Probability Surveys
Vol. 16 (2019) 228–276
ISSN: 1549-5787
https://doi.org/10.1214/18-PS318

Poisson approximation

S. Y. Novak

MDX University London
e-mail: S.Novak@mdx.ac.uk

Abstract: We overview results on the topic of Poisson approximation that are missed in existing surveys. The main attention is paid to the problem of Poisson approximation to the distribution of a sum of Bernoulli and, more generally, non-negative integer-valued random variables.

We do not restrict ourselves to a particular method, and overview the whole range of issues including the general limit theorem, estimates of the accuracy of approximation, asymptotic expansions, etc. Related results on the accuracy of compound Poisson approximation are presented as well.

We indicate a number of open problems and discuss directions of further research.

MSC 2010 subject classifications: Primary 60E15; secondary 60F05, 60G50, 60G51, 60G55, 60G70, 60J75, 62E17, 62E20.
Keywords and phrases: Poisson approximation, compound Poisson approximation, accuracy of approximation, asymptotic expansions, Poisson process approximation, total variation distance, long head runs, long match patterns.

Received November 2018.

Contents

1. Weak convergence to a Poisson law

Poisson approximation appears natural in situations where one deals with a large number of rare events. The topic has attracted a considerable body of research. It has important applications in insurance, extreme value theory, reliability theory, mathematical biology, etc. (cf. [7, 12, 54, 66, 82]). However, existing surveys are surprisingly sketchy, and miss not only a number of results obtained during the last three decades but even some classical results going back to 1930s.

The paper aims to fill the gap. We present a comprehensive list of results on the topic of Poisson approximation, and formulate a number of open problems. Related results on the topic of compound Poisson approximation are presented as well. The main attention is given to results that are missed in existing surveys.

1.1. Weak convergence to a Poisson law

We denote by $\mathbf{\Pi}(\lambda)$ a Poisson law with parameter λ.

The following Poisson limit theorem is due to Gnedenko [50] and Marcinkiewicz [73]. Hereinafter multiplication is superior to division.

Let $\{X_{n,1},...,X_{n,k_n}\}_{n\geq 1}$, where $\{k_n\}$ is a non-decreasing sequence of natural numbers, be a triangle array of *independent* random variables (r.v.s).

Random variables $\{X_{n,k}\}$ are called *infinitesimal* if

$$\lim_{n\to\infty}\max_{k\leq k_n}\mathbb{P}(|X_{n,k}|>\varepsilon)\to 0 \qquad (\forall\varepsilon>0). \tag{1}$$

Denote $B_\varepsilon = (-\varepsilon;\varepsilon)\cup(1-\varepsilon;1+\varepsilon)$,

$$S_n = X_{n,1}+...+X_{n,k_n}.$$

Theorem 1. [50, 73] *If $\{X_{n,k}\}$ are infinitesimal r.v.s, then*

$$\mathcal{L}(S_n)\Rightarrow\mathbf{\Pi}(\lambda) \qquad (\exists\lambda>0) \tag{2}$$

as $n\to\infty$ if and only if for any $\varepsilon\in(0;1)$, as $n\to\infty$,

$$\sum_k \mathbb{P}(|X_{n,k}-1|<\varepsilon)\to\lambda, \tag{3}$$

$$\sum_k \mathbb{P}(X_{n,k}\notin B_\varepsilon)\to 0, \quad \sum_k \mathbb{E}X_{n,k}\mathbb{1}\{|X_{n,k}|<\varepsilon\}\to 0, \tag{4}$$

$$\sum_k \left(\mathbb{E}X_{n,k}^2\mathbb{1}\{|X_{n,k}|<\varepsilon\} - \mathbb{E}^2X_{n,k}\mathbb{1}\{|X_{n,k}|<\varepsilon\}\right)\to 0. \tag{5}$$

The following corollary presents necessary and sufficient conditions for the weak convergence of a sum of independent and identically distributed (i.i.d.) non-negative integer-valued r.v.s to a Poisson random variable.

Let $\mathbb{N}$ denote the set of natural numbers, and let $\mathbf{Z}_+:=\mathbb{N}\cup\{0\}$.

Corollary 2. *If* $\{X_{n,1},...,X_{n,k_n}\}_{n\geq 1}$ *is a triangle array of independent random variables taking values in* $\mathbf{Z}_+$ *such that* $X_{n,i} \stackrel{d}{=} X_{n,1}$ $(1 \leq i \leq k_n)$, *then (2) holds if and only if*

$$k_n \mathbb{P}(X_{1,n}=1) \to \lambda \quad \text{and} \quad \mathbb{P}(X_{n,1} \geq 2)/\mathbb{P}(X_{n,1} \geq 1) \to 0. \tag{6}$$

Note that (6) yields $\mathbb{P}(X_{n,1} \geq 1) \sim \mathbb{P}(X_{n,1}=1)$ as $n \to \infty$.

The second relation in (6) means $X'_{n,1} \underset{p}{\to} 1$ as $n \to \infty$, where r.v. $X'_{n,1}$ has the distribution $\mathcal{L}(X'_{n,1}) = \mathcal{L}(X_{n,1}|X_{n,1} \neq 0)$.

In the case of Bernoulli $\mathbf{B}(p_{n,k})$ random variables relations (4) and (5) trivially hold, (3) means

$$\sum_k p_{n,k} \to \lambda \qquad (n \to \infty), \tag{3*}$$

while (1) states that $\max_k p_{n,k} \to 0$ as $n \to \infty$. The latter together with (3*) is equivalent to

$$\sum_k p_{n,k}^2 \to 0 \qquad (n \to \infty). \tag{1*}$$

Thus, conditions (1*) and (3*) are necessary and sufficient for the weak convergence (2).

Example 1.1. Let $\{X_{n,1},...,X_{n,n}\}$ be i.i.d. random variables with the distribution

$$\mathbb{P}(X=0) = 1-\lambda/n-1/n^{1.5}, \; \mathbb{P}(X=1) = \lambda/n, \; \mathbb{P}(X=n) = 1/n^{1.5} \qquad (\lambda>0).$$

Then (1) and (6) hold, hence $\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda)$. Note that $\mathbb{E}S_n \not\to \lambda$. □

The proof of Theorem 1 can be found in [52].

A compound Poisson limit theorem (weak convergence of $\mathcal{L}(S_n)$ to a compound Poisson law, where S_n is a sum of i.i.d. random variables that are equal to 0 with a large probability) has been given by Khintchin ([61], ch. 2.3).

1.2. Dependent Bernoulli random variables

The topic of Poisson approximation to the distribution of a sum of *dependent* Bernoulli r.v.s has applications in extreme value theory, reliability theory, etc. (cf. [7, 12, 66, 82]).

Let $\{X_{n,1},...,X_{n,n}\}_{n\geq 1}$ be a triangle array of 0-1 random variables such that sequence $X_{n,1},...,X_{n,n}$ is stationary for each $n \in \mathbb{N}$. For instance, in extreme value theory one often has

$$X_{n,k} = \mathbb{1}\{Y_k > u_n\},$$

where $\{Y_i, i \geq 1\}$ is a *stationary* sequence of random variables and $\{u_n\}$ is a sequence of "high" levels. The special case where $\{Y_i, i \geq 1\}$ is a moving average is related to the topic of the Erdös–Rényi partial sums (cf. [82], ch. 2).

Let $\mathcal{F}_{l,m}(\tau)$ be the σ–field generated by the events $\{X_{n,i}\}$, $l \le i \le m$. Set

$$\alpha_n(l) = \sup |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|, \quad \varphi_n(l) = \sup |\mathbb{P}(B|A) - \mathbb{P}(B)|,$$
$$\beta_n(l) = \sup \mathbb{E} \sup_B |\mathbb{P}(B|\mathcal{F}_{1,m}) - \mathbb{P}(B)|,$$

where the supremum is taken over $m \ge 1$, $A \in \mathcal{F}_{1,m}(\tau)$, $B \in \mathcal{F}_{m+l+1,n}$ such that $\mathbb{P}(A) > 0$. Conditions involving mixing coefficients $\alpha_n(\cdot), \beta_n(\cdot), \varphi_n(\cdot)$ are slightly weaker than those involving traditional mixing coefficients $\alpha(\cdot), \beta(\cdot), \varphi(\cdot)$.

Condition Δ is said to hold if $\alpha_n(l_n) \to 0$ for some sequence $\{l_n\}$ of natural numbers such that $1 \ll l_n \ll n$.

Class $\mathcal{R}$. If Δ holds, then there exists a sequence $\{r_n\}$ of natural numbers such that

$$n \gg r_n \gg l_n \gg 1, \; n r_n^{-1} \alpha_n^{2/3}(l_n) \to 0 \qquad (n \to \infty) \tag{7}$$

(for instance, one can take $r_n = \left[\sqrt{n \max\{l_n; n\alpha_n(l_n)\}}\,\right]$). We denote by $\mathcal{R}$ the class of all such sequences $\{r_n\}$.

Set

$$S_n = X_{n,1} + \ldots + X_{n,n}, \quad \lambda_n = \mathbb{E} S_n.$$

Let $\zeta_{r,n}$ be a r.v. with the distribution

$$\mathcal{L}(\zeta_{r,n}) = \mathcal{L}(S_r \,|\, S_r > 0). \tag{8}$$

In extreme value theory $\mathcal{L}(\zeta_{r,n})$ is known as the cluster size distribution.

Theorem 3. *Assume condition* Δ*. If, as* $n \to \infty$*,*

$$S_n \Rightarrow \pi_\lambda \qquad (\exists \lambda > 0), \tag{9}$$

then

$$\zeta_{r,n} \underset{p}{\to} 1 \qquad (n \to \infty) \tag{10}$$

for any sequence $\{r = r_n\}$ *obeying (7).*

If there exists the limit

$$\lim_{n \to \infty} \mathbb{P}(X_{n,1} = \ldots = X_{n,n} = 0) = e^{-\lambda} \qquad (\exists \lambda > 0) \tag{11}$$

and (10) holds for some $\{r = r_n\} \in \mathcal{R}$*, then* $S_n \Rightarrow \pi_\lambda$*.*

Theorem 3 generalises Corollary 2 to the case of dependent α-mixing r.v.s.

Condition (11) is an analogue of (3); it means that $\mathbb{P}(X_{n,i} \ne 0)$ are "properly small".

Condition (10) prohibits asymptotic clustering of rare events. In the case of independent r.v.s taking values in $\mathbf{Z}_+$ assumption (10) means $X'_{n,1} \underset{p}{\to} 1$ as $n \to \infty$, where r.v. $X'_{n,1}$ has the distribution $\mathcal{L}(X'_{n,1}) = \mathcal{L}(X_{n,1} | X_{n,1} \ne 0)$.

Remark 2.1. The following condition (D') has been widely used in extreme value theory (cf. [66, 82]):

$$\lim_{n\to\infty} n\sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0, X_{n,1}\neq 0\right) = 0 \qquad (D')$$

for any sequence $\{r=r_n\}$ such that $n\gg r_n\gg 1$. Condition (D') means that there is no asymptotic clustering of extremes. It was introduced by Loynes [71].

Closely related is the following condition

$$\lim_{n\to\infty} \sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0|X_{n,1}\neq 0\right) = 0. \qquad (\tilde{D}')$$

If conditions Δ and (11) hold, then (D') is *equivalent* to $(\tilde{D}')$.

Indeed, one can check that Δ and (11) yield

$$\mathbb{P}(S_r>0)\sim \lambda r/n \qquad (n\to\infty) \tag{12}$$

(cf. (16) below). Denote $p=\mathbb{P}(X_{n,1}\neq 0)$. Then

$$\lambda r/n \sim \mathbb{P}(S_r>0)\le rp,\ \ \lambda+o(1)\le np.$$

Hence $(D')\Rightarrow(\tilde{D}')$.

By Bonferroni's inequality,

$$\begin{aligned}\lambda r/n \ &\sim\ \mathbb{P}(S_r>0)\ \ge\ r\mathbb{P}(X_{n,1}\neq 0)-\mathbb{P}\Big(\cup_{1\le i<j\le r}\{X_{n,i}\neq 0, X_{n,j}\neq 0\}\Big)\\ &\ge\ rp - rp\sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0|X_{n,1}\neq 0\right).\end{aligned}$$

Therefore,

$$1 \ \ge\ \mathbb{P}(S_r>0)/rp\ge 1-\sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0|X_{n,1}\neq 0\right), \tag{13}$$

$$\lambda+o(1)\ \ge\ \Big(1-\sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0|X_{n,1}\neq 0\right)\Big)np. \tag{14}$$

Thus, np is bounded away from 0 and above, and (D') is equivalent to $(\tilde{D}')$.

Remark 2.2. Condition (10) is weaker than (D'): if conditions Δ and (11) hold, then (D') entails (10). Indeed, $\zeta_{r,n}\ge 1$ by construction. Note that

$$\begin{aligned}\mathbb{P}(S_r>1)\ &=\ \mathbb{P}\Big(\cup_{1\le i<j\le r}\{X_{n,i}\neq 0, X_{n,j}\neq 0\}\Big)\\ &\le\ r\sum_{i=1}^{r-1}\mathbb{P}\left(X_{n,i+1}\neq 0, X_{n,1}\neq 0\right).\end{aligned}$$

Thus, $\mathbb{P}(S_r > 1) = o(r/n)$ if (D') holds. In view of (12), $\mathbb{P}(\zeta_{r,n} > 1) \to 0$ as $n \to \infty$, i.e., (10) holds.

Remark 2.3. If conditions Δ and (D') hold, then (11) is equivalent to

$$\lim_{n\to\infty} n\mathbb{P}(X_{n,1} \neq 0) = \lambda. \tag{3'}$$

Indeed, this follows from (12), (13) and (16) (cf. [66], Theorem 3.4.1).

A generalisation of Corollary 2 to the case of stationary φ-mixing r.v.s has been given by Utev [112], Theorem 10.1, who has shown that conditions (3′) and (D') are necessary and sufficient for (9). Sufficient conditions for Poisson convergence without assuming stationarity have been provided by Sevastyanov [104]. A Poisson limit theorem in the case of a two-dimentional random field $\{X_{i,j}\}$ has been given by Banis [8].

Proof of Theorem 3. Let $\{r = r_n\}$ be an arbitrary sequence from $\mathcal{R}$. Condition Δ and Lemma 2.4.1 from [66] imply that for any $t \in \mathbb{R}$, as $n \to \infty$,

$$\mathbb{E}\exp(itS_n) = \exp\Big(\frac{n}{r}\mathbb{P}(S_r > 0)\mathbb{E}\left\{e^{itS_r} - 1 | S_r > 0\right\}\Big) + o(1), \tag{15}$$

$$\mathbb{P}(S_n = 0) = \mathbb{P}^{n/r}(S_r = 0) + o(1) = \exp\Big(-\frac{n}{r}\mathbb{P}(S_r > 0)\Big) + o(1) \tag{16}$$

(cf. (5.10) in [82]).

If (9) holds, then so does (11): $\mathbb{P}(S_n = 0) \to e^{-\lambda}$ as $n \to \infty$. Note that (11) and (16) yield (12). Since

$$\mathbb{E}e^{itS_n} \to \exp(\lambda(e^{it} - 1)) \qquad (\forall t \in \mathbb{R}) \tag{9*}$$

by the assumption, (15) and (12) entail $\mathbb{E}e^{it\zeta_{r,n}} \to e^{it}$, i.e., (10) holds.

On the other hand, if (10) and (11) hold for some $\{r = r_n\} \in \mathcal{R}$, then (12) is valid. Relations (12) and (15) yield (9*). □

2. Accuracy of Poisson approximation

The problem of evaluating the accuracy of Poisson approximation to the distribution of a sum

$$S_n = X_1 + \ldots + X_n$$

of independent 0-1 random variables has attracted a lot of attention among researchers (cf. [12, 82] and references wherein).

A natural task is to obtain a sharp estimate of the accuracy of Poisson approximation to the distribution of $\mathcal{L}(S_n)$. In this section we overview available estimates.

Historically, the accuracy of Poisson approximation was first studied in terms of the uniform distance (sometimes called the *Kolmogorov* distance).

The *uniform distance* $d_K(X;Y) \equiv d_K(F_X;F_Y)$ between the distributions of random variables X and Y with distribution functions (d.f.s) F_X and F_Y is defined as

$$d_K(F_X;F_Y) = \sup_x |F_X(x) - F_Y(x)|.$$

Many authors evaluated the accuracy of Poisson approximation to $\mathcal{L}(S_n)$ in terms of the *total variation distance*. Recall that the total variation distance $d_{TV}(X;Y)$ between the distributions of r.v.s X and Y is defined as

$$d_{TV}(X;Y) \equiv d_{TV}(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_{A\in\mathcal{A}} |\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)|\,,$$

where $\mathcal{A}$ is a Borel σ-field. Evidently, $d_K(X;Y) \le d_{TV}(X;Y)$. Note that

$$d_{TV}(X;Y) = \inf_{X',Y'} \mathbb{P}(X' \neq Y'),$$

where the infimum is taken over all random pairs (X',Y') such that $\mathcal{L}(X') = \mathcal{L}(X)$ and $\mathcal{L}(Y') = \mathcal{L}(Y)$ [45, 24].

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

$$d_G(X;Y) \equiv d_G(\mathcal{L}(X);\mathcal{L}(Y)) = \sup_{g\in\mathcal{L}} |\mathbb{E}g(X) - \mathbb{E}g(Y)|\,, \tag{17}$$

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

$$d_G(X;Y) = \inf_{X',Y'} \mathbb{E}|X' - Y'|, \tag{18}$$

where the infimum is taken over all random pairs (X',Y') such that $\mathcal{L}(X') = \mathcal{L}(X)$ and $\mathcal{L}(Y') = \mathcal{L}(Y)$ [113]. If X and Y take values in $\mathbf{Z}_+$, then [101, 39, 88]

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

Distance d_G was introduced by Gini [49]; Kantorovich [59] has introduced a class of distances that includes d_G. Barbour et al. [12] called d_G the "Wasserstein distance" after Dobrushin [45] attributed it to Vasershtein [114].

If distributions P_1 and P_2 have densities f_1 and f_2 with respect to a measure μ, set

$$d_H^2(P_1;P_2) := \frac{1}{2}\int \left(f_1^{1/2} - f_2^{1/2}\right)^2 d\mu = 1 - \int \sqrt{f_1 f_2}\, d\mu.$$

Then d_H denotes the Hellinger distance. It is known that

$$d_H^2 \le d_{TV} \le d_H \sqrt{2 - d_H^2}\,. \tag{19}$$

Denote

$$\chi^2(P_1;P_2) = \int_{\mathrm{supp}P_2} (dP_1/dP_2 - 1)^2\, dP_2.$$

By the Cauchy-Bunyakovski inequality,

$$2d_{TV}(P_1;P_2) \le \chi(P_1;P_2).$$

We denote by

$$d^2_{KL}(P_1;P_2) = \int_{\mathrm{supp}P_2} \ln(dP_1/dP_2)\, dP_1$$

the Kullback–Leibler divergence. According to a Pinsker-type inequality,

$$d_{TV} \le d_{KL}/\sqrt{2}\,. \tag{20}$$

Though d^2_{KL} is not a metric, it plays a role in statistics (cf. [53]) and in the theory of large deviations (cf. [82], p. 324, ex. 41).

The Dudley divergence

$$\rho_\varepsilon(P_1;P_2) = \inf_{X,Y} \mathbb{P}(|X-Y|>\varepsilon) \qquad (\varepsilon \ge 0),$$

where the infimum is taken over all random pairs (X,Y) such that $\mathcal{L}(X)=P_1$, $\mathcal{L}(Y)=P_2$, is a generalisation of the total variation distance: $d_{TV}(P_1;P_2) = \rho_0(P_1;P_2)$.

Certain other distances can be found in [70, 82, 89, 94]. Below we present estimates of the accuracy of Poisson approximation for $\mathcal{L}(S_n)$ in terms of d_K, d_{TV} and d_G distances. The main attention is given to results that are missed in existing surveys.

2.1. Independent Bernoulli r.v.s

We denote by $\mathbf{B}(n,p)$ the Binomial distribution with parameters n and p. Let $\mathbf{\Pi}(\lambda)$ denote the Poisson distribution with parameter λ; we denote by π_λ a Poisson $\mathbf{\Pi}(\lambda)$ random variable.

Let $X_1, X_2, ..., X_n$ be independent Bernoulli $\mathbf{B}(p_i)$ r.v.s. Denote $\lambda = \mathbb{E}S_n$,

$$p_i = \mathbb{P}(X_i=1) \quad (i \ge 1), \quad \lambda_k = \sum_{i=1}^n p_i^k \quad (k \ge 2), \quad \theta = \sum_{i=1}^n p_i^2/\lambda.$$

Many authors worked on the problem of evaluating the accuracy of Poisson approximation to $\mathcal{L}(S_n)$ in terms of the uniform distance d_K, the total variation distance d_{TV} and the Gini–Kantorovich distance d_G.

It seems natural to approximate $\mathbf{B}(n,p)$ by the Poisson distribution. For instance, in the case of identically distributed Bernoulli $\mathbf{B}(p)$ r.v.s $\{X_i\}$ one has

$$\mathbb{P}(S_n=k) \equiv \mathbb{P}(S_n=k|N_n=n) = \mathbb{P}(\pi_n(p)=k|\pi_n(1)=n), \tag{21}$$

where $N_n \equiv n$ is the total number of 0's and 1's among $X_1, X_2, ..., X_n$ and $\{\pi_n(t), t \in [0;1]\}$ is a Poisson jump process on $[0;1]$ with intensity rate n. Thus,

$$\mathbf{B}(n,p) = \mathcal{L}(\pi_n(p)|\pi_n(1)=n). \tag{21*}$$

Tsaregradskii [110] has shown that

$$d_K(F_X;F_Y) \le \int_{-\pi}^{\pi} \frac{|\mathbb{E}e^{itX} - \mathbb{E}e^{itY}|}{4|t|}\,dt \tag{22}$$

if X and Y are integer-valued r.v.s, and derived the estimate

$$d_K(\mathbf{B}(n,p);\mathbf{\Pi}(np)) \le p\pi^2 e^{2p(2-p)}/16(1-p) \qquad (p\in(0;1/2]). \tag{23}$$

Note that $\pi^2/16 \approx 0.617$. Inequality (23) seems to be the first estimate of the accuracy of Poisson approximation with explicit constant.

In the case of non-identically distributed Bernoulli $\mathbf{B}(p_i)$ random variables Franken [46] has shown that

$$d_K(S_n;\pi_\lambda) \le 0.6p_n^*$$

if $p_n^* := \max_{i\le n} p_i \le 1/4$. Shorgin [107] has proved that

$$d_K(S_n;\pi_\lambda) \le c_1\theta/(1-\sqrt{\theta}\,) \qquad (\theta<1)$$

where $c_1 = (1+\sqrt{\pi/2})/2 < 1.13$. According to Daley & Vere-Jones [38],

$$d_K(S_n;\pi_\lambda) \le 0.36\theta.$$

Roos [94] has shown that

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

Note that $1/2e \le 0.184$.

A good survey concerning estimates of $d_K(S_n;\pi_\lambda)$ is Zacharovas & Hwang [121].

Kontoyiannis et al. [63] have shown that

$$d_H^2(S_n;\pi_\lambda) \le \lambda^{-1}\sum_{i=1}^n p_i^3/(1-p_i).$$

Borisov & Vorozheikin [25] present sharp lower and upper bounds to $\chi^2(\mathbf{B}(n,p);\mathbf{\Pi}(np))$:

$$0 \le \chi^2(\mathbf{B}(n,p);\mathbf{\Pi}(np)) - p^2/2 - 2p^3/3n \le p^4/(1-p) + p^8(23-20p)/(1-p)^2.$$

Harremoës & Ruzankin [53] present lower and upper bounds to $d_{KL}^2(\mathbf{B}(n,p);\mathbf{\Pi}(np))$. In particular, they have shown that

$$2d_{KL}^2(\mathbf{B}(n,p);\mathbf{\Pi}(np)) = (-p - \ln(1-p))(1+O(1/n)).$$

Borisov [22] has shown that for any $\varepsilon \ge 0$

$$\rho_\varepsilon(\mathbf{B}(n,p);\mathbf{\Pi}(np)) \le Ce^{-c\varepsilon}\left((np^2)^{[\varepsilon]+1}\mathbb{1}\{np \ge 1\} + np^{[\varepsilon]+2}\mathbb{1}\{np < 1\}\right),$$

where c, C are absolute positive constants. Ruzankin [98] presents the bound $(k \in \mathbf{Z}_+)$

$$\rho_k(\mathbf{B}(n,p);\mathbf{\Pi}(np)) < \tfrac{7}{3}e^{-k^2/27np^3}\mathbb{1}\{k < np^2\} + e^{-np/15}\mathbb{1}\{k \ge np^2\},$$

$$\rho_k(\mathbf{B}(n,p);\mathbf{\Pi}(np)) < \exp\left(-\tfrac{1}{4}\sqrt{nk\ln^3(k/np^2e^2)}\right) \quad (np^2e^4 \le k \le n(1-p)/\ln(1/p)).$$

Many authors worked on the problem of evaluating the total variation distance $d_{TV}(S_n;\pi_\lambda)$ (cf. [12, 82] and references wherein).

Prokhorov [87] has established the existence of an absolute constant c such that

$$d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)) \le cp. \tag{24}$$

Kolmogorov [62] points out that

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

where C is an absolute constant. LeCam [67, 68] attributes inequality

$$d_{TV}(S_n;\pi_\lambda) \le \sum_{i=1}^n p_i^2 \tag{25}$$

to Khintchin [61]. Bound (25) is sharp: according to (2.10) in Deheuvels & Pfeifer [41],

$$d_{TV}(S_n;\pi_\lambda) \ge np^2(1+O(p))$$

in the case of i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s if $np \to 0$.

Note that (25) is a consequence of the property of d_{TV} and the following fact:

$$d_{TV}(\mathbf{B}(p);\mathbf{\Pi}(p)) = (1-e^{-p})p \le p^2. \tag{26}$$

Indeed, denote $\bar{X} = (X_1, ..., X_n)$, $\bar{\pi} = (\pi_{p_1}, ..., \pi_{p_n})$, where $\{\pi_{p_i}\}$ are independent Poisson $\mathbf{\Pi}(p_i)$ r.v.s. Then

$$d_{TV}(S_n;\pi_\lambda) \le d_{TV}(\bar{X};\bar{\pi}) \le \sum_{i=1}^n d_{TV}(X_i;\pi_{p_i}) \le \sum_{i=1}^n p_i^2. \tag{25$'$}$$

A similar argument due to D.J.Daley (cf. (5.5) in Serfling [103]) yields

$$d_K(S_n;\pi_\lambda) \le \sum_{i=1}^n p_i^2/2.$$

Set $\tilde{p}_i = -\ln(1-p_i)$ $(1 \le i \le n)$, and put $\mu = \sum_{i=1}^n \tilde{p}_i$. According to Serfling [102],

$$d_{TV}(S_n; \pi_\mu) \le \sum_{i=1}^n \tilde{p}_i^2/2.$$

Kerstan [60] has shown that

$$d_{TV}(S_n; \pi_\lambda) \le 1.05\theta. \tag{27}$$

Romanowska [91] has noticed that

$$d_{TV}(\mathbf{B}(n,p); \mathbf{\Pi}(np)) \le p/2\sqrt{1-p}\,. \tag{28}$$

The popular estimate

$$d_{TV}(S_n; \pi_\lambda) \le \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n p_i^2 \tag{29}$$

is effectively due to Barbour and Eagleson [9].

Presman [86] has established an estimate of $d_{TV}(S_n; \pi_\lambda)$ with the constant 0.83 at the leading term. In the case of i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s Presman's bound becomes

$$d_{TV}(\mathbf{B}(n,p); \mathbf{\Pi}(np)) \le 0.83p/(1-p)(1-1/n). \tag{30}$$

Xia [117] has derived an estimate with the constant 0.6844 at the leading term.

Roos [94] (see also Čekanavičius & Roos [33]) has obtained a bound with a correct constant $3/4e \approx 0.276$ at the leading term: if $\theta < 1$, then

$$d_{TV}(S_n; \pi_\lambda) \le 3\theta/4e(1-\sqrt{\theta}\,)^{3/2}\,. \tag{31}$$

Note that $\theta/(1-\sqrt{\theta}\,)^{3/2} \ge \theta(1+1.5\sqrt{\theta}+3.75\theta)$.

Roos [94] has shown also that

$$d_{TV}(S_n; \pi_\lambda) \sim 3\theta/4e$$

if $\theta \to 0$ and $\lambda \to 1$ as $n \to \infty$. Thus, constant $3/4e$ cannot be improved.

Denote

$$p_n^* = \max_{i \le n} p_i\,, \qquad \varepsilon = \min\Big\{1; (2\pi[\lambda - p_n^*])^{-1/2} + 2\delta/(1-p_n^*/\lambda)\Big\}\,,$$
$$\delta = \frac{1-e^{-\lambda}}{\lambda}\sum_{i=1}^n p_i^2\,, \qquad \delta^* = \frac{1-e^{-\lambda}}{\lambda}\sum_{i=1}^n p_i^3\,.$$

Note that $\delta^2 \le \delta^*$. The following inequality from [82], Theorem 4.12, sharpens the second-order term of the right-hand side (r.h.s.) of estimate (31):

$$d_{TV}(S_n; \pi_\lambda) \le 3\theta/4e + 2\delta^*\varepsilon + 2\delta^2. \tag{32}$$

In the case of $\mathcal{L}(S_n) = \mathbf{B}(n,p)$ estimate (32) becomes

$$d_{TV}(S_n;\pi_{np}) \leq 3p/4e + 2(1-e^{-np})p^2\varepsilon + 2(1-e^{-np})^2p^2, \tag{32*}$$

where $\varepsilon = \min\{1;(2\pi[(n-1)p])^{-1/2} + 2(1-e^{-np})p/(1-1/n)\}$. The second-order term in (32*) is of order $p^2 \wedge np^3$.

In applications one often has $\lambda \equiv \lambda(n) \to \infty$ as $n \to \infty$. Hence estimates with the "magic factor" $(1-e^{-\lambda})/\lambda$ attract special interest.

The possibility of the "super–magic" factor $e^{-\lambda}$ when one approximates $S_n \in A$ for a bounded A has been discussed in [82], ch. 4.5 (such approximations are of interest in extreme value theory). For instance, if $\{X_i\}$ are independent Bernoulli $\mathbf{B}(p_i)$ r.v.s and $A = \{0\}$, then

$$0 \leq \mathbb{P}(\pi_\lambda = 0) - \mathbb{P}(S_n = 0) \leq e^{-\lambda + p_n^*} \sum_{i=1}^n p_i^2/2. \tag{33}$$

Indeed, set $c_i = \prod_{j=1}^{i-1} e^{-p_j} \prod_{j=i+1}^n (1-p_j)$. Since $e^{-p_i} - 1 + p_i \leq p_i^2/2$ by Taylor's formula,

$$\mathbb{P}(\pi_\lambda = 0) - \mathbb{P}(S_n = 0) = \sum_{i=1}^n (e^{-p_i} - 1 + p_i)c_i \leq e^{-\lambda + p_n^*} \sum_{i=1}^n p_i^2/2.$$

In the case of the Binomial $\mathbf{B}(n,p)$ distribution (33) becomes

$$0 \leq \mathbb{P}(\pi_{np} = 0) - \mathbb{P}(S_n = 0) \leq \frac{1}{2} np^2 e^{-(n-1)p} \leq \frac{2n}{e^2(n-1)^2}\,. \tag{33*}$$

Note that $2/e^2 \approx 0.2707$.

Bound (32) is a consequence of inequality

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(\pi_\lambda \in A)| \leq |\mathbb{P}(\pi_\lambda + 1 \in A) - \mathbb{P}(\pi_\lambda^\star \in A)|\theta/2 + 2\delta^*\varepsilon + 2\delta^2. \tag{32*}$$

The first term on the r.h.s. of (32*) has the "super–magic" factor if A is finite.

Estimates in terms of the Gini-Kantorovich distance are available as well. Denote $\mu = -\sum_{i=1}^n \ln(1-p_i)$. If $p_n^* \leq 1/2$, then

$$d_G(S_n;\pi_\mu) \leq \sum_{i=1}^n p_i^2/2(1-p_i) \tag{34}$$

(Deheuvels et al. [43]). Witte [116] has shown that

$$d_G(S_n;\pi_\lambda) \leq \frac{-\sqrt{e\lambda}}{2\sqrt{2\pi}} \ln(1-2\theta e^{2p_n^*}). \tag{35}$$

According to [82], formula (4.53),

$$d_G(S_n;\pi_\lambda) \leq \left(1 \wedge \tfrac{4}{3}\sqrt{2/e\lambda}\right) \sum_{i=1}^n p_i^2. \tag{36}$$

Roos [94] has shown that

$$d_{G}(S_n;\pi_\lambda) \le \Big(1/\sqrt{2e} + 1.6\sqrt{\theta}(2-\theta)/(1-\sqrt{\theta})\Big)\,\theta\sqrt{\lambda}\,. \tag{37}$$

A recent survey is Zacharovas & Hwang [121].

Sharp non-Poisson approximation to the Binomial $\mathbf{B}(n,p)$ distribution function $\mathbb{P}(S_n \le \cdot)$ has been given by Zubkov & Serov [127].

Denote by Φ the standard normal d.f.. Let

$$\Lambda(x) = x\ln(x/p) + (1-x)\ln\left((1-x)/(1-p)\right) \qquad (0<x<1)$$

denote the rate function of the Bernoulli distribution $\mathbf{B}(p)$ (cf. [82], p. 322), and set

$$Z_{n,p}(k) = \Phi\Big(\mathrm{sgn}(k/n-p)\sqrt{2n\Lambda(k/n)}\,\Big).$$

Then [127]

$$Z_{n,p}(k) \le \mathbb{P}(S_n \le k) \le Z_{n,p}(k+1). \tag{38}$$

The following large deviations inequality is due to Bernstein [21], p. 168:

$$\mathbb{P}\left((S_n-np)/\sqrt{npq} > t(1+\gamma_{p,n})\right) < \exp(-t^2/2) \qquad (t>0),$$

where $q = 1-p,\ \gamma_{p,n} = t_{p,n}(q-p)/6 + t_{p,n}^2(p^3+q^3)/12,\ t_{p,n} = t/\sqrt{npq}\,.$

Asymptotics of $d_{TV}(S_n;\pi_\lambda)$. The asymptotics of $d_{TV}(S_n;\pi_\lambda)$ in the case of identically distributed Bernoulli $\mathbf{B}(p)$ r.v.s has been established by Prokhorov [87]:

$$d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)) = p/\sqrt{2\pi e}\Big(1 + O\big(1\wedge(p+1/\sqrt{np}\,)\big)\Big). \tag{39}$$

Kerstan [60], Deheuvels & Pfeifer [40, 42], Deheuvels et al. [43] and Roos [93] have generalised (39) to the case of non-identically distributed 0-1 r.v.s. Deheuvels & Pfeifer [40] present also the asymptotics of $d_{TV}(S_n;\pi_\lambda)$ in the case where $\lambda\to\mathrm{const}$ as $n\to\infty$.

The following result concerning the asymptotics of $d_{TV}(S_n;\pi_\lambda)$ uses the notation from [82], ch. 4. Given a non-negative integer-valued random variable Y, we denote by $Y^\star$ a random variable with the distribution

$$\mathbb{P}(Y^\star = k) = \mathbb{P}(Y=k)(k-\lambda)^2/\lambda \qquad (k\in\mathbf{Z}_+). \tag{40}$$

The next bound is a consequence of Theorem 11.

Theorem 4. *If $X_1, ..., X_n$ are independent Bernoulli r.v.s, $\mathcal{L}(X_i) = \mathbf{B}(p_i)$, then*

$$|d_{TV}(S_n;\pi_\lambda) - \theta d_{TV}(\pi_\lambda^\star;\pi_\lambda+1)/2| \le 2\delta^*\varepsilon + 2\delta^2. \tag{41}$$

One can check that

$$d_{TV}(\pi_\lambda^\star;\pi_\lambda+1) = \sqrt{2/\pi e} + O\big(1/\sqrt{\lambda}\,\big) \tag{42}$$

as $\lambda \to \infty$. Thus,

$$d_{TV}(S_n;\pi_\lambda) = \theta/\sqrt{2\pi e}\left(1 + O\left(\theta + 1/\sqrt{\lambda}\,\right)\right) \tag{43}$$

if $\lambda \to \infty$ and $\theta \to 0$ as $n \to \infty$.

Example 2.1. Let $p_i = 1/i$, $i \in \mathbb{N}$. Then $p_n^* = 1$, $\lambda = \lambda(n) \to \infty$, $\theta \to 0$ as $n \to \infty$, and (43) entails $d_{TV}(S_n;\pi_\lambda) \sim \theta/\sqrt{2\pi e}$. □

Shifted Poisson approximation. Shifted (translated) Poisson approximation to $\mathbf{B}(n,p)$ has been considered by a number of authors (see [16, 19, 31, 65, 83] and references therein). The accuracy of shifted Poisson approximation can be sharper than that of pure Poisson approximation. Another advantage of using shifted Poisson approximation is the possibility to derive a more general result (e.g., a uniform in p estimate of $d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np))$, cf. (45) below).

Let $X_1, ..., X_n$ be independent 0-1 r.v.s. Set $p_i = \mathbb{P}(X_i = 1)$, $q_i = 1 - p_i$,

$$\lambda = \mathbb{E}S_n, \;\; \sigma^2 = \mathrm{var}\, S_n, \;\; \lambda_2 = \lambda - \sigma^2.$$

Denote $[x] = \max\{k \in \mathbf{Z} \colon k \le x\}$, $\{x\} = x - [x]$. We define r.v.

$$Y = [\lambda_2] + \pi_{\lambda - [\lambda_2]}.$$

Note that $\mathrm{var}\,\pi_\lambda - \mathrm{var}\, S_n = \lambda_2$, while $\mathrm{var}\, Y - \mathrm{var}\, S_n = \{\lambda\} < 1$.

The following result is due to Čekanavičius & Vaitkus [31].

Theorem 5. *If* $\sigma^2 \ge 4$, *then*

$$d_{TV}(S_n;Y) \le 0.93\sigma^{-3}\lambda_2 + \{\lambda_2\}/(\sigma^2 + \{\lambda_2\}) + e^{-\sigma^2/4}. \tag{44}$$

Let $\{X_i\}$ be i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s. Then the right-hand side (r.-h.s.) of (44) is

$$O\left(\sqrt{p/n} + 1/np\right).$$

A similar bound in terms of the uniform distance has been established by Kruopis [65].

Set $q = 1 - p$, where $0 < p < 1$. Then

$$Y = [np^2] + \pi_{npq + \{np^2\}}, \quad \mathbb{E}Y = np, \quad \mathrm{var} Y = npq + \{np^2\}.$$

The following Theorem 6 presents a uniform in $p \in [0;1/2]$ bound to $d_{TV}(S_n;Y)$.

Theorem 6. [83] *As* $n > 4$,

$$\sup_{0 \le p \le 1/2} d_{TV}(S_n;Y) \le \frac{2/\sqrt{\pi e}}{\sqrt{n} - 2} + \frac{0.9n^{1/4}}{n-1} + \frac{2 + 1.8/n^{1/4}}{(\sqrt{n} - 1)^2}\;. \tag{45}$$

Theorem 6 can be compared with the Berry–Esseen inequality

$$d_K(\mathbf{B}(n,p);\mathcal{N}(np,npq)) \le C/\sqrt{np}$$

(see, e.g., [106]) as well as with the results by Meshalkin [75] and Pressman [85]. Estimate (45) is uniform in $p\in[0;1/2]$. Note that a uniform in $p\in[0;1/2]$ Berry–Esseen estimate would be infinite. Inequality (45) has advantages over Meshalkin's [75] and Pressman's [85] results as the constants in (45) are explicit (which matters in applications); besides, the structure of the approximating distribution $\mathcal{L}(Y)$ is simpler and does not assign mass to negative numbers. Bound (45) is preferable to (29) – (32) if $p>4e/\sqrt{n}$.

An estimate of the accuracy of shifted Poisson approximation to the distribution of a sum of Bernoulli $\mathbf{B}(p_i)$ r.v.s in terms of the Gini-Kantorovich distance has been given by Barbour & Xia [19] in the assumption that λ_2 is an integer.

Poisson approximation to the multinomial distribution. Results on the accuracy of Poisson approximation to the distribution of a sum of Bernoulli r.v.s can be generalised to the case of a multinomial distribution.

Let $\bar{S}_n$ be a random vector with multinomial distribution $\mathbf{B}(n,p_1,...,p_m)$:

$$\mathbb{P}(\bar{S}_n=\bar{l}\,) = \frac{n!}{l_1!...l_m!(n-l)!}p_1^{l_1}...p_m^{l_m}(1-p)^{n-l}\,, \tag{46}$$

where $l_i\in\mathbf{Z}_+$ $(\forall i)$, $\bar{l}=(l_1,..,l_m)$, $l=l_1+...+l_m\le n$, $p=p_1+...+p_m$.

Formula (46) describes, in particular, the joint distribution of the increments of the empirical d.f..

Note that

$$\bar{S}_n \stackrel{d}{=} \bar{\xi}_1+...+\bar{\xi}_n, \tag{47}$$

where $\bar{\xi},\bar{\xi}_1,...,\bar{\xi}_n$ are i.i.d. random vectors with the distribution

$$\mathbb{P}(\bar{\xi}=\bar{0}) = 1-p\,,\ \mathbb{P}(\bar{\xi}=\bar{e}_j) = p_j \qquad (1\le j\le m),$$

vector $\bar{e}_j$ has the j^{th} coordinate equal to 1 and the other coordinates equal to 0.

Let

$$\bar{\pi} = (\pi_1,...,\pi_m)$$

be a vector of independent Poisson r.v.s with parameters $np_1,...,np_m$, and let $\pi_n(\cdot)$ denote a Poisson jump process on $[0;1]$ with intensity rate n. Then $\bar{\pi}$ is a vector of increments of process $\pi_n(\cdot)$: $\pi_1 \stackrel{d}{=} \pi_n(p_1),...,\pi_m \stackrel{d}{=} \pi_n(p)-\pi_n(p{-}p_m)$. Note that

$$\mathbb{P}(\bar{S}_n=\bar{l}\,) = \mathbb{P}\left(\pi_n(p_1)=l_1,...,\pi_n(p)-\pi_n(p-p_m)=l_n \,|\, \pi_n(1)=1\right) \tag{46*}$$

(cf. (21)).

Arenbaev [6] has shown that

$$d_{TV}(\bar{S}_n;\bar{\pi}) = p/\sqrt{2\pi e}\,(1+O(1\wedge 1/\sqrt{np}\,)) \tag{48}$$

if $n\to\infty$ (the term $1/\sqrt{np}$ in (48) apparently needs to be replaced with $p+1/\sqrt{np}$, cf. (39)). Arenbaev ([6], formulas (5)–(9′)) has shown also that

$$d_{TV}(\bar{S}_n;\bar{\pi}) = d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)). \tag{49}$$

Using (49) and (32), we deduce

$$d_{TV}(\bar{S}_n;\bar{\pi}) \le 3p/4e + 4(1-e^{-np})p^2. \tag{50}$$

According to Deheuvels & Pfeifer [41],

$$|d_{TV}(\bar{S}_n;\bar{\pi}) - K_{n,\lambda}| \le \max\{16p^2; 5np^3\}, \tag{51}$$

where $K_{n,\lambda} = np^2e^{-np}\big((np)^{\alpha-np}(\alpha-np)/\alpha! - (np)^{\beta-np}(\beta-np)/\beta!\big)/2$,

$$\alpha = np+1/2+\sqrt{np+1/4}\,,\ \ \beta = np+1/2-\sqrt{np+1/4}\,.$$

The case of non-identically distributed random vectors $\bar{\xi}_1,...,\bar{\xi}_n$ has been treated by Roos [97]. A generalisation of (50) to the case of a stationary sequence of *dependent* r.v.s is given in [82], Theorem 6.8.

Open problem.
2.1. Improve the constants in (34)–(36).
2.2. Generalise Theorem 6 to the case of m-dependent r.v.s.

2.2. Dependent Bernoulli r.v.s

We present below generalisations of (25) and (29) to the case of dependent Bernoulli r.v.s.

Let $X_1,...,X_n$ be (possibly dependent) Bernoulli r.v.s. Chen [28] pioneered the use of Stein's method in deriving estimate of the accuracy of Poisson approximation, and obtained an estimate of the accuracy of Poisson approximation to the distribution of a sum of φ-mixing r.v.s.

Set $p_i = \mathbb{P}(X_i=1|X_1,...,X_{i-1})$. A generalisation of (25) has been given by Serfling [102]:

$$d_{TV}(S_n;\pi_\lambda) \le \sum_{i=1}^n(\mathbb{E}p_i)^2 + \sum_{i=1}^n\mathbb{E}|p_i-\mathbb{E}p_i|, \tag{25*}$$

$$d_K(S_n;\pi_\lambda) \le \frac{2}{\pi}\sum_{i=1}^n(\mathbb{E}p_i)^2 + \sum_{i=1}^n\mathbb{E}|p_i-\mathbb{E}p_i|. \tag{52}$$

Let $\{X_a, a\in J\}$ be a family of dependent Bernoulli $\mathbf{B}(p_a)$ random variables. Assign to each $a\in J$ a "neighborhood" $B_a\subset J$ such that $\{X_b, b\in J\setminus B_a\}$ are "almost independent" of X_a (for instance, if $\{X_b\}$ are m–dependent r.v.s and $J=\{1,...,n\}$, then $B_a=[a-m;a+m]\cap J$).

The idea of splitting the sample into "strongly dependent" and "almost independent" parts goes back to Bernstein [20] (see also [104]).

Denote

$$S=\sum_{a\in J}X_a,\;\;\lambda=\mathbb{E}S,$$

and let

$$\begin{aligned}\delta_1 &= \sum_{a\in J}\sum_{b\in B_a}\mathbb{E}X_a\mathbb{E}X_b\,,\;\;\delta_2=\sum_{a\in J}\sum_{b\in B_a\setminus\{a\}}\mathbb{E}X_aX_b,\\ \delta_3 &= \sum_{a\in J}\mathbb{E}\left|\mathbb{E}X_a-\mathbb{E}\Big\{X_a\Big|\sum\nolimits_{b\in J\setminus B_a}X_b\Big\}\right|.\end{aligned}$$

The following bound is cited from Arratia et al. [2] (see also Smith [108]).

Theorem 7. *There holds*

$$d_{_{TV}}(S;\pi_\lambda)\le\frac{1-e^{-\lambda}}{\lambda}\Big(\delta_1+\delta_2\Big)+\min\{1;1.4/\sqrt{\lambda}\,\}\delta_3\,. \tag{53}$$

According to Remark 10.2.4 in [12], the term $1.4/\sqrt{\lambda}$ in (53) can be replaced with $\sqrt{2/e\lambda}$.

In the case of independent random variables one can choose $B_a=\{a\}$, then (53) coincides with (29).

Theorem 7 has applications to the problem of Poisson approximation to the distribution of the number of long head runs in a sequence of Bernoulli r.v.s, and to the problem of Poisson approximation to the distribution of the number of long match patterns in two sequences (e.g., DNA sequences, see [12, 82] and references therein).

The topic concerning $\mathcal{L}(S_n)$ in the case of stationary dependent r.v.s $\{X_i\}$ has applications in extreme value theory [66, 82]. The case where the sequence $X_1,...,X_n$ is a moving average is related to the topic concerning the so-called Erdös–Rényi maximum of partial sums (cf. [82], ch. 2).

Estimates of the accuracy of Poisson approximation for some special types of dependence among $\{X_a,a\in J\}$ can be found in Barbour et al. [12]. An estimate of the accuracy of shifted Poisson approximation to the distribution of a sum of dependent Bernoulli $\mathbf{B}(p_i)$ r.v.s in terms of the total variation distance is given by Čekanavičius & Vaitkus [31]. A generalization of Theorem 7 to the case of compound Poisson approximation has been given by Roos [92].

Open problem.
2.3. Improve the constants in (53).

2.3. Independent integer-valued r.v.s

The topic of Poisson approximation to the distribution of a sum of integer-valued r.v.s has applications in extreme value theory, insurance, reliability theory, etc. (cf. [7, 12, 66, 82]). For instance, in insurance applications the sum $S_n = \sum_{i=1}^n Y_i \mathbb{1}\{Y_i > y_i\}$ of integer-valued r.v.s allows to account for the total loss from the claims exceeding excesses $\{y_i\}$. One would be interested if Poisson approximation to $\mathcal{L}(S_n)$ is applicable.

In extreme value theory one often deals with the number of extreme (rare) events represented by a sum $S_n = \xi_1 + ... + \xi_n$ of 0-1 r.v.s (indicators of rare events). The r.v.s $\xi_1, ..., \xi_n$ can be dependent. One way to cope with dependence is to split the sample into blocks, which can be considered almost independent (the so-called Bernstein's blocks approach [20]). The number of r.v.s in a block is an integer-valued r.v.; thus, the number of rare events is a sum of almost independent integer-valued r.v.s.

In all such situations one deals with a sum of non-negative integer-valued r.v.s that are non-zero with small probabilities, and Poisson or compound Poisson approximation to $\mathcal{L}(S_n)$ appears plausible. An estimate of the accuracy of Poisson approximation to the distribution of S_n can indicate whether Poisson approximation is applicable.

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 [10, 11, 82]. Inequality (25) and the Barbour-Eagleson estimate (29) have been generalised to the case of non-negative integer-valued r.v.s by Barbour [10]. Theorem 8 below presents another result of that kind (see [82], ch. 4.4).

Let $X_1, X_2, ..., X_n$ be independent non-negative integer-valued r.v.s,

$$S_n = X_1 + ... + X_n, \ \ \lambda = \mathbb{E} S_n,$$

π_λ denotes a Poisson $\mathbf{\Pi}(\lambda)$ r.v..

Franken [46] has shown that

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

Denote $\lambda^* = \sum_{i=1}^n \mathbb{P}(X_i = 1)$, $\lambda_2^* = \sum_{i=1}^n \mathbb{P}(X_i = 1)^2$. Kerstan [60] has proved that

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

An early survey on the topic is Witte [116].

Given a random variable Y that takes values in $\mathbf{Z}_+$, let Y^* denote a random variable with the distribution

$$\mathbb{P}(Y^* = m) = (m+1)\mathbb{P}(Y = m+1)/\mathbb{E} Y \qquad (m \geq 0). \tag{54}$$

Distribution (54) differs by a shift from the distribution introduced by Stein [109], p. 171. Note that $Y^* \stackrel{d}{=} Y$ if and only if $\mathcal{L}(Y)$ is Poisson.

Theorem 8. *As* $n \geq 1$,

$$d_{TV}(S_n;\pi_\lambda) \quad \leq \quad \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^{n} d_G(X_i;X_i^*)\mathbb{E}X_i\,, \tag{55}$$

$$d_G(S_n;\pi_\lambda) \quad \leq \quad \min\Big\{1;\tfrac{4}{3}\sqrt{2/e\lambda}\Big\}\sum_{i=1}^{n} d_G(X_i;X_i^*)\mathbb{E}X_i\,. \tag{56}$$

In the case of Bernoulli $\mathbf{B}(p_i)$ r.v.s one has $X_i^* \equiv 0$, and (55) coincides with (29).

In the case of i.i.d.r.v.s (55) becomes

$$d_{TV}(S_n;\pi_\lambda) \leq (1-e^{-\lambda})\mathbb{E}|X-X^*|\,.$$

Here X^* may be chosen independent of X, although one would prefer to define X and X^* on a common probability space in order to make $\mathbb{E}|X-X^*|$ smaller.

A generalisation of (32) to the case of independent integer-valued r.v.s has been given by Novak [83].

Example 2.2. Let $\xi, X_1, X_2, ...$ be i.i.d.r.v.s with geometric $\mathbf{\Gamma}_0(p)$ distribution:

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

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

Set $r = p/(1-p)$. Vervaat [115] has shown that $d_{TV}(S_n;\pi_\lambda) \leq r$, while Romanowska [91] has noticed that $d_{TV}(S_n;\pi_\lambda) \leq r/\sqrt{2}$. Roos [96] has shown that

$$d_{TV}(\mathbf{NB}(n,p);\mathbf{\Pi}(np)) \quad \leq \quad \min\{3r/4e; nr^2\}, \tag{57}$$

$$d_G(\mathbf{NB}(n,p);\mathbf{\Pi}(np)) \quad \leq \quad nr^2. \tag{58}$$

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

$$X_i^* \stackrel{d}{=} X_i + \xi, \tag{59}$$

and $\mathbb{E}|X-X^*| = p/(1-p)$. Note that

$$\lambda = n\mathbb{E}\xi = nr, \quad d_G(X;X^*) = \mathbb{E}\xi = r.$$

Theorem 8 entails

$$d_{TV}(S_n;\pi_\lambda) \quad \leq \quad (1-e^{-nr})r, \tag{60}$$

$$d_G(S_n;\pi_\lambda) \quad \leq \quad \min\Big\{1;\tfrac{4}{3}\sqrt{2/enp}\Big\}\,nr^2. \tag{61}$$

Inequality (60) has been established in [10], p. 758; estimate (61) is from [82], formula (4.53). □

Shifted Poisson approximation. A number of authors dealt with shifted Poisson approximation to the distribution of a sum S_n of integer-valued r.v.s (see [16, 83] and references therein). Let

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

where $[x]$ and $\{x\} = x - [x]$ denote the integer and the fractional parts of x.

Barbour & Čekanavičius [16] have shown that

$$d_{TV}(S_n; a+\pi_\mu) \le (1 \wedge \sigma^{-2})\Big(b + d_n \sum_{i=1}^n \psi_i\Big) + \mathbb{P}(S_n < a), \tag{62}$$

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

In the Binomial case (i.e., $\mathcal{L}(S_n) = \mathbf{B}(n,p)$) the r.-h.s. of (62) is $O\big(\sqrt{p/n} + 1/np\big)$. Further reading on the topic is [83].

An estimate of the accuracy of shifted Poisson approximation to the distribution of a random sum of i.i.d. integer-valued r.v.s has been presented by Röllin [90].

2.4. Dependent integer-valued r.v.s

Let $X_1, ..., X_n$ be (possibly dependent) non-negative integer-valued r.v.s. Set $p_i = \mathbb{P}(X_i = 1 | X_1, ..., X_{i-1})$. A generalisation of ($25^*$), (52) has been given by Serfling [102]:

$$d_{TV}(S_n; \pi_\lambda) \le \sum_{i=1}^n \left(\mathbb{E}^2 p_i + \mathbb{E}|p_i - \mathbb{E}p_i| + \mathbb{P}(X_i \ge 2)\right), \tag{25^+}$$

$$d_K(S_n; \pi_\lambda) \le \sum_{i=1}^n \Big(\tfrac{2}{\pi}\mathbb{E}^2 p_i + \mathbb{E}|p_i - \mathbb{E}p_i| + \mathbb{P}(X_i \ge 2)\Big). \tag{52^+}$$

Below we present a generalisation of Theorem 7.

Let $\{X_a, a \in J\}$ be a family of r.v.s taking values in $\mathbf{Z}_+$. Suppose one can choose the "neighborhoods" $\{B_a\}$ so that r.v.s $\{X_b, b \in J \backslash B_a\}$ are independent of X_a. We call this assumption the "local dependence" condition.

Let $\mathcal{L}(\pi_\lambda)$ denote a Poisson $\mathbf{\Pi}(\lambda)$ r.v.. Set

$$\delta_1^* \ = \ \sum_{a \in J} \sum_{b \in B_a \backslash \{a\}} \mathbb{E}X_a \mathbb{E}X_b \,, \quad \delta_4 = \sum_{a \in J} d_G(X_a; X_a^*)\mathbb{E}X_a,$$

and let $\delta_1, \delta_2, \delta_3$ be defined as in Theorem 7. Theorems 9 and 10 are from [82], ch. 4.

Theorem 9. *If* $\{X_b, b \in J \backslash B_a\}$ *are independent of* X_a, *then*

$$d_{TV}(S_n; \pi_\lambda) \le \frac{1 - e^{-\lambda}}{\lambda} (\delta_1^* + \delta_2 + \delta_4). \tag{63}$$

In Theorem 10 we drop the local dependence condition assumed in Theorem 9.

Theorem 10. *Denote* $\delta_5 = \sum_{a\in J} \mathbb{E}X_a(X_a-1)\mathbb{1}\{X_a\geq 2\}$. *Then*

$$d_{TV}(S_n;\pi_\lambda) \leq \frac{1-e^{-\lambda}}{\lambda}\Big(\delta_1+\delta_2+\delta_5\Big) + \min\{1;\sqrt{2/e\lambda}\,\}\delta_3. \tag{64}$$

Ruzankin [100] presents an estimate of the accuracy of Poisson approximation to $\mathbb{E}h(S_n)$, where h is an unbounded function.

Open problem.
2.4. Improve the constants in (63), (64).

2.5. Asymptotic expansions

Let $X_1,...,X_n$ be independent Bernoulli $\mathbf{B}(p_i)$ r.v.s, and let π_λ be a Poisson random variable.

Formal expansions of $\mathbb{P}(S_n\leq x)$ have been given by Uspensky [111], see also Franken [46]. Herrmann [55], Shorgin [107] and Barbour [10] present full asymptotic expansions with explicit estimates of the error terms. Kerstan [60], Kruopis [65], Čekanavičius [29] and Čekanavičius & Kruopis [30] present first-order asymptotic expansions. 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. [13] and Borisov & Ruzankin [23].

The formulation of the full asymptotic expansions is cumbersome and will be omitted. We present below first-order asymptotics of $\mathbb{E}h(S_n)$ for particular classes of functions h.

Of special interest are indicator functions $h(\cdot)=\mathbb{1}\{\cdot\in A\}$, $A\subset\mathbf{Z}_+$. Denote

$$\begin{aligned} Q_\lambda(A) &= \Big[\mathbb{P}(\pi_\lambda\in A)+\mathbb{P}(\pi_\lambda+2\in A)-2\mathbb{P}(\pi_\lambda+1\in A)\Big]\Big/2, \\ \varepsilon &= \min\Big\{1;(2\pi[\lambda-p_n^*])^{-1/2}+2\delta/(1-p_n^*/\lambda)\Big\}, \quad p_n^*=\max_{i\leq n}p_i. \end{aligned}$$

Let $\pi_\lambda^\star$ denote a random variable with distribution (40). Then

$$Q_\lambda(A) = [\mathbb{P}(\pi_\lambda^\star\in A)-\mathbb{P}(\pi_\lambda+1\in A)]/2\lambda$$

(see [82], ch. 4).

The following result from [82], ch. 4, sharpens (13) in [55] and the bound of Corollary 2.4 in [10] (Corollary 9.A.1 in [12]).

Theorem 11. *Let $X_1,...,X_n$ be independent Bernoulli r.v.s, $\mathcal{L}(X_i)=\mathbf{B}(p_i)$. Then*

$$\left|\mathbb{P}(S_n\in A)-\mathbb{P}(\pi_\lambda\in A)+Q_\lambda(A)\sum_{i=1}^n p_i^2\right| \leq 2\delta^*\varepsilon+2\delta^2\,, \tag{65}$$

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

Recall that $\lambda_k = \sum_{i=1}^n p_i^k \ \ (k \geq 2)$. Denote

$$\Delta h(\cdot) = h(\cdot + 1) - h(\cdot).$$

Theorem 12. [23] *If* $\mathbb{E}|h(\pi_\lambda)|\pi_\lambda^4 < \infty$, *then*

$$\begin{aligned} &\Big|\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda) + \lambda_2 \mathbb{E}\Delta^2 h(\pi_\lambda)/2\Big| \\ &\leq \frac{e^{p_n^*}}{(1-p_n^*)^2}\Big(\lambda_3 \mathbb{E}|\Delta^3 h(\pi_\lambda)|/3 + \lambda_2^2 \mathbb{E}|\Delta^4 h(\pi_\lambda)|/8\Big). \end{aligned} \tag{66}$$

Note that the assumption $\mathbb{E}|\Delta^k h(\pi_\lambda)| < \infty$ is equivalent to $\mathbb{E}\pi_\lambda^k |h(\pi_\lambda)| < \infty \ (k \in \mathbb{N})$, see Proposition 1 in [23]. Borisov & Ruzankin ([23], Lemma 2) have showed also that

$$\sup_k \mathbb{P}(S_n = k)/\mathbb{P}(\pi_\lambda = k) \leq (1-p_n^*)^2.$$

Deheuvels et al. ([43], Corollary 2.2) have shown that

$$\Big|d_G(S_n; \pi_\lambda) - \lambda_2 e^{-\lambda}\lambda^{[\lambda]}/[\lambda]!\Big| \leq 2(2\theta)^{3/2}\sqrt{\lambda}/(1-\sqrt{2\theta}) \qquad (\theta < 1/2).$$

Borisov & Vorozheikin [25] present asymptotic expansions of $\chi^2(\mathbf{B}(n,p); \mathbf{\Pi}(np))$.

Asymptotic expansions for $\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda)$, where $\{X_k\}$ are non-negative integer-valued random variables and function h is either bounded or grows at a polynomial rate, are presented in Barbour [10]. Asymptotic expansions for $\mathbb{E}h(S_n) - \mathbb{E}h(\pi_\lambda)$, where $\|h\|_1 = 1$, have been given by Barbour & Jensen [11]. A good survey is Zacharovas & Hwang [121].

Unit measure (signed measure) approximations. A number of authors evaluated the accuracy of unit measure (signed measure) approximation to the distribution of a sum S_n of independent Bernoulli r.v.s (see, e.g., [26, 18, 16]). In particular, Borovkov [26] has generalised inequality (25). Note that asymptotic expansion (65) is an example of a unit measure approximation.

Denote by P_n the distribution corresponding (with some abuse of notation) to $\pi_{\lambda+\lambda_2} + 2\pi_{-\lambda_2/2}$ (i.e., P_n is a convolution of $\mathbf{\Pi}(\lambda+\lambda_2)$ and a Poisson unit measure with parameter $-\lambda_2/2$ on $2\mathbf{Z}_+$). In the assumption that $\theta < 1/2$ Barbour & Xia ([18], Theorem 4.1) have shown that

$$d_{TV}(\mathcal{L}(S_n); P_n) \leq \lambda_3/\lambda(1-2\lambda_2)\sqrt{\lambda - \lambda_2 - p_n^*}\,.$$

Čekanavičius & Kruopis [30] present an estimate of the accuracy of unit measure approximation in terms of the Gini-Kantorovich distance:

$$d_G(\mathcal{L}(S_n); Q_n) \leq C\lambda_*^{-1}\lambda_2(1 + (\lambda/\lambda_2)^2),$$

where C is an absolute constant, $\lambda_* = \max\{1; \lambda - \lambda_2\}$ and Q_n (with some abuse of notation) corresponds to $\pi_{\lambda-\lambda_2/2} - \pi_{-\lambda_2/2}$ (Q_n is a convolution of $\mathbf{\Pi}(\lambda - \lambda_2/2)$ and a Poisson unit measure with parameter $-\lambda_2/2$ on $-\mathbf{Z}_+$). Note that $\sum_k kQ_n(k) = \mathbb{E}S_n$, $\sum_k (k-\lambda)^2 P_n(k) = \operatorname{var} S_n$.

Barbour & Čekanavičius [16] present a unit measure approximation to the distribution of a sum of independent integer-valued r.v.s.

2.6. Sum of a random number of random variables

Let $\nu, X, X_1, X_2, ...$ be independent non-negative random variables, where r.v. ν takes values in $\mathbf{Z}_+$, $X, X_1, X_2, ...$ are i.i.d. random variables.

Set

$$S_\nu = X_1 + ... + X_\nu \,.$$

A natural task is to evaluate the accuracy of Poisson approximation to $\mathcal{L}(S_\nu)$.

We consider first the case where $X, X_1, X_2, ...$ are Bernoulli $\mathbf{B}(p)$ r.v.s.

Denote $\bar\nu := \mathbb{E}\nu$. Then $\mathbb{E}S_\nu = p\bar\nu$.

Let $\mathcal{F}_2$ denote the class of functions $h : \mathbf{Z}_+ \to \mathbb{R}$ such that $||\Delta^2 h|| \le 1$, and set

$$d_2(X;Y) = \sup_{h\in\mathcal{F}_2} |\mathbb{E}h(X) - \mathbb{E}h(Y)|.$$

Logunov [70] points out that $d_{TV}(X;Y) \le d_2(X;Y)$, and shows that

$$d_2(S_\nu;\pi_{p\bar\nu}) \le p^2 d_*(\nu;\pi_{p\bar\nu}),$$

where $d_*(X;Y) = \sum_{k\ge1} k(k-1)|\mathbb{P}(X=k) - \mathbb{P}(Y=k)|/2$. Note that $d_*(X;Y) \ge d_G(X;Y)$.

Yannaros [120] has shown that

$$d_{TV}(\pi_\lambda;\pi_\mu) \le \min\{|\sqrt{\lambda} - \sqrt{\mu}|; |\lambda-\mu|\}. \tag{67}$$

The first term in (67) has been improved by Roos [96]:

$$d_{TV}(\pi_\lambda;\pi_\mu) \le \sqrt{\tfrac{2}{e}}\left|\sqrt{\lambda} - \sqrt{\mu}\right|. \tag{67*}$$

Note that the second term in the r.-h.s. of (67) is a consequence of the trivial inequality

$$d_{TV}(\pi_\lambda;\pi_\mu) \le 1 - \exp(-|\lambda-\mu|) \tag{67*}$$

that follows by defining π_λ and π_μ on a common probability space (cf. (4.10) in [82]).

It is easy to see that

$$\begin{aligned} d_{TV}(S_\nu;\pi_\lambda) &\le \sum\nolimits_k \mathbb{P}(\nu=k) d_{TV}(S_k;\pi_\lambda), \\ d_{TV}(S_k;\pi_\lambda) &\le d_{TV}(S_k;\pi_{kp}) + d_{TV}(\pi_{kp};\pi_\lambda). \end{aligned}$$

Using these inequalities and (67), Yannaros [120] has shown that

$$d_{TV}(S_\nu;\pi_{p\bar\nu}) \le \min\Big\{\frac{p}{2\sqrt{1-p}}\,;(1-\mathbb{E}e^{-p\nu})p\Big\} + \min\Big\{p\mathbb{E}|\nu-\bar\nu|; \sqrt{p\frac{\operatorname{var}\nu}{\bar\nu}}\Big\}. \tag{68}$$

The term $\min\big\{p/2\sqrt{1-p}\,;(1-\mathbb{E}e^{-p\nu})p\big\}$ in (68) is inherited from (28) and (29).

The right-hand side of (68) can be sharpened using (32), ($67^\star$) and (67^*):

$$d_{TV}(S_\nu;\pi_{p\bar\nu}) \le 3p/4e + 2(\delta^*+\delta^2) + \min\left\{\left(1-\mathbb{E}e^{-p|\nu-\bar\nu|}\right);\mathbb{E}|\sqrt{\nu}-\sqrt{\bar\nu}\,|\sqrt{2p/e}\right\}. \tag{69}$$

Note that $\mathbb{E}|\sqrt{\nu}-\sqrt{\bar\nu}\,| \le \min\{\bar\nu^{-1/2}\sqrt{\mathrm{var}\,\nu}\,;\bar\nu^{-3/2}\mathrm{var}\,\nu\}$.

Mixed Poisson distribution. A number of authors (see, e.g., Roos [96]) have evaluated the accuracy of Poisson approximation to the *mixed Poisson distribution*, i.e., the distribution of the r.v. π_ν, where $\mathcal{L}(\pi_t) = \mathbf{\Pi}(t)$, r.v. ν takes values in $[0;\infty)$:

$$\mathbb{P}(\pi_\nu = m) = \int_0^\infty \mathbb{P}(\pi_y = m)\mathbb{P}(\nu\in dy) \qquad (m\ge 0).$$

If $\{X_i\}$ are Poisson $\mathbf{\Pi}(\lambda)$ r.v.s, then $S_\nu \stackrel{d}{=} \pi_{\lambda\nu}$ is a mixed Poisson random variable.

Denote by $\mathbf{NB}(n,p)$ the negative Binomial distribution: $\mathcal{L}(S_n) = \mathbf{NB}(n,p)$ if

$$\mathbb{P}(S_n = i) = \binom{i+n-1}{i}(1-p)^n p^i \qquad (i\ge 0).$$

The negative Binomial distribution $\mathbf{NB}(t,p)$ is a mixed Poisson distribution with

$$\mathbb{P}(\nu\in dy)/dy = r^t y^{t-1}e^{-yr}/\Gamma(t) \qquad (y>0),$$

where $r = p/(1-p)$, $\Gamma(y) = \int_0^\infty x^{y-1}e^{-x}dx$.

Roos [96] presents estimates of the accuracy of Poisson approximation to the mixed Poisson distribution with a correct constant at the leading term.

Sum of 0-1 random variables till the stopping time. We now consider the situation where r.v. ν depends on $\{X_i\}$.

Let $X, X_1, X_2, ...$ be i.i.d. non-negative integer-valued r.v.s. Set $S_0 = 0$,

$$S_n = X_1 + ... + X_n \quad (n\ge 1),$$

and let $\mu(t)$ denote the stopping time:

$$\mu(t) = \max\{n\ge 0 : S_n \le t\}.$$

Theorems 13–14 below are cited from see [82], ch. 3. They provide estimates of the accuracy of Poisson approximation to the distribution of the number

$$N_t(x) = \sum_{j=1}^{\mu(t)} \mathbb{1}\{X_j \ge x\} + \mathbb{1}\{t - S_{\mu(t)} \ge x\} \tag{70}$$

of exceedances of a "high" level $x\in[0;t]$ till $\mu(t)$.

Note that

$$\{N_t(x) = 0\} = \{M_t < x\},$$

where

$$M_t = \max\{t - S_{\mu(t)}; \max_{1\le i\le \mu(t)} X_i\} \tag{71}$$

is the largest observation among $\{X_1, ..., X_{\mu(t)}, t - S_{\mu(t)}\}$.

Let $X_{k,t}$ denote the k^{th} largest element among $\{X_1, ..., X_{\mu(t)}, t - S_{\mu(t)}\}$. Then

$$\{X_{k,t} < x\} = \{N_t(x) < k\}.$$

The topic has applications in finance. For instance, suppose a bank has opened a credit line for a series of operations, and the total amount of credit is t units of money. The cost of the i-th operation is denoted by X_i. What is the probability that the bank will ever pay x or more units of money at once? that there will be a certain number of such payments? Information on the asymptotic properties of the distribution of random variables M_t and $N_t(x)$ can help to answer these questions.

Let $\{X_i^<, i\ge 1\}$, $\{X_j^>, j\ge 1\}$ be independent r.v.s with the distributions

$$\mathcal{L}(X^<) = \mathcal{L}(X|X < x), \ \mathcal{L}(X^>) = \mathcal{L}(X|X \ge x).$$

We set $p_x = \mathbb{P}(X \ge x)$,

$$S_0(k) = 0, \quad S_m(k) \;=\; \sum_{i=0}^{k} X_i^> + \sum_{i=k+1}^{m} X_i^< \quad (m \ge 1).$$

Let K_*, K^* denote the end-points of $\mathcal{L}(X)$, and set

$$\begin{aligned} \tau_k &= \tau_k' - k\,, \ \tau_k' = \min\{n : S_n(k) > t - x\}, \\ \lambda_k &\equiv \lambda_k(t, x, k) = p_x(t - x - k\mathbb{E}X^>)/\mathbb{E}X^<. \end{aligned}$$

In Theorems 13–14 we assume the following condition: *there exist constants* $D < \infty$ *and* $D_* \in (K_*; K^*)$ *such that*

$$\int_x^\infty \mathbb{P}(X \ge y)dy \le D\mathbb{P}(X \ge x) \qquad (x \ge D_*). \tag{72}$$

Condition (72) means the tail of $\mathcal{L}(X)$ is light (cf. (3.15) in [82]). Inequality (72) holds if function $g(x) = e^{cx}\mathbb{P}(X \ge x)$ is not increasing as $x > 1/c$ $(\exists c > 0)$. The equality in (72) for all $x \ge 0$ may be attained only if $\mathcal{L}(X)$ is exponential with $\mathbb{E}X = D$.

Theorem 13. *For any* $k \in \mathbf{Z}_+$, *as* $t \to \infty$,

$$\sup_{x\in B_+(t)} \left| \mathbb{P}(N_t(x) = k) - \mathbb{P}(\pi_{\lambda_k} = k) - \sum_{r=0}^{k-1} \big(\mathbb{P}(\pi_{\lambda_k} = r) - \mathbb{P}(\pi_{\lambda_{k-1}} = r)\big) \right| = O(1/t),$$

where $B_+(t) = (K_*; K^* \wedge t/(k+2))$.

Let $\pi(t, x)$ denote a Poisson r.v. with parameter $p_x t/\mathbb{E}X$.

Theorem 14. *For any* $k \in \mathbf{Z}_+$, *as* $t \to \infty$,

$$\sup_{K_* < x < K^*} |\mathbb{P}(N_t(x) = k) - \mathbb{P}(\pi(t,x) = k)| = O(t^{-1} \ln t).$$

One can show that $N_t(x)$ is "small" when x is "large":

$$\sup_{x \geq \sqrt{t}} \mathbb{P}(N_t(x) \geq 1) \leq q^{\sqrt{t}} \qquad (\exists q \in (0;1)).$$

Theorem 3.7 in [82] presents asymptotic expansions for $\mathbb{P}(N_x(t) = k)$. The asymptotic expansions for $\mathcal{L}(M_t)$ are available under a weaker moment assumption (cf. [82], ch. 3).

The number of intervals between consecutive jumps of a Poisson process. Consider a Poisson jump process $\{\pi_\lambda(s), s \geq 0\}$ with parameter $\lambda > 0$, and let η_i denote the moment of its i^{th} jump. Set $X_i = \eta_i - \eta_{i-1}$. Then $N_t(x)$ is the number of intervals between consecutive jumps with lengths greater or equal to x. If the points of jumps represent catastrophic/rare events, then $N_t(x)$ can be interpreted as the number of "long" intervals without catastrophes.

Let $\pi_{t,x}$ be a Poisson r.v. with parameter $t\lambda e^{-\lambda x}$. Then for any $k \in \mathbf{Z}_+$, as $t \to \infty$,

$$\sup_{0 < x < t} |\mathbb{P}(N_t(x) = k) - \mathbb{P}(\pi_{t,x} = k)| = O\big(t^{-1} \ln t\big) \tag{73}$$

(cf. (3.12) in [82]).

Open problems.
2.5. Will asymptotic expansions for $\mathcal{L}(N_x(t))$ hold under a weaker moment assumption?
2.6. Generalise the results of Theorems 13–14 to the case of

$$N_t(x) = \sum_{j=1}^{\mu(t)} Y_i \mathbb{1}\{X_j \geq x\} + Y_{\mu(t)+1} \mathbb{1}\{t - S_{\mu(t)} \geq x\},$$

where $\{(X_i, Y_i)_{i \geq 1}\}$ is a sequence of i.i.d. pairs of r.v.s, $Y_i > 0$.

3. Applications

Applications of the theory of Poisson approximation to meteorology, reliability theory and extreme value theory have been discussed in [7, 54, 66, 82]. In this section we present a number of results that are not fully covered in existing surveys.

3.1. Long head runs

Let $\{\xi_i, i \geq 1\}$ be a sequence of 0-1 random variables.

We say a head run (a series of 1's) starts at $i = 1$ if $\xi_1 = 1$; a series starts at $i > 1$ if $\xi_{i-1} = 0$, $\xi_i = 1$. If $\xi_{i-1} = 0$, $\xi_i = ... = \xi_{i+k-1} = 1$, we say the head run is of length $\geq k$.

For instance, if $n = 5$ and $\xi_1 = \xi_2 = \xi_3 = 1, \xi_4 = 0, \xi_5 = 1$, there is one series (head run) of length 3 and one series of length 1.

Denote

$$A_0 = \{\xi_1 = ... = \xi_k = 1\}, \;\; A_i = \{\xi_i = 0, \xi_{i+1} = ... = \xi_{i+k} = 1\} \quad (i>1).$$

Then

$$W_n(k) = \sum_{i=0}^{n-k} \mathbb{1}\{A_i\} \qquad (n \geq k \geq 1)$$

is the number of head runs of length $\geq k$ among $\xi_1, ..., \xi_n$ (NLHR).

Set

$$L_n = \max\{k : \xi_{i+1} = ... = \xi_{i+k} = 1 \; (\exists i \leq n-k)\}. \tag{74}$$

L_n is the *length of the longest head run* (LLHR) among $X_1, ..., X_n$. Obviously,

$$\{L_n < k\} = \{W_n(k) = 0\}.$$

The problem of approximating the distribution of LLHR is a topic of active research; it has applications in reliability theory and psychology (cf. [7, 82]).

Let $\{\xi_i, i \geq 1\}$ be i.i.d. Bernoulli $\mathbf{B}(p)$ r.v.s, $p \in (0;1)$, and let π_λ denote the Poisson $\mathbf{\Pi}(\lambda)$ r.v.. Theorem 7 with $B_i = [i-k; i+k]$ and

$$\lambda \equiv \lambda(n,k,p) = p^k(1+(n-k)(1-p))$$

yields the following

Corollary 15. *As* $n \geq k \geq 1$,

$$d_{TV}(W_n(k); \pi_\lambda) \leq (1-e^{-\lambda})(2k+1)p^k. \tag{75}$$

An open question is if estimate (75) can be improved. Note, for instance, that (75) does not yield (77) even for $j = 0$.

There is a close relation between $N_t(x)$ and $W_n(k)$. Let $\eta_0 = 0$,

$$\eta_i = \min\{k > \eta_{i-1} : \xi_k = 0\}, \;\; X_i = \eta_i - \eta_{i-1} \quad (i \geq 1).$$

Then

$$W_n(k) = \sum_{j=1}^{\mu(t)} \mathbb{1}\{X_j - 1 \geq k\} + \mathbb{1}\{n - \eta_{\mu(n)} \geq k\}. \tag{76}$$

Hence

$$W_{n-1}(k) = N_n(k+1).$$

Denote $\lambda_k = n(1-p)p^k$. Theorem 14 entails

Corollary 16. *For any* $j \in \mathbf{Z}_+$, *as* $n \to \infty$,

$$\max_{1 \le k \le n} |\mathbb{P}(W_n(k) = j) - \mathbb{P}(\pi_{\lambda_k} = j)| = O\left(n^{-1} \ln n\right). \qquad (77)$$

According to Theorem 3.13 in [82], the rate $n^{-1} \ln n$ in (77) cannot be improved.

The number of long non-decreasing runs. Let $\xi_i = \mathbb{1}\{Y_i \le Y_{i+1}\}$, where $\{Y_i\}$ are i.i.d.r.v.s with a continuous d.f.. Then NLHR $W_n(k)$ is the number of non-decreasing runs of length $\ge k$ (NLNR), and LLHR is the length of the longest non-decreasing run (LLNR) among $Y_1, ..., Y_{n+1}$. We denote LLNR by L_n^+ and NLNR by $W_n^+(k)$.

The topic concerning LLNR and NLNR has applications in finance. It is well known that prices of shares and financial indexes evolve in cycles of growth and decline. Knowing the asymptotics of L_n^+ and $W_n^+(k)$ can help evaluating the length of the longest period of continuous growth/decline of a particular financial instrument as well as the distribution of the number of such long periods.

Pittel [84] has proved a Poisson limit theorem for NLNR (see also Chryssaphinou et al. [35] concerning the case of a Markov chain).

We proceed with the case of i.i.d.r.v.s with a continuous d.f.. Note that $\mathcal{L}(\xi_i) = \mathbf{B}(1/2)$ and $\mathbb{P}(Y_1 \le ... \le Y_{k+1}) = 1/(k+1)!$. Set $\lambda_{n,k} = \mathbb{E}W_n^+(k)$. Then

$$\lambda_{n,k} = 1/(k+1)! + (n-k)/k!(k+2).$$

Theorem 7 with $B_i = [i-k-1; i+k+1]$ yields the following

Corollary 17. *As* $n \ge k \ge 1$,

$$d_{TV}(W_n^+(k); \pi_{\lambda_{n,k}}) \le (1-e^{-\lambda_{n,k}})(2k+3)/(k+1)!.$$

The accuracy of compound Poisson approximation to the distribution of the number of non-decreasing runs of fixed length has been evaluated by Barbour & Chryssaphinou [15], p. 982 (continuous d.f.) and Minakov [79] (discrete d.f.). Concerning the asymptotics of LLNR, see [36, 82] and references therein.

Open problem.
3.1. Improve the estimates of Corollary 15 and Corollary 17.
3.1. Derive (45)-type (i.e., uniform in k) estimates of the accuracy of (possibly shifted) Poisson approximation to $\mathcal{L}(W_n(k))$ and $\mathcal{L}(W_n^+(k))$.

3.2. Long match patterns

Closely related to the number of long head runs is the number of long match patterns (NLMP) between sequences of independent r.v.s. Information on the distribution of NLMP and the length of the longest match pattern (LLMP) can help recognising "valuable" fragments of DNA sequences (see [2, 3, 78, 80, 81]).

In this section we present results on the accuracy of Poisson approximation to the distribution of NLMP. Theorems 18, 20 and Lemma 22 below have been established by the author (see [82], ch. 4).

Let $X, X_1, ..., X_m$, $Y, Y_1, ..., Y_n$ be independent non-degenerate random variables taking values in a discrete state space $\mathbf{A}$. Denote $(k \in \mathbb{N})$

$$\begin{aligned} T_{ij} &= \mathbb{1}\{X_{i+1}=Y_{j+1}, \ldots, X_{i+k}=Y_{j+k}\}, \\ \tilde{T}_{ij} &= T_{ij}(k)\mathbb{1}\{X_i \neq Y_j\}, \\ T^*_{ij} &= \tilde{T}_{ij} \quad (i \geq 1, j \geq 1), \quad T^*_{ij} = T_{ij} \quad (i=0 \text{ or } j=0). \end{aligned}$$

Then

$$M^*_{m,n} = \max\left\{k \leq \min(m,n) : \max_{(i,j)\in J} T_{ij} = 1\right\}$$

is the *length of the longest match pattern* between $(X_1 \ldots X_m)$ and $(Y_1 \ldots Y_n)$.

LLMP $M^*_{m,n}$ is a 2–dimensional analog of LLHR L_n. If $\mathbf{A} = \{0,1\}$ and $Y_1 = ... = Y_n = 1$, then $M^*_{n,n} = L_n$.

Given $m \geq k$, $n \geq k$, let

$$J \equiv J(k,m,n) = \{(i,j) : 0 \leq i \leq m-k, 0 \leq j \leq n-k\}.$$

Denote by

$$W_{m,n} \equiv W_{m,n}(k) = \sum_{(i,j)\in J} T^*_{ij}$$

the *number of long match patterns* (patterns of length $\geq k$). Then

$$\{M^*_{m,n} < k\} = \{W_{m,n} = 0\}.$$

In the rest of this section we assume that r.v.s $X, X_1, ..., X_m$, $Y, Y_1, ..., Y_n$ are identically distributed. We set

$$\lambda \equiv \lambda_{k,m,n} = \mathbb{E}W_{m,n}, \quad m' = m-k+1, \quad n' = n-k+1.$$

Then $\lambda = (m'-1)(n'-1)(1-p)p^k + (m'+n'-1)p^k$.

Denote

$$p = \mathbb{P}(X=Y), \; p_j = \mathbb{P}(X=j), \; q_k = \sum_{j\in A} p_j^{k+1}, \; q = q_2,$$

and let

$$p_* = \max_{j\in A} p_j, \; c_+ = \log(1/q) - 1, \; c_* = \log(1/p_*),$$

where $\log$ is to the base $1/p$. Note that

$$p_*^2 < p \, , \; p^2 \leq q \leq p_* \, . \tag{78}$$

Taking into account Hölder's inequality, we conclude that

$$1 \geq c_+ \geq c_* > 1/2. \tag{79}$$

Note that $c_+ = c_* = 1$ if $\mathcal{L}(X)$ is uniform over a finite alphabet.

Let $\pi_{m,n}$ denote a Poisson random variable with parameter $\lambda_{k,m,n}$.

The following theorem shows that the distribution of the number of long match patterns can be well approximated by the Poisson law.

Theorem 18. *If $n \geq k$ and $m \geq k \geq 1$, then*

$$d_{TV}(W_{m,n}; \pi_{m,n}) \leq \frac{1-e^{-\lambda}}{\lambda} m'n'(2k+1)\left(2kq_{2k} + (m'+n'-1)(p^{2k}+q^k)\right). \quad (80)$$

Theorem 18 has been derived using Theorem 7 and Lemma 22.

Denote

$$\Delta_{m,n}(k) = |\mathbb{P}(M^*_{m,n} < k) - \exp(-\lambda)|.$$

Corollary 19. *For any constant $C \in \mathbb{R}$, as $m \to \infty$, $n \to \infty$,*

$$\max_{k \geq C + \log mn} \Delta_{m,n}(k) = O\left((m+n)(mn)^{-c_+}(\ln mn) + (mn)^{1-2c_*}(\ln mn)^2\right). \quad (81)$$

If $m \to \infty$ and $n \to \infty$ in such a way that $(\ln mn)/(\min\{m,n\}) \to 0$, then

$$\max_{1 \leq k \leq m \wedge n} \Delta_{m,n}(k) = O\Big((m+n)(mn)^{-c_+}(\ln mn)^{1+c_+} + (mn)^{1-2c_*}(\ln mn)^{1+2c_*}\Big). \quad (82)$$

It is easy to see that the accuracy of estimate (81) depends on the relation between m and n. If $\mathcal{L}(X)$ is uniform over a finite alphabet and $(\ln mn)/(m \wedge n) \to 0$, then Corollary 19 implies that

$$\max_{1 \leq k \leq m \wedge n} \left|\mathbb{P}(M^*_{m,n} < k) - e^{-\lambda}\right| = O\left(n^{-1}(\ln n)^2\right) \quad (83)$$

If $\mathcal{L}(X)$ is uniform over a finite alphabet and

$$c \leq m/n \leq 1/c$$

for some constant $c > 0$, then the right-hand side of (81) becomes $O(n^{-1} \ln n)$. We conject that the correct rate of convergence in (83) for the uniform $\mathcal{L}(X)$ is $O\left(n^{-1} \ln n\right)$.

The reason why (80) does not yield such a rate is the lack of factor $e^{-\lambda}$ on the right-hand side. Results obtained for LLHR by the method of recurrent inequalities do produce such a factor (cf. Theorem 3.12 in [82]).

In a more general situation one can consider NLMP with say r mismatches allowed. An estimate of the accuracy of Poisson approximation to the distribution of the number of long r-interrupted match patterns among $X_1, ..., X_m, Y_1, ..., Y_n$ (match patterns of length $\geq k$ with $\leq r$ "interruptions") can be found in [78, 82]. Neuhauser [81] considers a situation where $\mathcal{L}(X)$ may differ from $\mathcal{L}(Y)$ and only insertions and deletions (but no mismatches) are allowed to occur; she presents a logarithmic estimate of the rate of Poisson approximation to the distribution of the number of such long patterns.

The Zubkov–Mihailov statistic. Let now $Y_i = X_i$ $(\forall i)$, $m = n$. Denote

$$N_n^* \equiv N_n^*(k) = \sum_{(i,j)\in A(n,k)} T_{ij}^*\,,$$

where

$$A(n,k) = \{(i,j) : 0 \le i < j \le n-k\} \quad (n > k).$$

N_n^* is the number of long match patterns in one and the same sequence, $X_1, ..., X_n$.

Statistic N_n^* was introduced by Zubkov & Mihailov [126] who have shown that $\mathcal{L}(N_n^*)$ is asymptotically Poisson $\mathbf{\Pi}(\mu)$ if

$$n^2 p^k(1-p)/2 \to \mu > 0, \quad nk^t p_*^k \to 0 \quad (\forall\, t > 0).$$

Denote by

$$M_n^* = \max\{k \le n \colon \max_{(i,j)\in A(n,k)} T_{ij} = 1\}$$

the *length of the longest match pattern* among $X_1, ..., X_n$. Then $\{M_n^* < k\} = \{N_n^* = 0\}$.

The next theorem evaluates the accuracy of Poisson approximation to $\mathcal{L}(N_n^*)$.

Theorem 20. *If* $n > 3k \ge 3$, *then*

$$d_{TV}(N_n^*; \pi_{n,k}^*) \le \frac{1-e^{-\lambda^*}}{\lambda^*}\left((n^*)^3(2k+1)\,(p^{2k}+q^k) + 2(kn^*)^2 q_{2k}\right) + 2kn^* p^k,$$

where $\lambda^* \equiv \lambda_{n,k}^* = (n-3k+1)\,p^k(1 + (n-3k)(1-p)/2)$, $n^* = n-k$, $\mathcal{L}(\pi_{n,k}^*) = \mathbf{\Pi}(\lambda^*)$.

Theorem 20 has been derived using Theorem 7 and Lemma 22.

Denote

$$\Delta^*(n,k) = |\mathbb{P}(M_n^* < k) - \exp(-\lambda_{n,k}^*)|.$$

Corollary 21. *As* $n \to \infty$,

$$\max_{k \ge C+2\log n} \Delta^*(n,k) = O\left(n^{1-2c_+}\ln n + n^{2-4c_*}(\ln n)^2\right), \tag{84}$$

$$\max_{1 \le k < n/3} \Delta^*(n,k) = O\left(n^{1-2c_+}(\ln n)^{1+c_+} + n^{2-4c_*}(\ln n)^{1+2c_*}\right). \tag{85}$$

If $\mathcal{L}(X)$ is uniform over a finite alphabet, then the right-hand side of (84) is $O(n^{-1}\ln n)$, the right-hand side of (85) is $O(n^{-1}(\ln n)^2)$.

The key result behind Theorems 18 and 20 is the following

Lemma 22. *For all natural* i, j, i', j' *such that* $(i,j) \ne (i',j')$,

$$\mathbb{P}(T_{ij}^* = T_{i'j'}^* = 1) \le q_{2k}\,. \tag{86}$$

Denote by

$$\tau_k = \min\{n\colon N_n^*(k) \neq 0\}$$

the first instance a match pattern of length k appears in the sequence $\{X_i, i \geq 1\}$. Then

$$\{\tau_k > n\} = \{M_n^* < k\}.$$

The results concerning the asymptotics of τ_k can be derived from the corresponding results for M_n^*.

NLMP with a small number of mismatches has been considered by several authors (see [78, 82] and references therein).

A number of authors evaluated the accuracy of compound Poisson approximations to the distribution of NLMP (see [78, 82, 105] and references therein).

Open problems.
3.2. Derive uniform in k estimates of (possibly shifted) Poisson approximation to $\mathcal{L}(W_{m,n})$ and $\mathcal{L}(N_n^*)$.
3.3. Find the 2$^{\text{nd}}$-order asymptotic expansions for $\mathbb{P}(W_{m,n} \in \cdot)$ and $\mathbb{P}(N_n^* \in \cdot)$.
3.4. Check if the correct rate of convergence in (82) and (85) in the case of uniform $\mathcal{L}(X)$ is $O\left(n^{-1} \ln n\right)$.
3.5. Improve the estimate of the rate of convergence in the limit theorem for the length of the longest r-interrupted match pattern.

4. Compound Poisson approximation

The topic of compound Poisson (CP) approximation is vast. 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 [5, 68, 85, 87] and references therein). Recall that the class of infinitely divisible distributions coincides with the class of weak limits of compound Poisson distributions [61].

The topic has applications in extreme value theory, insurance, reliability theory, patterns matching, etc. (cf. [7, 12, 15, 66, 82]). For instance, in (re)insurance applications the sum $S_n = \sum_{i=1}^n Y_i \mathbb{1}\{Y_i > x_i\}$ of integer-valued r.v.s allows to account for the total loss from the claims $\{Y_i\}$ that exceed excesses $\{x_i\}$. If the probabilities $\mathbb{P}(Y_i > x_i)$ are small, $\mathcal{L}(S_n)$ can be accurately approximated by a Poisson or a compound Poisson law.

In extreme value theory one deals with the number of extreme (rare) events represented by a sum of 0-1 r.v.s (indicators of rare events). The indicators can be dependent. A well-known approach consists of grouping observations into blocks which can be considered almost independent [20]. The number of r.v.s in a block is an integer-valued r.v., hence the number of rare events is a sum of almost independent integer-valued r.v.s. In all such situations the block sums are non-zero with small probabilities. More information concerning applications can be found in [7, 12, 48, 66].

This section concentrates on results concerning compound Poisson (CP) approximation that can be derived from the results concerning pure Poisson approximation. The main attention is given to results that are missed in existing surveys.

4.1. CP limit theorem

Compound Poisson (CP) distribution is the distribution of a r.v.

$$\sum_{i=1}^{\pi_\lambda} \zeta_i \,,$$

where $\zeta_0 = 0$, r.v.s $\pi_\lambda, \zeta, \zeta_1, \zeta_2, ...$ are independent, $\mathcal{L}(\zeta) = \mathbf{\Pi}(\lambda)$, $\zeta_i \stackrel{d}{=} \zeta$ $(i \geq 1)$.

We denote $\mathcal{L}(\sum_{i=1}^{\pi_\lambda} \zeta_i)$ by $\mathbf{\Pi}(\lambda, \zeta) \equiv \mathbf{\Pi}(\lambda, \mathcal{L}(\zeta))$.

Typically $\zeta \neq 0$ $w.p.\,1$. The requirement $\zeta \neq 0$ $w.p.\,1$ may be omitted. Indeed, denote $p = \mathbb{P}(\zeta \neq 0)$. Then by Khintchin's formula ([61], ch. 2),

$$\zeta \stackrel{d}{=} \tau_p \zeta', \tag{87}$$

where τ_p and ζ' are independent r.v.s, $\mathcal{L}(\zeta') = \mathcal{L}(\zeta|\zeta \neq 0)$, $\mathcal{L}(\tau_p) = \mathbf{B}(p)$. Note that

$$\mathbf{\Pi}(t, \tau_p \zeta') = \mathbf{\Pi}(tp, \zeta')$$

(cf. (6.26) in [82]).

Let $\{X_{n,1}, ..., X_{n,n}\}_{n \geq 1}$ be a triangle array of stationary dependent 0-1 random variables, i.e., sequence $X_{n,1}, ..., X_{n,n}$ is stationary for each $n \in \mathbb{N}$. Set

$$S_n = X_{n,1} + ... + X_{n,n}.$$

Let $\zeta_{r,n}$ be a r.v. with distribution (8). The following Theorem 23 generalises Theorem 3 to the case of CP approximation. It states that under certain assumptions weak convergence of the cluster size distribution (see (89) below) is necessary and sufficient for the CP limit theorem for S_n.

In Theorem 23 below we will assume (11) and the following condition:

$$\limsup_{n \to \infty} n \mathbb{P}(X_{n,1} \neq 0) < \infty. \tag{88}$$

Note that relation (11) does not imply (88) — for example, consider the case $X_{n,1} \equiv X$. Denzel & O'Brien [44] present an example of an α–mixing sequence such that (11) holds though (88) does not.

Theorem 23. *Assume conditions (11), (88) and* Δ. *If*

$$\zeta_{r,n} \Rightarrow \zeta \qquad (n \to \infty) \tag{89}$$

for a sequence $\{r = r_n\} \in \mathcal{R}$*, then*

$$S_n \Rightarrow \sum_{i=0}^{\pi(\lambda)} \zeta_i. \tag{90}$$

The limit in (90) does not depend on the choice of a sequence $\{r_n\} \in \mathcal{R}$*.*

If S_n *converges weakly to a random variable* Y*, then* $\mathcal{L}(Y)$ *is compound Poisson* $\mathbf{\Pi}(\lambda, \zeta)$*, where* $\lambda = -\ln \mathbb{P}(Y = 0)$*. If* $\lambda > 0$*, then (89) holds for some random variable* ζ *and sequence* $\{r = r_n\} \in \mathcal{R}$*.*

Theorem 23 is effectively Theorem 5.1 from [82].

4.2. Accuracy of CP approximation

Let $\{X_i\}$ be independent r.v.s that are non-zero with small probabilities (cf. [68, 76, 86, 122]). Set $S_n := X_1 + ... + X_n$, and denote

$$p_i = \mathbb{P}(X_i \neq 0) \quad (i \geq 1), \;\; \lambda = p_1 + ... + p_n \,.$$

According to Khintchin's formula (87),

$$X_i \stackrel{d}{=} \tau_i X_i', \tag{87*}$$

where τ_i and X_i' are independent r.v.s, $\mathcal{L}(X_i') = \mathcal{L}(X_i | X_i \neq 0)$, $\mathcal{L}(\tau_i) = \mathbf{B}(p_i)$. Hence

$$S_n \stackrel{d}{=} \tau_1 X_1' + ... + \tau_n X_n'.$$

Let $\zeta_1, ..., \zeta_n$ be independent compound Poisson $\mathbf{\Pi}(p_i, X_i')$ random variables. Set $Z_n = \sum_{i=1}^n \zeta_i$. Note that Z_n is a compound Poisson random variable:

$$\mathcal{L}(Z_n) = \mathbf{\Pi}(\lambda, X_\eta'),$$

where r.v. η is independent of $X_1', ..., X_n'$, $\mathbb{P}(\eta = j) = p_j / \lambda \;\; (1 \leq j \leq n)$.

A simple estimate of the accuracy of CP approximation to $\mathcal{L}(S_n)$ follows from the property of d_{TV} and (26):

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

(cf. LeCam [68], Theorem 1).

Zaitsev [122] has derived an estimate of the accuracy of compound Poisson approximation that can be sharper if λ is "large". The following Theorem 24 presents Zaitsev's result.

Theorem 24. *There exists an absolute constant* C *such that*

$$d_K(S_n; Z_n) \leq C p_n^* \,. \tag{91}$$

Inequality (91) has been generalised to the multidimensional situation by Zaitsev [123].

We consider now the situation where

$$X_i' \stackrel{d}{=} X' \qquad (\forall i).$$

In such a situation an estimate of the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ follows from the estimate of the accuracy of pure Poisson approximation to $\mathcal{L}(\tau_1+...+\tau_n)$.

Indeed, denote

$$\nu_n = \tau_1 + ... + \tau_n, \quad Y = \sum_{i=1}^{\pi_\lambda} X_i',$$

where Poisson $\mathbf{\Pi}(\lambda)$ r.v. π_λ is independent of $X_1', X_2', ...$. Then

$$S_n \stackrel{d}{=} \sum_{i=1}^{\nu_n} X_i'. \tag{92}$$

It is easy to check (see, e.g., Presman [86]) that

$$d_{TV}(S_n; Y) \equiv d_{TV}\Big(\sum_{i=1}^{\nu_n} X_i'; \sum_{i=1}^{\pi_\lambda} X_i'\Big) \le d_{TV}(\nu_n; \pi_\lambda). \tag{93}$$

Kolmogorov ([62], formula (30)) has applied (93) without formulating it explicitly. Presman [86] was probably the first to formulate (93) explicitly and present its proof.

Presman [86] has evaluated $d_{TV}(\nu_n; \pi_\lambda)$ (and hence $d_{TV}(S_n; Y)$) using (93) and (30). Michel [76] has applied (93) and the Barbour–Eagleson estimate (29). An application of (93) and (32) yields

$$d_{TV}(S_n; Y) \le 3\theta/4e + 2\delta^*\varepsilon + 2\delta^2. \tag{94}$$

According to [82], Lemma 5.4,

$$d_G(S_n; Y) \le d_G(\nu; \pi_\lambda)\mathbb{E}|X'|. \tag{95}$$

A combination of (36) and (95) entails

$$d_G(S_n; Y) \le \Big(1 \wedge \tfrac{4}{3}\sqrt{2/e\lambda}\Big)\lambda_2\mathbb{E}|X'|. \tag{96}$$

Further results on the accuracy of compound Poisson approximation can be found in [14, 32, 33, 95, 125, 119].

Open problem.
4.1 Evaluate constant C in (91).

4.3. CP approximation to $\mathbf{B}(n,p)$

Below we present an estimate of the accuracy of compound Poisson approximation to the Binomial law related to the topic of pure Poisson approximation.

Let $X, X_1, ...$ be independent Bernoulli $\mathbf{B}(p)$ r.v.s. Presman [85] has shown that

$$\sup_p d_{TV}(\mathbf{B}(n,p); F_{n,p}) = O(n^{-2/3}), \tag{97}$$

where the shifted compound Poisson distribution $F_{n,p}$ is constructed via *Poisson* distributions (a similar result in terms of d_K is due to Meshalkin [75]).

We present Presman's result in Theorem 25 below (see also [5], ch. 4).

Denote by $\lceil x \rceil$ the integer number that is the nearest to x from above, and let

$$\gamma = \lceil 3np^2 - 2np^3 \rceil, \quad \beta = \gamma - 3np^2 + 2np^3 \in [0;1), \quad q = 1-p.$$

Let η_1, η_2, η_3 be independent r.v.s with distributions

$$\mathcal{L}(\eta_1) = \mathbf{\Pi}(pq^2 - \beta/n), \ \mathcal{L}(\eta_2) = \mathbf{\Pi}(p^2q + \beta/3n), \ \mathcal{L}(\eta_3) = \mathbf{\Pi}(\beta/6n).$$

Set

$$Y := \gamma/n + \eta_1 - \eta_2 + 2\eta_3.$$

Note that $Y - \gamma/n$ is a CP r.v.. One can check that

$$\mathbb{E}Y = p, \ \mathbb{E}(Y-p)^2 = pq, \ \mathbb{E}(Y-p)^3 = pq(q-p).$$

Let $F_{n,p} := \mathcal{L}(Y_1 + ... + Y_n)$, where $\{Y_i\}$ are independent copies of Y.

Theorem 25. *There exists an absolute constant* C *such that*

$$d_{TV}(\mathbf{B}(n,p); F_{n,p}) \le C\varepsilon_{n,p} \qquad (0 \le p \le 1/2), \tag{98}$$

where $\varepsilon_{n,p} = \min\{np^2; p; \max\{1/(np)^2; 1/n\}\}$.

Bound (97) follows after noticing that $\sup_{0\le p\le 1/2} \varepsilon_{n,p} = O(n^{-2/3})$. Clearly, it suffices to consider only $p \in [0;1/2]$: if $\mathcal{L}(S_n) = \mathbf{B}(n,p)$, then $\mathcal{L}(n - S_n) = \mathbf{B}(n, 1-p)$.

Dependent 0-1 r.v.s. Let $X, X_1, ...$ be a stationary sequence of 0-1 r.v.s. The following Theorem 26 is an application of (93) in the case of dependent r.v.s.

Let $\pi, \zeta_1^{(r)}, \zeta_2^{(r)}, \ldots$ be independent random variables, where $1 \le r \le n$, $\pi_{n,r}$ is a Poisson $\mathbf{\Pi}(kq)$ r.v., $\zeta_0^{(r)} = 0$,

$$\begin{aligned} \mathcal{L}(\zeta_i^{(r)}) &= \mathcal{L}(S_r | S_r > 0) \quad (i \ge 1), \\ q &= \mathbb{P}(S_r \ne 0), \quad k = [n/r]. \end{aligned}$$

Denote $p = \mathbb{P}(X=1)$,

$$Y_n = \sum_{i=0}^{\pi_{n,r}} \zeta_i^{(r)}.$$

The distribution of $S_n = X_1 + ... + X_n$ can be approximated by a CP distribution $\mathcal{L}(Y_n)$.

Theorem 26. *If* $n > r > l \geq 0$, *then*

$$d_{TV}(S_n; Y_n) \leq \kappa_{n,r} rp + (2kl + r')p + nr^{-1}\gamma_n(l), \tag{99}$$

$$d_G(S_n; Y_n) \leq rp \min\Big\{np\,; \tfrac{4}{3}\sqrt{2np/e}\Big\} + (2kl + r')p + n\gamma_n(l), \tag{100}$$

where $r' = n - rk$, $\kappa_{n,r} = \min\{1 - e^{-np}\,; 3/4e + (1 - e^{-np})rp\}$ *and* $\gamma_n(l) = \min\{4\alpha(l)\sqrt{r}\,; \beta_n(l)\}$.

Theorem 26 is effectively Theorem 5.2 from [82].

If the random variables $\{X_i\}$ are independent, then (99) with $r = 1$, $l = 0$ yields (29) and (32).

If the random variables $\{X_i\}$ are m–dependent, then one can choose $l = m$, $r = \lceil \sqrt{mn}\, \rceil$, the smallest integer greater than or equal to $\sqrt{mn}$, and get the estimate $d_{TV}(S_n; Y_n) \leq 4p\lceil \sqrt{mn}\, \rceil$.

Further reading on the topic of the accuracy of compound Poisson approximation to the distribution of a sum of dependent r.v.s includes [92, 34] and references therein.

Open problem.
4.2. Evaluate constant C in (98).

5. Poisson process approximation

The topic of point process approximation is vast; an interested reader is referred to [38, 74]. This section concentrates on the results concerning Poisson process approximation that are closely related to the results on Poisson approximation to the distribution of a sum of 0-1 random variables. The main attention is given to results that are missed in existing surveys.

Point process counting locations of rare events. Let $\{\xi_i, i \geq 1\}$ be Bernoulli r.v.s (e.g., $\xi_i = \mathbb{1}\{X_i > u_n\}$, where u_n is a "high" level). Then

$$S_n(\cdot) = \sum_{i=1}^{n} \xi_i \mathbb{1}\{\, i/n \in \cdot\} \tag{101}$$

can be called a "Bernoulli process".

$S_n(\cdot)$ counts *locations* of extreme/rare events represented by r.v.s $\{\xi_i\}$. A typical example of a rare event is an exceedance of a high threshold.

For instance, let $X, X_1, X_2, ...$ be a stationary sequence of random variables, and let $\{u_n\}$ be a sequence of levels. Set $\xi_i = \mathbb{1}\{X_i > u_n\}$. Then $S_n(\cdot) =$

$N_n(\cdot, u_n)$, where

$$N_n(B, u_n) = \sum_{i=1}^{n} \mathbb{1}\{\, i/n \in B, X_i > u_n\} \qquad (B \subset (0;1]). \tag{101*}$$

Process $N_n(\cdot, u_n)$ counts *locations* of exceedances of level u_n.

Let $\{r = r_n\}$ be a sequence obeying (7). We denote by $\zeta_{r,n}$ a r.v. with distribution (8).

Theorem 27. *Assume (11), (88) and mixing condition* Δ. *If (10) holds, then*

$$N_n(\cdot, u_n) \Rightarrow N(\cdot), \tag{102}$$

where $N(\cdot)$ *is a Poisson point process with intensity rate* λ.

Theorem 27 is a particular case of Theorem 7.2 in [82]. The necessity part of Theorem 27 is given by Theorem 3: if (102) holds, then so does (10). Leadbetter et al. [66], Theorem 5.2.1, present a version of Theorem 27 with condition (D') instead of (10).

Denote by Ξ_n a Poisson point process with intensity measure

$$\lambda(\cdot) = \sum_{i=1}^{n} p_i \mathbb{1}\{i/n \in \cdot\},$$

where $p_i = \mathbb{P}(\xi_i = 1)$.

The accuracy of Poisson process approximation to $\mathcal{L}(S_n(\cdot))$ has been evaluated by Brown [27] and Kabanov et al. [57], Theorem 3.2: if $\{\xi_i\}$ are independent, then

$$d_{TV}(S_n(\cdot); \Xi_n(\cdot)) \leq \sum_{i=1}^{n} p_i^2. \tag{25''}$$

Arratia et al. [2] have generalised (25″) to the case of dependent Bernoulli r.v.s.

Ruzankin [99] and Xia [118] present estimates of the accuracy of Poisson process approximation in terms of a d_G-type distance.

In the general case (when the limiting distribution of $\zeta_{r,n}$ is not degenerate) the limiting distribution of $N_n(\cdot, u_n)$ is necessarily compound Poisson (Hsing et al. [56], see also [82], ch. 7).

Excess process. Let $X, X_1, X_2, ...$ be a stationary sequence of r.v.s. If one is interested in the joint distribution of exceedances of several levels among $X_1, ..., X_n$, a natural tool is the excess process $N_n^{\varepsilon}(\cdot)$.

Set

$$N_n^{\varepsilon}(t) = \sum_{i=1}^{n} \mathbb{1}\{X_i > u_n(t)\} \qquad (t > 0).$$

Given $T > 0$, we call $\{N_n^{\varepsilon}(t), t \in [0; T]\}$ the *excess process.*

Process $N_n^\varepsilon(\cdot)$ describes variability in the *heights* of the observations.

Note that $N_n^\varepsilon(\cdot)$ is the "tail empirical process" for $Y_{n,1},...,Y_{n,n}$, where $Y_{n,i} = u_n^{-1}(X_i)$:

$$N_n^\varepsilon(t) = \sum_{i=1}^n \mathbb{1}\{Y_{n,i} < t\}. \quad (103)$$

There is considerable amount of research on the topic of tail empirical processes (see, e.g., [37, 72] and references therein).

We present necessary and sufficient conditions for the weak convergence of the excess process to a Poisson process in Theorem 28 below (cf. [82], ch. 7).

Suppose there is a sequence $\{u_n(\cdot), n \geq 1\}$ of functions on $[0;\infty)$ such that function $u_n(\cdot)$ is strictly decreasing for all large enough n, $u_n(0) = \infty$,

$$\limsup_{n\to\infty} n\mathbb{P}(X > u_n(t)) < \infty \qquad (0 < t < \infty), \quad (104)$$

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n(t)) = e^{-t} \qquad (t \geq 0), \quad (105)$$

where $M_n = \max\{X_1, ..., X_n\}$ is the sample maximum. Conditions (104) and (105) mean that $u_n(\cdot)$ is a "proper" normalising sequence for the sample maximum.

First, we recall the definitions of mixing (weak dependence) conditions.

Given $0 < t_1 < ... < t_k < \infty$, where $k \geq 1$, and a sequence $\{u_n(\cdot)\}_{n\geq 1}$, we denote

$$\tau = (t_1, ..., t_k), \;\; u_n(\tau) = (u_n(t_1), ..., u_n(t_k)).$$

Let $\mathcal{F}_{l,m}(\tau)$ be the σ–field generated by the events $\{X_i > u_n(t_j)\}$, $l \leq i \leq m, 1 \leq j \leq k$; mixing (weak dependence) coefficient $\alpha_n(l_n) := \alpha(l_n, u_n(\tau))$ is defined as above.

Condition $\Delta(\{u_n(\tau)\})$ is said to hold if $\alpha_n(l_n) \to 0$ for some sequence $\{l_n\}$ such that $l_n \to \infty$, $l_n/n \to 0$ as $n \to \infty$.

Condition Δ holds if $\Delta(\{u_n(\tau)\})$ is in force $(\forall\, 0 < t_1 < \ldots < t_k < \infty,\; k \geq 1)$.

Class $\mathcal{R}(\tau)$. If $\Delta\{u_n(\tau)\}$ holds, then there exists a sequence $\{r_n\}$ such that (7) holds (for instance, one can take $r_n = \left[\sqrt{n \max\{l; n\alpha_n(l_n)\}}\,\right]$). We denote by $\mathcal{R}(\tau)$ the class of all such sequences.

The next condition describes the joint distribution of exceedances of several levels.

We say that *condition* C'_τ holds if there exists a sequence $\{r_n\} \in \mathcal{R}(\tau)$ such that for every $1 \leq i < j \leq k$ and every $t_i < t_j$ from $\{t_1, ..., t_k\}$

(a) $\mathbb{P}(N_r[u_n(t_{i-1}); u_n(t_i)) = 1) \sim (t_i - t_{i-1})r/n$,
$\mathbb{P}(N_r[u_n(t_{i-1}); u_n(t_i)) = j) = o(r/n) \quad (j \geq 2)$,

(b) $\mathbb{P}(N_r(u_n(t_i)) > 0, N_r[u_n(t_i); u_n(t_j)) > 0) = o(r/n)$.

Condition C' holds if C'_τ *is valid for all* $0<t_1<...<t_k<\infty,\ k\geq 1.$

Theorem 28. *Assume mixing condition condition* Δ*, and let* $\pi(\cdot)$ *denote a Poisson process with intensity rate 1. Then*

$$N_n^\varepsilon(\cdot) \Rightarrow \pi(\cdot) \tag{106}$$

if and only if condition C' *holds.*

Example 5.1. Let $X, X_1, X_2, ...$ be i.i.d.r.v.s with the distribution function (d.f.) F. Denote $K^* = \sup\{x\colon F(x)<1\}$, and assume that

$$\mathbb{P}(X\geq x)/\mathbb{P}(X>x) \to 1 \tag{G}$$

as $x \to K^*$ (Gnedenko's condition [51]). Set $u_n(t) = F_c^{-1}(t/n)$, where

$$F_c(\cdot) := \mathbb{P}(X>\cdot).$$

Then excess process $\{N_n^\varepsilon(\cdot), t\in[0;1]\}$ converges weakly to a pure Poisson process N with intensity rate 1 (cf. [82], ch. 8). Process N admits the representation

$$N \stackrel{d}{=} \sum_{j=1}^{\pi(1)} \gamma_j(\cdot),$$

where $\gamma_j(t) \stackrel{d}{=} \mathbb{1}\{\xi<t\}$ and r.v. ξ has uniform $\mathbf{U}[0;1]$ distribution. □

The accuracy of approximation $N_n^\varepsilon(\cdot) \approx N(\cdot)$ can be evaluated as well (cf. Deheuvels & Pfeifer [41], Kabanov & Liptser [58], Novak [82], ch. 8).

Given $T>0$, let $\pi(np)$ denote a Poisson $\mathbf{\Pi}(np)$ r.v., where $p = \mathbb{P}(X > u_n(T))$. Let $\eta, \eta_1, \eta_2, ...$ be independent of $\pi(np)$ i.i.d. processes with the distribution

$$\mathcal{L}(\eta(\cdot)) = \mathcal{L}(\mathbb{1}\{X>u_n(\cdot)\}|X>u_n(T)) \equiv \mathcal{L}(\mathbb{1}\{Y_{n,1}<\cdot\}|Y_{n,1}<T)$$

$(i\geq 1)$. One can check that $N_n^\varepsilon(\cdot) \stackrel{d}{=} \sum_{i=1}^{\nu_n} \eta_i(\cdot)$, where r.v. ν_n is independent of $\{\eta_i\}$, $\mathcal{L}(\nu_n) = \mathbf{B}(n,p)$.

Let $\{X_i\}$ be independent r.v.s. An application of (93) and (32) yields

$$d_{TV}\Big(N_n^\varepsilon(\cdot); \sum_{i=1}^{\pi(np)} \eta_i(\cdot)\Big) \leq 3p/4e + 2(1-e^{-np})(1+\varepsilon)p^2, \tag{107}$$

where $\varepsilon = \min\big\{1; (2\pi[(n-1)p])^{-1/2} + 2(1-e^{-np})p/(1-1/n)\big\}$ (cf. [82], Theorem 8.3, where we applied (29) instead of (32)).

Note that $\sum_{i=1}^{\pi(np)} \eta_i(\cdot)$ is a Poisson process. Inequality (107) is an improvement of a particular result by Major [72].

Let $X, X_1, X_2, ...$ be i.i.d.r.v.s, and let $T > 0$. According to (49) (see also (6.5) in [41]), the total variation distance between $\{\sum_{i=1}^n \mathbb{1}\{Y_{n,i} < t\}, t \in [0;T]\}$ and the approximating Poisson process coincides with $d_{TV}(\mathbf{B}(n,p); \mathbf{\Pi}(np))$, where $p = \mathbb{P}(Y_{n,1} \le T)$.

In a general situation excess process $\{N_n^\varepsilon(\cdot)\}$ may converge weakly to a process of more complex structure:

$$\{N_n^\varepsilon(t),\, t \le T\} \Rightarrow \left\{\sum_{j=1}^{\pi} \gamma_j(t/T),\, t \le T\right\}, \tag{108}$$

where π is a Poisson r.v., $\{\gamma_j(\cdot)\}$ are independent jump processes.

Process $\left\{\sum_{j=1}^{\pi} \gamma_j(\cdot)\right\}$ can be called *Poisson cluster process* or *compound Poisson process of the second order* (regarding the standard CP process as a "compound Poisson process of the first order").

Necessary and sufficient conditions for the weak convergence of the excess process to a compound Poisson process or a Poisson cluster processes are presented in [82], ch. 7, 8.

General point process of exceedances. Consider now a two–dimensional point process N_n^* that counts locations of rare events (e.g., exceedances of "high" thresholds) as well as their "heights": for any Borel set $A \subset (0;1] \times [0;\infty)$ we set

$$N_n^*(A) := \sum_{i=1}^{n} \mathbb{1}\{\, \big(i/n, u_n^{-1}(X_i)\big) \in A \,\}. \tag{109}$$

If $\{X_i\}$ are i.d.d.r.v.s, or if $\{X_i, i \ge 1\}$ is a strictly stationary sequence obeying certain mixing conditions, then $N_n^*(\cdot)$ converges weakly to a pure Poisson point process (Adler [1]). Theorem 29 below presents Adler's result.

We will assume a multilevel version of the "declustering" *condition* (D'):

$$\lim_{n\to\infty} n \sum_{i=1}^{r} \mathbb{P}(X_{i+1} > u_n(t), X_1 > u_n(t)) = 0 \tag{D'_+}$$

for any sequence $\{r = r_n\} \in \mathcal{R}(t)$, $0 < t < \infty$.

Theorem 29. *If conditions* Δ *and* (D'_+) *hold, then* N_n^* *converges weakly to a pure Poisson point process* N^* *on* $(0;1] \times [0;\infty)$ *with the Lebesgue intensity measure.*

Example 5.2. Let $Y, Y_1, Y_2, ...$ be a sequence of i.i.d.r.v.s with exponential $\mathbf{E}(1)$ distribution, and set

$$X_i = Y_i + Y_{i+1}\,.$$

Evidently, $\{X_i, i \ge 1\}$ is a stationary sequence of 1–dependent r.v.s.

Let $u \equiv u_n(t) = \ln[t^{-1} n \ln n]$, $t > 0$. Then $\mathbb{P}(X > u_n(t)) \sim t/n$, and condition (D'_+) holds. According to Theorem 29, $N_n^* \Rightarrow N^*$, the Poisson point process with the Lebesgue intensity measure (cf. [82], ch. 7). □

Adler's result has been generalised to the case of compound Poisson approximation: necessary and sufficient conditions for the weak convergence of N_n^* to a compound Poisson point process can be found in [82], ch. 7. Necessary and sufficient conditions for the weak convergence of N_n^* to a Poisson cluster process are given in [82], ch. 8.

An estimate of the accuracy of approximation $N_n^*(\cdot) \approx \sum_{j=1}^{\pi(T)} \gamma_j(\cdot)$ in terms of the $d_G(X;Y)$-type distance has been established in [17].

Open problem.
5.1. Improve the estimate of the accuracy of approximation $N_n^* \approx N^*$ presented in [17].

Acknowledgements

The author is grateful to P. S. Ruzankin and the referee for helpful remarks.

References

[1] Adler R.J. (1978) Weak convergence results for extremal processes generated by dependent random variables. — Ann. Probab., v. 6, No 4, 660–667. MR0494408

[2] Arratia R., Goldstein L. and Gordon L. (1989) Two moments suffice for Poisson approximation. — Ann. Probab., v. 17, No 1, p. 9–25. MR0972770

[3] Arratia R., Gordon L. and Waterman M.S. (1990) The Erdös–Rényi law in distribution, for coin tossing and sequence matching. — Ann. Statist., v. 18, No 2, p. 539–570. MR1056326

[4] Arak T.V. (1981) On the convergence rate in Kolmogorov's uniform limit theorem. — Theory Probab. Appl., v. 26, 219–239, 437–451. MR0616618

[5] Arak T.V. and Zaitsev A.Yu. (1984) Uniform limit theorems for sums of independent random variables. — Proc. Steklov Inst. Math., v. 174, 3–214. MR0871856

[6] Arenbaev N.K. (1976) Asymptotic behavior of the multinomial distribution. — Theory Probab. Appl., v. 21, 805–810. MR0478288

[7] Balakrishnan N. and Koutras M.V. (2001) *Runs and scans with applications*. New York: Wiley. MR1882476

[8] Banis R. (1985) A Poisson limit theorem for rare events of a discrete random field. — Litovsk. Mat. Sb., v. 25, No 1, 3–8. MR0795851

[9] 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. MR0706618

[10] Barbour A.D. (1987) Asymptotic expansions in the Poisson limit theorem. — Ann. Probab., v. 15, No 2, 748–766. MR0885141

[11] Barbour A.D. and Jensen J.L. (1989) Local and tail approximations near the Poisson limit. — Scand. J. Statist., v. 16, 75–87. MR1003970

[12] Barbour A.D., Holst L. and Janson S. (1992) *Poisson Approximation.* Oxford: Clarendon Press. MR1163825

[13] Barbour A.D., Chen L.H.Y. and Choi K.P. (1995) Poisson approximation for unbounded functions: independent summands. — Statistica Sinica, v. 5, No 2, 749–766. MR1347617

[14] Barbour A.D. and Utev S.A. (1999) Compound Poisson approximation in total variation. — Stochastic Process. Appl., v. 82, 89–125. MR1695071

[15] Barbour A.D. and Chryssaphinou O. (2001) — Ann. Appl. Probab., v. 11, No. 3, 964–1002. MR1865030

[16] 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. MR1905850

[17] Barbour A.D., Novak S.Y. and Xia A. (2002) Compound Poisson approximation for the distribution of extremes. — Adv. Appl. Probab., v. 34, No 1, 223–240. MR1895339

[18] Barbour A.D. and Xia A. (1999) Poisson perturbations. — ESAIM Probab. Statist., v. 3, 131–150. MR1716120

[19] Barbour A.D. and Xia A. (2006) On Stein's factors for Poisson approximation in Wasserstein distance. — Bernoulli, v. 6, 943–954. MR2274850

[20] 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. MR1512353

[21] Bernstein S.N. (1946) *Probability Theory*. Moscow: Nauka.

[22] Borisov I.S. (1996) Poisson approximation of the partial sum process in Banach spaces. — Sibirsk. Math. J., v. 37, No 4, 627–634. MR1643350

[23] 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. MR1944003

[24] Borisov I.S. (2003) A remark on a theorem of Dobrushin and couplings in the Poisson approximation in Abelian groups. — Theory Probab. Appl., v. 48, No 3, 521–528. MR2141351

[25] Borisov I.S. and Vorozheikin I.S. (2008) Accuracy of approximation in the Poisson theorem in terms of χ^2 distance. — Sibir. Math. J., v. 49, No 1, 5–17. MR2400567

[26] Borovkov K. A. (1988) On the problem of improving Poisson approximation. — Theory Probab. Appl., v. 33, No 2, 343–347. MR0954584

[27] Brown T.C. (1983) Some Poisson approximations using compensators. — Ann. Probab. 11, 726–744. MR0704559

[28] Chen L.H.Y. (1975) Poisson approximation for dependent trials. — Ann. Probab., v. 3, 534–545. MR0428387

[29] Čekanavičius V. (1998) Poisson approximations for sequences of random variables. — Statist. Probab. Lett., v. 39, 101–107. MR1652516

[30] Čekanavičius V. and Kruopis J. (2000) Signed Poisson approximation: a possible alternative to normal and Poisson laws. — Bernoulli, v. 6, No 4, 591–606. MR1777684

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

[32] Čekanavičius V. and Wang Y.H. (2003) Compound Poisson approximation for sums of discrete nonlattice random variables. — Adv. Appl. Probab., v. 35, 228–250. MR1975512

[33] Čekanavičius V. and Roos B. (2006) An expansion in the exponent for compound binomial approximations. — Liet. Matem. Rink., v. 46, 67–110. MR2251442

[34] Čekanavičius V. and Vellaisamy P. (2010) Compound Poisson and signed compound Poisson approximations to the Markov binomial law. — Bernoulli, v. 16, No 4, 1114–1136. MR2759171

[35] Chryssaphinou O., Papastavridis S. and Vaggelatou E. (2001) Poisson approximation for the nonoverlapping appearances of several words in Markov chains. — Combinator. Probab. Computing, v. 10, no. 4, 293–308. MR1860437

[36] Csaki E. and Földes A. (1996) On the length of the longest monotone block. — Studia Scientiarum Mathematicarum Hungarica, v. 31, 35–46. MR1367699

[37] Dabrowski A., Ivanoff G. and Kulik R. (2009) Some notes on Poisson limits for empirical point processes. — Canadian J. Statist., v. 37, 347–360. MR2547203

[38] Daley D.J. and Vere-Jones D. (2008) *An introduction to the theory of point processes, v. II. General Theory and Structure*. New York: Springer. MR2371524

[39] Dall' Aglio G. (1956) Sugli estremi dei momenti delle funzioni di ripartizione dopia. — Ann. Scuola Normale Superiore Di Pisa, ser. Cl. Sci., v. 3, No 1, 33–74. MR0081577

[40] Deheuvels P. and Pfeifer D. (1986) A semigroup approach to Poisson approximation. — Ann. Probab., v. 14, No 2, 663–676. MR0832029

[41] Deheuvels P. and Pfeifer D. (1988) Poisson approximation of distributions and point processes. — J. Multivar. Anal., v. 25, 65–89. MR0935295

[42] Deheuvels P. and Pfeifer D. (1988) On a relationship between Uspensky's theorem and Poisson approximation. — Ann. Inst. Statist. Math., v. 40, 671–681. MR0996692

[43] Deheuvels P., Pfeifer D. and Puri M.L. (1989) A new semigroup technique in Poisson approximation. — Semigroup Forum, v. 38, 189–201. MR0976202

[44] Denzel G.E. and O'Brien G.L. (1975) Limit theorems for extreme values of chain–dependent processes. — Ann. Probab., v. 3, No 5, 773–779. MR0386068

[45] Dobrushin R. L. (1970) Prescribing a system of random variables by conditional distributions. — Theory Probab. Appl., v. 15, No 3, 458–486. MR0298716

[46] Franken P. (1964) Approximation der verteilungen von summen unabhängiger nichtnegativer ganzzahliger zufallsgrössen durch Poissonsche verteilunged. — Mathematische Nachrichten, v. 27, 303–340. MR0171294

[47] Galambos J. (1987) *The asymptotic theory of extreme order statistics*. Melbourne: R.E.Krieger Publishing Co. MR0936631

[48] Gerber H.U. (1979) *An introduction to mathematical risk theory*. Philadelphia: Huebner Foundation. MR0579350

[49] Gini C. (1914) Di unà 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.

[50] Gnedenko B.V. (1938) Uber die konvergenz der verteilungsgesetze von summen voneinander unabhangiger summanden. — C.R. Acad. Sci. URSS, v. 18, 231–234.

[51] Gnedenko B.V. (1943) Sur la distribution du terme maximum d'une série aléatoire. — Ann. Math., v. 44, 423–453. MR0008655

[52] Gnedenko B.V. and Kolmogorov A.N. (1954) *Limit distributions for sums of independent random variables*. Addison-Wesley: Cambridge, MA. MR0062975

[53] Harremoës P. and Ruzankin P.S. (2004) Rate of convergence to Poisson law in terms of information divergence. — IEEE Trans. Inform Theory, v. 50, No 9, 2145–2149. MR2097199

[54] Haight F.A. (1967) *Handbook of the Poisson distribution*. New York: Wiley. MR0208713

[55] Herrmann H. (1965) Variationsabstand zwischen der Verteilung einer Summe unabhängiger nichtnegativer ganzzahliger Zufallsgrössen und Poissonschen Verteilungen. — Mathematische Nachrichten, v. 29, No 5, 265–289. MR0190979

[56] Hsing T., Hüsler J. and Leadbetter M.R. (1988) On the exceedance point process for stationary sequence. — Probab. Theory Rel. Fields, v. 78, 97–112. MR0940870

[57] 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. MR0700211

[58] Kabanov Yu.M. and Liptser R.Sh. (1983) On convergence in variation of the distributions of multivariate point processes. — Z. Wahrscheinlichkeitstheor. verw. Geb., v. 63, 475–485. MR0705618

[59] 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. MR0096552

[60] Kerstan J. (1964) Verallgemeinerung eines satzes von Prochorow und Le Cam. — Z. Wahrsch. Verw. Gebiete, v. 2, 173–179. MR0165555

[61] Khintchin A.Y. (1933) *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete.* Berlin: Springer.

[62] Kolmogorov A.N. (1956) Two uniform limit theorems for sums of independent random variables. — Theory Probab. Appl., v. 1, No 4, 384–394.

[63] Kontoyiannis I., Harremoes P. and Johnson O.T. (2005) Entropy and the law of small numbers. — IEEE Trans. Inform. Theory, v. 51, No 2, 466–472. MR2236061

[64] Kozulyaev P.A. (1939) Asymptotic analysis of a fundamental formula of Probability Theory. — Acad. Notes Moscow Univ., v. 15, 179–182.

[65] Kruopis J. (1986) Precision of approximations of the generalized binomial distribution by convolutions of Poisson measures. — Lithuanian Math. J., v. 26, 37–49. MR0847204

[66] Leadbetter M.R., Lindgren G. and Rootzen H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer Verlag. MR0691492

[67] LeCam L. (1960) An approximation theorem for the Poison binomial distribution. — Pacif. J. Math., v. 19, 1 3, p. 1181–1197. MR0142174

[68] 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. MR0199871

[69] Liapunov A.M. (1901) Nouvelle forme du théor'eme sur la limite des probabilites. — Mem. Acad. Imp. Sci. St.–Peterburg, v. 12, 1–24.

[70] Logunov P.L. (1990) Estimates for the convergence rate to the Poisson distribution for random sums of independent indicators. — Theory Prob. Appl., 35, 587–590. MR1091218

[71] Loynes R.M. (1965) Extreme values in uniformly mixing stationary stochastic processes. — Ann. Math. Statist., v. 36, 993–999. MR0176530

[72] Major P. (1990) A note on the approximation of the uniform empirical process. — Ann. Probab., v. 18, No 1, 129–139. MR1043940

[73] Marcinkiewicz J. (1938) Sur les fonctions independantes II. — Func. Mat., v. 30, 349–364.

[74] Matthes K., Kerstan J. and Mecke J. (1978) *Infinitely divisible point processes*. New York: Wiley. MR0517931

[75] Meshalkin L.D. (1960) On the approximation of polynomial distributions by infinitely-divisible laws. — Theory Probab. Appl., v. 5, No 1, 114–124. MR0132586

[76] Michel R. (1987) An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. — ASTIN Bulletin, v. 17, 165–169.

[77] Mihailov V.G. (2001) Estimate of the accuracy of compound Poisson approximation for the distribution of the number of matching patterns. — Theory Probab. Appl., v. 46, No 4, 667–675. MR1971829

[78] Mihailov V.G. (2008) Poisson-type limit theorem for the number of pairs of almost matching patterns. — Theory Probab. Appl., v. 53, No 1, 59–71. MR2760565

[79] Minakov A.A. (2015) Compound Poisson approximation of the distribution of the number of monotone runs of fixed length. — Appl. Discrete Math., v. 28, No 2, 21–29.

[80] Mori T.F. (1990) More on the waiting time till each of some given patterns occurs as a run. — Can. J. Math., v. 42, No 5, 915–932. MR1081003

[81] Neuhauser C. (1994) A Poisson approximation for sequence comparisons with insertions and deletions. — Ann. Statist., v. 22, No 3, 1603–1629. MR1311992

[82] Novak S.Y. (2011) *Extreme value methods with applications to finance.* London: Chapman & Hall/CRC Press. ISBN 9781439835746 MR2933280

[83] Novak S.Y. (2019) On the accuracy of Poisson approximation. — Extremes, v. 22.

[84] Pittel B.G. (1981) Limiting behavior of a process of runs. — Ann. Probab., v. 9, no. 1, 119–129. MR0606801

[85] Presman E.L. (1983) Approximation of binomial distributions by infinitely divisible ones. — Theory Probab. Appl., v. 28, 393–403. MR0700218

[86] Presman E.L. (1985) The variation distance between the distribution of a sum of independent Bernoulli variables and the Poisson law. — Theory Probab. Appl., v. 30, No 2, 391–396. MR0792634

[87] Prokhorov Y.V. (1953) Asymptotic behavior of the binomial distribution. — Uspehi Matem. Nauk, v. 8, No 3(55), 135–142. MR0056861

[88] Prokhorov Y.V. (1956) Convergence of random processes and limit theorems in probability theory. — Theory Probab. Appl., v. 1, 157–214. MR0084896

[89] Rachev S.T. (1984) The Monge–Kantorovich problem on mass transfer and its stochastic applications. — Theory Probab. Appl., v. 29, No 4, 625–653. MR0773434

[90] Röllin A. (2005) Approximation of sums of conditionally independent variables by the translated Poisson distribution. — Bernoulli, v. 11, 1115–1128. MR2189083

[91] 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. MR0467889

[92] Roos M. (1994) Stein's method for compound Poisson approximation: the local approach. — Ann. Appl. Probab., v. 4, No 4, 1177–1187. MR1304780

[93] Roos B. (1999) Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. — Bernoulli, v. 5, No 6, 1021–1034. MR1735783

[94] Roos B. (2001) Sharp constants in the Poisson approximation. — Statist. Probab. Letters, v. 52, 155–168. MR1841404

[95] Roos B. (2003) Kerstan's method for compound Poisson approximation. — Ann. Probab., v. 31, No 4, 1754–1771. MR2016599

[96] Roos B. (2003) Improvements in the Poisson approximation of mixed Poisson distributions. — J. Statist. Plan. Inference, v. 113, 467–483. MR1965122

[97] Roos B. (2015) Refined total variation bounds in the multivariate and compound Poisson approximation. — arXiv:1509.04167v1. MR3647296

[98] Ruzankin P.S. (2001) On the Poisson approximation of the binomial distribution. — Siberian Math. J., v. 42, No 2, 353–363. MR1833166

[99] Ruzankin P.S. (2004) On the rate of Poisson process approximation to a Bernoulli process. — J. Appl. Probab., v. 41, No 1, 271–276. MR2036288

[100] Ruzankin P.S. (2010) Approximation for expectations of unbounded functions of dependent integer-valued random variables. — J. Appl. Prob., v. 47, 594–600. MR2668509

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

[102] Serfling R.J. (1975) A general Poisson approximation theorem. — Ann. Probab., v. 3, 726–731. MR0380946

[103] Serfling R.J. (1978) Some elementary results on Poisson approximation in a sequence of Bernoulli trials. — SIAM Rev., v. 20, No 3, 567–579. MR0482958

[104] Sevastyanov B.A. (1972) Limit Poisson law in a scheme of dependent random variables. — Theory Probab. Appl., v. 17, No 4, 733–737. MR0310943

[105] Schbath S. (2000) An overview on the distribution of word counts in Markov chains. — J. Comput. Biology, v. 7, 193–201.

[106] Shevtsova I.G. (2011) On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands. — arXiv:1111.6554v1. MR2848430

[107] Shorgin S.Y. (1977) Approximation of a generalized binomial distribution. — Theory Probab. Appl., v. 22, No 4, 846–850. MR0458544

[108] Smith R.L. (1988) Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. — Stoch. Proc. Appl., v. 30, No 2, 317–327. MR0978362

[109] Stein C. (1992) A way of using auxiliary randomization. — In: *Probability Theory.* Proc. Singapore Probab. Conf., pp. 159–180. Berlin: de Gruyter. MR1188718

[110] Tsaregradskii I.P. (1958) On uniform approximation of the binomial distribution with infinitely divisible laws. — Theory Probab. Appl., v. 3, No 4, 470–474. MR0102121

[111] Uspensky J.V. (1931) On Ch.Jordan's series for probability. — Ann. Math., v. 32, No 2, 306–312. MR1502999

[112] Utev S.A. (1992) *Extremal problems, characterisation, and limit theorems of Probability Theory.* — DSc Thesis. Novosibirsk: Novosibirsk Inst. Math., 280 pp.

[113] Vallander S.S. (1973) Calculation of the Wasserstein distance between probability distributions on the line. — Theory Probab. Appl., v. 18, No 4, 824–827. MR0328982

[114] Vasershtein L.N. (1969) Markov processes on a countable product of spaces describing large automated systems. — Probl. Inform. Trans., v. 14, 64–73. MR0314115

[115] Vervaat W. (1969) Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. — Statist. Neerlandica, v. 23, 79–86. MR0242235

[116] Witte H.-J. (1990) A unification of some approaches to Poisson approximation. — J. Appl. Probab., v. 27, No 3, pp. 611–621. MR1067026

[117] Xia A. (1997) On using the first difference in the Stein–Chen method. — Ann. Appl. Probab., v. 7, No 4, 899–916. MR1484790

[118] Xia A. (2005) Stein's method and Poisson process approximation. — In: *An introduction to Stein's method* (A.D. Barbour and L.H.Y. Chen, eds.) Singapore: World Scientific, 115–181. MR2235450

[119] Xia A. (2015) Stein's method for conditional compound Poisson approximation. — Statist. Probab. Lett., v. 100, 19–26. MR3324070

[120] Yannaros N. (1991) Poisson approximation for random sums of Bernoulli random variables. — Statist. Probab. Lett., v. 11, 161–165. MR1092977

[121] Zacharovas V. and Hwang H.-K. (2010) A Charlier-Parseval approach to Poisson approximation and its applications. — Lith. Math. J., v. 50, No 1, 88–119. MR2607681

[122] Zaitsev A.Yu. (1983) On the accuracy of approximation of distributions of sums of independent random variables, which are non-zero with a small probability, by accompanying laws. — Theory Probab. Appl., v. 28, No 4, 657–669. MR0726889

[123] Zaitsev A.Yu. (1988) A multidimensional variant of Kolmogorov's second uniform limit theorem. — Theory Prob. Appl., v. 34, 108–128. MR0993957

[124] Zaitsev A.Yu. (1991) An example of a distribution whose set of n-fold convolutions is uniformly separated from the set of infinitely divisible laws in the sense of the variation distance. — Theory Probab. Appl., v. 36, 419–425. MR1119511

[125] Zaitsev A.Yu. (2005) Approximation of a sample by a Poisson point process. — J. Math. Sciences, v. 128, No 1, 2556–2563. MR2038866

[126] Zubkov A.M. and Mihailov V.G. (1979) On the repetitions of s–tuples in a sequence of independent trials. — Theory Probab. Appl., v. 24, No 2, p. 267–279. MR0532442

[127] Zubkov A.M. and Serov A.A. (2012) A complete proof of universal inequalities for the distribution function of the binomial law. — Theory Probab. Appl., v. 57, No 3, 539–544. MR3196787