l > 0$, $0 \le l \le 4$, such that the following inequalities are satisfied:

$$\begin{gathered} \max\nolimits_{1 \le i \le n} \sigma_i^{-2} \mathbb{E}|Z_i - \mathbb{E}Z_i|^{r+1} \le \tau_0, \\ \max_{1 \le i \le n} \max_{0 \le m \le r} \frac{1}{\sigma_i^2} \left(\frac{|\kappa_{m+1}^{(i)}| \mu_{[r+1-m]}^{+(i)}}{m!(r+1-m)!} \right) \le \tau_1, \end{gathered} \tag{5.3}$$

where $\kappa_s^{(i)} = \kappa_s(Z_i)$ is as before and $\mu_{[s]}^{+(i)} := \mathbb{E}|Z_i(Z_i - 1)\cdots(Z_i - s + 1)|$; and also

$$\begin{gathered} \max\left\{1, \max_{1 \le i \le n} \max_{0 \le m \le r} \left(\frac{|\kappa_{m+1}^{(i)}|}{m!\sigma_i^2} \right) \right\} \le \tau_2, \qquad \max_{1 \le i \le n} \sigma_i^2 \le \tau_3, \\ \max_{1 \le i \le n} \sigma_i^{-2} \mathbb{E}\left\{ \prod_{l=0}^{r} |Z_i - l| \min(n^{1/2}, |Z_i - r - 1|) \right\} \le \tau_4. \end{gathered} \tag{5.4}$$

The τ_l merely act as a convenient shorthand, for instance, for defining some unpleasant constants, which appear in the error bounds:

$$
\begin{aligned}
&C_1 := 4(r+1)\tau_1\left(\frac{8r}{v_*}\right)^{r/2}, \\
&C_2 := 2r(r+1)\{4(r-1)\}^{2(r-1)}(\tau_2\tau_3)^{\lfloor(4r-2)/3\rfloor}, \\
&C_3 := 5(1+\tau_0)2^{(r+1)/2}R_{r+1}, \\
&C_4 := 2^{r+2}S^2(2\tau_2)^{\lfloor 2S/3\rfloor+1}\left(\frac{4S}{e}\right)^S, \\
&C_5 := 2^{r+1}\tau_2\left\{\frac{4}{\sqrt{v_*}} + \sqrt{\frac{8}{e}} + 2S\sqrt{3}\left(\frac{2S}{e}\right)^{S/2}(2\tau_2)^{r-1}\right\}, \\
&n_0 := 2^{2r+2}C_5^2, \\
&C_6 := 2 + S\left(\frac{2S}{e}\right)^{S/2}(2\tau_2)^{r-1}, \\
&C_7 := \left(\frac{8}{r!}\right)\left(\frac{8r}{v_*}\right)^{(r-1)/2}\max\left\{1, \left(\frac{8r}{v_*}\right)^{1/2}\right\},
\end{aligned}
\tag{5.5}
$$

where $S := \max\{1, 3(r-1)\}$ and R_s denotes the constant from the sth Rosenthal (1970) inequality [Petrov (1995), Theorem 2.9 and (2.35)]:

$$R_s \le \max\{s^s, s(es^{1/2})^s\{\Gamma(s/2)\}^2/\Gamma(s)\}. \tag{5.6}$$

We shall also need the quantities

$$d_{n,s} := \max_{1\le i\le n} \|\mathcal{L}(W_i) * (E_1 - E)^{*s}\|$$

for $1 \le s \le 4r$, when, as for (4.9), it follows from (3.3), (3.4) and (5.2) that

$$d_{n,s} \le \left\{\frac{2}{\sqrt{(\lfloor n/s\rfloor - 1)v_*}}\right\}^s \le \left(\frac{8s}{v_*}\right)^{s/2} n^{-s/2} \tag{5.7}$$

if $n \ge 16r$.

To state the main theorem, we need to define an approximating measure. We base our argument on the Stein equation (2.2) for suitably chosen λ_l. However, these choices need not be such that simple bounds like (2.17) on the corresponding $\|\pi_{(\boldsymbol{\lambda},\gamma)}\|$ are valid, making a direct SCP approximation awkward. To avoid such problems, we instead introduce a perturbation of the Poisson distribution $\mathrm{Po}(\mu)$ as our approximation.

To do so, first define the real numbers b_u, $u \geq 1$, to be the coefficients in the power series expansion

$$(5.8) \qquad B(z) := 1 + \sum_{u \geq 1} b_u z^u = \exp\left\{ \sum_{j=2}^{r+1} \frac{\kappa_j z^j}{j!} \right\},$$

where $\kappa_j = \kappa_j(W)$ as before, and let $C_u(j; a)$ denote the uth Charlier polynomial [Chihara (1978), (1.9), page 171]. Then set

$$(5.9) \qquad \tilde{\nu}_r\{j\} := \mathrm{Po}(\mu)\{j\} \left\{ 1 + \sum_{u=1}^{S} (-1)^u b_u C_u(j; \mu) \right\},$$

where $S = \max\{1, 3(r-1)\}$, as above.

The measure $\tilde{\nu}_r$, in general a signed measure, nonetheless has a number of good properties. First of all, $\tilde{\nu}_r(\mathbb{Z}_+) = 1$ and $\tilde{\nu}_r(j) = 0$ for all $j < 0$. Then it is shown in (5.26) that the b_u are at worst of order $O(\mu^{\lfloor u/3 \rfloor})$ as μ increases, while, in any region $\mu - C\sqrt{\mu} \leq j \leq \mu + C\sqrt{\mu}$, $|C_u(j; \mu)| = O(\mu^{-u/2})$, by Lemmas 6.1 and 6.2. Recall that, because of the "centering" of the random variables Z_i, the resulting mean μ satisfies $|\mu - \sigma^2| \leq 1$, so that, when interpreting these asymptotics, μ should be understood as being equivalent to $\mathrm{Var}\, W$. Thus $\tilde{\nu}_r$ is just a rather small perturbation of $\mathrm{Po}(\mu)$ in the region where the latter concentrates its mass; it is, indeed, a natural refinement of the centered Poisson approximation of Section 3 (the case $r = 1$: note that $b_1 = 0$). Finally, the measure $\tilde{\nu}_r$ is completely explicit, in the sense that everything is visible as a polynomial modification of the Poisson density in (5.9), except for the constants b_u, which are derived from the factorial cumulants of W using (5.8).

THEOREM 5.1. *If* Z_i, $1 \leq i \leq n$, *are independent integer-valued random variables with finite* $(r+1)$*th moments*, *which also satisfy* (5.1) *and* (5.2), *and if* $W = \sum_{i=1}^n Z_i$, *then*

$$\|\mathcal{L}(W) - \tilde{\nu}_r\| \leq 3Kn^{-r/2}(1 + \tau_4),$$

where

$$K := K(r, \tau_0, \tau_1, \tau_2, \tau_3, v_*) := \max\{C_1 + C_2 + C_3 + C_4, C_7, C_6(n_0)^{r/2}\}.$$

REMARK. (1) The way in which τ_4 enters the error estimate is highlighted because, even with independent and identically distributed random variables, it grows with n unless the $(r+2)$th moment is finite; see (5.4).

(2) For triangular arrays of integer random variables $(Z_i^{(n)},\ 1 \leq i \leq n,\ n \geq 1)$ in which, within each row, the random variables are independent and satisfy conditions (5.1) and (5.2) with $v_*^{(n)} = v_*$, and for which we can take $\tau_l^{(n)} = \tau_l$, $0 \leq l \leq 3$, uniformly for all n, Theorem 5.1 implies that $\|\mathcal{L}(W) - \tilde{\nu}_r\| \leq 3Kn^{-r/2}(1 + \tau_4^{(n)})$ for $K := K(r, \tau_0, \tau_1, \tau_2, \tau_3, v_*)$ as before. If, in addition, for some $0 < \delta \leq 1$,

$$\max_{1\le i\le n}(\sigma_i^{(n)})^{-2}\mathbb{E}\left\{|Z_i^{(n)}-r-1|^{\delta}\prod_{l=0}^{r}|Z_i^{(n)}-l|\right\}\le\tau_{4\delta}<\infty \tag{5.10}$$

for all n, it follows that $\|\mathcal{L}(W)-\tilde{\nu}_r\|\le 3K(1+\tau_{4\delta})n^{-(r-1+\delta)/2}$ for the same value of K, an error bound of order $O(n^{-(r-1+\delta)/2})$ under an $(r+1+\delta)$th moment assumption. In particular, for $\delta=1$, the error is actually of order $O(n^{-r/2})$.

PROOF. The first step in the proof is to show that $|\mathbb{E}(\mathcal{A}h)(W)|$ is suitably small for all bounded h: $\mathbb{Z}\to\mathbb{R}$, where $\mathcal{A}$ is a Stein operator of the general form $\mathcal{A}_{(\boldsymbol{\lambda},0)}$ — the "centering" has already been accounted for in (5.1) — with $t=r+1$ and with $\lambda_l=0$ for all $l<0$. We parametrize $\mathcal{A}$ in a different way, writing it as

$$(\mathcal{A}_r h)(j):=\sum_{m=0}^{r}\beta_m\Delta^m h(j+1)-jh(j) \tag{5.11}$$

for parameters β_m, $0\le m\le r$, to be chosen. We first show that $\beta_m:=\kappa_{m+1}/m!$ is a good choice. Note that β_1 is fixed, by the centering, to satisfy $|\beta_1|<1$, so that, including the centering, we are again fitting $r+1$ parameters to the first $r+1$ factorial cumulants.

As before, we shall use Newton's expansion, this time in the form

$$\begin{aligned} h(j+l)&=\sum_{s=0}^{m}\binom{l}{s}\Delta^s h(j)+\sum_{s=0}^{l-1}\binom{l-s-1}{m}\Delta^{m+1}h(j+s), && l\ge 0,\\ h(j+l)&=\sum_{s=0}^{m}\binom{l}{s}\Delta^s h(j)-\sum_{s=1}^{-l}\binom{l+s-1}{m}\Delta^{m+1}h(j-s), && l<0, \end{aligned} \tag{5.12}$$

where $\binom{0}{0}$ is taken to be 1. Thus, as in the previous proofs, we obtain

$$\left|\mathbb{E}\Delta^m h(W+1)-\sum_{s=0}^{r-m}\frac{\mu_{[s]}^{(i)}}{s!}\mathbb{E}\Delta^{m+s}h(W_i+1)\right|\le\frac{\mu_{[r-m+1]}^{+(i)}}{(r-m+1)!}\|\Delta h\|\,d_{n,r} \tag{5.13}$$

for any bounded h: $\mathbb{Z}\to\mathbb{R}$ and any $1\le i\le n$, $0\le m\le r$, where the symbol $\mu_{[s]}^{(i)}:=\mathbb{E}\{Z_i(Z_i-1)\cdots(Z_i-s+1)\}$ denotes the sth factorial moment of Z_i. Similarly, recalling the remark following Theorem 4.1,

$$\begin{aligned} &\left|\mathbb{E}\{Z_i h(W)\}-\sum_{j\in\mathbb{Z}}j\mathbb{P}[Z_i=j]\sum_{s=0}^{r}\binom{j-1}{s}\mathbb{E}\Delta^s h(W_i+1)\right|\\ &\quad=\left|\mathbb{E}\{Z_i h(W)\}-\sum_{s=0}^{r}\frac{\mu_{[s+1]}^{(i)}}{s!}\mathbb{E}\Delta^s h(W_i+1)\right|\\ &\quad\le\frac{1}{(r+1)!}\mathbb{E}\left\{\prod_{l=0}^{r}|Z_i-l|\min\bigl(2(r+1)d_{n,r-1},|Z_i-r-1|d_{n,r}\bigr)\right\}\|\Delta h\|. \end{aligned} \tag{5.14}$$

Combining these two estimates, we obtain

$$\left|\mathbb{E}\left\{\sum_{m=0}^{r}\beta_m^{(i)}\Delta^m h(W+1) - Z_i h(W)\right\}\right| \le \rho_{r,n,i} d_{n,r}\|\Delta h\|, \tag{5.15}$$

with

$$\begin{aligned}\rho_{r,n,i} = {} & \sum_{m=0}^{r}\frac{|\beta_m^{(i)}|\mu_{[r-m+1]}^{+(i)}}{(r-m+1)!} \\ & + \frac{1}{(r+1)!}\mathbb{E}\left\{\prod_{l=0}^{r}|Z_i - l|\min\bigl(2(r+1)\{d_{n,r-1}/d_{n,r}\}, |Z_i - r - 1|\bigr)\right\},\end{aligned} \tag{5.16}$$

if the $\beta_m^{(i)}$ satisfy the equations

$$\sum_{t=0}^{m}\frac{\beta_t^{(i)}\mu_{[m-t]}^{(i)}}{(m-t)!} = \frac{\mu_{[m+1]}^{(i)}}{m!}, \qquad 0 \le m \le r, \tag{5.17}$$

which is the case if $\beta_m^{(i)} = \kappa_{m+1}^{(i)}/m!$. Hence, by the additivity of the factorial cumulants, and from (5.3) and (5.7), the random variable W satisfies

$$\begin{aligned}|\mathbb{E}(\mathcal{A}_r h)(W)| &= \left|\mathbb{E}\left\{\sum_{m=0}^{r}\beta_m\Delta^m h(W+1) - Wh(W)\right\}\right| \\ &\le \bigl\{(r+1)\tau_1 + (2/r!)\max\{1, (8r/v_*)^{-1/2}\}\tau_4\bigr\} \\ &\quad \times (8r/v_*)^{r/2}\sigma^2 n^{-r/2}\|\Delta h\|,\end{aligned} \tag{5.18}$$

where $\beta_m = \kappa_{m+1}/m!$. Note, in particular, that $\beta_0 = \mu$ and that $|\beta_1| \le 1$.

If it were the case that

$$\theta' := \mu^{-1}\sum_{m=1}^{r} m2^{m-1}|\beta_m| < 1/2 \tag{5.19}$$

were satisfied, then (2.4) would be satisfied, and a perturbation argument as for Theorem 2.1 could be used to approximate $\mathcal{L}(W)$ by ν_r, the signed measure with Stein operator $\mathcal{A}_r$ of (5.11) having $\beta_m = \kappa_{m+1}/m!$; see Theorem 5.2. Our aim here is to show that an estimate of order $O(n^{-r/2}(1+\tau_4))$ is still valid, even if $\theta > 1/2$. Since we then have no control over the solutions of the Stein equation $\mathcal{A}_r h = f - \nu_r(f)$, and, consequently, little control over ν_r, we take a more indirect approach.

Instead of ν_r itself, we consider measures of the form

$$\tilde{\nu}(j) = \mathrm{Po}(\mu)\{j\}\left\{1 + \sum_{u\ge 1}(-1)^u b_u C_u(j;\mu)\right\}, \qquad j \in \mathbb{Z}_+, \tag{5.20}$$

to approximate $\mathcal{L}(W)$. The Charlier polynomials are a convenient choice for the expansion, because of the property that, if $Y \sim \mathrm{Po}(\mu)$ and $h\colon \mathbb{Z}_+ \to \mathbb{R}$ is polynomially bounded, then

$$\mathbb{E}\{(-1)^u C_u(Y;\mu)h(Y+1)\} = \mathbb{E}\{\Delta^u h(Y+1)\}; \tag{5.21}$$

it then also follows that $\tilde{\nu}(\mathbb{Z}_+) = 1$ and that

$$\mathbb{E}\{(-1)^u C_u(Y;\mu)Yh(Y)\} = \mathbb{E}\{\mu\Delta^u h(Y+1) + u\Delta^{u-1}h(Y+1)\}. \tag{5.22}$$

Applying these formulas and recalling that $\beta_0 = \mu$, it follows that

$$\begin{aligned}\tilde{\nu}\{\mathcal{A}_r h\} &:= \sum_{j\ge 0}\tilde{\nu}(j)(\mathcal{A}_r h)(j)\\ &= \mathbb{E}\left\{\sum_{u\ge 0} b_u\left(\sum_{t=1}^{r}\beta_t\Delta^{u+t}h(Y+1) - u\Delta^{u-1}h(Y+1)\right)\right\},\end{aligned} \tag{5.23}$$

where we have defined $b_0 = 1$. The coefficient of $\Delta^s h(Y+1)$ in the sum is given by

$$-(s+1)b_{s+1} + \sum_{t=1}^{s} b_{s-t}\beta_t \tag{5.24}$$

if we define $\beta_t = 0$ for $t > r$. Thus, for any $S \ge 1$, the coefficients of $\Delta^s h(Y+1)$, $0 \le s \le S-1$, are all zero if we take b_t, $1 \le t \le S$, to be successively defined by the relation

$$(s+1)b_{s+1} = \sum_{t=1}^{s} b_{s-t}\beta_t, \qquad 0 \le s \le S-1, \tag{5.25}$$

that is, as defined by (5.8). In particular, $b_1 = 0$, $2b_2 = \beta_1 = \kappa_2$, and since, from (5.4), $\max_{1\le t\le r}|\beta_t| \le \tau_2\sigma^2$ and $|\kappa_2| \le 1$, it follows inductively from (5.25) that

$$|b_{s+1}| \le \frac{1}{s+1}\left(\sum_{t=2}^{s\wedge r}|b_{s-t}|\tau_2\sigma^2 + |b_{s-1}|\right) \le (\tau_2\sigma^2)^{\lfloor (s+1)/3\rfloor}, \qquad s \ge 0. \tag{5.26}$$

If $\tilde{\nu}_{r,S}$ denotes the measure of the form (5.20) with this choice of b_s, $1 \le s \le S$, and with $b_u = 0$ for $u > S$, then it follows from (5.23), (5.24) and (5.26) that

$$
\begin{aligned}
|\tilde{\nu}_{r,S}(\mathcal{A}_r h)| &\le \sum_{s=S}^{S+r}\sum_{t=1}^{s} |b_{s-t}\beta_t|\,|\mathbb{E}\Delta^s h(Y+1)| \\
&\le \sum_{s=S}^{S+r}\left\{(\tau_2\sigma^2)^{\lfloor (s-1)/3\rfloor} + \sum_{t=2}^{r}(\tau_2\sigma^2)^{1+\lfloor (s-t)/3\rfloor}\right\} \\
&\quad \times \left(\frac{s-1}{\mu}\right)^{(s-1)/2}\|\Delta h\| \\
&\le r(S+r-1)^{(S+r-1)/2}(\tau_2\tau_3)^{\lfloor (S+r+1)/3\rfloor} \\
&\quad \times \left(\sum_{s=S}^{S+r} n^{(1-s)/2+\lfloor (s+1)/3\rfloor}\right)\|\Delta h\|,
\end{aligned}
\tag{5.27}
$$

provided that $n \ge 3$, where we have used Lemma 6.2(1) for the bounds on $|\mathbb{E}\Delta^s h(Y+1)|$ as well as the inequalities $\mu^{-1} \le 2\sigma^{-2} \le n^{-1}$ and $\sigma^2 \le n\tau_3$. Taking $S = \max\{1, 3(r-1)\}$ and writing $\tilde{\nu}_r = \tilde{\nu}_{r,S}$, this gives

$$
|\tilde{\nu}_r(\mathcal{A}_r h)| \le 16^{r-1} r(r+1)(r-1)^{2(r-1)}(\tau_2\tau_3)^{\lfloor (4r-2)/3\rfloor} n^{1-r/2}\|\Delta h\|. \tag{5.28}
$$

Thus, through (5.18) and (5.28), we have shown that both $|\mathbb{E}(\mathcal{A}_r h)(W)|$ and $|\tilde{\nu}_r(\mathcal{A}_r h)|$ are of order $n^{1-r/2}\|\Delta h\|$, under "typical" conditions.

The usual Stein argument would now move by way of the solutions to the equation $\mathcal{A}_r h = f - \tilde{\nu}_r\{f\}$, but without the condition $\theta < 1/2$ we have no control over them. We avoid the difficulty by instead taking for h the solution to the equation

$$
\begin{aligned}
(\mathcal{A}_1 h)(j) &:= \sum_{m=0}^{1} \beta_m \Delta^m h(j+1) - jh(j) \\
&= \kappa_2 h(j+2) + (\mu - \kappa_2)h(j+1) - jh(j) \\
&= f(j) - \nu_1(f), \qquad j \ge 0,
\end{aligned}
\tag{5.29}
$$

with $h(j) = 0$ for $j \le 0$, for which $\theta = |\kappa_2|/\mu < 1/2$ in $\mu > 2$, and hence

$$
\begin{aligned}
\|h\| &\le \frac{2}{1 - 2\mu^{-1}|\kappa_2|}\mu^{-1/2}\|f\|, \\
\|\Delta h\| &\le \frac{2}{\mu - 2|\kappa_2|}\|f\| \le 4\sigma^{-2}\|f\| \quad \text{in } n \ge 3,
\end{aligned}
\tag{5.30}
$$

because of assumption (5.2). Thus, for $f\colon \mathbb{Z} \to \mathbb{R}$ bounded and with $f(j) = 0$ for $j < 0$, we can compute

$$
\begin{aligned}
\mathbb{E}\{(\mathcal{A}_r h)(W)\} &= \kappa_2 \mathbb{P}[W = -1]h(1) + \sum_{j \ge 0} \mathbb{P}[W = j]\{f(j) - \nu_1(f)\} \\
&\quad + \sum_{m=2}^{r} \beta_m \mathbb{E}\Delta^m h(W+1),
\end{aligned}
$$

$$\tilde{\nu}_r(\mathcal{A}_r h) = \sum_{j\ge 0} \tilde{\nu}_r\{j\}\{f(j) - \nu_1(f)\} + \sum_{m=2}^{r} \beta_m \tilde{\nu}_r\big(\Delta^m h(\cdot + 1)\big),$$

and hence

$$\begin{aligned} &\mathbb{E}\{(\mathcal{A}_r h)(W)\} - \tilde{\nu}_r(\mathcal{A}_r h) \\ &\quad = \mathbb{E} f(W) - \tilde{\nu}_r(f) + \nu_1(f)\mathbb{P}[W < 0] + \kappa_2 \mathbb{P}[W = -1]\Delta h(0) \\ &\qquad + \sum_{m=2}^{r} \beta_m \big\{\mathbb{E}\Delta^m h(W+1) - \tilde{\nu}_r\big(\Delta^m h(\cdot + 1)\big)\big\}. \end{aligned} \tag{5.31}$$

On the other hand, from (5.18), (5.28) and (5.30),

$$|\mathbb{E}\{(\mathcal{A}_r h)(W)\} - \tilde{\nu}_r(\mathcal{A}_r h)| \le (C_1 + C_7\tau_4 + C_2)n^{-r/2}\|f\|, \tag{5.32}$$

giving

$$\begin{aligned} |\mathbb{E} f(W) - \tilde{\nu}_r(f)| &\le (C_1 + C_7\tau_4 + C_2)n^{-r/2} + (\|\nu_1\| + 2)\mathbb{P}[W < 0] \\ &\qquad + 4\tau_2 \sum_{m=2}^{r} \big\|\big(\mathcal{L}(W) - \tilde{\nu}_r\big) * (E_1 - E)^{*(m-1)}\big\| \end{aligned} \tag{5.33}$$

for all f such that $\|f\| \le 1$. From an $(r+1)$th moment Chebyshev inequality and Rosenthal's inequality, we deduce that

$$\begin{aligned} \mathbb{P}[W < 0] &\le \mu^{-(r+1)}\mathbb{E}|W - \mu|^{r+1} \\ &\le R_{r+1}\mu^{-(r+1)}\left\{\sigma^{r+1} + \sum_{i=1}^{n} \mathbb{E}|Z_i - \mathbb{E}Z_i|^{r+1}\right\} \\ &\le 2^{r+1}R_{r+1}(1+\tau_0)\sigma^{-(r+1)} \\ &\le 2^{(1+r)/2}R_{r+1}(1+\tau_0)n^{-(1+r)/2} \end{aligned} \tag{5.34}$$

in view of (5.1)–(5.3) and (5.5), whereas $\|\nu_1\| \le (1 - 2|\kappa_2|/\mu)^{-1} \le 3$ in $n \ge 3$ from Barbour and Xia (1999), (2.17). Hence we conclude from (5.33) that

$$\begin{aligned} \|\mathcal{L}(W) - \tilde{\nu}_r\| &\le (C_1 + C_7\tau_4 + C_2 + C_3)n^{-r/2} \\ &\qquad + 2^{r+1}\tau_2\big\|\big(\mathcal{L}(W) - \tilde{\nu}_r\big) * (E_1 - E)\big\|. \end{aligned} \tag{5.35}$$

To bound the final term in (5.35), we express it in terms of the approximation error for a similar problem with a smaller value of n, and use induction. So let $\tilde{\nu}_r^{(1)}$ and $\tilde{\nu}_r^{(2)}$ be defined in the same way as $\tilde{\nu}_r$, but using the sets of random variables $(Z_i,\ 1 \le i \le \lfloor n/2 \rfloor)$ and $(Z_i, \lfloor n/2\rfloor + 1 \le i \le n)$, respectively, whose sums are denoted by $W^{(1)}$ and $W^{(2)}$. Then $\kappa_1^{(1)} + \kappa_1^{(2)} = \mu$ and $\kappa_l^{(1)} + \kappa_l^{(2)} = \kappa_l$ for $2 \le l \le r$, and also $|\kappa_2^{(1)}|, |\kappa_2^{(2)}| < 1$. Note that, for each of these sets of random variables, the quantities τ_l, $0 \le l \le 4$, and v_* can be left unchanged.

Now, since

$$\sum_{k=0}^{j} \frac{s^k}{k!} \frac{t^{j-k}}{(j-k)!} C_l(k; s) C_m(j-k; t) = \frac{(s+t)^j}{j!} C_{l+m}(j; s+t),$$

and since $b_u = \sum_{m=0}^{u} b_m^{(1)} b_{u-m}^{(2)}$ because of (5.8) and the additivity of (factorial) cumulants, it follows that

$$\begin{aligned} \|\tilde{\nu}_r^{(1)} * \tilde{\nu}_r^{(2)} - \tilde{\nu}_r\| &\le \sum_{j\ge 0} \mathrm{Po}(\mu)\{j\} \sum_{u=S+1}^{2S} \sum_{m=0}^{u} |b_m^{(1)} b_{u-m}^{(2)} C_u(j; \mu)| \\ &\le \sum_{u=S+1}^{2S} 2S(\tau_2\sigma^2)^{\lfloor u/3 \rfloor} \mathbb{E}|C_u(Y; \mu)|, \end{aligned} \tag{5.36}$$

where $Y \sim \mathrm{Po}(\mu)$. Since, by Lemma 6.2(2), $\mathbb{E}|C_u(Y; \mu)| \le (2u/e\mu)^{u/2}$, it thus follows that

$$\|\tilde{\nu}_r^{(1)} * \tilde{\nu}_r^{(2)} - \tilde{\nu}_r\| \le 2S^2 (2\tau_2)^{\lfloor 2S/3 \rfloor} (4S/e)^S n^{-r/2}, \tag{5.37}$$

once again using $\sigma^2/\mu \le 2$. Hence the measure $\tilde{\nu}_r$ is close to the convolution $\tilde{\nu}_r^{(1)} * \tilde{\nu}_r^{(2)}$, and, of course, $\mathcal{L}(W) = \mathcal{L}(W^{(1)}) * \mathcal{L}(W^{(2)})$, so that the approximation of the distributions $\mathcal{L}(W^{(1)})$ and $\mathcal{L}(W^{(2)})$ by $\tilde{\nu}_r^{(1)}$ and $\tilde{\nu}_r^{(2)}$ can be used to show the closeness of $\mathcal{L}(W)$ and $\tilde{\nu}_r$.

In view of (5.35), what we actually need to bound is

$$\begin{aligned} \|(\mathcal{L}(W) - \tilde{\nu}_r) * (E_1 - E)\| \le{}& \|\mathcal{L}(W^{(1)}) - \tilde{\nu}_r^{(1)}\| \, \|\mathcal{L}(W^{(2)}) * (E_1 - E)\| \\ &+ \|\mathcal{L}(W^{(2)}) - \tilde{\nu}_r^{(2)}\| \, \|\tilde{\nu}_r^{(1)} * (E_1 - E)\| \\ &+ 2\|\tilde{\nu}_r - \tilde{\nu}_r^{(1)} * \tilde{\nu}_r^{(2)}\|. \end{aligned} \tag{5.38}$$

Now the last term is covered by (5.37), and $\|\mathcal{L}(W^{(2)}) * (E_1 - E)\| \le 4\{nv_*\}^{-1/2}$, as in (3.4). Furthermore,

$$\|\tilde{\nu}_r^{(1)} * (E_1 - E)\| \le 2\|\tilde{\nu}_r^{(1)} - \mathrm{Po}(\mu^{(1)})\| + \|\mathrm{Po}(\mu^{(1)}) * (E_1 - E)\|, \tag{5.39}$$

where the latter term is at most $2 \max_j \mathrm{Po}(\mu^{(1)})\{j\} \le \sqrt{(8/en)}$ and

$$\|\tilde{\nu}_r^{(1)} - \mathrm{Po}(\mu^{(1)})\| \le \sum_{u=2}^{S} |b_u^{(1)}| \mathbb{E}|C_u(Y^{(1)}; \mu^{(1)})| \tag{5.40}$$

for $Y^{(1)} \sim \mathrm{Po}(\mu^{(1)})$; bounding the elements in the latter sum using (5.26) and Lemma 6.2, we get

$$\|\tilde{\nu}_r^{(1)} - \mathrm{Po}(\mu^{(1)})\| \le S\sqrt{3}(2S/e)^{S/2}(2\tau_2)^{r-1} n^{-1/2} \tag{5.41}$$

for $n \ge 3$, since then $\mu^{(1)} \ge n/3 \ge 1$.

Thus we have expressions to cope with all the elements of (5.38), except for the factors $\|\mathcal{L}(W^{(l)}) - \tilde{\nu}_r^{(l)}\|$, $l = 1, 2$, for which we use the induction hypothesis to conclude that

$$\begin{aligned} \|\mathcal{L}(W^{(l)}) - \tilde{\nu}_r^{(l)}\| &\le 3K(1+\tau_4)\lfloor n/2\rfloor^{-r/2} \\ &\le 3 \times 2^r K(1+\tau_4)n^{-r/2}, \qquad l = 1, 2. \end{aligned} \tag{5.42}$$

Combining (5.35) with (5.37)–(5.42), we have shown that

$$\begin{aligned} \|\mathcal{L}(W) - \tilde{\nu}_r\| &\le (C_1 + C_7\tau_4 + C_2 + C_3 + C_4)n^{-r/2} \\ &\quad + 3 \times 2^r K(1+\tau_4)n^{-r/2}C_5 n^{-1/2} \\ &\le 3K(1+\tau_4)n^{-r/2} \end{aligned} \tag{5.43}$$

if $3 \times 2^r C_5 n^{-1/2} \le 2$, which is the case for $n \ge n_0 := 2^{2r+2}C_5^2$. On the other hand, for smaller n, the argument of (5.40) and (5.41) applied to $\tilde{\nu}_r$ shows that

$$\|\mathcal{L}(W) - \tilde{\nu}_r\| \le 2 + S(2S/e)^{S/2}(2\tau_2)^{r-1}n^{-1/2} \le C_6,$$

and the statement of the theorem is now immediate. □

Although the statement of the theorem is quite explicit, it cannot be claimed that the estimates given are likely to be very precise. On the other hand, the influence of the two key ingredients, the moments of the summands and the extent to which they avoid being lattice with some span larger than 1, can be clearly seen in the bounds.

If condition (5.2) is not satisfied, Theorem 5.1 cannot be applied directly. However, as noted at (5.19), if $\mu > 0$ and

$$\theta' := \mu^{-1}\sum_{m=1}^{r} m2^{m-1}|\kappa_{m+1}|/m! < 1/2 \tag{5.44}$$

is satisfied, a perturbation argument along the lines of Theorem 2.1 can be used, now approximating $\mathcal{L}(W)$ by the SCP measure ν_r which has the Stein operator $\mathcal{A}_r$ of (5.11) for $\beta_m = \kappa_{m+1}/m!$; this means that the nonzero λ_l's are given by

$$\lambda_{l+1} := \frac{1}{l+1}\sum_{m=l}^{r}(-1)^{m-l}\binom{m}{l}\frac{\kappa_{m+1}}{m!}, \qquad 0 \le l \le r. \tag{5.45}$$

Before formulating the theorem, we derive a bound for $d_{n,s}$ to replace that in (5.7), since v_* may now be zero. Let v_i, v^* and V be as in (3.3) and (3.4), and split $V - v_i$ into s partial sums

$$\sum_{\substack{l=m_j+1 \\ l\ne i}}^{m_{j+1}} v_l, \qquad 0 \le j \le s-1,$$

each not smaller than $s^{-1}V - v^*$; this can be done just by taking $m_0 = 0$ and

$$m_{j+1} := \min\Biggl\{m > m_j : \sum_{\substack{l=m_j+1 \\ l \neq i}}^{m} v_l \geq s^{-1}V - v^*\Biggr\}.$$

Then, applying Proposition 4.6 of Barbour and Xia (1999), it follows that

$$\|\mathcal{L}(W_i) * (E_1 - E)^{*s}\| \leq 2^s\{1 \wedge (s^{-1}V - v^*)_+^{-s/2}\}.$$

Using $1 \wedge (x - \frac{1}{2})^{-1} \leq 1 \wedge (3x^{-1}/2)$, we thus obtain

$$d_{n,s} \leq w_r^s, \qquad 1 \leq s \leq r, \tag{5.46}$$

where $w_r := 2 \wedge \{6r/V\}^{1/2}$ and V is as for (3.4).

THEOREM 5.2. *If* $W = \sum_{i=1}^n Z_i$, *where* Z_i, $1 \leq i \leq n$, *are independent integer-valued random variables with finite* $(r+1)$*th moments*, *and if* $\mathbb{E}W = \mu > 0$ *and* (5.44) *holds*, *then*

$$\|\mathcal{L}(W) - \nu_r\| \leq \left(\frac{2}{1 - 2\theta'}\right)\Biggl\{(1 \wedge \mu^{-1}) w_r^r \sum_{i=1}^n \rho'_{r,n,i} + \mathbb{P}[W < 0]\Biggr\}, \tag{5.47}$$

where ν_r *is the SCP with* λ_l*'s defined in* (5.45), w_r *is as for* (5.46) *and*

$$\rho'_{r,n,i} := \sum_{m=0}^{r} \frac{|\kappa_{m+1}^{(i)}| \mu_{[r+1-m]}^{+(i)}}{m!(r+1-m)!} + \frac{1}{(r+1)!}\mathbb{E}\Biggl\{\left(\prod_{l=0}^{r} |Z_i - l|\right) \min\{2(r+1)w_r^{-1}, |Z_i - r - 1|\}\Biggr\}.$$

In general,

$$\mathbb{P}[W < 0] \leq \left(\frac{3}{2}\right)^{r+1} R_{r+1}\Biggl\{\sigma^{-(r+1)} + \sigma^{-2(r+1)} \sum_{i=1}^n \mathbb{E}|Z_i - \mathbb{E}Z_i|^{r+1}\Biggr\}, \tag{5.48}$$

where R_{r+1} *is the* $(r+1)$*th Rosenthal constant*.

PROOF. Using (5.46) in place of (5.7), it follows as for (5.15) that

$$|\mathbb{E}(\mathcal{A}_r h)(W)| \leq w_r^r \|\Delta h\| \sum_{i=1}^n \rho'_{r,n,i}.$$

Then, because $\theta \leq \theta'$ and (5.44) is satisfied, Corollary 2.3 can be applied, giving (5.47). To bound $\mathbb{P}[W < 0]$, note that $|\kappa_2| \leq \mu/2$ because $\theta < 1/2$, and hence that $\mu \geq 2\sigma^2/3$; (5.48) now follows from (5.34). □

In particular, for $Z_i \sim \mathrm{Be}(p_i)$, we have $\mathbb{P}[W < 0] = 0$ and

$$\beta_m^{(i)} = (-p_i)^{m+1}, \qquad \rho'_{r,n,i} = p_i^{r+2}, \qquad V = \sum_{i=1}^{n} \{p_i \wedge (1 - p_i)\} \geq \sigma^2. \tag{5.49}$$

The condition $\theta < 1/2$ is satisfied, for instance, if $\max_{1 \leq i \leq n} p_i \leq 3/16$. If this is the case, then we obtain a bound on $\|\mathcal{L}(W) - \nu_r\|$ of order

$$O\left\{(1 \wedge \sigma^{-(r+2)}) \sum_{i=1}^{n} p_i^{r+2}\right\}.$$

For $r = 0$, this is just the usual bound for Poisson approximation. For $r = 1$ and equal p_i's, the bound is the same as in Presman (1983), Proposition 1.

6. Auxiliary results. Let the Charlier polynomials $C_m(j; \mu)$ be defined as in Chihara (1978), (1.9), page 171.

LEMMA 6.1. *If* $\mu \geq 1$, *then*

$$|C_m(j; \mu)| \leq \{|1 - j/\mu| + m/\sqrt{\mu}\}^m \leq 2^{m-1}\{|1 - j/\mu|^m + (m/\sqrt{\mu})^m\}.$$

PROOF. By induction. The claim is true for $m = 0, 1$. For general m, use the recurrence relation for Charlier polynomials, as in Chihara [(1978), (1.4), page 170], together with the equation

$$C_m(j; \mu) - C_{m-1}(j; \mu) = -j\mu^{-1}C_{m-1}(j - 1; \mu),$$

to give

$$|C_{m+1}(j; \mu)| \leq |1 - j/\mu|\,|C_m(j; \mu)| + jm\mu^{-2}|C_{m-1}(j - 1; \mu)|,$$

and then apply the inequalities

$$|1 - (j - 1)/\mu| + (m - 1)/\sqrt{\mu} \leq |1 - j/\mu| + m/\sqrt{\mu},$$

$$jm\mu^{-2} \leq (m/\sqrt{\mu})\{|1 - j/\mu| + m/\sqrt{\mu}\}.$$ □

LEMMA 6.2. *For any* $m \geq 1$, *we have*

(1) $\|\mathrm{Po}(\mu) * (E_1 - E)^{*m}\| \leq \left(\dfrac{2m}{e\mu}\right)^{m/2}$,

(2) $\mathrm{Po}(\mu)\{|C_m(\cdot; \mu)|\} \leq \left(\dfrac{2m}{e\mu}\right)^{m/2}$.

PROOF. For part (1), observe that

$$\|\operatorname{Po}(\mu) * (E_1 - E)^{*m}\| \le \left\|\operatorname{Po}\left(\frac{\mu}{m}\right) * (E_1 - E)\right\|^m$$
$$\le \left(2 \max_{j \ge 0} \operatorname{Po}\left(\frac{\mu}{m}\right)\{j\}\right)^m \le \left(\frac{2m}{e\mu}\right)^{m/2},$$

by Barbour, Holst and Janson (1992), Proposition A.2.7. For part (2), take $u = m$ and $h(j+1) := (-1)^m \operatorname{sgn}\{C_m(j;\mu)\}$ in (5.21), which, with part (1), completes the proof. □

The remainder of this section discusses how to bound $\eta(\boldsymbol{\lambda})$ for use with Corollary 2.3, when it is not the case that $\lambda_l = 0$ whenever $l < 0$.

A first observation is that, if $\lambda > 0$ and $\theta < 1/2$, then

$$\begin{aligned}\lambda_1 + 2\lambda_2 &\ge 2 \sum_{l \in L_t} l(l-1)|\lambda_l| - \sum_{l \in L_t'} |l\lambda_l| \\ &= 4|\lambda_2| + \sum_{l \in L_t'} |l|\{2|l-1| - 1\}|\lambda_l|,\end{aligned} \tag{6.1}$$

where $L_t' := L_t \setminus \{2\}$. If $\lambda_2 > 0$, then (6.1) implies that

$$\lambda_1 \ge 2\lambda_2 + 3 \sum_{l \in L_t'} |l\lambda_l| > 0,$$

and, because

$$|\lambda - \lambda_1 - 2\lambda_2| \le \sum_{l \in L_t'} |l\lambda_l|, \tag{6.2}$$

it follows that

$$2\lambda_1/3 \le \lambda \le 2\lambda_1. \tag{6.3}$$

If $\lambda_2 \le 0$, (6.1) implies that $\lambda_1 \ge 6|\lambda_2| + 3\sum_{l \in L_t'} |l\lambda_l|$, and (6.2) then gives

$$2\lambda_1/3 \le \lambda \le 4\lambda_1/3. \tag{6.4}$$

Thus the condition $\theta < 1/2$ always implies that λ_1 is positive, and accounts for a substantial fraction of λ. A variant on this theme is given in the following lemma.

LEMMA 6.3. *If $\lambda > 0$ and $\theta \le 1/K$ for any $K \ge 2$, then*

$$(K-1) \sum_{l \in L_t} |\lambda_l| \le \lambda_1, \qquad (K-1) \sum_{l \in L_t} |l\lambda_l| \le \lambda_1, \qquad \tfrac{1}{2}(K-1) \sum_{l \in L_t} l^2 |\lambda_l| \le \lambda_1.$$

PROOF. The definition of θ gives

$$K \sum_{l \in L_t} l(l-1)|\lambda_l| \le \lambda_1 + \sum_{l \in L_t} l\lambda_l,$$

implying, in particular, that

$$\lambda_1 \ge \sum_{l \in L_t} \{Kl(l-1) - |l|\}|\lambda_l|.$$

The proof is now immediate. □

Lemma 6.3 suggests that, for small enough θ, the measure $\pi_{(\boldsymbol{\lambda},\gamma)}$ should be close enough to $\mathrm{Po}(\lambda_1)$ to make $\eta(\boldsymbol{\lambda})$ exponentially small with λ_1. To show this, we start by letting Y denote a random variable with a compound Poisson distribution, having probability generating function

$$\mathbb{E}\{z^Y\} = \exp\left\{ \sum_{1 \le |l| \le t} |\lambda_l|(z^l - 1) \right\}. \tag{6.5}$$

In the next result, we bound $|\pi_{(\boldsymbol{\lambda},\gamma)}|$ using the distribution of Y.

LEMMA 6.4. *For all* $k \in \mathbb{Z}$,

$$|\pi_{(\boldsymbol{\lambda},\gamma)}(k)| \le \exp\left\{ \sum_{1 \le |l| \le t} (|\lambda_l| - \lambda_l) \right\} \mathbb{P}[Y = k].$$

PROOF. The coefficient C_k of z^k in the expansion of the exponential generating function $\exp\{\sum_{1 \le |l| \le t} |\lambda_l| z^l\}$ is clearly at least as big in modulus as C'_k from the generating function $\exp\{\sum_{1 \le |l| \le t} \lambda_l z^l\}$, and $\pi_{(\boldsymbol{\lambda},\gamma)}(k) = C'_k \exp\{-\sum_{1 \le |l| \le t} \lambda_l\}$, whereas $\mathbb{P}[Y = k] = C_k \exp\{-\sum_{1 \le |l| \le t} |\lambda_l|\}$. □

We now need bounds for the distribution of Y. To state them, set

$$\Lambda_1 := \mathbb{E}Y = \sum_{1 \le |l| \le t} l|\lambda_l|, \qquad \Lambda_2 := \lambda_1 + 2 \sum_{l \in L_t} l^2 |\lambda_l|. \tag{6.6}$$

LEMMA 6.5. *For t as in the definition of* $\boldsymbol{\lambda}$, *if* $\lambda_1 \ge 0$, *we have*

$$\mathbb{P}[Y < 0] \le \exp\left\{ -\frac{1}{2t}\left(\Lambda_1 - \frac{\Lambda_2}{4t} \right) \right\}.$$

PROOF. For any $\phi > 0$, we have

$$\begin{aligned}\mathbb{P}[Y < 0] &\le \mathbb{E}(e^{-\phi Y})\\ &= \exp\biggl\{-\Lambda_1\phi + \lambda_1(e^{-\phi} - 1 + \phi) + \sum_{l \in L_t} |\lambda_l|(e^{-l\phi} - 1 + l\phi)\biggr\}\\ &\le \exp\{-\Lambda_1\phi + \tfrac{1}{2}\Lambda_2\phi^2\},\end{aligned}$$

provided that $e^{t\phi} \le 2$. Now take $\phi = 1/2t$. □

Lemmas 6.4 and 6.5 can often be combined directly to show that $\eta(\boldsymbol{\lambda})$ is exponentially small. In the next lemma, we demonstrate this under two different sets of assumptions.

LEMMA 6.6. *Let t be as in the definition of $\boldsymbol{\lambda}$.*

(1) *If $\lambda > 0$ and $\theta \le 1/(8t+1)$, then $\eta(\boldsymbol{\lambda}) \le \exp\{-\lambda_1(3 - 2t^{-1})/8t\}$.*
(2) *If $\lambda_l \ge 0$ for all l, $\lambda > 0$ and $\theta \le 1/2$, then $\eta(\boldsymbol{\lambda}) \le \exp\{-\lambda/(8t)\}$.*

PROOF. From Lemma 6.3, it follows that

$$\begin{gathered}\sum_{1 \le |l| \le t} (|\lambda_l| - \lambda_l) = 2\sum_{\substack{l \in L_t\\ \lambda_l < 0}} |\lambda_l| \le \lambda_1/8t,\\ \Lambda_1 \ge \lambda_1 - \sum_{l \in L_t} |l\lambda_l| \ge \lambda_1(1 - 1/8t), \qquad \Lambda_2 \le \lambda_1(1 + 1/2t).\end{gathered} \tag{6.7}$$

Then, from Lemmas 6.4 and 6.5, we have

$$\eta(\boldsymbol{\lambda}) \le \exp\biggl\{\sum_{1 \le |l| \le t} (|\lambda_l| - \lambda_l)\biggr\} \exp\biggl\{-\frac{1}{2t}\biggl(\Lambda_1 - \frac{\Lambda_2}{4t}\biggr)\biggr\},$$

and part (1) of the lemma is obtained by substituting from (6.7).

Under the conditions of part (2), observe that $\Lambda_1 = \lambda$ and

$$\Lambda_2 = \lambda_1 + 2\lambda\theta + 2(\lambda - \lambda_1) = 2\lambda(1 + \theta) - \lambda_1 \le 3\lambda,$$

by (6.4). Hence

$$\Lambda_1 - \Lambda_2/(4t) \ge \lambda\bigl(1 - 3/(4t)\bigr) \ge \lambda/4,$$

and Lemma 6.5 completes the proof. □

LEMMA 6.7. *Let Z be an integer-valued random variable having $\sigma^2 := \operatorname{Var} Z < \infty$ and with $\mu := \mathbb{E}Z$ satisfying $|\mu| \le 1/2$. Then*

$$|\mu| \le \mathbb{E}|Z| \le 2\sigma^2.$$

PROOF. Since Z is integer valued, we have

$$|\mu| \le \mathbb{E}|Z| \le \mathbb{E}Z^2 = \sigma^2 + \mu^2 \le \sigma^2 + \tfrac{1}{2}|\mu|.$$

This shows first that $|\mu| \le 2\sigma^2$ and then that $\mathbb{E}|Z| \le 2\sigma^2$ also. □

Acknowledgment. The authors thank the referee for the unusual amount of work invested in our paper, both in the original and in the revision, and for the numerous constructive suggestions, including his elegant proofs of Lemma 6.2(2) and Corollary 4.5. These have greatly improved the organization and transparency of the results.

REFERENCES

ARAK, T. V. (1981). On the convergence rate in Kolmogorov's uniform limit theorem. *Theory Probab. Appl.* **26** 219–239, 437–451.

ARAK, T. V. and ZAITSEV, A. Yu. (1988). Uniform limit theorems for sums of independent random variables. *Proc. Steklov Inst. Math.* **174**.

BARBOUR, A. D., CHEN, L. H. Y. and LOH, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation.* Oxford Univ. Press.

BARBOUR, A. D. and UTEV, S. (1999). Compound Poisson approximation in total variation. *Stochastic Process. Appl.* **82** 89–125.

BARBOUR, A. D. and XIA, A. (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3** 131–150.

BOROVKOV, K. A. and PFEIFER, D. (1996). On improvements of the order of approximation in the Poisson limit theorem. *J. Appl. Probab.* **33** 146–155.

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

ČEKANAVIČIUS, V. (1998). Poisson approximations for sequences of random variables. *Statist. Probab. Lett.* **39** 101–107.

ČEKANAVIČIUS, V. and VAITKUS, P. (1999). A centred Poisson approximation via Stein's method. Preprint 99–22, Dept. Math. Informatics, Vilnius Univ.

ČEKANAVIČIUS, V. and MIKALAUSKAS, M. (1999). Signed Poisson approximations for Markov chains. *Stochastic Process. Appl.* **82** 205–227.

CHIHARA, T. S. (1978). *An Introduction to Orthogonal Polynomials.* Gordon and Breach, New York.

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

IBRAGIMOV, I. A. and LINNIK, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Grøningen.

KENDALL, M. G. and STUART, A. (1963). *The Advanced Theory of Statistics* **1**, 2nd ed. Griffin, London.

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

MESHALKIN, L. D. (1961). On the approximation of distributions of sums by infinitely divisible laws. *Teor. Veroyatnost. i Primenen.* **6** 257–275.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

PETROV, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford Univ. Press.

PIPIRAS, V. (1970). Asymptotic expansions for distribution functions of sums of independent lattice random variables. *Litovsk. Mat. Sb.* **10** 517–536 (in Russian).

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

ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303.

TSAREGRADSKII, I. P. (1958). On uniform approximation of the binomial distribution by infinitely divisible distributions. *Theory Probab. Appl.* **3** 434–438.

ZAITSEV, A. Yu. (1991). An example of a distribution whose set of n-fold convolutions is uniformly separated from the set of infinitely divisible laws in the sense of the variation distance. *Theory Probab. Appl.* **36** 419–425.

ABTEILUNG ANGEWANDTE MATHEMATIK
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT ZÜRICH
CH-8057 ZÜRICH
SWITZERLAND
E-MAIL: adb@amath.unizh.ch

DEPARTMENT OF MATHEMATICAL STATISTICS
VILNIUS UNIVERSITY
NAUGARDUKO 24
VILNIUS 2600
LITHUANIA