Journal of Theoretical Probability
https://doi.org/10.1007/s10959-023-01310-4

On Poisson Approximation

S. Y. Novak[1]

Received: 9 June 2023 / Revised: 3 November 2023 / Accepted: 1 December 2023

Abstract
The problem of evaluating the accuracy of Poisson approximation to the distribution of a sum of independent integer-valued random variables has attracted a lot of attention in the past six decades. Among authors who contributed to the topic are Prokhorov, Kolmogorov, LeCam, Shorgin, Barbour, Hall, Deheuvels, Pfeifer, Roos, and many others. From a practical point of view, the problem has important applications in insurance, reliability theory, extreme value theory, etc. From a theoretical point of view, the topic provides insights into Kolmogorov's problem concerning the accuracy of approximation of the distribution of a sum of independent random variables by infinitely divisible laws. The task of establishing an estimate of the accuracy of Poisson approximation with a correct (the best possible) constant at the leading term remained open for decades. We present a solution to that problem in the case where the accuracy of approximation is evaluated in terms of the point metric. We generalise and sharpen the corresponding inequalities established by preceding authors. A new result is established for the intensively studied topic of compound Poisson approximation to the distribution of a sum of integer-valued r.v.s.

Keywords Poisson approximation · Compound Poisson approximation · Point metric

Mathematics Subject Classification (2020) E15

1 Introduction

The task of approximating the distribution of a sum of independent random variables lies at the heart of the probability theory. The central role is played by the normal approximation. However, in situations where one deals with rare events Poisson or compound Poisson approximation may be preferable (cf. [4, 5, 35, 36, 40]).

✉ S. Y. Novak
S.Novak@mdx.ac.uk

1 Middlesex University, London, UK

Published online: 28 February 2024

Interest to the topic of Poisson/compound Poisson approximation arises in connection with applications in extreme value theory, insurance, reliability theory, etc. (cf. [6, 7, 29, 33]). The theory of Poisson/compound Poisson approximation underpins the extreme value theory [29, 33].

Let $X_1, X_2, \ldots$ be integer-valued non-negative random variables (r.v.s). Denote $S_0 = 0$,

$$S_n = X_1 + \cdots + X_n, \quad \lambda \equiv \lambda(n) = \mathbb{E} S_n \quad (n > 1).$$

For example, given r.v.s $\xi_1, \xi_2, \ldots$, one may deal with exceedances of a chosen "threshold" x. Set $X_i = \mathbb{1}\{\xi_i > x\}$. Then

$$S_n \equiv S_n(x) = \sum_{i=1}^{n} \mathbb{1}\{\xi_i > x\}$$

denotes the number of exceedances of threshold x. In particular,

$$\left\{\max_{1 \le i \le n} \xi_i \le x\right\} = \{S_n(x) = 0\}, \quad \{\xi_{k,n} \le x\} = \{S_n(x) < k\} \quad (k \in \mathbb{N}), \tag{1}$$

where $\xi_{k,n}$ denotes the k-th largest element among $\xi_1, \ldots, \xi_n$. Thus, results concerning the distribution of sample extremes can be derived from the corresponding results concerning $\mathcal{L}(S_n)$.

In applications, indicators $\mathbb{1}\{\xi_1 > x\}, \mathbb{1}\{\xi_2 > x\}, \ldots$ may be dependent. A well-known approach (Bernstein's blocks method [14]) consists of grouping observations into blocks which can be considered almost independent. The number of r.v.s in a block is an integer-valued random variable; hence, the number of rare events is a sum of almost independent integer-valued r.v.s.

In (re)insurance applications, the sum $\sum_{i=1}^{n} Y_i \mathbb{1}\{Y_i > x\}$ of integer-valued r.v.s allows to account for the total loss from the claims $\{Y_i\}$ exceeding threshold x [22]. More information concerning applications can be found in [6, 7, 22, 29, 35, 36].

The distribution of a sum $S_n(x)$ of integer-valued non-negative random variables can often be approximated by a Poisson or compound Poisson law. In early 1950s, Kolmogorov has formulated the task of evaluating the accuracy of approximation of the distribution of a sum S_n of independent and identically distributed (i.i.d.) r.v.s by infinitely divisible distributions (Kolmogorov's uniform approximation problem). The topic has attracted a lot of attention among researchers (see, e.g., [4, 19, 35, 36, 40, 49] and references therein).

From a theoretical point of view, the question concerning the accuracy of Poisson/compound Poisson approximation is a particular case of Kolmogorov's problem. Besides, it was noticed that estimates of the accuracy of approximation to the Binomial distribution can provide important insights in other areas of probability theory [32, 38]. In a sense, the Binomial distribution plays the role of a "testing stone" [4].

Note that there is a strong connection between the topics of Poisson and compound Poisson approximation [35, 36, 38, 48]; the latter plays a special role in approximating $\mathcal{L}(S_n)$ by infinitely divisible laws since the class of infinitely divisible distributions coincides with the class of weak limits of compound Poisson distributions [26, Theorem 26].

In a range of situations, both normal and (compound) Poisson approximations can be applicable (cf. [4, 40]). Due to the complex structure of the compound Poisson distribution, in applications one would prefer pure Poisson approximation where possible.

One can choose between possible types of approximation by comparing estimates of the accuracy of approximation. Obviously, one would make a choice according to the sharpest estimate, thus the need of sharp estimates of the accuracy of approximation with explicit constants.

1.1 Notation

Let $\mathcal{S}$ denote the class of measurable functions taking values in [0; 1]. Then

$$d_{TV}(X;Y) \equiv d_{TV}(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_{h\in\mathcal{S}} (\mathbb{E}h(X) - \mathbb{E}h(Y))$$

denotes the total variation distance between the distributions of r.v.s X and Y.

We denote by

$$d_K(X;Y) \equiv d_K(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_x |F_X(x) - F_Y(x)|$$

the uniform (Kolmogorov's) distance between the distributions of random variables X and Y with distribution functions (d.f.s) F_X and F_Y.

The Gini–Kantorovich distance between the distributions of r.v.s X and Y with finite first moments (known also as the Kantorovich–Wasserstein distance) is

$$d_G(X;Y) \equiv d_G(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_{g\in\mathcal{L}} |\mathbb{E}g(X) - \mathbb{E}g(Y)|,$$

where $\mathcal{L} = \{g : |g(x)-g(y)| \le |x-y|\}$ is the set of Lipschitz functions. Note that

$$d_G(X;Y) = \inf_{X',Y'} \mathbb{E}|X' - Y'|,$$

where the infimum is taken over all random pairs (X', Y') such that $\mathcal{L}(X') = \mathcal{L}(X)$ and $\mathcal{L}(Y') = \mathcal{L}(Y)$.

Set $\|f\| = \sup_k |f(k)|$. We denote by $\|\cdot\|_1$ the L_1-norm of a function. Given a discrete r.v. Y, we denote

$$\mathbb{P}_Y = \mathbb{P}(Y = \cdot).$$

In the case of discrete distributions, it is natural to exploit the point metric

$$d_o(X;Y) \equiv d_o(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_k |\mathbb{P}(X=k) - \mathbb{P}(Y=k)| = \|\mathbb{P}_X - \mathbb{P}_Y\|,$$

where $\|\cdot\|$ denotes the sup-norm.

Initially, estimates of the accuracy of Poisson approximation to the Binomial distribution $\mathbf{B}(n, p)$ were established in terms of the uniform distance d_K and the total variation distance d_{TV} [27, 40, 47]. Metrics $d_{TV}, \ d_K, \ d_o$ have obvious merits. For instance, they are shift- and scale invariant. The $O(n^{-2/3})$ estimate of the accuracy of approximation in Kolmogorov's problem holds in terms of d_K [1–3] (and hence in terms of d_o) but generally not in terms of d_{TV} [49]. Bounds in terms of d_o can be sharper than those in terms of the uniform and the total variation distances.

In extreme value theory, one is interested in the distribution of the k-th largest sample element $X_{n:k}$. Metric d_o appears more suitable than d_K and d_{TV} if one evaluates probabilities like $\mathbb{P}(S_n(x) < k)$ for "small" k, cf. (1). A similar observation can be made concerning the length of the longest head run (LLHR), the length of the longest match pattern (LLMP), etc. (see [35]). Metric d_o is shift and scale invariant. Another advantage of using d_o in comparison with $d_{TV}, \ d_K$ is the higher rate of approximation (cf. (4), (8)).

1.2 Independent Bernoulli r.v.s

Hereinafter, $\{X_i\}$ are independent non-negative integer-valued random variables; multiplication is superior to the division. W.l.o.g. we may assume that $\mathbb{E}X_i > 0 \ (\forall i)$.

Let π_λ denote a Poisson $\Pi(\lambda)$ r.v. Many authors worked on the problem of evaluating the accuracy of approximation

$$S_n \approx \pi_\lambda$$

in the case where $\{X_i\}$ are 0-1 r.v.s (see, e.g., [4, 7, 35, 36] and references therein). The problem goes back to Prokhorov [40]. It has attracted a lot of attention among specialists (see, e.g., [35, 36] and references therein).

Estimates of the accuracy of Poisson approximation to the distribution of a sum of independent 0-1 random variables in terms of the uniform distance d_K and the total variation distance d_{TV} have been derived by Kolmogorov [27, Lemma 5], Tsaregradskii [47], LeCam [30], Kerstan [25], Romanowska [41], Shorgin [46], Barbour and Eagleson [12] and other authors (see, e.g., [20, 34, 39, 45]). Concerning estimates in terms of some other metrics, see [15, 21, 35–37] and references therein.

In the case of independent 0-1 r.v.s estimates with correct (the best possible) constants at the leading terms have been found by Roos [45]:

$$d_{TV}(S_n; \pi_\lambda) \le 3\theta/4e(1-\sqrt{\theta}\,)^{3/2}, \tag{2}$$

$$d_K(S_n; \pi_\lambda) \le \theta/2e + 1.2\theta\sqrt{\theta}/(1-\sqrt{\theta}), \tag{3}$$

$$d_o(S_n; \pi_\lambda) \le \theta(3/2e)^{3/2}/2\lambda^{1/2} + \frac{\theta\sqrt{\theta}}{3\sqrt{\lambda}}\frac{6-4\sqrt{\theta}}{(1-\sqrt{\theta})^2}, \tag{4}$$

where $\theta = \sum_{i=1}^n p_i^2/\lambda, \ p_i = \mathbb{P}(X_i = 1) \ (i \ge 1)$; constants at the leading terms in (2)–(4) cannot be improved. Note that $3/4e \approx 0.276, \ 1/2e \le 0.184, \ \frac{1}{2}(3/2e)^{3/2} \le 0.205$.

Estimate (3) has a sharper constant at the leading term than that in (2), estimate (4) has a sharper rate of decay if $\lambda \equiv \lambda(n) \to \infty$ as $n \to \infty$.

Estimates of the accuracy of shifted (translated) Poisson approximation to the distribution of a sum of independent 0-1 r.v.s have been given by Barbour and Xia [9], Čekanavičius and Vaitkus [18], Barbour and Čekanavičius [11], Röllin [43], Novak [34, 37]. Kruopis [28] has evaluated the accuracy of shifted Poisson approximation to $\mathcal{L}(S_n)$ in terms of d_{o}:

$$d_{\mathrm{o}}(S_n; Y) \le \sum_{j=1}^{n} p_j^2 \min\left\{1; \sqrt{\frac{e}{\pi}}\left(\sum_{j=1}^{n} p_j(1-p_j)\right)^{-3/2}\right\}, \tag{5}$$

where $Y = \lfloor \lambda_2 + 0.5 \rfloor + \pi_{\lambda - \lfloor \lambda_2 + 0.5 \rfloor}$, $\lambda_2 = \sum_{j=1}^{n} p_j^2$. Note that $\mathbb{E}Y = \lambda$, $|\mathrm{var}\, Y - \mathrm{var}\, S_n| \le 1/2$. In the case of the Binomial distribution $\mathbf{B}(n, p)$, the right-hand side (r.h.s) of (5) is of order $\sqrt{p/n}$.

1.3 Integer-valued r.v.s

Let $\{X_i\}_{i \ge 1}$ be independent non-negative integer-valued r.v.s. The problem of evaluating the accuracy of Poisson approximation to the distribution of a sum of independent non-negative *integer-valued* r.v.s has been considered, e.g. in [8, 11, 13, 21, 25, 33, 37]; an overview of the results on the topic can be found in [35, 36].

Röllin [43, formula (2.13)], assuming that third moments are finite, states that

$$d_{\mathrm{o}}(S_n; [\lambda - \sigma^2] + \pi_{\sigma^2 + \{\lambda - \sigma^2\}}) \le \left(2 + d' \sum_{i=1}^{n} \psi_i\right) / \sigma^2, \tag{6}$$

where $\sigma^2 = \mathrm{var}\, S_n$, $\psi_i = \sigma_i^2 \mathbb{E}X_i(X_i - 1) + |\mathbb{E}X_i - \sigma_i^2| \mathbb{E}(X_i - 1)(X_i - 2) + \mathbb{E}|X_i(X_i - 1)(X_i - 2)|$, $d' = \max_{i \le n} \|\mathbb{P}_{S_{n,i}+2} - 2\mathbb{P}_{S_{n,i}+1} + \mathbb{P}_{S_{n,i}}\|_1/2$, $S_{n,i} = S_n - X_i$, $[x] = \max\{k \in \mathbf{Z} : k \le x\}$, $\{x\} = x - [x]$. In the case of the Binomial distribution $\mathbf{B}(n, p)$ bound (6) is of order $1/np$.

Tsaregradskii [47] has shown that the rate of the accuracy of compound Poisson approximation to the Binomial distribution $\mathbf{B}(n, p)$ in terms of the uniform distance is $O(1/np)$. Presman [38] has shown that

$$d_{TV}(\mathbf{B}(n, p); P_{n,p}) \le C \min\left\{np^2; p; \max\{(np)^{-2}; n^{-1}\}\right\} \qquad (0 \le p \le 1/2),$$

where C is an absolute constant and $P_{n,p}$ is a particular compound Poisson distribution. Hence,

$$\sup_{p \le 1/2} d_{TV}(\mathbf{B}(n, p); P_{n,p}) = O(n^{-2/3}). \tag{7}$$

According to Čekanavičius [16], there exists an absolute constant C such that

$$d_o(\mathbf{B}(n,p); P_{n,p}) \le C \max\{(np)^{-2}; n^{-1}\}(np)^{-1/2} \qquad (n^{-2/3} \le p \le 1/2). \tag{8}$$

In the case of i.i.d.r.v.s $X, X_1, X_2, \ldots$ obeying $\mathbb{E}X^4 < \infty$ Čekanavičius [17] has shown that the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ is $C_X n^{-3/2}$, where C_X depends on $\mathcal{L}(X)$.

The problem of establishing an estimate of the accuracy of Poisson approximation to the distributions of a sum of independent non-negative integer-valued r.v.s in terms of the point metric with a correct constant at the leading term remained open for a long while. In particular, an open question was whether $\frac{1}{2}(3/2e)^{3/2}$ would remain the best possible constant at the leading term in the case of integer-valued r.v.s. We give below the affirmative answer to that question. We generalise and sharpen the corresponding results from [16, 21, 43, 45].

Theorem 1 presents an estimate of the accuracy of Poisson approximation to $\mathcal{L}(S_n)$ in terms of the point metric with a correct constant at the leading term. Theorem 2 provides a first-order asymptotic expansion. Theorem 3 presents an estimate of the accuracy of shifted Poisson approximation. Theorem 4 provides an estimate of the accuracy of compound Poisson approximation in terms of the point metric. Theorems 1–3 only assume that the second moments are finite, Theorem 4 does not impose moment requirements, and the constants are explicit.

2 Results

Let $\{X_i\}_{i\ge1}$ be independent non-negative integer-valued r.v.s.

First, we present an estimate of the accuracy of pure Poisson approximation.

2.1 Poisson Approximation

If random variables X, ξ, η have finite second moments, let

$$\kappa_X = \mathbb{E}X - \mathrm{var}X, \ \ \gamma_{\xi,\eta} = \mathbb{E}|\xi(\xi-1) - \eta(\eta-1)|.$$

Denote $X_0 = 0$, $\varepsilon^*_\lambda = 1 \wedge 1/\sqrt{2\pi[\lambda]}$. If $i \in \{0, 1, \ldots, n\}$, let $S_{n,i} = S_n - X_i$, $\lambda_i = \mathbb{E}S_{n,i}$,

$$u_i = 1 - d_{TV}(X_i; X_i+1), \ \ U = \sum_{i=1}^n u_i, \ \ u^* = \max_{1\le i\le n} u_i, \ \ U_* = U - u^*,$$

$$\varepsilon_{i,n} = \varepsilon^o_{i,n} \wedge \left(\varepsilon^*_{\lambda_i} + 2\varepsilon^+_{i,n}\right), \ \ \varepsilon^o_{i,n} = 1 \wedge \sqrt{2/\pi}\,\left(1/4 + U - u_i\right)^{-1/2},$$

$$\varepsilon^+_{i,n} = \frac{1-e^{-\lambda}}{\lambda_i} \sum_{j=1}^n d_G(X_j; X^*_j)\mathbb{E}X_j,$$

$$r^*_{i,n} = \min\left\{\varepsilon_{i,n}; \ \tfrac{8}{\pi}(U_*^2 + (1-2u^*)U_* + 1/4)^{-1/2}\right\},$$

where X_i^* denotes a random variable with the distribution

$$\mathbb{P}(X_i^* = m) = (m+1)\mathbb{P}(X_i = m+1)/\mathbb{E}X_i \qquad (m \geq 0) \tag{9}$$

(cf. [33, Ch. 4.4]). Note that $U_*^2 + (1-2u^*)U_* + 1/4 \geq 0$. If $u^* \leq 1/4$, then $U_*^2 + (1-2u^*)U_* + 1/4 \geq U^2$, and $r_{i,n}^* \leq \min\{\varepsilon_{i,n}; 8/\pi U\}$.

Set

$$\varepsilon_1^{\text{o}} = \lambda^{-1} \sum_{i=1}^{n} \min\{4d_G(X_i; X_i^*)\varepsilon_{i,n}; \gamma_{X_i, X_i^*} r_{i,n}^*\}\mathbb{E}X_i, \quad \varepsilon_+^{\text{o}} = 2\sum_{i=1}^{n} |\kappa_{X_i}|\varepsilon_{i,n}/\lambda,$$

$$\varepsilon_2^{\text{o}} = 2\lambda^{-1} \sum_{i=1}^{n} |\kappa_{X_i}|\mathbb{E}X_i r_{i,n}^*, \quad \varepsilon_3^{\text{o}} = 2\lambda^{-1}|\kappa_{S_n}|(\varepsilon_1^{\text{o}} + \varepsilon_+^{\text{o}}).$$

Theorem 1 *If $X_1, \ldots, X_n$ are independent non-negative integer-valued random variables with finite second moments, then*

$$d_{\text{o}}(S_n; \pi_\lambda) \leq c_{\text{o}}|\kappa_{S_n}|\lambda^{-3/2} + (1-e^{-\lambda})(\varepsilon_1^{\text{o}} + \varepsilon_2^{\text{o}} + \varepsilon_3^{\text{o}}) \qquad (n > 1), \tag{10}$$

where $c_{\text{o}} = \frac{1}{2}(3/2e)^{3/2}$.

If $X, X_1, \ldots, X_n$ are i.i.d.r.v.s, set $\theta^* = |\kappa_X|/\mathbb{E}X$,

$$\begin{aligned} \varepsilon_X^{\text{o}} &= \tilde{\varepsilon}_1^{\text{o}} + 16|\kappa_X|/\pi(n-1)u_1 + 2\theta^*(2\theta^*\varepsilon_{1,n} + \tilde{\varepsilon}_1^{\text{o}}), \\ \tilde{\varepsilon}_1^{\text{o}} &= 2d_G(X; X^*)\varepsilon_{1,n} \wedge 8\gamma_{X,X^*}/\pi(n-1)u_1. \end{aligned}$$

Then (10) becomes

$$d_{\text{o}}(S_n; \pi_\lambda) \leq c_{\text{o}}|\kappa_X|/(\mathbb{E}X)^{3/2}\sqrt{n} + (1-e^{-\lambda})\varepsilon_X^{\text{o}}. \tag{10*}$$

According to [45], constant c_{o} in (10), (10*) cannot be improved.

Note that the moment assumption in Theorem 1 can be relaxed at a cost of adding an extra term if one uses truncation at some levels $\{K_i\}$ (i.e. switches from $\{X_i\}$ to $\{X_i'\}$, where $X_i' = X_i \mathbb{1}\{X_i \leq K_i\}$) since $d_{TV}((X_1, \ldots, X_n); (X_1', \ldots, X_n')) \leq \sum_{i=1}^n \mathbb{P}(X_i > K_i)$.

Example 1 Let $\{X_i\}$ be independent Bernoulli $\mathbf{B}(p_i)$ random variables, where $p_i \in [0; 1/2]$ $(i \geq 1)$. Then $\lambda = \sum_{i=1}^n p_i$, $\kappa_{S_n} = \sum_{i=1}^n p_i^2$, $\varepsilon_1^{\text{o}} = 0$, and (10) yields

$$\begin{aligned} d_{\text{o}}(S_n; \pi_\lambda) \leq\; & c_{\text{o}} \sum_{i=1}^{n} p_i^2 \left(\sum_{i=1}^{n} p_i\right)^{-3/2} \\ & + 2\lambda^{-1}(1-e^{-\lambda})\left(\sum_{i=1}^{n} p_i^3 r_{i,n}^* + \sum_{i=1}^{n} p_i^2 \varepsilon_+^{\text{o}}\right). \end{aligned} \tag{11}$$

The constant at the leading term in (11) is sharper than that in (5).

In the case of the Binomial distribution $\mathbf{B}(n,\,p)$, one has $\kappa_X = p^2$, $u_1 = p$,

$$\varepsilon_2^{\mathrm{o}} \le 16p/\pi(n-1), \quad \varepsilon_+^{\mathrm{o}} = 2p\varepsilon_{1,n}, \quad \varepsilon_3^{\mathrm{o}} = 4p^2\varepsilon_{1,n}, \quad \varepsilon_{1,n} \le \tilde{\varepsilon}_n,$$

where $\tilde{\varepsilon}_n = \min\Big\{\sqrt{2/\pi}\Big/\Big(1/4+(n-1)p\Big)^{1/2};\ 1/\sqrt{2\pi[(n-1)p]} + 2(1-e^{-np})p/(1-1/n)\Big\}$. Then

$$d_{\mathrm{o}}(\mathbf{B}(n,\,p);\,\Pi(np)) \le \tfrac{1}{2}(3/2e)^{3/2}\sqrt{p/n} \\ +4(1-e^{-np})p\left(\sqrt{2p/\pi(n-1)} + 4/\pi(n-1)\right). \qquad (12)$$

The rate of the second-order term in (12) is sharper than that in (4) if $p \equiv p(n) \to 0$ as $n \to \infty$. Note that $\sup_k \mathbb{P}(S_n = k) = O(1/\sqrt{np}\,)$, cf. [31, 42].

Example 2 Let $X, X_1, X_2, \ldots$ be independent geometric $\Gamma_{\mathbf{0}}(p)$ r.v.s:

$$\mathbb{P}(X = m) = (1-p)p^m \quad (m \ge 0,\ 0 \le p < 1).$$

Then S_n is a Negative Binomial $\mathbf{NB}(n,\,p)$ r.v.:

$$\mathbb{P}(S_n = j) = \frac{\Gamma(n+j)}{\Gamma(n)\,j!}(1-p)^n p^j \qquad (j \ge 0,\ 0 \le p < 1),$$

where $\Gamma(y) = \int_0^\infty x^{y-1}e^{-x}dx$. It is easy to see that $\mathbb{P}(X_i^* = m) = (m+1)p^m(1-p)^2$. Hence,

$$X_i^* \stackrel{d}{=} X_i + X.$$

Set $r = p/(1-p)$. Note that $\lambda \equiv \mathbb{E}S_n = nr$.

It is easy to check that $\kappa_X = -r^2$, $\gamma_{X_1,X_1^*} = 4r^2$, $u_1 = p$, $\theta^* = r$,

$$\varepsilon_X^{\mathrm{o}} \le \tilde{\varepsilon}_1^{\mathrm{o}} + 32p/\pi q(n-1) + 2r\left(2r\sqrt{2/\pi(1/4+(n-1)p)} + \tilde{\varepsilon}_1^{\mathrm{o}}\right), \\ \tilde{\varepsilon}_1^{\mathrm{o}} \le 32p/\pi q^2(n-1), \quad \varepsilon_{1,n} \le \varepsilon_{n,p}^{\star},$$

where $\varepsilon_{n,p}^{\star} = \sqrt{2/\pi}\Big/\Big(1/4+(n-1)p\Big)^{1/2} \wedge \Big(1\Big/\sqrt{2\pi[(n-1)p]} + 2r/(1-1/n)\Big)$. Theorem 1 yields

$$d_{\mathrm{o}}(S_n;\,\pi_\lambda) \le \tfrac{1}{2}(3/2e)^{3/2}\sqrt{r/n} + (1-e^{-nr})\varepsilon_X^{\mathrm{o}}. \qquad (13)$$

Estimate (13) has a correct constant at the leading term, $\varepsilon_X^{\mathrm{o}} = O(p\sqrt{p/n} + p/n)$.

Example 3 Let $X, X_1, \ldots$ be i.i.d.r.v.s with the distribution

$$\mathbb{P}(X=0)=1-p+p^2/2,\ \mathbb{P}(X=1)=p-p^2,\ \mathbb{P}(X=2)=p^2/2,$$

where $p\in[0;\,1/2]$. Then $\mathbb{E}X = \mathrm{var}X = p$, $\mathcal{L}(X^*) = \mathbf{B}(p)$.

Note that $\mathbb{E}X=\mathbb{E}X^* = p$,

$$\kappa_X=0,\quad \varepsilon_2=\varepsilon_3=0,\quad u_1 = p-p^2/2,\quad \gamma_{X,X^*}=d_G(X;X^*)=p^2.$$

Hence, $\varepsilon_1^{\mathrm{o}} \le 8p/\pi(1-p/2)(n-1)$, and Theorem 1 yields

$$d_{\mathrm{o}}(S_n;\pi_\lambda)\le 8(1-e^{-np})p/\pi(1-p/2)(n-1) \qquad (n>1). \tag{14}$$

While the rate of the accuracy of approximation in (12) is $\sqrt{p/n}$, the rate is p/n in (14).

2.2 A First-Order Asymptotic Expansion

For any function f, denote $\Delta f(\cdot) = f(\cdot+1) - f(\cdot)$. The following theorem provides a first-order asymptotic expansion in terms of d_{o}.

Theorem 2 *Let $X_1, \ldots, X_n$ be independent non-negative integer-valued random variables with finite second moments. If $h(\cdot)=\mathbb{1}\{\cdot=k\}$, where $k\in\mathbb{N}$, then*

$$\left|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) + \mathbb{E}\Delta^2 h(\pi_\lambda)\kappa_{S_n}/2\right| \le (1-e^{-\lambda})\left(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}+\varepsilon_3^{\mathrm{o}}\right). \tag{15}$$

Let $\pi_\lambda^\star$ denote a random variable with the distribution

$$\mathbb{P}(\pi_\lambda^\star=k) = \mathbb{P}(\pi_\lambda=k)(k-\lambda)^2/\lambda \qquad (k\in\mathbb{Z}_+), \tag{16}$$

where $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ is the set of non-negative integer numbers. Note that

$$\lambda\mathbb{E}\Delta^2 h(\pi_\lambda) = \mathbb{E}h(\pi_\lambda^\star) - \mathbb{E}h(\pi_\lambda+1) = \mathbb{P}(\pi_\lambda^\star=k) - \mathbb{P}(\pi_\lambda+1=k) \tag{17}$$

(cf. [34, Remark 1]). Thus,

$$\begin{aligned}&\left|\mathbb{P}(S_n=k)-\mathbb{P}(\pi_\lambda=k) + (\mathbb{P}(\pi_\lambda^\star=k)-\mathbb{P}(\pi_\lambda+1=k))\kappa_{S_n}/2\lambda\right|\\ &\quad\le (1-e^{-\lambda})\left(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}+\varepsilon_3^{\mathrm{o}}\right).\end{aligned}$$

In particular,

$$|d_{\mathrm{o}}(S_n;\pi_\lambda) - d_{\mathrm{o}}(\pi_\lambda^\star;\pi_\lambda+1)|\kappa_{S_n}|/2\lambda| \le (1-e^{-\lambda})\left(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}+\varepsilon_3^{\mathrm{o}}\right).$$

According to Roos [44], $\left|\lambda\|\Delta^2\mathbb{P}_{\pi_\lambda}\| - 1/\sqrt{2\pi\lambda}\right| \le C/\lambda$. Therefore,

$$|d_{\rm o}(S_n;\pi_\lambda) - |\kappa_{S_n}|\lambda^{-3/2}/2\sqrt{2\pi}| = O(\varepsilon_1^{\rm o}+\varepsilon_2^{\rm o}+\varepsilon_3^{\rm o}+|\kappa_{S_n}|\lambda^{-2}). \tag{18}$$

If $X_1,\ldots,X_n$ are i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s, then for any $k\in\mathbb{N}$

$$\|\mathbb{P}(S_n=k) - \mathbb{P}(\pi_{np}=k) + (\mathbb{P}(\pi_{np}^\star=k)-\mathbb{P}(\pi_{np}+1=k))p/2\| \le (1-e^{-np})\varepsilon_X^{\rm o},$$

where $\varepsilon_X^{\rm o} = 4p\left(\sqrt{2p/\pi(n-1)} + 4/\pi(n-1)\right)$; hence,

$$\left|d_{\rm o}(S_n;\pi_\lambda) - \tfrac{1}{2}\sqrt{p/2\pi n}\right| = O\left(p\sqrt{p/n} + 1/np\right). \tag{18$^{\rm o}$}$$

2.3 Shifted Poisson Approximation

Set $\mu=\mathrm{var}\,S_n+\{\kappa_{S_n}\}$. The next theorem deals with shifted Poisson approximation

$$\mathcal{L}(S_n) \approx \mathcal{L}([\kappa_{S_n}]+\pi_\mu).$$

Note that $[\kappa_{S_n}]+\mathbb{E}\pi_\mu = \mathbb{E}S_n$, $|\mathrm{var}\,\pi_\mu - \mathrm{var}\,S_n| < 1$. W.l.o.g. we may assume that $\mu>0$.

Given a r.v. Y, we denote $\bar{Y} = Y-\mathbb{E}Y$. Let $\sigma^2=\mathrm{var}\,S_n$, $U_i=U-u_i$,

$$\hat\varepsilon_\mu = \min\left\{2\mu^{-1}(1-e^{-\mu})\varepsilon_{0,n};\,\bar\varepsilon_\mu\right\},\quad \bar\varepsilon_\mu = 2\varepsilon_\mu^*\sqrt{2/e\mu} + 2\mu^{-1}(1-e^{-\mu})\varepsilon_\mu^\star,$$

$$\varepsilon_\mu^\# = \mu^{-1}\sum_{i=1}^n \min\{4\mathbb{E}|\bar X_i X_i|\varepsilon_{i,n}^{\rm o};\,\mathbb{E}|\bar X_i X_i||X_i-1-2\tilde X_i|r_{i,n}^*\}.$$

Theorem 3 *Let $X_1,\ldots,X_n$ be independent non-negative integer-valued random variables with finite second moments. Then*

$$d_{\rm o}(S_n;[\kappa_{S_n}]+\pi_\mu) \le |\{\kappa_{S_n}\}|\hat\varepsilon_\mu + (1-e^{-\mu})\varepsilon_\mu^\#. \tag{19}$$

Example 1 (continued). Let $\mathcal{L}(S_n) = \mathbf{B}(n,p)$, where $p\in[0;1/2]$. Set $q=1-p$. Clearly,

$$\mu = npq+\{np^2\},\ \ \kappa_{S_n} = np^2,\ \ \delta_{X_i}^{(\mu)} = 0,\ \ \varepsilon_{0,n} \le \sqrt{2/\pi np},\ \ r_{1,n}^* \le 8/\pi(n-1)p.$$

Therefore, $\mu\le np$, $\hat\varepsilon_\mu \le 2\mu^{-1}\varepsilon_{0,n}$, and (19) yields $(n>1)$

$$d_{\rm o}(S_n;[np^2]+\pi_\mu) \le 2\sqrt{2/\pi}(1-e^{-np})(npq)^{-3/2} + 16(1-e^{-np})q/\pi(n-1). \tag{20}$$

Using (12) as $p < 1/\sqrt{n}$ and (20) as $1/\sqrt{n} \le p \le 1/2$, we derive

$$\sup_{0\le p\le 1/2} d_o(S_n; [np^2] + \pi_\mu) \le \frac{2\sqrt{2/\pi}}{n^{3/4}(1-1/\sqrt{n}\,)^{3/2}} + \frac{16}{\pi(n-1)} + \frac{4\sqrt{2/\pi}}{(n-1)^{5/4}}. \tag{21}$$

Indeed, (12) entails

$$d_o(S_n; \pi_{np}) \le c_o n^{-3/4} + 4\sqrt{2/\pi}(n-1)^{-5/4} + 16/\pi(n-1)^{3/2}$$

if $p < 1/\sqrt{n}$. If $p \ge 1/\sqrt{n}$, then (20) yields

$$d_o(S_n; [np^2] + \pi_\mu) \le 2\sqrt{2/\pi}\, n^{-3/4}(1-1/\sqrt{n}\,)^{-3/2} + 16/\pi(n-1),$$

and (21) follows. A uniform in $p \in [0; 1/2]$ estimate of the accuracy of Poisson approximation to the Binomial distribution $\mathbf{B}(n, p)$ in terms of the point metric seems to be new.

3 Compound Poisson Approximation

The topic of compound Poisson approximation to the distribution of a sum of random variables has attracted a lot of attention in the past decades (see, e.g., [19] and references therein).

The topic has applications in extreme value theory, insurance, reliability theory, patterns matching, etc. (cf. [6, 10, 22, 33]). In order to decide if a particular compound Poisson approximation to $\mathcal{L}(S_n)$ is applicable, one would require an estimate of the accuracy of compound Poisson approximation indicating the distance between two distributions is "small", hence the need of sharp bounds with explicit constants.

From a theoretical point of view, the interest to the topic arises in connection with Kolmogorov's problem concerning the accuracy of approximation of the distribution of a sum of independent r.v.s by infinitely divisible laws (see [4, 19] and references therein). From a practical point of view, the problem has important applications in insurance, reliability theory, extreme value theory, etc., cf. [19], and references therein.

Estimates of the accuracy of compound Poisson approximation have been derived mainly in terms of the uniform distance and the total variation distance. Very few estimates of the accuracy of compound Poisson approximation are available in terms of the point metric. However, in situations where one needs to evaluate $\mathbb{P}(S_n < k)$ for a "small" k the point metric may be advantages as estimates in terms of the point metric are expected to have better rate of approximation than estimates in terms of the uniform of the total variation distances, cf. (2)–(4).

By Khintchine's formula (see [24, Ch. 2]), the distribution of any random variable X obeys

$$X \stackrel{d}{=} \tau_p X', \tag{22}$$

where τ_p and X' are independent r.v.s, $\mathcal{L}(X') = \mathcal{L}(X|X \neq 0)$, τ_p is a Bernoulli $\mathbf{B}(p)$ r.v., $p = \mathbb{P}(X \neq 0)$. Since $X_1, \ldots, X_n$ are independent r.v.s,

$$S_n \stackrel{d}{=} \tau_1 X'_1 + \cdots + \tau_n X'_n, \tag{23}$$

where $\tau_1, X'_1, \ldots, \tau_n, X'_n$ are independent r.v.s, $p_i = \mathbb{P}(X_i \neq 0)$,

$$\mathcal{L}(X'_i) = \mathcal{L}(X_i|X_i \neq 0), \ \ \mathcal{L}(\tau_i) = \mathbf{B}(p_i) \quad (\forall i). \tag{24}$$

Assume that $X', X'_1, \ldots, X'_n$ are i.i.d.r.v.s. Then

$$S_n \stackrel{d}{=} \sum_{i=1}^{\nu_n} X'_i, \tag{25}$$

where r.v. $\nu_n = \tau_1 + \cdots + \tau_n$ is independent of $\{X'_i\}$.

If $\mathcal{L}(X')$ is degenerate (i.e. $X' = c$, where c is a constant), then $S_n \stackrel{d}{=} c\nu_n$, and the problem of compound Poisson approximation to $\mathcal{L}(S_n)$ reduces to the problem of Poisson approximation to $\mathcal{L}(\nu_n)$.

Assume that $\mathcal{L}(X')$ is not degenerate. According to Kolmogrov [27], formula (30),

$$d_{TV}\Bigg(\sum_{i=1}^{\nu_n} X'_i; \sum_{i=1}^{\pi_\lambda} X'_i\Bigg) \le d_{TV}(\nu_n; \pi_\lambda), \tag{26}$$

where π_λ denotes a Poisson $\Pi(\lambda)$ r.v.. Thus, an estimate of the accuracy of "accompanying" compound Poisson approximation (the terminology of Gnedenko and Kolmogorov [23]) in terms of the total variation distance follows from an estimate of the accuracy of pure Poisson approximation.

The following theorem presents an estimate of the accuracy of "accompanying" compound Poisson approximation to $\mathcal{L}(S_n)$ in terms of the point metric.

Recall (24), where $p_i = \mathbb{P}(X_i \neq 0)$. Denote $\bar{p} = \sum_{i=1}^n p_i/n$.

Theorem 4 *If $X, X_1, X_2, \ldots$ are independent integer-valued random variables, $X', X'_1, X'_2, \ldots$ are i.i.d.r.v.s, where $\mathcal{L}(X') = \mathcal{L}(X|X \neq 0)$, then*

$$d_{\mathrm{o}}\Bigg(S_n; \sum_{i=1}^{\pi_{n\bar{p}}} X'_i\Bigg) \le c_{\mathrm{o}} \sum_{i=1}^n p_i^2 \Bigg(\sum_{i=1}^n p_i\Bigg)^{-3/2} + 2\frac{1-e^{-n\bar{p}}}{n\bar{p}}\Bigg(\sum_{i=1}^n p_i^3 r^*_{i,n} + \sum_{i=1}^n p_i^2 \varepsilon^{\mathrm{o}}_+\Bigg). \tag{27}$$

If $X, X_1, \ldots$ are i.i.d.r.v.s, then $\bar{p} = p$, where $p = \mathbb{P}(X \neq 0)$, and (27) becomes

$$d_{\mathrm{o}}\Bigg(S_n; \sum_{i=1}^{\pi_{np}} X'_i\Bigg) \le c_{\mathrm{o}}\sqrt{p/n} + 4(1-e^{-np})p\,(p\tilde{\varepsilon}_n + 4/\pi(n-1)). \tag{27*}$$

Bound (27) appears suitable for evaluating probabilities $\mathbb{P}(S_n < k)$ when k is "small". A feature of the bound is that Theorem 4 does not impose moment requirements.

The rate of the accuracy of approximation in (27*) is at least $n^{-1/2}$, the rate is $o(n^{-1/2})$ if $p \equiv p(n) \to 0$ as $n \to \infty$.

If X is a Bernoulli r.v., then $X' \equiv 1$, $\sum_{i=1}^{\pi_{np}} X'_i = \pi_{np}$ is a Poisson r.v., and (27) coincides with (12). Thus, the constant at the leading term in (27) cannot in general be improved.

The advantage of employing the "accompanying" compound Poisson distribution is the simplicity of the approximating distribution. An open question is if the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ can be improved using more complex approximating laws? Relation (29) suggests the following hypothesis. Let $X, X_1, X_2, \ldots$ be i.i.d.r.v.s. Denote $P_X = \mathcal{L}(X)$, and let $\mathbb{CP}$ denote the class of (shifted) compound Poisson r.v.s. An open question is if there exists an absolute constant C_o such that

$$\sup_{P_X} d_o(S_n; \mathbb{CP}) \le C_o n^{-5/6}.$$

Example 4 Suppose that $X, X_1, \ldots$ are i.i.d. random variables,

$$\mathbb{P}(X=0) = 1-p, \; \mathbb{P}(X=1) = \mathbb{P}(X=2) = p/2.$$

Then $\mathbb{P}(X'=1) = \mathbb{P}(X'=1) = 1/2$. Hence, $\sum_{i=1}^{\pi_{np}} X'_i \overset{d}{=} \pi' + 2\pi''$, where π', π'' are independent Poisson $\Pi(np/2)$ r.v.s. Note that $\pi' + 2\pi''$ is a compound Poisson r.v. Theorem 4 yields

$$d_o(S_n; \pi' + 2\pi'') \le c_o\sqrt{p/n} + 4\sqrt{2/\pi}\, p^{3/2}/\sqrt{n-1} + 16p/\pi(n-1). \tag{28}$$

If $p \equiv p(n) \to 0$ as $n \to \infty$, which is typically the case when one deals with rare events, then the r.h.s. of (28) is $o(n^{-1/2})$.

We now present a uniform in $p \in [0; 1/2]$ estimate of the accuracy of compound Poisson approximation to the Binomial distribution in terms of the point metric.

Let $P_{n,p}$ denote the compound Poisson distribution from (7).

Proposition 5 *There exists an absolute constant C such that*

$$\sup_{0 \le p \le 1/2} d_o(\mathbf{B}(n, p); \mathbf{\Pi}(np)) \wedge d_o(\mathbf{B}(n, p); P_{n,p}) \le C n^{-5/6}. \tag{29}$$

Bound (29) provides a better rate of approximation than (21). However, (21) offers a simpler approximating distribution, and the constants are explicit.

4 Proofs

First, we present the proof of Theorem 2, then the proof of Theorem 1.

We start with a particular lemma, which is needed in the proof of Theorem 2.

For any function $h \in \mathcal{S}$, we denote by $g \equiv g_h$ the solution of the Stein equation

$$g(n+1) - \lambda^{-1} n g(n) = h(n) - \mathbb{E}h(\pi_\lambda) \qquad (n \in \mathbb{Z}_+). \tag{30}$$

Note that it is possible to write down the Stein equation in a different manner (see, e.g., [33, Remark 4.1]). The way shown in (30) is in line with the general approach to the characterisation of discrete distributions (cf. [33, Ch. 12]).

Lemma 6 *Let* $h(\cdot) \equiv h_k(\cdot) = \mathbb{1}\{\cdot = k\}$ *for a particular* $k \in \mathbb{Z}_+$. *If* g_h *is given by* (30), *then*

$$\|g_h\| \le 1 - e^{-\lambda}. \tag{31}$$

Proof of Lemma 6 It is known that the solution of Eq. (30) is

$$g(m) = (\mathbb{E}h(\pi_\lambda)\mathbb{1}\{\pi_\lambda < m\} - \mathbb{E}h(\pi_\lambda)\mathbb{P}(\pi_\lambda < m))/\mathbb{P}(\pi_\lambda = m-1) \quad (m \ge 1) \tag{32}$$

(see, e.g., [7, 33]). The value of $g(0)$ is irrelevant (we can set $g(0)=0$).

If $h(\cdot) = \mathbb{1}\{\cdot = k\}$, where $k \in \mathbb{Z}_+$, then

$$g(m) = (\mathbb{P}(\pi_\lambda = k < m) - \mathbb{P}(\pi_\lambda = k)\mathbb{P}(\pi_\lambda < m))/\mathbb{P}(\pi_\lambda = m-1) \quad (m \ge 1). \tag{33}$$

Denote

$$G(n) = \mathbb{P}(\pi_\lambda > n)/\mathbb{P}(\pi_\lambda = n), \quad G_*(n) = \mathbb{P}(\pi_\lambda \le n)/\mathbb{P}(\pi_\lambda = n) \quad (n \in \mathbb{Z}_+).$$

It is known (see, e.g., [33, p. 82]) that function G is decreasing; function G_* is increasing. Therefore, (33) yields

$$\begin{aligned} -g_{h_k}(m) &= \mathbb{P}(\pi_\lambda = k)G_*(m-1) \le \mathbb{P}(\pi_\lambda = k)G_*(k-1) = \mathbb{P}(\pi_\lambda \le k-1)\lambda/k \\ &\le \mathbb{P}(\pi_\lambda \le k) - e^{-\lambda} \le 1 - e^{-\lambda} \quad (m \le k), \\ g_{h_k}(m) &= \mathbb{P}(\pi_\lambda = k)G(m-1) \le \mathbb{P}(\pi_\lambda = k)G(k) = \mathbb{P}(\pi_\lambda > k) \le 1 - e^{-\lambda} \qquad (m > k), \end{aligned}$$

and (31) follows (Röllin [43] mentions without proof that $\|g_h\| \le 1$). □

Proof of Theorem 2 We will use Stein's method. The details of the method have been presented in many publications (see, e.g., [7, 33]).

According to (30),

$$\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) = \mathbb{E}g(S_n+1) - \lambda^{-1}\mathbb{E}S_n g(S_n).$$

Below we evaluate $|\mathbb{E}g(S_n+1) - \lambda^{-1}\mathbb{E}S_n g(S_n)|$.

Let $h(\cdot)=\mathbb{1}\{\cdot=k\}$, where $k\in\mathbb{Z}_+$. Then

$$g(m) = \mathbb{P}(\pi_\lambda = k)\frac{\mathbb{P}(\pi_\lambda > m-1)}{\mathbb{P}(\pi_\lambda = m-1)} \quad (k<m),$$
$$g(m) = -\mathbb{P}(\pi_\lambda = k)\frac{\mathbb{P}(\pi_\lambda \le m-1)}{\mathbb{P}(\pi_\lambda = m-1)} \quad (k\ge m).$$

If $\|\Delta g\|=0$, then g is a constant, and $\lambda\mathbb{E}g(S_n+1) - \mathbb{E}S_n g(S_n) = 0$. Therefore, without loss of generality we may assume that $\|\Delta g\|>0$.

Recall that $S_{n,i} = S_n - X_i$. Set

$$g_i(\cdot) = \mathbb{E}g(S_{n,i}+1+\cdot).$$

It is known that

$$\mathbb{E}X_i f(X_i) = \mathbb{E}X_i \mathbb{E}f(X_i^*+1) \tag{34}$$

for any function f such that $\mathbb{E}|X_i f(X_i)| < \infty$ (cf. [33, Ch. 4]). Therefore,

$$\begin{aligned}\lambda\mathbb{E}g(S_n+1) - \mathbb{E}S_n g(S_n) &= \lambda\mathbb{E}g(S_n+1) - \sum_{i=1}^n \mathbb{E}X_i g(S_{n,i}+X_i)\\ &= \sum_{i=1}^n \mathbb{E}X_i \left(\mathbb{E}g(S_{n,i}+X_i+1) - \mathbb{E}g(S_{n,i}+X_i^*+1)\right)\\ &= \sum_{i=1}^n \mathbb{E}X_i \left(\mathbb{E}g_i(X_i) - \mathbb{E}g_i(X_i^*)\right).\end{aligned} \tag{35}$$

Given a function $f:\mathbb{Z}_+\to\mathbb{R}$, we denote ($\ell\ge 0$, $m\ge 0$, $k\ge 0$)

$$\begin{aligned}R_f(m,k,\ell) &= f(m) - f(k) - (m-k)\Delta f(\ell), \qquad (36)\\ c_1(f) &= \sup_{i,j}|\Delta f(i) - \Delta f(j)|,\ \ c_2(f) = \|\Delta^2 f\|,\\ \delta_{m,k}^{(\ell)} &= \min\{c_1(f)|m-k|;\, c_2(f)|(m-\ell)(m-\ell-1) - (k-\ell)(k-\ell-1)|/2\}.\end{aligned}$$

According to Proposition 4 in [35, 36],

$$|f(m) - f(k) - (m-k)\Delta f(\ell)| \le \delta_{m,k}^{(\ell)}. \tag{37}$$

Clearly,

$$g_i(X_i) - g_i(X_i^*) = (X_i - X_i^*)\Delta g_i(0) + R_{g_i}(X_i, X_i^*, 0). \tag{38}$$

From (35), (38), (37),

$$\left|\lambda \mathbb{E} g(S_n+1) - \mathbb{E} S_n g(S_n) - \sum_{i=1}^{n} \mathbb{E} X_i \mathbb{E}(X_i - X_i^*) \mathbb{E} \Delta g(S_{n,i}+1)\right| \\ \leq \sum_{i=1}^{n} \mathbb{E} X_i \mathbb{E} \delta^{(0)}_{X_i, X_i^*}. \tag{39}$$

It is known (see Barbour and Eagleson [12] or [7, Remark 1.1.2]) that

$$\|\Delta g_h\| \leq 1-e^{-\lambda}, \quad \|g_h\| \leq \sqrt{2\lambda/e} \wedge \lambda. \tag{40}$$

Thus, $|c_1(g_i)| \leq 2\|\Delta g_i\| \leq 2(1-e^{-\lambda})$. Using (51), we get

$$\|\Delta g_i\| \leq \|g_i\| \sum_k |\Delta \mathbb{P}(S_{n,i}=k)| = 2\|g_i\|\, d_{TV}(S_{n,i};\, S_{n,i}+1). \tag{41}$$

Recall that

$$d_{TV}(S_{n,i};\, S_{n,i}+1) \leq \varepsilon_{i,n}. \tag{42}$$

Taking into account Proposition 6 and (42), we derive

$$|c_1(g_i)| \leq 4(1-e^{-\lambda})\varepsilon_{i,n}.$$

Note that

$$|\mathbb{E} \Delta^2 g(S_{n,i}+\ell)| \leq 2\|\Delta g\|\varepsilon_{i,n}.$$

Taking into account (40), $\|\Delta^2 g_i\| \leq 2(1-e^{-\lambda})\varepsilon_{i,n}$. By (51), (56) and (31),

$$|\Delta^2 g_i(\cdot)| \leq \|g\| \sum_k |\Delta^2 \mathbb{P}(S_{n,i}=k)| \leq \tfrac{16}{\pi}(1-e^{-\lambda}) \Big/ \sqrt{U_*^2+(1-2u^*)U+1/4}.$$

Therefore,

$$\|\Delta^2 g_i(\cdot)\|/2 \leq (1-e^{-\lambda}) r^*_{i,n}. \tag{43}$$

Combining these estimates, we get

$$\mathbb{E} \delta^{(0)}_{X_i, X_i^*} \leq (1-e^{-\lambda}) \min\{4\mathbb{E}|X_i - X_i^*|\varepsilon_{i,n};\, \gamma_{X_i, X_i^*} r^*_{i,n}\}. \tag{44}$$

Here random variables X_i, X_i^* can be defined on a common probability space in such a way that $\mathbb{E}|X_i - X_i^*| = d_G(X_i, X_i^*)$. Notice that

$$\kappa_{X_i} = \mathbb{E} X_i \mathbb{E}(X_i - X_i^*) \tag{45}$$

(cf. (34)). Taking into account (39), (44), (45), we derive

$$\Big|\lambda\mathbb{E}g(S_n+1)-\mathbb{E}S_ng(S_n)-\sum_{i=1}^{n}\kappa_{X_i}\mathbb{E}\Delta g(S_{n,i}+1)\Big|\le\lambda(1-e^{-\lambda})\varepsilon_1^{\mathrm{o}}. \tag{46}$$

By (51), $\mathbb{E}\Delta g(Y)=-\sum_k g(k)\Delta\mathbb{P}_Y(k)$. Using (42) and (31), we get

$$|\mathbb{E}\Delta g(S_{n,i}+1)|\le 2\|g\|d_{TV}(S_{n,i};\,S_{n,i}+1)\le 2(1-e^{-\lambda})\varepsilon_{i,n}.$$

This and (46) entail

$$d_{\mathrm{o}}(S_n;\pi_\lambda)\le(1-e^{-\lambda})(\varepsilon_1^{\mathrm{o}}+\varepsilon_+^{\mathrm{o}}). \tag{47}$$

Now we replace $\mathbb{E}\Delta g(S_{n,i}+1)$ in (46) with $\mathbb{E}\Delta g(S_n+1)$.
If $m>k$, then

$$f(m)-f(k)=\sum_{i=k}^{m-1}\Delta f(i) \tag{48}$$

for any function f. In particular,

$$\mathbb{E}\Delta g(S_{n,i}+X_i+1)-\mathbb{E}\Delta g(S_{n,i}+1)=\mathbb{E}\sum_{\ell=0}^{X_i-1}\Delta^2 g_i(\ell). \tag{49}$$

According to [35], Lemma 5, for any bounded function f

$$|\mathbb{E}\Delta f(S_{n,i})|\le\min\Big\{2\|f\|\varepsilon_{i,n};\,(\|\Delta f\|\wedge 2\|f\|\varepsilon^*_{\lambda_i})+2\|\Delta f\|\varepsilon^+_{i,n}\Big\}. \tag{50}$$

An application of (50) with $f=\Delta g_i$ yields

$$\Big|\mathbb{E}\Delta g(S_{n,i}+X_i+1)-\mathbb{E}\Delta g(S_{n,i}+1)\Big|\le 2\|\Delta g\|\varepsilon_{i,n}\mathbb{E}X_i,$$

where $\|\Delta g\|\le 1-e^{-\lambda}$ by (40). Note that (43) and (49) yield

$$|\mathbb{E}\Delta g(S_{n,i}+X_i+1)-\mathbb{E}\Delta g(S_{n,i}+1)|\le 2(1-e^{-\lambda})r^*_{i,n}\mathbb{E}X_i.$$

Therefore,

$$\Big|\sum_{i=1}^{n}\Big(\mathbb{E}\Delta g(S_n+1)-\mathbb{E}\Delta g(S_{n,i}+1)\kappa_{X_i}\Big)\Big|\le 2(1-e^{-\lambda})\sum_{i=1}^{n}|\kappa_{X_i}|\mathbb{E}X_i r^*_{i,n}.$$

We have shown that

$$\Big|\mathbb{E}h(S_n)-\mathbb{E}h(\pi_\lambda)-\lambda^{-1}\kappa_{S_n}\mathbb{E}\Delta g(S_n+1)\Big|\le(1-e^{-\lambda})(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}).$$

It remains to evaluate $\mathbb{E}\Delta g(S_n+1) - \mathbb{E}\Delta g(\pi_\lambda+1)$.

Recall that $h(\cdot)=\mathbb{1}\{\cdot=k\}$, where $k\in\mathbb{Z}_+$. It is known that $\Delta g(i)\le 0$ if $i\ne k$, while $0\le\Delta g(k)\le 1-e^{-\lambda}$ (cf. [12] or the proof of Proposition 6). Note that $\sum_{i\ne k}\Delta g(i) = -\Delta g(k)$. Therefore,

$$\left|\sum_{i\ne k}\Delta g(i)\,(\mathbb{P}(S_n+1=i)-\mathbb{P}(\pi_\lambda+1=i))\right|\le(1-e^{-\lambda})d_{\mathrm{o}}(S_n;\pi_\lambda),$$
$$|\Delta g(k)\,(\mathbb{P}(S_n+1=k)-\mathbb{P}(\pi_\lambda+1=k))|\le(1-e^{-\lambda})d_{\mathrm{o}}(S_n;\pi_\lambda).$$

An application of (47) yields

$$|\mathbb{E}\Delta g(S_n+1)-\mathbb{E}\Delta g(\pi_\lambda+1)|\le 2(1-e^{-\lambda})d_{\mathrm{o}}(S_n;\pi_\lambda)\le 2(1-e^{-\lambda})(\varepsilon_1^{\mathrm{o}}+\varepsilon_+^{\mathrm{o}}).$$

Thus,

$$\left|\mathbb{E}h(S_n)-\mathbb{E}h(\pi_\lambda)-\lambda^{-1}\kappa_{S_n}\mathbb{E}\Delta g(\pi_\lambda+1)\right|\le(1-e^{-\lambda})(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}+\varepsilon_3^{\mathrm{o}}).$$

Note that

$$\mathbb{E}\Delta g_{h_A}(\pi_\lambda+1)=\big(\mathbb{P}(\pi_\lambda+1\in A)-\mathbb{P}(\pi_\lambda^\star\in A)\big)/2$$

for any indicator function $h_A(\cdot)=\mathbb{1}\{\cdot\in A\}$ (cf. (4.31) in [33]). Since every function $h\in\mathcal{S}$ can be represented as $h(\cdot)=\sum_k c_k\mathbb{1}\{\cdot=k\}$, where $\{c_k\}$ are constants,

$$\mathbb{E}\Delta g_h(\pi_\lambda+1)=\big(\mathbb{E}h(\pi_\lambda+1)-\mathbb{E}h(\pi_\lambda^\star)\big)/2\qquad(\forall h\in\mathcal{S}).$$

Therefore,

$$\begin{aligned}\big|\mathbb{E}h(S_n)-\mathbb{E}h(\pi_\lambda)-\big(\mathbb{E}h(\pi_\lambda+1)-\mathbb{E}h(\pi_\lambda^\star)\big)\kappa_{S_n}/2\lambda\big|&\\ \le(1-e^{-\lambda})\left(\varepsilon_1^{\mathrm{o}}+\varepsilon_2^{\mathrm{o}}+\varepsilon_3^{\mathrm{o}}\right).&\end{aligned}\tag{15*}$$

This and (17) lead to (15). □

Proof of Theorem 1 If $\lambda=0$, then $S_n=\pi_\lambda=0$, and (10) trivially holds. Therefore, w.l.o.g. we may assume that $\lambda>0$.

Given a r.v. Y, denote

$$\Delta\mathbb{P}_Y(\cdot)=\mathbb{P}(Y+1=\cdot)-\mathbb{P}(Y=\cdot).$$

In particular, $\Delta f(\cdot,\cdot)$ means the increment of the first argument.

Clearly, for any bounded function h

$$\mathbb{E}\Delta h(Y)=-\sum_j h(j)\Delta\mathbb{P}_Y(j).\tag{51}$$

Given an arbitrary $k \in \mathbb{Z}_+$, let $h(\cdot) = \mathbb{1}\{\cdot = k\}$. We apply (15).

Note that

$$\Delta^2 h(\cdot) = \mathbb{1}\{\cdot = k-2\} - 2\mathbb{1}\{\cdot = k-1\} + \mathbb{1}\{\cdot = k\}.$$

Therefore,

$$\mathbb{E}\Delta^2 h(\pi_\lambda) = \Delta^2 \mathbb{P}_{\pi_\lambda}(k).$$

According to Lemma 3 in Roos [45],

$$\|\Delta^2 \mathbb{P}_{\pi_\lambda}\| \le (3/2\lambda e)^{3/2}. \tag{52}$$

Bounds (15) and (52) entail

$$|\mathbb{P}(S_n = k) - \mathbb{P}(\pi_\lambda = k)| \le \tfrac{1}{2}|\kappa_{S_n}|(3/2\lambda e)^{3/2} + (1 - e^{-\lambda})(\varepsilon_1^{\mathrm{o}} + \varepsilon_2^{\mathrm{o}} + \varepsilon_3^{\mathrm{o}}).$$

The proof is complete. □

The proof of Theorem 3 requires the following

Proposition 7 *For any bounded function f and any integer-valued random variable Y*

$$|\mathbb{E}\Delta f(Y)| \le 2\|f\|\, d_{TV}(Y; Y+1), \tag{53}$$
$$|\mathbb{E}\Delta^2 f(Y)| \le \|f\|\, \|\Delta^2 \mathbb{P}_Y\|_1. \tag{54}$$

As a consequence,

$$|\mathbb{E}\Delta f(S_n)| \le 2\|f\|\, \varepsilon_{0,n}, \tag{55}$$
$$|\mathbb{E}\Delta^2 f(S_n)| \le \tfrac{16}{\pi}\|f\| \Big/ \sqrt{U^2 + (1-2u^*)U + 1/4}. \tag{56}$$

If $2u^* \le 1$, then (56) yields

$$|\mathbb{E}\Delta^2 f(S_n)| \le 16\|f\| \big/ \pi U.$$

If $2u^* > 1$, then (56) entails

$$|\mathbb{E}\Delta^2 f(S_n)| \le 16\|f\| \big/ \pi(U - 1/2).$$

Proof of Proposition 7 Relation (53) follows from (51): $\mathbb{E}\Delta f(Y) = -\sum_k f(k)\Delta \mathbb{P}_Y(k)$, hence

$$|\mathbb{E}\Delta f(Y)| \le \|f\|\, \|\Delta \mathbb{P}_Y\|_1 = 2\|f\|\, d_{TV}(Y; Y+1).$$

Similarly,

$$|\mathbb{E}\Delta^2 f(Y)| = \Big|\sum_k f(k)\Delta^2 \mathbb{P}_Y(k)\Big| \le \|f\|\,\|\Delta^2 \mathbb{P}_Y\|_1.$$

Bound (55) is an immediate consequence of (53) and (42). Relation (56) will follow from (54) and the following inequality:

$$\|\Delta^2 \mathbb{P}(S_n = \cdot)\|_1 \le 16/\pi\sqrt{U^2 + (1-2u^*)U + 1/4}. \tag{57}$$

We proceed with the proof of (57); the argument is similar to that behind (4.9) in [11].

Let I_x denote the distribution concentrated at point x, $I \equiv I_0$, and let $*$ denote the convolution of measures. Then

$$\Delta \mathbb{P}_Y = \mathbb{P}_Y * (I_1 - I), \;\; \Delta^2 \mathbb{P}_Y(\cdot) = (I_1 - I)^{*2} * \mathbb{P}_Y.$$

Let $Q_1,\ Q_2$ be two measures. By the property of the total variation norm,

$$\|Q_1 * Q_2\|_1 \le \|Q_1\|_1\,\|Q_2\|_1.$$

Set $J = \{1,\ldots,n\}$. If $J = C \cup D$, we denote $S' = \sum_{i\in C} X_i$, $S'' = \sum_{i\in D} X_i$. Since $\mathbb{P}_{S_n} = \mathbb{P}_{S'} * \mathbb{P}_{S''}$, we have

$$\begin{aligned}\|\Delta^2 \mathbb{P}_{S_n}\|_1 &= \|(I_1 - I)^{*2} * \mathbb{P}_{S_n}\|_1 = \|(I_1 - I)*\mathbb{P}_{S'} * (I_1 - I)*\mathbb{P}_{S''}\|_1 \\ &\le \|(I_1 - I)*\mathbb{P}_{S'}\|_1\,\|(I_1 - I)*\mathbb{P}_{S''}\|_1 = 4d_{TV}(S'; S'+1)d_{TV}(S''; S''+1).\end{aligned}$$

We will exploit this bound and (59).

If $U = 0$, then (57) trivially holds. Therefore, we may assume that $U > 0$. Set J can be split into $C \cup D$ so that sets C, D are non-empty,

$$U_C > U/2 - u^*, \quad U_D \ge U/2, \tag{58}$$

where $U_C = \sum_{i\in C} u_i,\ U_D := \sum_{i\in D} u_i$.

Indeed, r.v.s $X_1,\ldots,X_n$ (and hence numbers $u_1,\ldots,u_n$) can be rearranged without affecting S_n. Therefore, we may assume that $u_0 := 0 \le u_1 \le \cdots \le u_n = u^*$. Denote

$$\Sigma_k = u_1 + \cdots + u_k, \quad \nu = \min\{k \ge 1 : \Sigma_k > U/2\}.$$

If $\Sigma_{n-1} \le U/2$, then $\nu = n$, hence $u_n > U/2$. One can choose $C = \{1,\ldots,n{-}1\},\ D = \{n\}$. Then

$$U_C = U - u^* \ge (U/2 - u^*)_+,\ U_D = u_n > U/2.$$

If $\Sigma_{n-1} > U/2$, then $\nu \le n-1$ and $u^* \le U/2$. One can choose $C = \{1, \ldots, \nu-1\}$, $D = \{\nu, \ldots, n\}$. Then

$$U_C = \Sigma_\nu - u_\nu \ge \Sigma_\nu - u^* > U/2 - u^*, \quad U_D = U - \Sigma_{\nu-1} \ge U/2,$$

and (58) holds.

According to Mattner and Roos [31, Corollary 1.6],

$$d_{TV}(S_n; S_n+1) \le \sqrt{2/\pi} \Big/ (1/4+U)^{1/2}. \tag{59}$$

Inequality (59) can be applied to $S',\ S''$:

$$d_{TV}(S'; S'+1) \le \sqrt{2/\pi} \Big/ \Big(1/4+U_C\Big)^{1/2}, \quad d_{TV}(S''; S''+1) \le \sqrt{2/\pi} \Big/ \Big(1/4+U_D\Big)^{1/2}.$$

One can check that $(1/4+U_C)(1/4+U_D) \ge U^2/4+(1-2u^*)U/4+1/16$. Hence,

$$\|\Delta^2 \mathbb{P}_{S_n}\|_1 \le \tfrac{8}{\pi}(U_C+1/4)^{-1/2}(U_D+1/4)^{-1/2} \le \tfrac{16}{\pi}(U^2+(1-2u^*)U+1/4)^{-1/2}.$$

This leads to (57) and hence to (56). The proof is complete. □

Proof of Theorem 3 Recall that $\mathbb{Z}$ denotes the set of integer numbers. Set

$$a = [\kappa_{S_n}], \quad b = \{\kappa_{S_n}\}.$$

Then

$$\mu = \sigma^2 + b = \lambda - a \ge 0. \tag{60}$$

We need to evaluate $|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\mu + a)|$, where $h(\cdot) = \mathbb{1}\{\cdot = k\}$.

W.l.o.g. we may assume that $\mu > 0$: if $\mu = 0$, then (60) yields $\sigma^2 = \{\lambda\} = 0$; hence, every X_i is a constant, $a = \lambda$, and (19) trivially holds.

In the case of shifted Poisson approximation, the basic equation is

$$f(k+1) - \mu^{-1}(k-a)f(k) = h(k) - \mathbb{E}h(\pi_\mu + a) \qquad (k \ge a) \tag{61}$$

(cf. (12.26) in [33]). The solution $f \equiv f_h$ of Eq. (61) is

$$f(k) = 0 \quad (k \le a), \quad f(k) = g_{\tilde{h}}(k-a) \quad (k \ge a),$$

where $\tilde{h}(m) = h(m+a)$ $(m \ge 0)$ and $g \equiv g_{\tilde{h}}$ is given by (32) with λ replaced with μ.

According to (61),

$$\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\mu + a) = \mathbb{E}f(S_n+1) - \mu^{-1}\mathbb{E}(S_n - a)f(S_n).$$

Below we evaluate $|\mu\mathbb{E}f(S_n+1) - \mathbb{E}(S_n-a)f(S_n)|$.

First, we show that

$$d_o(S_n;\pi_\mu+a) \le \mu^{-1}|b|\,|\mathbb{E}\Delta f(S_n)| + (1-e^{-\mu})\varepsilon^{\#}_{\mu}. \tag{62}$$

Let $\{\tilde X_j\}$ denote independent copies of $\{X_j\}$. Set $\bar X_i = X_i - \mathbb{E}X_i$,

$$f_i(\cdot) = \mathbb{E}f(S_{n,i}+\cdot),\; r_i = R_{f_i}(X_i, 0, \tilde X_i),$$

cf. (36). Because of (60),

$$\begin{aligned}\mu f(S_n+1) - (S_n-a)f(S_n) &= \mu f(S_n+1) - (\bar S_n+\mu)f(S_n)\\ &= \mu\Delta f(S_n) - \bar S_n f(S_n).\end{aligned}$$

Since $r_i = f_i(X_i) - f_i(0) - X_i\Delta f_i(\tilde X_i)$,

$$\begin{aligned}\mathbb{E}\bar S_n f(S_n) &= \sum_{i=1}^n \mathbb{E}\bar X_i\left(f(S_n) - f(S_{n,i})\right)\\ &= \sigma^2\mathbb{E}\Delta f(S_n) + \sum_{i=1}^n \mathbb{E}\bar X_i r_i.\end{aligned}$$

Recall that $\mu = \sigma^2+b$. Hence,

$$\mu\mathbb{E}f(S_n+1) - \mathbb{E}(S_n-a)f(S_n) = b\mathbb{E}\Delta f(S_n) - \sum_{i=1}^n \mathbb{E}\bar X_i r_i.$$

Note that $|r_i| \le \delta^{\tilde X_i}_{X_i,0}$, $c_1(f_i) \le 2\|\Delta f_i\| \le 4(1-e^{-\mu})\varepsilon^{o}_{i,n}$ by (37), (41), (42), (31). Therefore,

$$\mathbb{E}|\bar X_i r_i| \le 4(1-e^{-\mu})\varepsilon^{o}_{i,n}\mathbb{E}|\bar X_i X_i|.$$

According to (55), $\|\Delta^2 f_i\| \le 2\|\Delta g\|\varepsilon_{i,n}$, while (56) yields

$$\|\Delta^2 f_i\| \le \tfrac{16}{\pi}\|g\|\Big/\sqrt{U_i^2+(1-2u^*)U_i+1/4}.$$

In view of (40) and (31), $\max\{\|g\|;\,\|\Delta g\|\} \le 1-e^{-\mu}$. Therefore,

$$c_2(f_i) \equiv \|\Delta^2 f_i\| \le 2(1-e^{-\mu})r^*_{i,n}. \tag{63}$$

Note that $(m-\ell)(m-\ell-1) - \ell(\ell+1) = m(m-1) - 2m\ell$. Hence,

$$\begin{aligned}\mathbb{E}|\bar X_i r_i| &\le \mathbb{E}|\bar X_i||X_i(X_i-1)-2X_i\tilde X_i|\|\Delta^2 f_i\|/2\\ &\le (1-e^{-\mu})\mathbb{E}|\bar X_i||X_i(X_i-1)-2X_i\tilde X_i|r^*_{i,n}\end{aligned}$$

by (37), (40), (55), (31), (63). Thus,

$$\mathbb{E}|\bar{X}_i r_i| \le (1-e^{-\mu})\min\{4\mathbb{E}|\bar{X}_i X_i|\varepsilon^{\mathrm{o}}_{i,n};\, \mathbb{E}|\bar{X}_i X_i||X_i-1-2\tilde{X}_i|r^*_{i,n}\},$$

i.e. (62) holds.

It is known (cf. [35, 36]) that

$$|\mathbb{E}\Delta f(S_n)| \le \mu\bar{\varepsilon}_\mu.$$

According to (51), (41), (42), (31),

$$|\mathbb{E}\Delta f(S_n)| \le 2(1-e^{-\mu})\varepsilon_{0,n}.$$

Thus, (19) holds. The proof is complete. □

Remark 1 Estimates of Theorems 1–3 work best if the maximum span of $\mathcal{L}(X_1), \ldots, \mathcal{L}(X_n)$ is 1. If the maximum span of $\mathcal{L}(X_i)$ is >1 for a particular i, then $d_{TV}(X_i;\, X_i+1)$ may be equal to 1, reducing $U=\sum_{i=1}^n u_i$ and hence increasing the bounds.

The following inequality is employed in the proof of Theorem 4.

Lemma 8 *Let* $\tilde{\nu}, \nu, X, X_1, X_2, \ldots$ *be independent r.v.s taking values in* $\mathbb{Z}_+$, $X_i \stackrel{d}{=} X\ (\forall i)$. *Denote* $S_0=0,\ S_k = X_1+\cdots+X_k\ (k\in\mathbb{N})$. *Then*

$$d_{\mathrm{o}}(S_{\tilde{\nu}};\, S_\nu) \le d_{\mathrm{o}}(\tilde{\nu};\, \nu)/\mathbb{P}(X\neq 0). \tag{64}$$

Relation (64) is an analogue of (26) in terms of the point metric.

Proof of Lemma 8 W.l.o.g. we may assume that $p:=\mathbb{P}(X\neq 0)\neq 0$. For any $m\in\mathbb{Z}_+$

$$\mathbb{P}(S_{\tilde{\nu}}=m) - \mathbb{P}(S_\nu=m) = \sum_{k\ge 0}(\mathbb{P}(\tilde{\nu}=k)-\mathbb{P}(\nu=k))\,\mathbb{P}(S_k=m).$$

Therefore,

$$|\mathbb{P}(S_{\tilde{\nu}}=m) - \mathbb{P}(S_\nu=m)| \le d_{\mathrm{o}}(\nu_n;\, \nu)\sum_{k\ge 0}\mathbb{P}(S_k=m). \tag{65}$$

Denote

$$f(m) = \sum_{k\ge 0}\mathbb{P}(S_k=m) \qquad (m\in\mathbb{Z}_+).$$

Estimate (64) will follow if we show that $f(m)\le 1/p$ for any $m\in\mathbb{Z}_+$.

Clearly, $f(0)=\sum_{k\ge 0} q^m = 1/p$, where $q=1-p$. Let $m\in\mathbb{N}$. By Khintchine's formula (23),

$$S_k \stackrel{d}{=} S'_{\nu_k},$$

where Binomial $\mathbf{B}(k,\, p)$ r.v. ν_k is independent of $\{X_i'\}$. Hence

$$f(m) = \sum_{k\geq 0} \mathbb{P}(S_{\nu_k}' = m) = \sum_{k\geq 0} \sum_{j=0}^{k} \binom{k}{j} p^j q^{k-j} \mathbb{P}(S_j' = m)$$
$$= \sum_{j\geq 0} p^j \mathbb{P}(S_j' = m) \sum_{k\geq j} \binom{k}{j} q^{k-j} = \sum_{j\geq 0} \mathbb{P}(S_j' = m)/p,$$

where $0^0 := 1$; we have used the fact that $\sum_{k\geq j} \binom{k}{j} p^j q^{k-j} = 1/p$.

Denote by $\eta(\cdot)$ the renewal process

$$\eta(m) = \max\{k \in \mathbb{N} \colon S_k \leq m\} \qquad (m \in \mathbb{Z}_+). \tag{66}$$

Then $\{S_k' \leq m\} = \{k \leq \eta(m)\}$, $\mathbb{E}\eta(m) = \sum_{k\geq 1} \mathbb{P}(S_k' \leq m)$,

$$\sum_{k\geq 0} \mathbb{P}(S_k' = m) = \mathbb{E}\eta(m) - \mathbb{E}\eta(m-1) \qquad (m \in \mathbb{N}).$$

Since X' takes values in $\mathbb{N}$, we have

$$|\eta(m) - \eta(m-1)| \leq 1.$$

Indeed, $S_{\eta(m-1)} \leq m-1,\;\; S_{\eta(m-1)+1} \geq m$ by (66). If $S_{\eta(m-1)+1} > m$, then $\eta(m) = \eta(m-1)$. Therefore,

$$0 \leq \eta(m) - \eta(m-1) \leq \mathbb{1}\{S_{\eta(m-1)+1} = m\}.$$

Thus, $\sum_{j\geq 0} \mathbb{P}(S_j' = m) \leq 1$ for all $m \in \mathbb{Z}_+$, and (65) entails (64). □

Proof of Theorem 4 According to (25),

$$S_n \stackrel{d}{=} S_{\nu_n}',$$

where $\nu_n = \tau_1 + \cdots + \tau_n$, $\tau_1, X_1', \ldots, \tau_n, X_n'$ are independent r.v.s, $\mathcal{L}(X_i') = \mathcal{L}(X_i | X_i \neq 0)$, $\mathcal{L}(\tau_i) = \mathbf{B}(p_i)$ $(\forall i)$, $p_i = \mathbb{P}(X_i \neq 0)$,

$$S_0' := 0, \quad S_k' = X_1' + \cdots + X_k' \quad (k \in \mathbb{N}).$$

Thus, $d_{\mathrm{o}}(S_n; S_{\pi_{n\bar{p}}}') = d_{\mathrm{o}}(S_{\nu_n}'; S_{\pi_{n\bar{p}}}')$. Since $\mathbb{P}(X' \neq 0) = 1$, inequality (64) yields

$$d_{\mathrm{o}}(S_{\nu_n}'; S_{\pi_{n\bar{p}}}') \leq d_{\mathrm{o}}(\nu_n; \pi_{n\bar{p}}). \tag{64*}$$

An application of (11) leads to (27). □

Proof of Proposition 5 If $p \le n^{-2/3}$, then (12) yields

$$d_o(\mathbf{B}(n, p); \Pi(np)) \le \frac{1}{2}(3/2e)^{3/2} n^{-5/6} + 4\sqrt{2/\pi}\,(n-1)^{-3/2} + \frac{16}{\pi}(n-1)^{-5/3}.$$

If $p > n^{-2/3}$, then we apply (8) to get

$$d_o(\mathbf{B}(n, p); P_{n,p}) \le C n^{-5/6}.$$

Combining these bounds, we derive (29). □

Acknowledgements The author is grateful to the reviewer for many helpful remarks.

Data Availability Not applicable.

Declarations

Competing interests The author declares no competing interests.

References

1. Arak, T.V.: On the convergence rate in Kolmogorov's uniform limit theorem. I. Theory Probab. Appl. **26**(2), 219–239 (1981)
2. Arak, T.V.: On the convergence rate in Kolmogorov's uniform limit theorem. II. Theory Probab. Appl. **26**(3), 437–451 (1981)
3. Arak, T.V.: An improvement of the lower bound for the rate of convergence in Kolmogorov's uniform limit theorem. Theory Probab. Appl. **27**, 767–772 (1982)
4. Arak, T.V., Zaitsev, A.Y.: Uniform limit theorems for sums of independent random variables. Proc. Steklov Inst. Math. **174**, 3–214 (1986)
5. Arenbaev, N.K.: Asymptotic behavior of the multinomial distribution. Theory Probab. Appl. **21**, 805–810 (1976)
6. Balakrishnan, N., Koutras, M.V.: Runs and Scans with Applications. Wiley, New York (2001)
7. Barbour, A.D., Holst, L., Janson, S.: Poisson Approximation. Clarendon Press, Oxford (1992)
8. Barbour, A.D.: Asymptotic expansions in the Poisson limit theorem. Ann. Probab. **15**(2), 748–766 (1987)
9. Barbour, A.D., Xia, A.: Poisson perturbations. ESAIM Probab. Stat. **3**, 131–150 (1999)
10. Barbour, A.D., Chryssaphinou, O.: Compound Poisson approximation: a user's guide. Appl. Probab. **11**(3), 964–1002 (2001)
11. Barbour, A.D., Čekanavičius, V.: Total variation asymptotics for sums of independent integer random variables. Ann. Probab. **30**(2), 509–545 (2002)
12. Barbour, A.D., Eagleson, G.K.: Poisson approximation for some statistics based on exchangeable trials. Adv. Appl. Probab. **15**(3), 585–600 (1983)
13. Barbour, A.D., Jensen, J.L.: Local and tail approximations near the Poisson limit. Scand. J. Stat. **16**, 75–87 (1989)

14. Bernstein, S.N.: Sur l'extensiori du theoreme limite du calcul des probabilites aux sommes de quantites dependantes. Math. Ann. **97**, 1–59 (1926)
15. Borisov, I.S., Vorozheikin, I.S.: Accuracy of approximation in the Poisson theorem in terms of χ^2 distance. Sib. Math. J. **49**(1), 5–17 (2008)
16. Čekanavičius, V.: Remarks on infinitely divisible approximations to the Binomial law. In: Grigelionis, B. et al. (eds) Probability Theory and Mathematical Statistics. Proceedings 7th Vilnius Conference, VSP/TEV, pp. 135–146 (1999)
17. Čekanavičius, V.: On compound Poisson approximations under moment restrictions. Theory Probab. Appl. **44**(1), 18–28 (1999)
18. Çekanavičius, V., Vaitkus, P.: A centered Poisson approximation via Stein's method. Lith. Math. J. **41**(4), 319–329 (2001)
19. Čekanavičius, V., Novak, S.Y.: Compound Poisson approximation. Probab. Surv. **19**, 1–80 (2022)
20. Deheuvels, P., Pfeifer, D.: A semigroup approach to Poisson approximation. Ann. Probab. **14**(2), 663–676 (1986)
21. Franken, P.: Approximation der verteilungen von summen unabhängiger nichtnegativer ganzzahliger zufallsgrössen durch Poissonsche verteilunged. Math. Nachr. **27**, 303–340 (1964)
22. Gerber, H.U.: An Introduction to Mathematical Risk Theory. Huebner Foundation, Philadelphia (1979)
23. Gnedenko, B.V., Kolmogorov, A.N.: Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Cambridge (1954)
24. Khintchine, A.Y.: Asymptotische Gesetze der Wahrscheinlichkeitsrechnung Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin (1933)
25. Kerstan, J.: Verallgemeinerung eines satzes von Prochorow und Le Cam. Z. Wahrsch. Verw. Gebiete **2**, 173–179 (1964)
26. Khintchine, A.Y.: Limit laws for sums of random variables. ONTI, Moscow (1938). (**in Russian**)
27. Kolmogorov, A.N.: Two uniform limit theorems for sums of independent random variables. Theory Probab. Appl. **1**(4), 384–394 (1956)
28. Kruopis, J.: Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures. Lith. Math. J. **26**(1), 37–49 (1986)
29. Leadbetter, M.R., Lindgren, G., Rootzen, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983)
30. LeCam, L.: An approximation theorem for the Poisson binomial distribution. Pacif. J. Math. **19**(3), 1181–1197 (1960)
31. Mattner, L., Roos, B.: A shorter proof of Kanter's Bessel function concentration bound. Probab. Theory Relat. Fields **139**, 407–421 (2007)
32. Meshalkin, L.D.: On the approximation of polynomial distributions by infinitely-divisible laws. Theory Probab. Appl. **5**(1), 114–124 (1960)
33. Novak, S.Y.: Extreme Value Methods with Applications to Finance. Chapman & Hall/CRC Press, London. ISBN 9781439835746 (2011)
34. Novak, S.Y.: On the accuracy of Poisson approximation. Extremes **22**, 729–748 (2019)
35. Novak, S.Y.: Poisson approximation. Probab. Surv. **16**, 228–276 (2019)
36. Novak S.Y.: Poisson approximation. Addendum. Probab. Surv. **18**, 272–275 (2021)
37. Novak, S.Y.: Poisson approximation in terms of the Gini–Kantorovich distance. Extremes **24**(1), 67–84 (2021)
38. Presman, E.L.: Approximation of binomial distributions by infinitely divisible ones. Theory Probab. Appl. **28**, 393–403 (1983)
39. Presman, E.L.: On Poisson approximation in total variation for a sum of independent Bernoulli random variables. Theory Probab. Appl. **30**(2), 391–396 (1985)
40. Prokhorov, Y.V.: Asymptotic behavior of the binomial distribution. Uspehi Matem. Nauk **8**(3), 135–142 (1953)
41. Romanowska, M.: A note on the upper bound for the distribution in total variation between the binomial and the Poisson distribution. Stat. Neerlandica **31**, 127–130 (1977)
42. Rogozin, B.A.: On evaluating the concentration functions. Theory Probab. Appl. **6**(1), 94–97 (1961)
43. Röllin, A.: Approximation of sums of conditionally independent variables by the translated Poisson distribution. Bernoulli **11**, 1115–1128 (2005)
44. Roos, B.: Asymptotic and sharp bounds in the Poisson approximation to the Poisson-Binomial distribution. Bernoulli **5**(6), 1021–1034 (1999)
45. Roos, B.: Sharp constants in the Poisson approximation. Stat. Probab. Lett. **52**, 155–168 (2001)

46. Shorgin, S.Y.: Approximation of a generalized binomial distribution. Theory Probab. Appl. **22**(4), 846–850 (1977)
47. Tsaregradskii, I.P.: On uniform approximation of the binomial distribution with infinitely divisible laws. Theory Probab. Appl. **3**(4), 470–474 (1958)
48. Zaitsev, A.Y.: On the accuracy of approximation of distributions of sums of independent random variables which are nonzero with a small probability by means of accompanying laws. Theory Probab. Appl. **28**(4), 657–669 (1983)
49. Zaitsev, A.Y.: An example of a distribution whose set of n-fold convolutions is uniformly separated from the set of infinitely divisible laws in distance in variation. Theory Probab. Appl. **36**(2), 419–425 (1991)