On the accuracy of Poisson approximation

S.Y.Novak

MDX University London, UK

Extremes, v. 22, 2019.

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. From a practical point of view, it 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.

The task of establishing an estimate with the best possible constant at the leading term remained open for decades. The paper presents a solution to that problem. A first-order asymptotic expansion is established as well.

We generalise and sharpen the corresponding inequalities of Prokhorov, LeCam, Barbour, Hall, Deheuvels, Pfeifer, and Roos. A new result is established for the intensively studied topic of Poisson approximation to the binomial distribution.

Key words: Poisson approximation, Kolmogorov's problem, total variation distance.
AMS Subject Classification: 60E15.

1 Introduction

Let $X_1, X_2, ..., X_n$ be integer-valued random variables (r.v.s). Denote

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

The task of approximating the distribution of a sum of independent random variables lies at the hart of the probability theory. The central role plays the normal approximation. However, in many situations Poisson approximation is preferable (cf. [2, 42]).

Interest to the topic of Poisson approximation arises in connection with applications in extreme value theory, insurance, reliability theory, etc. (cf. [3, 4, 24, 32, 38]).

Poisson approximation appears naturally in extreme value theory and other situations where one deals with the distribution of a large number of rare events [4, 32, 38].

Let $X_{k,n}$ denote the k-th largest sample element, and let $N_n(x) = \sum_{i=1}^n \mathbb{1}\{X_i > x\}$ be the number of exceedances of threshold x. Then

$$\{X_{k,n} \leq x\} = \{N_n(x) < k\}.$$

In applications indicators $\mathbb{1}\{X_i > x\}$ can be dependent. A well-known approach consists of grouping observations into blocks which can be considered almost independent [12]. 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 reinsurance 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 a threshold x [22]. More information concerning applications can be found in [3, 4, 22, 32].

In 1950s Kolmogorov has formulated a question about the accuracy of approximation of the distribution of a sum 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., [1, 39, 42] and references therein).

From a theoretical point of view, the question about the accuracy of Poisson approximation is a particular case of Kolmogorov's problem. Besides, there is a connection between the topics of Poisson and compound Poisson approximation [41, 56, 57].

In a range of situations both normal and (compound) Poisson approximations can be applicable (cf. [1, 2, 42]). Due to the complex structure of the compound Poisson distribution, in applications one often would prefer normal or pure Poisson approximation.

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.

The problem of evaluating the accuracy of normal approximation was raised by Liapunov [35]. It lead to a vast literature with contributions from many renowned authors (see, e.g., [1, 38, 49] and references therein).

The problem of evaluating the accuracy of Poisson approximation to the binomial distribution goes back to Prokhorov [42]. The problem attracted a lot of attention among specialists (see, e.g., [8, 9, 13, 39, 41, 45, 50, 55] and references therein).

Let π_λ denote a Poisson $\mathbf{\Pi}(\lambda)$ r.v.. In the case of independent 0-1 r.v.s $\{X_i\}$ we set

$$p_i = \mathbb{P}(X_i = 1) \;\; (i \geq 1), \quad p_n^* = \max_{i \leq n} p_i, \quad \theta = \sum_{i=1}^n p_i^2/\lambda.$$

Many authors worked on the problem of evaluating the total variation distance $d_{TV}(S_n;\pi_\lambda)$ when $\{X_i\}$ are 0-1 r.v.s (see, e.g., [4, 39] and references wherein).

In the case of independent and identically distributed (i.i.d.) Bernoulli $\mathbf{B}(p)$ r.v.s Prohorov [42] has established the existence of an absolute constant c such that

$$d_{TV}(S_n;\pi_{np}) \leq cp. \tag{1}$$

LeCam [33] has shown that

$$d_{TV}(S_n;\pi_\lambda) \leq 4.5 p_n^*, \quad d_{TV}(S_n;\pi_\lambda) \leq 8\theta \;\text{ if } p_n^* \leq 1/4.$$

Kolmogorov ([30], Lemma 5) points out that

$$d_{TV}(S_n;\pi_\lambda) \leq C \sum_{i=1}^n p_i^2, \tag{2}$$

where C is an absolute constant. LeCam [33, 34] attributes inequality

$$d_{TV}(S_n;\pi_\lambda) \leq \sum_{i=1}^n p_i^2 \tag{3}$$

to Khintchin [28]. Kerstan [27] has shown that

$$d_{TV}(S_n;\pi_\lambda) \le 1.05\sum_{i=1}^n p_i^2/\lambda \tag{4}$$

if $p_n^* := \max_{i\le n} p_i \le 1/4$. Romanowska [43] has noticed that

$$d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)) \le p/2\sqrt{1-p}\,. \tag{5}$$

We set $(1-e^{-x})/x = 1$ if $x = 0$. Estimate

$$d_{TV}(S_n;\pi_\lambda) \le \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n p_i^2 \tag{6}$$

is a straightforward consequence of Lemma 4 in Barbour & Eagleson [8]. More estimates of the accuracy of Poisson approximation can be found [19, 39, 41, 44, 55], see also references therein.

An estimate of $d_{TV}(S_n;\pi_\lambda)$ with correct (the best possible) constant at the leading term has been found by Roos [45]:

$$d_{TV}(S_n;\pi_\lambda) \le 3\theta/4e(1-\sqrt{\theta}\,)^{3/2} \tag{7}$$

(see also [17]). In the case of $\mathcal{L}(S_n) = \mathbf{B}(n,p)$ the right-hand side (r.-h.s.) of (7) is

$$3p/4e(1-\sqrt{p}\,)^{3/2} \ge \tfrac{3}{4e}p(1+1.5\sqrt{p}+3.75p).$$

It is shown in [45] that constant $3/4e$ at the main term in (7) cannot be improved.

The rate of the second-order term of the right-hand side of estimate (7) has been improved by Novak [38], Theorem 4.12:

$$d_{TV}(S_n;\pi_\lambda) \le 3\theta/4e + 2\delta^*\varepsilon + 2\delta^2, \tag{8}$$

where $\varepsilon = \min\{1;(2\pi[\lambda-p_n^*])^{-1/2} + 2\delta/(1-p_n^*/\lambda)\}$, $p_n^* = \max_{i\le n} p_i$,

$$\delta = \Big(1-e^{-\lambda}\Big)\sum_{i=1}^n p_i^2/\lambda,\ \ \delta^* = \Big(1-e^{-\lambda}\Big)\sum_{i=1}^n p_i^3/\lambda.$$

Note that $\delta^2 \le \delta^*$. If $\mathcal{L}(S_n) = \mathbf{B}(n,p)$, then (8) becomes

$$d_{TV}(S_n;\pi_{np}) \le 3p/4e + 2(1-e^{-\lambda})p^2\varepsilon + 2(1-e^{-\lambda})^2p^2, \tag{8*}$$

where $\lambda = np$, $\varepsilon = \min\{1;(2\pi[(n-1)p])^{-1/2} + 2(1-e^{-\lambda})p/(1-1/n)\}$. The second-order term in (8^*) is of order $p^2 \wedge np^3$.

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 [5, 10, 6, 21, 27, 38]. W.l.o.g. we may assume that $\mathbb{E}X_i > 0$ $(\forall i)$.

Franken [21] has shown that

$$d_K(S_n;\pi_\lambda) \le \frac{2}{\pi}\sum_{i=1}^n (\mathbb{E}^2 X_i + \mathbb{E}X_i(X_i-1)),$$

where $d_K(X;Y) = \sup_x |F_X(x) - F_Y(x)$ denotes the uniform distance between the distributions of random variables X and Y with distribution functions (d.f.s) F_X and F_Y.

Denote $\lambda^* = \sum_{i=1}^n \mathbb{P}(X_i=1)$, $\lambda_2^* = \sum_{i=1}^n \mathbb{P}(X_i=1)^2$. Kerstan [27] has proved that

$$d_{TV}(S_n;\pi_{\lambda^*}) \le \sum_{i=1}^n \mathbb{P}(X_i \ge 2) + \min\{\lambda_2^*; 1.05\lambda_2^*/\lambda^*\}.$$

Inequalities (3) and (6) have been generalised to the case of non-negative integer-valued r.v.s:

$$d_{TV}(S_n;\pi_\lambda) \le \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n \mathbb{E}|X_i - X_i^*|\mathbb{E}X_i, \tag{9}$$

where X_i^* denotes a random variable with the distribution

$$\mathbb{P}(X_i^*=m) = (m+1)\mathbb{P}(X_i=m+1)/\mathbb{E}X_i \qquad (m \ge 0) \tag{10}$$

(see [38], ch. 4). Distribution (10) differs by a shift from the distribution introduced by Stein [51], p. 171. Hereinafter random variables X_i and X_i^* can be considered defined on a common probability space (this includes the case where X_i^* is independent of X_i).

Note that $X^* \stackrel{d}{=} X$ if and only if $\mathcal{L}(X)$ is Poisson. If $\{X_i\}$ are Bernoulli $\mathbf{B}(p_i)$ r.v.s, then $X_i^* \equiv 0$, and (9) entails (6).

In Theorem 1 below we derive a (8)-type bound for a sum of independent non-negative integer-valued random variables with a correct constant at the leading term.

A number of authors approximated $\mathcal{L}(S_n)$ by unit measures (signed measures) in order to achieve a higher rate of the accuracy of approximation (cf. [14, 11, 16, 6]). We shall understand by unit measures only those unit measures that are not probability distributions. Such a unit measure μ obey $\mu(A) < 0$ $(\exists A \subset \mathbf{Z})$, where $\mathbf{Z}$ denotes the set of integer numbers. An example of a unit measure is Poisson measure $\mathbf{\Pi}_s$ with $s<0$: $\mathbf{\Pi}_s(k) = e^{-s}s^k/k!$ as $k \in \mathbf{Z}_+$, where $\mathbf{Z}_+ = \{0,1,2,...\}$ is the set of non-negative integer numbers.

Theorem 2 below presents an estimate of the accuracy of approximation to $\mathcal{L}(S_n)$ by a particular unit measure.

One can consider also shifted (translated) Poisson approximation. Let

$$\sigma^2 = \mathrm{var} S_n,\ a = [\lambda - \sigma^2],\ b = \{\lambda - \sigma^2\},\ \mu = \sigma^2 + b, \tag{11}$$

where $[x] = \max\{k \in \mathbf{Z}: k \le x\}$ and $\{x\} = x - [x]$.

Barbour & Čekanavičius [6] have shown that

$$d_{TV}(S_n; a+\pi_\mu) \le (1 \wedge \sigma^{-2})\Big(b + d_n \sum_{i=1}^n \psi_i\Big) + \mathbb{P}(S_n < a), \tag{12}$$

where $d_n = \max_{i\le n} d_{TV}(S_{n,i}; S_{n,i}+1)$, $S_{n,i} = S_n - X_i$, $\sigma_i^2 = \mathrm{var} X_i$,

$$\psi_i = \sigma_i^2 \mathbb{E} X_i(X_i-1) + |\mathbb{E}X_i - \sigma_i^2| \mathbb{E}(X_i-1)(X_i-2) + \mathbb{E}|X_i(X_i-1)(X_i-2)|.$$

If $\{X_i\}$ are i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s, then $\mu = npq + \{np^2\}$, and (12) yields

$$d_{TV}(S_n; a+\pi_\mu) \le (1 \wedge 1/npq)\left(\{np^2\} + 2np^2 q d_n\right) + \mathbb{P}(S_n < [np^2]), \qquad (12^*)$$

where $q = 1-p$. Note that $d_n \le 1/\sqrt{(n-1)p}$ if $p \le 1/2$ (see Proposition 4.6 in [11]).

Let $\{X_i\}$ be independent Bernoulli $\mathbf{B}(p_i)$ r.v.s, $\lambda_2 = \lambda - \sigma^2$, a, b, μ, σ^2 are defined in (11). Čekanavičius & Vaitkus [18] have shown that

$$d_{TV}(S_n; a+\pi_\mu) \le \{\lambda_2\}/(\sigma^2 + \{\lambda_2\}) + 0.93\sigma^{-3}\lambda_2 + \mathbb{P}(S_n < a) \qquad (13)$$

and $\mathbb{P}(S_n < a) \le e^{-\sigma^2/4}$. If $p_i = p$ $(\forall i)$, then (13) becomes

$$d_{TV}(S_n; [np^2] + \pi_{npq+\{np^2\}}) \le \{np^2\}/npq + 0.93\sqrt{p/nq} + e^{-npq/4}. \qquad (13^*)$$

Theorem 3 below removes the terms $\mathbb{P}(S_n < a)$, $e^{-\sigma^2/4}$, and sharpens the constants at the main terms in (12) and (13); the moment assumption is weaker than that in (12).

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

$$d_{TV}(X;Y) = \sup_{h\in\mathcal{S}} [\, \mathbb{E}h(X) - \mathbb{E}h(Y)\,] \qquad (*)$$

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

Asymptotic expansions for $\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda)$ have been given by Barbour [5]. Barbour & Jensen [10] have considered the case $h \in \ell_1$. Asymptotic expansions in the case of independent 0-1 r.v.s and unbounded functions h are given by Barbour et al. [7] and Borisov & Rouzankin [13].

The problem of establishing an estimate of the accuracy of Poisson approximation to the distributions of a sum of independent integer-valued non-negative r.v.s in terms of the total variation distance with a correct constant at the leading term remained open for a long while. In particular, an open question was whether $3/4e$ would remain the best possible constant.

Below we give the affirmative answer to that question. We generalise and sharpen the corresponding results from [5, 6, 7, 13, 18]. An estimate of the total variation distance with a correct constant at the leading term in the case of integer-valued r.v.s seems to be established for the first time.

2 Results

Let $\{X_i\}_{i\ge 1}$ be independent non-negative integer-valued r.v.s.

Recall the definitions of two distances we use below. The equivalent definition of the total variation distance between the distributions of r.v.s X and Y is

$$d_{TV}(X;Y) \equiv d_{TV}(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_{A\subset \mathbf{Z}_+} \Big(\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)\Big).$$

The Gini–Kantorovich distance between the distributions of r.v.s X and Y with finite first moments is defined as

$$d_G(X;Y) = \inf_{\tilde X,\tilde Y} \mathbb{E}|\tilde X - \tilde Y|, \qquad (14)$$

where the infimum is taken over all random pairs $(\tilde X, \tilde Y)$ with the marginal distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ respectively. Barbour et al. [4] call it the "Wasserstein distance" after Dobrushin [20] attributed it to Vasershtein [54].

Distance d_G was introduced by Gini [23] and Kantorovich [29] (to be precise, Kantorovich has introduced a class of distances that includes (14)). If X and Y take values in $\mathbf{Z}_+$, then [47]

$$d_G(X;Y) = \sum_{k\ge 1} |\mathbb{P}(X<k) - \mathbb{P}(Y<k)|\,.$$

Taking into account (14), estimate (9) can be rewritten as

$$d_{TV}(S_n;\pi_\lambda) \le \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n d_G(X_i;X_i^*)\mathbb{E}X_i. \qquad (9^*)$$

Given a random variable X and a random pair (ξ,η) with finite second moments, let

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

We set $\varepsilon^*_\lambda = 1\wedge 1/\sqrt{2\pi[\lambda]}$, $\lambda_i = \lambda - \mathbb{E}X_i$, $u_i = 1 - d_{TV}(X_i;X_i+1)$, $U = \sum_{i=1}^n u_i$, $X_0 := 0$,

$$\begin{aligned}
\varepsilon_1 &= \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n \min\{2\mathbb{E}|X_i - X_i^*|; \gamma_{X_i,X_i^*}\varepsilon_{i,n}\}\mathbb{E}X_i,\\
\varepsilon_2 &= 2\lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n \mathbb{E}X_i|\kappa_{X_i}|\varepsilon_{i,n}, \quad \varepsilon_3 = 2\lambda^{-1}(1-e^{-\lambda})|\kappa_{S_n}|\varepsilon^+_{0,n},\\
\varepsilon_{i,n} &= 1\wedge \sqrt{2/\pi}\Big/\Big(1/4 + U - u_i\Big)^{1/2} \wedge \Big(\varepsilon^*_{\lambda_i} + 2\varepsilon^+_{i,n}\Big), \;\; \varepsilon^+_{i,n} = \frac{1-e^{-\lambda}}{\lambda_i}\sum_{j=1}^n d_G(X_j;X_j^*)\mathbb{E}X_j.
\end{aligned}$$

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

$$d_{TV}(S_n;\pi_\lambda) \le 3\theta^*/4e + \varepsilon_1 + \varepsilon_2 + \varepsilon_3, \qquad (15)$$

where $\theta^* = \Big|\sum_{i=1}^n \kappa_{X_i}\Big|/\lambda$.

In the case of independent and identically distributed r.v.s (15) becomes

$$d_{TV}(S_n;\pi_\lambda) \le \tfrac{3}{4e}|\kappa_X|/\mathbb{E}X + (1-e^{-\lambda})\varepsilon_X, \qquad (15^*)$$

where $\varepsilon_X = \min\{2\mathbb{E}|X - X^*|; \varepsilon_{1,n}\gamma_{X,X^*}\} + 2|\kappa_X|\varepsilon_{1,n} + 2(1-e^{-\lambda})|\kappa_X|d_G(X;X^*)/\mathbb{E}X$.

Theorem 1 generalises inequality (8) to the case of integer-valued r.v.s: if $\{X_i\}$ are 0-1 r.v.s, then $\varepsilon_1 = 0$, and (15) entails (8). Constant $3/4e$ in (15) cannot be improved.

The moment assumption 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 $X_i' = X_i 1\!\!\mathrm{I}\{X_i \le K_i\}$) since $d_{TV}((X_1,...,X_n);(X_1',...,X_n')) \le \sum_{i=1}^n \mathbb{P}(X_i > K_i)$.

Note that the terms ε_2 and ε_3 in (15) vanish if $\kappa_{X_i} = 0$ (equivalently, $\mathbb{E}X_i = \mathbb{E}X_i^*$) for all i (cf. Example 2).

Quantity $1-(\mathrm{var}\ S_n)/\mathbb{E}S_n$ appears naturally in the theory of Poisson approximation (cf. [4], p. 49). Theorems 1 and 2 highlight that individual quantities κ_{X_i} play central role in the bound to the accuracy of Poisson approximation when one deals with a sum of integer-valued r.v.s.

Example 1. Let $X, X_1, X_2, ...$ be independent geometric $\mathbf{\Gamma_0}(p)$ r.v.s:

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

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

It is easy to see that $\mathbb{P}(X_i^* = m) = (m+1)p^m(1-p)^2$. Hence

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

With $\lambda := np/(1-p)$, $r := p/(1-p)$, (9) entails the estimate

$$d_{TV}(S_n;\pi_\lambda) \le \left(1-e^{-np/(1-p)}\right)r, \tag{16}$$

which is due to Barbour [5]. It is easy to check that

$$\begin{aligned} \varepsilon_1 &\le 2\left(1-e^{-nr}\right)\min\{r;2r^2\},\ \varepsilon_{1,n} \le \varepsilon^\star_{n,p}\,, \\ \varepsilon_2 &= 2\left(1-e^{-nr}\right)r^2\varepsilon_{n,p}\,,\ \varepsilon_3 = 2\left(1-e^{-nr}\right)^2 r^2, \end{aligned}$$

where

$$\varepsilon^\star_{n,p} = 1 \wedge \sqrt{2/\pi}\Big/\left(1/4+(n-1)p\right)^{1/2} \wedge \left(1/\sqrt{2\pi[(n-1)p]} + 2r/(1-1/n)\right).$$

If $p \le 1/2$, then (15) yields

$$d_{TV}(S_n;\pi_\lambda) \le 3r/4e + \left(1-e^{-nr}\right)\left(2+6\varepsilon^\star_{n,p}\right)r^2, \tag{17}$$

which is sharper than (16) if $p < \left(1-3/4e(1-e^{-\lambda})\right)\Big/\left(3-3/4e(1-e^{-\lambda})+6\varepsilon^\star_{n,p}\right)$.

Estimate (17) has the correct constant $3/4e$ at the leading term. The best estimate of $d_{TV}(\mathbf{NB}(n,p);\mathbf{\Pi}(np))$ is due to Roos [46]:

$$d_{TV}(\mathbf{NB}(n,p);\mathbf{\Pi}(np)) \le \min\{3r/4e; nr^2\}. \tag{18}$$

Example 2. Let $X_1, X_2, ..., X_n$ be independent r.v.s with the distributions

$$\mathbb{P}(X_i = 0) = 1-p_i+p_i^2/2,\ \mathbb{P}(X_i = 1) = p_i - p_i^2,\ \mathbb{P}(X_i = 2) = p_i^2/2.$$

Note that $\mathbb{E}X_i = \mathrm{var}X_i = p_i$. One can check that $d_{TV}(X_i;\pi_{p_i}) \le \frac{2}{3}p_i$. Therefore,

$$d_{TV}(S_n;\pi_\lambda) \le \sum_{i=0}^n d_{TV}(X_i;\pi_{p_i}) \le \frac{2}{3}\sum_{i=1}^n p_i^3. \tag{19}$$

This inequality is due to Deheuvels & Pfeifer [19].

From (10), X_i^* is a Bernoulli $\mathbf{B}(p_i)$ random variable. Note that $\mathbb{E}X_i = \mathbb{E}X_i^*$, $\kappa_{X_i} = 0$ $(\forall i)$ and $\varepsilon_2 = \varepsilon_3 = 0$. One can check that

$$\varepsilon_1 = \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n p_i^3 \varepsilon_{i,n}, \;\; d_{TV}(X_i;X_i^*) = d_G(X_i;X_i^*) = \gamma_{X_i,X_i^*} = p_i^2.$$

Theorem 1 yields

$$d_{TV}(S_n;\pi_\lambda) \le \frac{1-e^{-\lambda}}{\lambda}\sum_{i=1}^n p_i^3 \min\left\{1;\left(\frac{1}{\sqrt{2\pi[\lambda-p_i]}} + 2\frac{1-e^{-\lambda}}{\lambda-p_i}\sum_{j=1}^n p_j^3\right)\right\}, \tag{20}$$

which is sharper than (19) if $2\lambda > 3(1-e^{-\lambda})$. □

Asymptotic expansions to the Binomial distribution $\mathbf{B}(n,p)$ have been given by Uspensky [53] (see also Franken [21]). Herrmann [25], Shorgin [50] and Barbour [5] present full asymptotic expansions with explicit estimates of the error terms, the latter for a sum of independent non-negative integer-valued r.v.s. Asymptotic expansions for $\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda)$ in the case of independent 0-1 r.v.s $\{X_k\}$ and unbounded function h have been given by Barbour et al. [7] and Borisov & Ruzankin [13].

The formulation of the full asymptotic expansions is cumbersome. Considerable attention has been given to first-order asymptotic expansions (see, e.g., Kerstan [27], Kruopis [31], Čekanavičius & Kruopis [16]), Barbour & Čekanavičius [6], Barbour et al. [4]).

The next theorem presents a first-order asymptotic expansion to $\mathcal{L}(S_n)$ in the case of a sum of independent non-negative integer-valued random variables with finite second moments.

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

$$\mathbb{P}(\pi_\lambda^\star = k) = \mathbb{P}(\pi_\lambda = k)(k-\lambda)^2/\lambda \qquad (k \in \mathbf{Z}_+). \tag{21}$$

Theorem 2 *Let $X_1,...,X_n$ be independent non-negative integer-valued random variables with finite second moments. For any function $h \in \mathcal{S}$*

$$\left|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) - (\mathbb{E}h(\pi_\lambda+1) - \mathbb{E}h(\pi_\lambda^\star))\sum_{i=1}^n \kappa_{X_i}/2\lambda\right| \tag{22}$$
$$\le \varepsilon_1 + \varepsilon_2 + \varepsilon_3\,.$$

Theorem 2 sharpens the corresponding bounds in Barbour [5], Barbour et al. [7], Barbour et al. [4], corollary 9.A.1, Barbour & Čekanavičius [6] and Borisov & Rouzankin [13]; the moment assumption is weaker than those in [5, 6].

If $X, X_1, ..., X_n$ are i.i.d.r.v.s, then (22) becomes

$$\Big|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) - (\mathbb{E}h(\pi_\lambda+1) - \mathbb{E}h(\pi^\star_\lambda))\,\kappa_X/2\mathbb{E}X\Big| \le \varepsilon_1 + \varepsilon_2 + \varepsilon_3. \qquad (22^*)$$

The right-hand side of (22*) is typically of order $(\mathbb{E}X)^2$, while the right-hand side of (15*) is typically of order $\mathbb{E}X$.

A straightforward consequence of (22) is the following relation:

$$|d_{TV}(S_n;\pi_\lambda) - d_{TV}(\pi^\star_\lambda;\pi_\lambda+1)\theta^*/2| \le \varepsilon_1 + \varepsilon_2 + \varepsilon_3\,. \qquad (23)$$

Using Stirling's formula, one can check that

$$d_{TV}(\pi^\star_\lambda;\pi_\lambda+1) = \sqrt{2/\pi e} + O\Big(1/\sqrt{\lambda}\Big) \qquad (\lambda \to \infty)$$

(cf. (4.59) in [38]). Hence

$$d_{TV}(S_n;\pi_\lambda) = \theta^*/\sqrt{2\pi e} + O\Big(\theta^*/\sqrt{\lambda} + \varepsilon_1 + \varepsilon_2 + \varepsilon_3\Big). \qquad (23^*)$$

Inequality (23*) generalises the corresponding results of Prokhorov [42] and Deheuvels & Pfeifer [19] to the case of non-negative integer-valued r.v.s.

Remark 1. Note that

$$\mathbb{E}h(\pi_\lambda+1) - \mathbb{E}h(\pi^\star_\lambda) = -\lambda\mathbb{E}\Delta^2 h(\pi_\lambda).$$

Thus, (22) can be rewritten as

$$\Big|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) + \mathbb{E}\Delta^2 h(\pi_\lambda)\sum_{i=1}^n \kappa_{X_i}/2\Big| \le \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \qquad (h\in\mathcal{S}). \qquad (22^\star)$$

Remark 2. Denote by P_n the unit measure

$$P_n(\cdot) = \mathbb{P}(\pi_\lambda = \cdot) + \Big(\mathbb{P}(\pi_\lambda+1 = \cdot) - \mathbb{P}(\pi^\star_\lambda = \cdot)\Big)\sum_{i=1}^n \kappa_{X_i}/2\lambda.$$

Theorem 2 states that

$$\|\mathcal{L}(S_n) - P_n\| \le 2(\varepsilon_1 + \varepsilon_2 + \varepsilon_3),$$

where $\|\cdot\|$ denotes the total variation norm.

The next theorem deals with the approximation of S_n by $[\kappa_{S_n}]+\pi_\mu$ (shifted Poisson approximation), where

$$\sigma^2 = \operatorname{var} S_n,\ \ \mu = \operatorname{var} S_n + \{\kappa_{S_n}\}.$$

Note that

$$[\kappa_{S_n}] + \mathbb{E}\pi_\mu = \mathbb{E}S_n,\ \ |\operatorname{var}\pi_\mu - \operatorname{var} S_n| < 1.$$

Let $\{\tilde{X}_j\}$ denote independent copies of $\{X_j\}$, $\bar{Y} := Y - \mathbb{E}Y$, $\varepsilon_n = \varepsilon_{0,n}$,

$$\hat{\varepsilon}_\mu \;=\; \min\Big\{\mu^{-1}(1-e^{-\mu});\bar{\varepsilon}_\mu\Big\}, \qquad \bar{\varepsilon}_\mu = 2\varepsilon^*_\mu\sqrt{2/e\mu} + 2\mu^{-1}(1-e^{-\mu})\varepsilon^\star_\mu,$$

$$\varepsilon^\star_\mu \;=\; \mu^{-1}(1-e^{-\mu})|\{\kappa_{S_n}\}| + \varepsilon^\#_n,\ \ \varepsilon^\#_n = 2\mu^{-1}(1-e^{-\mu})\sum_{i=1}^n \sigma_i^2\varepsilon_{i,n}\mathbb{E}X_i + \sum_{i=1}^n \delta^\mu_{X_i,\tilde{X}_i},$$

$$\delta^\mu_{X_i,\tilde{X}_i} \;=\; \mu^{-1}(1-e^{-\mu})\min\Big\{2\mathbb{E}|\bar{X}_i||X_i - \tilde{X}_i|;\,\mathbb{E}|\bar{X}_i||X_i(X_i-1) - \tilde{X}_i(\tilde{X}_i-1)|\varepsilon_{i,n}\Big\}\,.$$

Theorem 3 *If* $X_1, ..., X_n$ *are independent non-negative integer-valued random variables with finite second moments, then*

$$d_{TV}(S_n; [\kappa_{S_n}] + \pi_\mu) \le |\{\kappa_{S_n}\}| \hat{\varepsilon}_\mu + \varepsilon_n^{\#}. \tag{24}$$

Inequality (24) is sharp in the following sense: if all $\{X_j\}$ are constants, then (24) becomes the equality. If $\{X_j\}$ are i.i.d. non-degenerate r.v.s, then typically $\mu \to \infty$ as $n \to \infty$, $\hat{\varepsilon}_\mu \sim 2/\mu\sqrt{\pi e}$, $\varepsilon_n^{\#} \sim 2(1/\sqrt{2\pi\mu} + 2d_G(X; X^*))\mathbb{E}X$.

Estimate (24) has advantages over the Berry–Esseen bound (cf. (25) below).

Example 3. Let $\mathcal{L}(S_n) = \mathbf{B}(n, p)$, and set $q = 1-p$. Clearly,

$$\mu = npq + \{np^2\}, \ \kappa_{S_n} = np^2, \ \delta^{\mu}_{X_i, \tilde{X}_i} = 0.$$

Estimate (24) entails

$$d_p^{(n)} \equiv d_{TV}\Big(S_n; [np^2] + \pi_{npq+\{np^2\}}\Big) \le \{np^2\}\bar{\varepsilon}_\mu + 2p\varepsilon_{1,n}, \tag{24*}$$

where $\varepsilon_{1,n} \le \min\Big\{\frac{\sqrt{2/\pi}}{\sqrt{1/4+(n-1)p}}; \frac{1}{\sqrt{2\pi[(n-1)p]}} + 2p\frac{1-e^{-np}}{1-1/n}\Big\}$.

The term $2p\varepsilon_{1,n}$ appears here because of estimate (27). $\mathbf{B}(n, p)$ is known to be unimodal. For the unimodal distribution we can apply (27′), i.e., replace $\varepsilon_{i,n}$ with $\varepsilon'_{i,n} = \max_k \mathbb{P}(S_{n,i} = k)$:

$$d_p^{(n)} \le \{np^2\}\bar{\varepsilon}_\mu + 2p\varepsilon_n^{o}, \tag{24+}$$

where $\varepsilon_n^{o} := \min\{\varepsilon_{1,n}; \varepsilon'_{1,n}\}$. Note that

$$\begin{aligned} \varepsilon^\star_\mu &\le \{np^2\}/npq + 2p\varepsilon^o_n, \ \varepsilon^*_\mu \le 1/\sqrt{2\pi[npq]}\ , \\ \bar{\varepsilon}_\mu &\le 2/[npq]\sqrt{e\pi}\ + 2\varepsilon^\star_\mu/npq \le 2/[npq]\sqrt{e\pi}\ + 2\{np^2\}/(npq)^2 + 4\varepsilon^o_n/nq. \end{aligned}$$

Thus,

$$d_p^{(n)} \le \frac{2\{np^2\}}{\sqrt{e\pi}\,[npq]} + 2\frac{\{np^2\}^2}{(npq)^2} + 4\frac{\{np^2\}}{nq}\varepsilon^o_n + 2p\varepsilon^o_n. \tag{24o}$$

In particular, if $p = p(n)$ obeys $n^{-1} \ll p \ll n^{-1/3}$, then

$$\bar{\varepsilon}_\mu \le 2/np\sqrt{\pi e}(1+o(1)), \ \varepsilon_{1,n} \sim 1/\sqrt{2\pi np}\ , \ \varepsilon^\star_\mu \le \{np^2\}/np(1+o(1)),$$

and hence

$$d_p^{(n)} \le \left(\frac{2}{\sqrt{\pi e}}\frac{\{np^2\}}{np} + \frac{\sqrt{2p}}{\sqrt{\pi n}}\right)(1+o(1)). \tag{24⋆}$$

Constants in the r.-h.s. of (24⋆) are better than those in (12*), (13*).

Note that one only needs to consider $p \le 1/2$ since $\mathcal{L}(n - S_n) = \mathbf{B}(n, 1-p)$ (cf. [42]). Evaluating $d_p^{(n)}$ separately for $p < 1/\sqrt{n}$ and $p \in [1/\sqrt{n}; 1/2]$ yields

$$\sup_{0 \le p \le 1/2} d_p^{(n)} \le \frac{2/\sqrt{\pi e}}{\sqrt{n} - 2} + \frac{0.9n^{1/4}}{n-1} + \frac{2 + 1.8/n^{1/4}}{(\sqrt{n} - 1)^2} \qquad (n > 4). \tag{25}$$

Indeed, $d_p^{(n)} \leq 3/4e\sqrt{n} + 4/n$ by (8^*) if $p < 1/\sqrt{n}$. If $1/\sqrt{n} \leq p \leq 1/2$, we denote by $f(p)$ the r.-h.s. of (24^o) with ε_n^o replaced by $\varepsilon'_{1,n}$. Note that $\varepsilon'_{1,n} \leq 0.45/\sqrt{(n-1)p}$ (see (2.22) in [18]). Function f on $[1/\sqrt{n}\,; 1/2]$ first declines, then grows. Hence $\sup_{1/\sqrt{n} \leq p \leq 1/2} f(p) = \max\{f(1/\sqrt{n}); f(1/2)\} = f(1/\sqrt{n})$, yielding (25).

Bound (25) is preferable to (6) – (8^*) if $p \geq 8\sqrt{e}/3\sqrt{\pi n}\,(1+o(1))$. Inequality (25) has advantages over Meshalkin's [37] and Presman's [40] results as only estimates with explicit constants matter in applications; besides, the structure of the approximating distribution in (25) is simpler and does not involve r.v.s that allow for negative values.

For all small enough p the right-hand side of estimate (25) is sharper than that of the Berry–Esseen inequality. Note that estimate (25) has another advantage over the Berry–Esseen inequality: a uniform in $p \in (0;1/2]$ Berry–Esseen bound is infinite. □

Given a function $f: \mathbf{Z}_+ \to \mathbb{R}$, we denote

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

$(\ell \geq 0, m \geq 0, k \geq 0)$.

Proposition 4 *For any function* $f: \mathbf{Z}_+ \to \mathbb{R}$ *and any* $\ell \geq 0, m \geq 0, k \geq 0$,

$$|R_f(m,k,\ell)| \leq \delta_{m,k}^{\ell}. \tag{26}$$

Relation (26) is a discrete analogue of Teylor's formula.

Lemma 5 *For any bounded function* $\tilde{f}$

$$|\mathbb{E}\Delta \tilde{f}(S_{n,i})| \leq \min\left\{2\|\tilde{f}\|\varepsilon_{i,n}\,; (\|\Delta \tilde{f}\| \wedge 2\|\tilde{f}\|\varepsilon_{\lambda_i}^*) + 2\|\Delta \tilde{f}\|\varepsilon_{i,n}^+\right\}. \tag{27}$$

Acknowledgements. The author is grateful to the reviewers for helpful remarks.

References

[1] Arak T.V. and Zaitsev A.Yu. (1986) Uniform limit theorems for sums of independent random variables. — Proc. Steklov Inst. Math., v. 174, 3–214.

[2] Arenbaev N.K. (1976) Asymptotic behavior of the multinomial distribution. — Theory Probab. Appl., v. 21, 805–810.

[3] Balakrishnan N. and Koutras M.V. (2001) *Runs and scans with applications.* New York: Wiley.

[4] Barbour A.D., Holst L. and Janson S. (1992) *Poisson approximation.* Oxford: Clarendon Press.

[5] Barbour A.D. (1987) Asymptotic expansions in the Poisson limit theorem. — Ann. Probab., v. 15, No 2, 748–766.

[6] Barbour A.D. and Čekanavičius V. (2002) Total variation asymptotics for sums of independent integer random variables. — Ann. Probab., v. 30, No 2, 509–545.

[7] Barbour A.D., Chen L.H.Y., Choi K.P. (1995). Poisson approximation for unbounded functions, I: Independent summands. Statistica Sinica, v. 5, No 2, 749–766.

[8] Barbour A.D. and Eagleson G.K. (1983) Poisson approximation for some statistics based on exchangeable trials. — Adv. Appl. Probab., v. 15, No 3, 585–600.

[9] Barbour A.D. and Hall P. (1984) On the rate of Poisson convergence. — Math. Proceedings Cambridge Phil. Soc., v. 95, 473–480.

[10] Barbour A.D. and Jensen J.L. (1989) Local and tail approximations near the Poisson limit. — Scand. J. Statist., v. 16, 75–87.

[11] Barbour A.D. and Xia A. (1999) Poisson perturbations. — ESAIM Probab. Statist., v. 3, 131–150.

[12] Bernstein S.N. (1926) Sur l'extensiori du theoreme limite du calcul des probabilites aux sommes de quantites dependantes. – Math. Annalen, v. 97, 1–59.

[13] Borisov I.S. and Ruzankin P.S. (2002) Poisson approximation for expectations of unbounded functions of independent random variables. — Ann. Probab., v. 30, No 4, 1657–1680.

[14] Borovkov K. A. (1988) Renement of Poisson approximation. — Theory Probab.Appl., v. 33, No 2, 343–347.

[15] Brown T.C. (1983) Some Poisson approximations using compensators. — Ann. Probab. 11, 726–744.

[16] Čekanavičius V. and Kruopis J. (2000) Signed Poisson approximation: a possible alternative to normal and Poisson laws. — Bernoulli, v. 6, No 4, 591–606.

[17] Čekanavičius V. and Roos B. (2006) An expansion in the exponent for compound binomial approximations. — Liet. Matem. Rink., v. 46, 67–110.

[18] Čekanavičius V. and Vaitkus P. (2001) A centered Poisson approximation via Stein's method. — Lithuanian Math. J., v. 41, No 4, 319–329.

[19] Deheuvels P. and Pfeifer D. (1986) A semigroup approach to Poisson approximation. — Ann. Probab., v. 14, No 2, 663–676.

[20] Dobrushin R.L. (1970) Prescribing a system of random variables by conditional distributions. — Theory Probab. Appl., v. 15, No 3, 458–486.

[21] Franken P. (1964) Approximation der verteilungen von summen unabhängiger nichtnegativer ganzzahliger zufallsgrössen durch Poissonsche verteilunged. — Mathematische Nachrichten, v. 27, 303–340.

[22] Gerber H.U. (1979). An introduction to mathematical risk theory. — Philadelphia: Huebner Foundation.

[23] Gini C. (1914) Di un misura delle relazioni tra le graduatorie di due caratteri. — In: *Appendix to Hancini A. Le Elezioni Generali Politiche del 1913 nel Comune di Roma.* Rome: Ludovico Cecehini.
[24] Haight F.A. (1967) Handbook of the Poisson distribution. New York: Wiley.
[25] Herrmann H. (1965) Variationsabstand zwischen der Verteilung einer Summe unabhängiger nichtnegativer ganzzahliger Zufallsgrössen und Poissonschen Verteilungen. — Mathematische Nachrichten, v. 29, No 5, 265–289.
[26] Kabanov Yu.M., Liptser R.Sh. and Shiryaev A.N. (1983) Weak and strong convergence of the distributions of point processes. — Theory Probab. Appl., v. 28, No 2, 288–319.
[27] Kerstan J. (1964) Verallgemeinerung eines satzes von Prochorow und Le Cam. — Z. Wahrsch. Verw. Gebiete, v. 2, 173–179.
[28] Khintchin A.Y. (1933) *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete.* — Berlin: Springer.
[29] Kantorovich L.V. (1942) On the translocation of mass.— Doklady USSR Acad. Sci., v. 37, No 7-8, 227–229. Trans: Management Sci. (1958) v. 5, No 1, 1–4.
[30] Kolmogorov A.N. (1956) Two uniform limit theorems for sums of independent random variables. — Theory Probab. Appl., v. 1, No 4, 384–394.
[31] Kruopis J. (1986) Precision of approximations of the generalized binomial distribution by convolutions of Poisson measures. — Lithuanian Math. J., v. 26, 37–49.
[32] Leadbetter M.R., Lindgren G., Rootzen H. (1983) *Extremes and related properties of random sequences and processes.* — New York: Springer.
[33] LeCam L. (1960) An approximation theorem for the Poison binomial distribution. — Pacif. J. Math., v. 19, No 3, p. 1181–1197.
[34] LeCam L. (1965) On the distribution of sums of independent random variables. — In: Proc. Intern. Res. Sem. Statist. Lab. Univ. California, pp. 179–202. New York: Springer.
[35] Liapunov A.M. (1901) Nouvelle forme du theor'eme sur la limite des probabilites. — Mem. Acad. Imp. Sci. St.–Peterburg, v. 12, 1–24.
[36] Mattner L. and Roos B. (2007) A shorter proof of Kanters Bessel function concentration bound. Probab. — Theory and Related Fields, v. 139, 407–421.
[37] Meshalkin L.D. (1960) On the approximation of polynomial distributions by infinitely-divisible laws. — Theory Probab. Appl., v. 5, No 1, 114–124.
[38] Novak S.Y. (2011) *Extreme value methods with applications to finance.* — London: Chapman & Hall/CRC Press. ISBN 9781439835746
[39] Novak S.Y. (2018) Poisson approximation. — https://arxiv.org/abs/1901.01847
[40] Presman E.L. (1983) Approximation of binomial distributions by infinitely divisible ones. — Theory Probab. Appl., v. 28, 393–403.
[41] Presman E.L. (1985) On Poisson approximation in total variation for a sum of independent Bernoulli random variables. — Theory Probab. Appl., v. 30, No 2, 391–396.
[42] Prokhorov Y.V. (1953) Asymptotic behavior of the binomial distribution. — Uspehi Matem. Nauk, v. 8, No 3(55), 135–142.
[43] Romanowska M. (1977) A note on the upper bound for the distribution in total variation between the binomial and the Poisson distribution. — Statist. Neerlandica, v. 31, 127–130.
[44] Roos B. (1999) Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. — Bernoulli, v. 5, No 6, 1021–1034.
[45] Roos B. (2001) Sharp constants in the Poisson approximation. — Statist. Probab. Letters, v. 52, 155–168.
[46] Roos B. (2003) Improvements in the Poisson approximation of mixed Poisson distributions. — J. Statist. Plan. Inference, v. 113, 467–483.

[47] Salvemini T. (1943) Sul calcolo degli indici di concordanza tra due caratteri quantitativi. — Atti della VI Riunione della Soc. Ital. di Statistica.

[48] Serfling R.J. (1975) A general Poisson approximation theorem. — Ann. Probab., v. 3, 726–731.

[49] Shevtsova I.G. (2013) On absolute constants in the Berry–Esseen inequality and its structural and non-uniform refinements. — Informatics Appl., v. 7, No 1, 124–125.

[50] Shorgin S.Y. (1977) Approximation of a generalized binomial distribution. — Theory Probab. Appl., v. 22, No 4, 846–850.

[51] Stein C. (1992) A way of using auxiliary randomization. — In: *Probability Theory.* Proc. Singapore Probab. Conf., pp. 159–180. Berlin: de Gruyter.

[52] Tsaregradskii I.P. (1958) On uniform approximation of the binomial distribution with infinitely divisible laws. — Theory Probab. Appl., v. 3, No 4, 470–474.

[53] Uspensky J.V. (1931) On Ch.Jordan's series for probability. — Ann. Math., v. 32, No 2, 306–312.

[54] Vasershtein (1969) Markov processes on a countable product of spaces describing large automated systems. — Probl. Inform. Trans., v. 14, 64–73.

[55] Xia A. (1997) On using the first difference in the Stein–Chen method. — Ann. Appl. Probab., v. 7, No 4, 899–916.

[56] Xia A. (2015) Stein's method for conditional compound Poisson approximation. — Statist. Probab. Lett., v. 100, 19–26.

[57] Zaitsev A.Yu. (1983) 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., v. 28, No 4, 657–669.