Probability Surveys
Vol. 19 (2022) 271–350
ISSN: 1549-5787
https://doi.org/10.1214/22-PS8

Compound Poisson approximation*

V. Čekanavičius

Vilnius University, Lithuania

and

S. Y. Novak

MDX University London, UK

Abstract: We overview the results on the topic of compound Poisson approximation to the distribution of a sum $S_n = X_1 + \cdots + X_n$ of (possibly dependent) random variables. We indicate a number of open problems and discuss directions of further research.

MSC2020 subject classifications: Primary 60E15, 60F05, 60G50, 60G51, 60G55, 60G70, 60J75; secondary 62E17, 62E20 .
Keywords and phrases: Compound Poisson approximation, signed compound Poisson measure, Kolmogorov's problem, total variation distance, Gini–Kantorovich distance.

Received August 2021.

Contents

*S. Y. Novak was supported by the Engineering and Physical Sciences Research Council [grant number EP/W010607/1].

Compound Poisson (CP) approximation appears naturally in situations where one deals with a large number of rare events. It has important applications in insurance, extreme value theory, reliability theory, mathematical biology, etc. (cf. [10, 13, 101, 127, 144]). The topic is an integral part of Kolmogorov's problem concerning infinitely divisible approximation to the distribution of a sum of independent r.v.s. It has attracted a considerable body of research.

However, existing surveys are surprisingly sketchy and typically present only results related to Stein's method, cf. [17, 18, 39]. A number of results obtained during the last three decades and even some classical results appear missed in existing surveys.

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

1. Preliminaries

1.1. Notation

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

Given r.v.s $X_1, ..., X_n$, we will denote

$$S_n = X_1 + \cdots + X_n \,.$$

Let $\mathbf{\Pi}(\lambda)$ denote a Poisson distribution with parameter λ; we usually denote by π_λ a Poisson $\mathbf{\Pi}(\lambda)$ random variable (r.v.).

Random variable Y has a *compound Poisson distribution* $\mathbf{\Pi}(\lambda, X) \equiv \mathbf{\Pi}(\lambda, \mathcal{L}(X))$ if

$$Y \stackrel{d}{=} X_0 + ... + X_{\pi_\lambda} \,, \tag{1.1}$$

where Poisson $\mathbf{\Pi}(\lambda)$ random variable π_λ is independent of $\{X_i\}_{i\geq 1}$, $X_0 = 0$, random variables $X, X_1, X_2, ...$ are independent, $X_i \stackrel{d}{=} X$ $(i\geq 1)$.

The characteristic function (ch.f.) of $\mathbf{\Pi}(\lambda, X)$ is

$$\exp(\lambda(\varphi_X(t)-1)),$$

where φ_X is a ch.f. of $\mathcal{L}(X)$. We call $\mathcal{L}(X)$ a compounding or multiplicity distribution.

If $\mathbb{E}|X|<\infty$, then $\mathbb{E}Y=\lambda\mathbb{E}X$. If $\mathbb{E}X^2<\infty$, then $\operatorname{var} Y=\lambda\mathbb{E}X^2$.

Given a set of non-negative numbers $\{\lambda_j\}_{j\geq 1}$ such that $\lambda := \sum_{j\geq 1}\lambda_j<\infty$, denote

$$Z = \sum_{j=1}^{\infty} j\pi_{\lambda_j}, \tag{1.2}$$

where $\{\pi_{\lambda_j}\}$ are independent Poisson $\mathbf{\Pi}(\lambda_j)$ variables ($\pi_{\lambda_j}\equiv 0$ if $\lambda_j=0$). Then Z is a compound Poisson random variable with characteristic function

$$\mathbb{E}\exp(\mathrm{i}tZ) = \exp\left(\sum_{j=1}^{\infty}\lambda_j(\mathrm{e}^{\mathrm{i}tj}-1)\right). \tag{1.3}$$

In other words, $Z \stackrel{d}{=} X_0 + ... + X_{\pi_\lambda}$, where $\mathbb{P}(X=j) = \lambda_j/\lambda$.

A compound Poisson distribution with a geometric multiplicity distribution is called sometimes a Pólya-Aeppli distribution, cf. [113], p. 410.

Random variable $S_{r,p}$ has a *Negative Binomial* $\mathbf{NB}(r,p)$ distribution with parameters $p\in(0,1)$ and $r>0$ if

$$\mathbb{P}(S_{r,p}=j) = \frac{\Gamma(r+j)}{\Gamma(r)\, j!}(1-p)^r p^j \qquad (j\geq 0), \tag{1.4}$$

where

$$\Gamma(y) = \int_0^\infty x^{y-1}e^{-x}dx.$$

The characteristic function of $\mathbf{NB}(r,p)$ is

$$(1-p)^r/(1-p\mathrm{e}^{\mathrm{i}t})^r.$$

Hence

$$\mathbb{E}S_{r,p} = rp/(1-p),\ \operatorname{var} S_{r,p} = rp/(1-p)^2.$$

It is known that the Negative Binomial distribution is a particular compound Poisson distribution.

If $n\in\mathbb{N}$, then $S_{n,p} \stackrel{d}{=} \xi_1 + ... + \xi_n$, where $\xi_1, ..., \xi_n$ are independent r.v.s with geometric $\mathbf{\Gamma}_0(p)$ distribution.

Random variables $\{X_{n,1}, ..., X_{n,n}\}$ are called *infinitesimal* if

$$\lim_{n\to\infty}\max_{1\leq j\leq n}\mathbb{P}(|X_{n,j}|>\varepsilon) = 0 \qquad (\forall\varepsilon>0). \tag{1.5}$$

Exponent of a measure. Let $I \equiv I_{\rm o}$ denote the distribution concentrated at 0, i.e.,

$$I(\{0\})=1,\ I(\mathbb{R}\backslash\{0\})=0.$$

In the multivariate case I denotes the distribution concentrated at $\bar{0} = (0,\ldots,0)$. Similarly, I_a is the distribution concentrated at a.

Given a finite measure Q, we denote

$$\exp(Q) = \sum_{j\geq 0} Q^{*j}/j!$$

Here powers Q^{*j} are understood in the convolution sense, $Q^{*0} = I$, where I is a degenerate distribution concentrated at zero.

Note that for any $a\in\mathbb{R}$

$$\exp(aI) = e^a I, \quad \exp(Q-aI) = e^{-a}\exp(Q).$$

Poisson distribution $\mathbf{\Pi}(\lambda)$ can be presented as

$$\mathbf{\Pi}(\lambda) = \exp\left(\lambda(I_1-I)\right) = e^{-\lambda}\exp(\lambda I_1),$$

compound Poisson distribution $\mathbf{\Pi}(\lambda, X)$ with $P_X := \mathcal{L}(X)$ can be presented as

$$\mathbf{\Pi}(\lambda, X) = \exp\left(\lambda(P_X-I)\right) = e^{-\lambda}\exp(\lambda P_X).$$

Measure Q is called a *unit measure* or a *signed measure* if $Q(\mathbb{R}) = 1$ but there exists a measurable set A such that $Q(A) < 0$.

The definition of $\mathbf{\Pi}(\lambda, X)$, where $\lambda \geq 0$, can be extended to the case of a signed compound Poisson (SCP) measure $\mathbf{\Pi}(-\lambda, X)$. Though probabilistic interpretation requires introduction of generalized "random variables", the structure of $\mathbf{\Pi}(-\lambda, X)$ is the same as that of $\mathbf{\Pi}(\lambda, X)$:

$$\mathbf{\Pi}(-\lambda, X) = \exp\left(-\lambda(P_X-I)\right) = e^{\lambda}\exp(-\lambda P_X).$$

With some abuse of notation we denote by $\pi_{-\lambda}$ a signed Poisson "random variable" meaning we use a signed Poisson measure $\exp\left(-\lambda(I_1-I)\right)$.

Accompanying distribution. Given a r.v. X, let $\pi_1, X_0 = 0, X_1, X_2, ...$ be independent r.v.s, where π_1 is a Poisson $\mathbf{\Pi}(1)$ r.v., $X_i \stackrel{d}{=} X$ $(i\geq 1)$. Set

$$\tilde{X} = \sum_{j=1}^{\pi_1} X_j\,. \tag{1.6}$$

Then $\mathcal{L}(\tilde{X})$ is called an "accompanying distribution", $\tilde{X}$ is called an accompanying X r.v. (terminology of Gnedenko [92]). The definition is valid for random elements taking values in a general measurable space as well.

Clearly,

$$\mathcal{L}(\tilde{X}) = \mathbf{\Pi}(1, X) = \exp\left(\mathcal{L}(X)-I\right), \quad \mathbb{E}e^{\mathrm{i}t\tilde{X}} = \exp\left(\mathbb{E}e^{\mathrm{i}tX}-1\right).$$

If $\mathbb{E}|X|<\infty$, then $\mathbb{E}\tilde{X} = \mathbb{E}X$. If $\mathbb{E}X^2<\infty$ and $\mathbb{E}X=0$, then $\operatorname{var}\tilde{X} = \operatorname{var}X$.

If

$$X \stackrel{d}{=} \tau X', \tag{1.7}$$

where τ is independent of X', $\mathcal{L}(\tau) = \mathbf{B}(p)$, then (cf. (6.26) in [144])

$$\mathcal{L}(\tilde{X}) = \mathbf{\Pi}(p, X') = \exp(p(\mathcal{L}(X')-I)), \quad \mathbb{E}e^{\mathrm{i}t\tilde{X}} = \exp\Big(p(\mathbb{E}e^{\mathrm{i}tX'}-1)\Big). \tag{1.8}$$

Given a sequence $\{X_1, ..., X_n\}$ or a triangular array $\{X_1 \equiv X_{n,1}, ..., X_n \equiv X_{n,n}\}_{n\geq 1}$ of random variables, recall that $S_n = X_1 + \cdots + X_n$. By

$$\tilde{S}_n = \tilde{X}_1 + \cdots + \tilde{X}_n \tag{1.9}$$

we denote the sum of independent accompanying random variables. Clearly,

$$\mathcal{L}(\tilde{S}_n) = \exp\left(\sum_{i=1}^{n}(\mathcal{L}(X_i)-I)\right). \tag{1.9*}$$

This presentation can be combined with (1.7), cf. (3.1).

If $X, X_1, ..., X_n$ are identically distributed r.v.s, then $\mathcal{L}(\tilde{S}_n) = \mathbf{\Pi}(n, X)$.

A sequence of random variables $\{X_k\}_{k\geq 1}$ is called m-dependent if $X_1, \ldots, X_s$ and $X_t, X_{t+1}, \ldots, X_n$ are independent for arbitrary s, t such that $1\leq s<t<\infty$, $t-s>m$. Observe that by grouping consecutive summands one can present the sum of m-dependent random variables as a sum of 1-dependent ones.

A sequence of random variables $X_1, X_2, \ldots X_n$ is *strictly stationary* if for arbitrary integer numbers $r, k, i_1<i_2<\ldots<i_r$ the distribution of $X_{k+i_1}, \ldots, X_{k+i_r}$ does not depend on k. In particular, r.v.s $X_1, X_2, \ldots X_n$ are identically distributed.

For any $x\in\mathbb{R}, k\in\mathbb{N}$,

$$x^{(k)} = x(x-1)\cdots(x-k+1)$$

is called the k^{th} *factorial* of x. We set $x^{(0)} = 1$.

If X is a random variable, then $\mathbb{E}X^{(k)}$ is called the k^{th} factorial moment of X. Factorial moments appear in Taylor's expansion of the factorial moment generating function

$$\mathbb{E}(1+t)^X = 1 + t\mathbb{E}X + \frac{t^2}{2!}\mathbb{E}X^{(2)} + \frac{t^3}{3!}\mathbb{E}X^{(3)} + \cdots$$

The so-called *factorial cumulants* (factorial semi-invariants) $\varkappa_k$ come from Taylor's expansion of the logarithm of the factorial moment generating function

$$\ln\mathbb{E}(1+t)^X = t\varkappa_1 + \frac{t^2}{2!}\varkappa_2 + \frac{t^3}{3!}\varkappa_3 + \cdots, \tag{1.10}$$

see [113], p. 53–55.

We denote by

$$Q_Y^+(h) \equiv Q_{\mathcal{L}(Y)}^+(h) = \sup_x \mathbb{P}(x \le Y \le x+h)$$

the concentration function of $\mathcal{L}(Y)$. Sometimes we may use the following variant of the concentration function:

$$Q_Y(h) \equiv Q_{\mathcal{L}(Y)}(h) = \sup_x \mathbb{P}(x < Y \le x+h).$$

We may use the same symbol C to denote different absolute constants (with or without indexes). Symbols $C(F)$, C_F, C_X denote constants that depend on the distribution function (d.f.) F of $\mathcal{L}(X)$.

As usual, $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = 1$.

We write $f(n) = O(g(n))$ if $f(n)/g(n) \le C < \infty$ for all large enough n.

For any $x \in \mathbb{R}$ let $[x]$ and $\{x\}$ denote the integer and the fractional parts of x. Below multiplication is superior to division.

1.2. Metrics

Historically, the accuracy of 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 F_X and F_Y is defined as

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

(in the multi-dimensional case F_X, F_Y denote multivariate distribution functions). Note that a version of multivariate Kolmogorov's distance based on comparing values of distributions on convex polyhedra has been proposed in [96].

In the case of integer-valued r.v.s it is natural to evaluate the accuracy of approximation in terms of a stronger *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, $d_K(X;Y) \le d_{TV}(X;Y)$.

According to Dobrushin's theorem (see [77, 32]),

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

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

The total variation distance can be expressed as

$$d_{TV}(X;Y) = \sup_f |\mathbb{E}f(X) - \mathbb{E}f(Y)|$$

where the supremum is over the class of measurable functions taking values in $[0;1]$ (see, e.g., [68], ch. 1.3, or [144], ch. 14.4).

If X and Y take values in the set $\mathbb{Z}$ of integer numbers, then

$$d_{TV}(X,Y) = \frac{1}{2}\sum_j |\mathbb{P}(X=j) - \mathbb{P}(Y=j)|.$$

The Gini–Kantorovich distance $d_G(X;Y) \equiv d_G(\mathcal{L}(X);\mathcal{L}(Y))$ between the distributions of r.v.s X and Y with finite first moments (known also as the Kantorovich–Wasserstein distance) is given by

$$d_G(X;Y) = \sup_{g\in\mathcal{L}_1} |\mathbb{E}g(X) - \mathbb{E}g(Y)|\,, \tag{1.11}$$

where $\mathcal{L}_1 = \{g\colon |g(x)-g(y)| \le |x-y|\}$ is the set of Lipschitz functions. Note that

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

where the infimum is taken over all random pairs (X',Y') such that $\mathcal{L}(X') = \mathcal{L}(X)$, $\mathcal{L}(Y') = \mathcal{L}(Y)$ [158, 183]. If X and Y take values in $\mathbb{Z}_+$, then [172, 74]

$$d_G(X;Y) = \sum_{i\ge1} |\mathbb{P}(X\ge i) - \mathbb{P}(Y\ge i)|.$$

Distance d_G was introduced by Gini [89]; Kantorovich [114] has introduced a class of distances that includes d_G. A generalization of d_G is distance

$$d_t(X;Y) \equiv d_t(\mathcal{L}(X);\mathcal{L}(Y)) = \inf_{X',Y'} \mathbb{E}^{1/t}|X' - Y'|^t \qquad (t>1).$$

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

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

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

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,

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

Let

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

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

$$d_{TV} \le d_{KL}/\sqrt{2}\,. \tag{1.12}$$

Though d^2_{KL} is not a metric, it plays a role in statistics (cf. [103]) and in the theory of large deviations (cf. [144], formula (14.40), and ex. 41 on p. 324).

Given $\varepsilon \ge 0$, the *Dudley* divergence is defined as

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

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

Lévy's metric is defined as

$$d_L(X;Y) = \inf\{\varepsilon>0 : \mathbb{P}(X<x-\varepsilon)-\varepsilon \le \mathbb{P}(Y<x) \le \mathbb{P}(X<x+\varepsilon)+\varepsilon \quad (\forall x\in\mathbb{R})\}.$$

It is weaker than Kolmogorov's distance: $d_L(X;Y) \le d_K(X;Y)$. Convergence in d_L entails weak convergence of distributions.

Certain other distances can be found in [132, 144, 161, 165]. For the relations between metrics see, e.g., [88, 179].

2. Compound Poisson limit theorem

Compound Poisson limit theorem plays important role in the theory of sums of r.v.s. 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 [6, 129, 153, 156] and references therein). Recall that the class of infinitely divisible distributions coincides with the class of weak limits of compound Poisson distributions (Khintchine [117], Theorem 26).

The topic has applications in extreme value theory, insurance, reliability theory, patterns matching, etc. (cf. [10, 13, 17, 127, 144]). 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 [26]. 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 that are non-zero with small probabilities.

In molecular biology long match patterns between DNA sequences may indicate "valuable" fragments. A natural question is if such long patterns appear by chance. Information on the distribution of the number of long match patterns (NLMP) between sequences of independent r.v.s can help answering that question. The distribution of NLMP can often be approximated by a Poisson or a compound Poisson law.

More information concerning applications can be found in [10, 13, 85, 127].

2.1. Basic properties of a compound Poisson distribution

Recall that *compound Poisson* (CP) distribution $\mathbf{\Pi}(\lambda,\zeta) \equiv \mathbf{\Pi}(\lambda,\mathcal{L}(\zeta))$ is the distribution of a random variable

$$\sum_{i=0}^{\pi_\lambda} \zeta_i, \tag{1.1$^+$}$$

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

Typically $\zeta \neq 0 \ w.p.\, 1$. The requirement $\zeta \neq 0 \ w.p.\, 1$ may be omitted. Indeed, denote $p = \mathbb{P}(\zeta \neq 0)$. By Khintchine's formula ([115], ch. 2), any random variable ζ obeys

$$\zeta \stackrel{d}{=} \tau_p \zeta', \tag{2.1}$$

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

$$\mathbf{\Pi}(\lambda, \tau_p \zeta') = \mathbf{\Pi}(\lambda p, \zeta'), \tag{2.2}$$

i.e., (1.1$^+$) can be rewritten as

$$\sum_{i=0}^{\pi_\lambda} \zeta_i \stackrel{d}{=} \sum_{i=0}^{\pi_{\lambda p}} \zeta_i' \tag{1.1*}$$

(cf. (1.8)). Therefore, one usually deals with $\mathbf{\Pi}(t,\zeta)$, where $\mathbb{P}(\zeta \neq 0) = 1$.

Denote $P_X = \mathcal{L}(X)$. Recall that $\mathbf{\Pi}(\lambda, X) = \mathbf{\Pi}(\lambda, P_X)$ can be presented as

$$\exp\left(\lambda(P_X - I)\right) = \mathrm{e}^{-\lambda} \exp\left(\lambda P_X\right) = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \lambda^k P_X^{*k} / k! \tag{2.3}$$

If Y is a compound Poisson r.v. and c is a constant, then cY is obviously a compound Poisson r.v..

A sum of two independent compound Poisson random variables is a compound Poisson random variable.

Note that compound Poisson distribution is infinitely divisible. A random variable X with support on $[0;\infty]$ and $\mathbb{P}(X=0) > 0$ is infinitely divisible if and only if it is compound Poisson [102]. $\mathbf{\Pi}(\lambda,\zeta)$ is not absolutely continuous since there is an atom at zero.

2.2. Compound Poisson limit theorem for independent summands

In this section we present compound Poisson limit theorems for a sum $S_n = X_1 + ... + X_n$ of independent random variables.

Let $\{X_{n,1}, ..., X_{n,n}\}_{n\geq 1}$ be a triangle array of independent random variables. In the sequel we often write $X_1, ..., X_n$ instead of $X_{n,1}, ..., X_{n,n}$.

The topic of compound Poisson approximation to the distribution of a sum S_n of random variables representing "rare" events plays important role in insurance, extreme value theory, reliability theory, mathematical biology, etc. (see, e.g., [85, 147] and references therein); it is an integral part of the topic of infinitely divisible approximation within the framework of Kolmogorov's problem.

For instance, in extreme value theory one is interested in the distribution of the k-th largest sample element $X_{n:k}$.

Given $x \in \mathbb{R}$, let $N_n(x) = \sum_{i=1}^n \mathbb{1}\{X_i > x\}$ denote the number of exceedances of x. Then

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

In particular,

$$\big\{\max_{1\leq i\leq n} X_i \leq x\big\} = \{N_n(x) = 0\},$$

$\{X_{n:2} \leq x\} = \{N_n(x) \leq 1\}$, etc. Thus, results concerning the distribution of sample extremes can be derived from the corresponding results concerning $N_n(x)$.

The distribution of $N_n(x)$ can often be approximated by a compound Poisson law.

Random variables that are zero with large probabilities.

Khintchine ([115], ch. 2.3) was probably the first to prove a compound Poisson limit theorem. Below we present Khintchine's result.

Suppose that r.v.s $X, X_1, ..., X_n$ are independent and identically distributed (i.i.d.) in each row ($\mathcal{L}(X)$ may depend on n). Denote

$$p \equiv p(n) = \mathbb{P}(X \neq 0).$$

Theorem 2.1. [115] *Suppose that there exists $\lambda > 0$ and a random variable X' such that $p \sim \lambda/n$ and*

$$\mathcal{L}(X|X\neq 0) \Rightarrow \mathcal{L}(X') \tag{2.4}$$

as $n \to \infty$. Then

$$\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, X') \qquad (n \to \infty). \tag{2.5}$$

To be precise, Khintchine [115] assumed $\mathcal{L}(X|X\neq 0) = \mathcal{L}(X')$ but the argument holds in the situation presented above.

The proof of (2.5) is based on Khintchine's formula (2.1). As a consequence,

$$S_n \stackrel{d}{=} \tau_1 X_1' + ... + \tau_n X_n', \tag{2.6}$$

where $\tau_1, X_1', ..., \tau_n, X_n'$ are independent r.v.s,

$$\mathcal{L}(X_i') = \mathcal{L}(X|X\neq 0), \;\; \mathcal{L}(\tau_i) = \mathbf{B}(p) \quad (\forall i).$$

Since $\{X_i'\}$ are i.i.d.r.v.s,

$$S_n \stackrel{d}{=} \sum_{i=1}^{\nu_n} X_i', \tag{2.7}$$

where Binomial $\mathbf{B}(n,p)$ r.v.

$$\nu_n = \tau_1 + ... + \tau_n \tag{2.8}$$

is independent of $\{X_1', ..., X_n'\}$.

Estimates of the accuracy of compound Poisson approximation follow from the estimates of the accuracy of pure Poisson approximation thanks to the following observation ([119], formula (30)): since $\{X_i'\}$ are identically distributed r.v.s,

$$d_{TV}\Big(\sum_{i=1}^{\nu_n} X_i'; \sum_{i=1}^{\pi_\lambda} X_i'\Big) \leq d_{TV}(\nu_n; \pi_\lambda). \tag{2.9}$$

A similar result is valid in terms of d_G, cf. [144], Lemma 5.4. A (2.9)-type bound is valid for unit measure approximations, cf. [68], Section 2.3.

Note that weak convergence $\nu_n \Rightarrow \pi_\lambda$ is equivalent to the convergence $d_{TV}(\nu_n; \pi_\lambda) \to 0$ as $n \to \infty$. Thus, Khintchine's compound Poisson limit theorem is a consequence of the Poisson limit theorem: if $p \sim \lambda/n$, where $\lambda > 0$, then weak convergence $\nu_n \Rightarrow \pi_\lambda$ together with (2.4) entails (2.5).

The following proposition states that (2.4) is necessary for (2.5) assuming $p \sim \lambda/n$. Proposition 2.2 is a consequence of Theorem 2.3 below.

Proposition 2.2. *Suppose that* $\exists \lim_{n\to\infty} np := \lambda$. *If there exist a r.v.* S *such that* $S_n \Rightarrow S$ *as* $n \to \infty$, *then there exist a r.v.* X' *such that* $\mathcal{L}(S) = \mathbf{\Pi}(\lambda, X')$. *If* $\lambda > 0$, *then (2.4) holds.*

Example 2.1. Let $\{X, X_1, ..., X_n\}_{n\geq 1}$ be a triangular array of i.i.d.r.v.s,

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

Set $S_n = X_1 + ... + X_n$. Then $X' \stackrel{d}{=} 1+\eta$ and

$$\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, 1+\eta) \qquad (n \to \infty),$$

where η is a Bernoulli $\mathbf{B}(1/2)$ r.v.. □

Non-i.i.d. r.v.s that are zero with large probabilities.

We consider now a special case where $\{X_i\}$ are non-i.d. r.v.s but $\{X_i'\}$ are. Denote

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

According to Khintchine's formula (2.1),

$$X_i \stackrel{d}{=} \tau_i X_i',$$

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

We consider the case where $\{X'_i\}$ are i.i.d.r.v.s. Clearly, (2.7) still holds, where $\nu_n = \tau_1 + ... + \tau_n$ is independent of $\{X'_1, ..., X'_n\}$. If $\nu_n \Rightarrow \mathbf{\Pi}(\lambda)$ as $n \to \infty$, then (cf. (2.9))

$$S_n \Rightarrow \mathbf{\Pi}(\lambda, X') \qquad (n \to \infty).$$

Thus, a compound Poisson limit theorem is a consequence of a Poisson limit theorem.

If $\{X'_i\}$ are non-i.d. r.v.s, then

$$S_n \stackrel{d}{=} \tau_1 X'_1 + ... + \tau_n X'_n$$

can be approximated by a compound Poisson random variable

$$\tilde{S}_n = \tilde{X}_1 + \cdots + \tilde{X}_n \tag{1.9*}$$

where $\{\tilde{X}_i\}$ are independent compound Poisson $\mathbf{\Pi}(p_i, X'_i)$ r.v.s, cf. (3.6), (3.8).

In the assumption that $\{X_{n,1}, ..., X_{n,n}\}_{n\geq 1}$ is a triangular array of independent integer-valued random variables satisfying the infinitesimality assumption (1.5), Grigelionis [100] presents the following sufficient condition for the weak convergence (2.5). Denote $p_i \equiv p_i(n) = \mathbb{P}(X_{n,i} \neq 0)$. If there exists $\lambda > 0$ and a random variable X' such that

$$\lim_{n\to\infty} \sum_{j=1}^{n} p_i = \lambda, \ \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}(X_{n,j} = k) = \lambda \mathbb{P}(X' = k) \quad (\forall k \neq 0),$$

then $\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, X')$ as $n \to \infty$.

Wang [189] presents sufficient conditions for convergence of an unbounded function of a sum S_n of non-negative integer-valued r.v.s to a corresponding function of a compound Poisson random variable, see also Chen & Roos [38].

2.3. Compound Poisson limit theorem for dependent r.v.s

Let $\{X, X_1, ..., X_n\}_{n\geq 1}$ be a triangle array of dependent r.v.s, strictly stationary in each row ($\mathcal{L}(X)$ may depend on n).

Recall the definitions of mixing (weak dependence) coefficients:

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

where the supremum is taken over $m \geq 1$, $A \in \mathcal{F}_{1,m}$, $B \in \mathcal{F}_{m+l+1,n}$ such that $\mathbb{P}(A) > 0$, $\mathcal{F}_{l,m}$ denotes the σ–field generated by $\{X_i\}_{l\leq i\leq m}$.

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

If condition Δ holds, then there exists a sequence $\{r = r_n\}$ of natural numbers such that

$$n \gg r_n \gg l_n\,, \quad n r_n^{-1} \alpha_n^{2/3}(l_n) \to 0 \tag{2.10}$$

as $n \to \infty$ — for instance, one can take $r_n = \big[\max\{\sqrt{nl_n}\,; n\sqrt{\alpha_n(l_n)}\,; 1\}\big]$. We denote by $\mathcal{R}$ the class of all such sequences $\{r_n\}$. In Theorem 2.3 below we assume condition Δ.

In applications $\{X_i\}$ are typically non-negative r.v.s; they usually represent "rare" events. Therefore, in this section we assume that $X_i \geq 0$ $(\forall i)$.

Denote

$$p \equiv p(n) = \mathbb{P}(X \neq 0).$$

A common approach is to assume that there exists the limit

$$\lim_{n\to\infty} \mathbb{P}(S_n = 0) = e^{-\lambda} \qquad (\exists \lambda > 0). \tag{2.11}$$

Condition (2.11) means that r.v.s are "properly" normalised. If $\{X_i\}$ are independent, then (2.11) is equivalent to

$$np \sim \lambda \qquad (n \to \infty). \tag{2.12}$$

The same is true if $\{X_i\}$ are dependent but condition (D') holds (see, e.g., [147]).

Note that (2.11) and Δ yield

$$\lim_{n\to\infty} p(n) = 0$$

(cf. ex. 15 in [144], p. 11).

Weaker than (2.12) is assumption

$$\limsup_{n\to\infty} np < \infty. \tag{2.13}$$

Note that (2.11) does not imply (2.13) — Denzel & O'Brien [75] present an example of an α–mixing sequence such that (2.11) holds though (2.13) does not. On the other hand, (2.13) follows from (2.11) if the sequence $\{X_1, ..., X_n\}$ is φ–mixing (cf. ex. 16 in [144], p. 11).

Theorem 2.3. *Assume conditions Δ, (2.11) and (2.13). If there exists a sequence $\{r = r_n\} \in \mathcal{R}$ such that*

$$\mathcal{L}(S_r | S_r \neq 0) \Rightarrow \mathcal{L}(\zeta) \qquad (n \to \infty), \tag{2.14}$$

then

$$\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, \zeta). \tag{2.15}$$

The limit in (2.15) does not depend on the choice of a sequence $\{r_n\} \in \mathcal{R}$.

If S_n converges weakly to a random variable S, then there exists $\lambda \geq 0$ and a random variable ζ such that $\mathcal{L}(S) = \mathbf{\Pi}(\lambda, \zeta)$, where $\lambda = -\ln \mathbb{P}(S = 0)$. If $\lambda > 0$, then (2.11) holds, and there exist a sequence $\{r_n\} \in \mathcal{R}$ such that (2.14) holds true.

Theorem 2.3 is essentially Theorem 5.1 from [144]. It states that conditions (2.11), (2.14) are necessary and sufficient for the weak convergence of $\mathcal{L}(S_n)$ to a compound Poisson law.

A random variable ζ taking values in $\mathbb{N}$ is called the *limiting cluster size* if (2.14) holds. The notion of the limiting cluster size plays important role in extreme value theory (see, e.g., [144], ch. 5, 6). If $\{X_i\}$ are 0-1 random variables, then the class of limiting cluster size distributions coincides with the family of all integer-valued distributions [34].

Condition (2.14) suggests the following estimator of $\mathcal{L}(\zeta)$:

$$\hat{P}_n(\zeta = \cdot) = \sum_{j=0}^{[n/r]} \mathbb{1}\{S_{r,j} = \cdot\} \Big/ \sum_{j=0}^{[n/r]} \mathbb{1}\{S_{r,j} \neq 0\},$$

where $S_{r,j} = \sum_{m=jr+1}^{(j+1)r\wedge n} X_m$, see Hsing [108]. Results concerning consistent estimation of $\mathcal{L}(\zeta)$ can be found, e.g., in [108, 109, 163].

Example 2.2. Let $\{\xi_i\}$ be a sequence of i.i.d.r.v.s. Suppose that $\{u_n\}$ is a sequence of threshold levels such that $\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\le i\le n} \xi_i \le u_n\Big) = e^{-\lambda}$ $(\exists\lambda > 0)$. Set

$$X_i = \mathbb{1}\{\max\{\xi_i;\xi_{i+1}\} > u_n\}. \tag{2.16}$$

Then (2.14) holds with $\zeta \equiv 2$, $\{r\} \in \mathcal{R}$. Hence $S_n \Rightarrow 2\pi(\lambda)$, i.e., $\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, 2)$. □

3. Accuracy of CP approximation: rare events

Compound Poisson approximation appears naturally in situations where rare events form clusters and the number of clusters is asymptoticaly Poisson (which is typically the case).

This section presents results on the accuracy of compound Poisson to the distribution of a sum of r.v.s representing *rare* events.

3.1. Independent random variables

The task of evaluating the accuracy of compound Poisson approximation to the distribution of a sum of r.v.s that are non-zero with small probabilities (i.e., are rare) have been approached by many authors, cf. [129, 136, 144, 154, 195], etc.

Let $\{X_i\}$ be independent r.v.s that are non-zero with small probabilities, i.e., $\{X_i\}$ represent rare events. Set $S_n = X_1 + ... + X_n$,

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

According to Khintchine's formula (2.1),

$$X_i \stackrel{d}{=} \tau_i X_i', \tag{2.1*}$$

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

Relation (2.1*) can be rewritten as

$$\mathcal{L}(X_i) = (1 - p_i)I + p_i \mathcal{L}(X_i').$$

Hence (2.6) holds:

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

Some authors call $\mathcal{L}(S_n)$ a "compound Binomial distribution," cf. [63].

Concerning the sum $\tilde{S}_n$ of independent accompanying random variables (cf. (1.8), (1.9)), note that

$$\mathcal{L}(\tilde{S}_n) = \exp\left(\sum_{i=1}^{n} p_i (\mathcal{L}(X_i') - I) \right) = \mathbf{\Pi}(\lambda, \mathcal{L}(X_\eta')), \tag{3.1}$$

where $\lambda = p_1 + ... + p_n$, r.v. η is independent of $X_1', ..., X_n'$, $\mathbb{P}(\eta = j) = p_j / \lambda$ $(1 \leq j \leq n)$.

The i.i.d. case.

Consider the situation where $X_i \stackrel{d}{=} X \quad (\forall i)$. Denote

$$\nu_n = \tau_1 + ... + \tau_n .$$

Then

$$S_n \stackrel{d}{=} \sum_{i=1}^{\nu_n} X_i' , \quad \tilde{S}_n \stackrel{d}{=} \sum_{i=1}^{\pi_\lambda} X_i' , \tag{3.2}$$

where Poisson $\mathbf{\Pi}(\lambda)$ r.v. π_λ is independent of $\{X_i'\}$.

Estimates of the accuracy of compound Poisson approximation follow from the estimates of the accuracy of pure Poisson approximation thanks to (2.9):

$$d_{TV}\left(S_n ; \sum_{i=1}^{\pi_\lambda} X_i' \right) \equiv d_{TV}\Big(\sum_{i=1}^{\nu_n} X_i' ; \sum_{i=1}^{\pi_\lambda} X_i' \Big) \leq d_{TV}(\nu_n ; \pi_\lambda). \tag{2.9*}$$

Inequality (2.9*) and Theorem 4.12 in [144] yield

$$d_{TV}(S_n ; \tilde{S}_n) \leq 3\theta / 4\mathrm{e} + 4\delta^* , \tag{3.3}$$

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

According to [144], Lemma 5.4,

$$d_G(S_n ; \tilde{S}_n) \leq d_G(\nu_n ; \pi_\lambda) \mathbb{E}|X'|. \tag{3.4}$$

A combination of (3.4) and formula (4.53) in [144] entails

$$d_G(S_n ; \tilde{S}_n) \leq \Big(\lambda \wedge \tfrac{4}{3} \sqrt{2\lambda / \mathrm{e}} \Big) \theta \mathbb{E}|X'|. \tag{3.5}$$

Other estimates of the accuracy of pure Poisson approximation can be applied too, cf. [147].

The case of i.i.d. $\{X_i'\}$.
We now consider the situation where

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

while $\{\tau_i\}$ are not required to be i.i.d.. In such a situation (3.2) still holds, and estimates of the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ follows from the estimates of the accuracy of pure Poisson approximation thanks to (2.9*), see [154, 136, 24, 147]. In particular, (3.3) and (3.5) hold.

The task of deriving estimates of the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ appears demanding if r.v.s $\{X_i'\}$ are not identically distributed.

The non-i.i.d. case.

If r.v.s $\{X_i'\}$ are independent but not identically distributed, it appears natural to approximate each X_j by an "accompanying" r.v. $\tilde{X}_j$, and the sum $\tau_1 X_1' + ... + \tau_n X_n'$ by a compound Poisson random variable

$$\tilde{S}_n = \tilde{X}_1 + ... + \tilde{X}_n \,,$$

where $\{\tilde{X}_j\}$ are independent compound Poisson $\mathbf{\Pi}(1, X_i) = \mathbf{\Pi}(p_i, X_i')$ random variables, cf. (1.8). Recall that $\tilde{S}_n$ is a compound Poisson random variable.

A simple estimate of the accuracy of compound Poisson approximation to $\mathcal{L}(S_n)$ follows from the property of d_{TV} and a well-known fact that $d_{TV}(\mathbf{B}(p); \mathbf{\Pi}(p)) \leq p^2$:

$$d_{TV}(S_n; \tilde{S}_n) \leq \sum_{i=1}^{n} d_{TV}(X_i; \tilde{X}_i) \leq \sum_{i=1}^{n} d_{TV}(\tau_i; \pi_{p_i}) \leq \sum_{i=1}^{n} p_i^2 \,. \tag{3.6}$$

Similar estimates can be derived in terms of some other distances.

The explicit proof of estimate (3.6) has been given by Le Cam [128], who attributed the idea of the proof to Khintchine [115].

Let $X_1, ..., X_n$ be independent non-negative r.v.s, and let

$$P_{\rm o} = \exp\left(\frac{1}{2}\sum_{i=1}^{n} p_i(1-p_i)(V_i - I)\right).$$

If all $p_i < 1$, then

$$d_K(S_n; \tilde{S}_n) \leq \frac{\pi^2}{8} \sum_{i=1}^{n} \frac{c_i p_i^2}{1-p_i} Q_{P_{\rm o}}(\mathbb{E}X_i') \tag{3.7}$$

[168], where $c_i = 1$ if $V_i = I$, $c_i = 2$ otherwise. A numerical example in [105] shows that inclusion of concentration function in the estimate can significantly improve the estimate.

Zaitsev [195] has derived an estimate of the accuracy of compound Poisson approximation that can be sharper than (3.6) if λ is "large". Denote

$$p_n^* := \max_{i \leq n} p_i \,.$$

Theorem 3.1. [195] *There exists an absolute constant C such that*

$$d_K(S_n;\tilde{S}_n) \le Cp_n^* \,. \tag{3.8}$$

Inequality (3.8) is instrumental in the work on the so-called second Kolmogorov's problem, see [6, 94, 197, 200, 204].

The following estimate is due to Roos [169]. Denote

$$\bar{P}(\cdot) = \frac{1}{\lambda}\sum_{j=1}^{n} p_i \mathbb{P}(X_i' \in \cdot),$$

Assume that $\mathcal{L}(X_i')$ is absolutely continuous with respect to $\bar{P}$ for every i. Let f_i denote the density of $\mathcal{L}(X_i')$ with respect to $\bar{P}$, and set $\rho_i = \int f_i^2 d\bar{P}$. Then

$$d_{TV}(S_n;\tilde{S}_n) \le 8.8 \sum_{i=1}^{n} p_i^2 \min\{1;\rho_i/\lambda\}. \tag{3.9}$$

By Khintchine's formula (see [115] or formula (14.5) in [144]), any distribution can be presented as

$$\mathcal{L}(X) = (1-p)U + pV, \tag{3.10}$$

where $0 \le p \le 1$, U and V are two distributions. One can choose $U = \mathcal{L}(X|a \le X \le b)$, where $[a;b]$ is a finite interval. By shifting U, one can ensure that the mean of the shifted distribution is zero. The derivation of estimates of the accuracy of compound Poisson and infinitely divisible approximations in [119, 129, 157] is based on (3.10).

Let $\{X_i\}$ be independent r.v.s with distributions $\{P_i\}$ obeying

$$P_i = (1-p_i)U_i + p_iV_i \qquad (i = 1,\ldots,n). \tag{3.10*}$$

In other words,

$$X_i \stackrel{d}{=} (1-\tau_i)X_i^o + \tau_i X_i' \,, \tag{3.10*}$$

where $\mathcal{L}(\tau_i) = \mathbf{B}(p_i)$, $\mathcal{L}(X_i^o) = U_i$, $\mathcal{L}(X_i') = V_i$, random variables $\{\tau_i\,, X_i^o\,, X_i'\}$ are independent. Note that $\mathcal{L}(S_n) = \prod_{i=1}^{*n} P_i$. Set

$$G = \prod_{i=1}^{*n}((1-p_i)I + p_iV_i), \quad G^* = \exp\left(\sum_{i=1}^{n} p_i(V_i - I)\right).$$

Then $G = \mathcal{L}(\sum_{i=1}^{n} \tau_i X_i')$, while $G^* = \prod_{i=1}^{*n} \mathbf{\Pi}(p_i, V_i)$ is the accompanying distribution.

Recall that

$$\mathcal{L}(\tilde{S}_n) = \exp\Big(\sum_{i=1}^{n}(P_i - I)\Big)$$

is the distribution of a sum of accompanying $\{X_i\}$ r.v.s; $Q_X(\cdot)$ denotes the concentration function of $\mathcal{L}(X)$.

Let $g(x)$ be a non-negative even function that is positive for $x\neq 0$ and does not decrease for $x\geq 0$. Assume that function $x/g(x)$ is non-decreasing for $x>0$, and suppose that

$$\mathbb{E}X_i^o = 0, \quad \sigma^2 = \sum_{i=1}^n (1-p_i)\mathbb{E}X_i^{o2}, \quad \beta = \sum_{i=1}^n (1-p_i)\mathbb{E}X_i^{o2} g(X_i^o) < \infty.$$

Theorem 3.2. [200] *Assume that* $\sigma>0$. *There exists an absolute constant* C *such that*

$$d_K(S_n;\tilde{S}_n) \leq C\left(p_n^* + \alpha \min\{Q_G(\sigma); Q_{G^*}(\sigma)\}\right), \tag{3.11}$$

where $\alpha = \min\left\{1; \beta/\sigma^2 g(\sigma)\right\}$.

Estimate (3.11) generalizes (3.8).

Integer-valued random variables.

Let $\{X_i\}$ be non-negative integer-valued random variables. Then

$$X = \sum_{j=1}^\infty j\,\mathbb{I}\{X=j\}, \quad S_n = \sum_{j=1}^\infty j \sum_{i=1}^n \mathbb{I}\{X_i=j\}.$$

Denote

$$\lambda_j = \sum_{i=1}^n \mathbb{P}(X_i=j) \quad (j\geq 1), \quad \lambda = \sum_{j\geq 1}\lambda_j = \sum_{i=1}^n \mathbb{P}(X_i \geq 1)\,.$$

Set

$$Z = \sum_{j=1}^\infty j\pi_{\lambda_j}\,.$$

Erhardsson ([81], Example 3.7), has shown that

$$d_{TV}(S_n;Z) \leq \mathcal{M}(\lambda)\sum_{j=1}^n \mathbb{E}^2 X_j \tag{3.12}$$

if $\mathbb{E}X_i^2<\infty$ $(\forall i)$, where

$$\mathcal{M}(\lambda) \leq \min\left\{1; \lambda_1^{-1}\right\} \mathrm{e}^{\lambda}\,. \tag{3.13}$$

If $i\lambda_i \geq (i+1)\lambda_{i+1}$ $(\forall i)$, then [14]

$$\mathcal{M}(\lambda) \leq \min\left\{1; \frac{1}{\lambda_1-2\lambda_2}\left(\frac{1}{4(\lambda_1-2\lambda_2)} + \ln^+(2(\lambda_1-2\lambda_2))\right)\right\}, \tag{3.14}$$

where $\ln^+$ denotes a positive part of the natural logarithm.

Denote $\theta' = \sum_{i\geq 2} i(i-1)\lambda_i/\sum_{i\geq 1} i\lambda_i$. If $\theta' < 1/2$, then [16]

$$\mathcal{M}(\lambda) \leq 1\Big/(1-2\theta')\sum_{i\geq 1} i\lambda_i\,. \tag{3.15}$$

Results similar to (3.6), (3.12) have been presented in [15, 36]. In particular, Boutsikas & Vaggelatou [36] show that

$$d_{TV}(S_n;Z) = O(p+n^2p^4)$$

if $X, X_1, \ldots$ are i.i.d.r.v.s, $\mathbb{E}X^2 < \infty$, $p < \ln 2$ and

$$\sum_{k\in\mathbb{Z}} |\mathbf{\Pi}(np,X')\{k\} - 2\mathbf{\Pi}(np,X')\{k-1\} + \mathbf{\Pi}(np,X')\{k-2\}| = O(1/np),$$

where $p = \mathbb{P}(X \neq 0)$, $\mathcal{L}(X') = \mathcal{L}(X|X\neq 0)$.

Assume now that r.v.s $X_1,\ldots,X_n$ take values in a finite set $\{0,1,\ldots,N\}$ of natural numbers. A lower bound to $d_{TV}(S_n;\tilde{S}_n)$ has been established by Barbour et al. [14]:

$$d_{TV}(S_n;\tilde{S}_n) \geq \frac{1}{32N^2} \min\left(1;(\mathbb{E}S_n)^{-1}\right) \sum_{j=1}^{n} \mathbb{E}^2 X_j \,.$$

Concerning lower bounds to the accuracy of Poisson approximation to the Binomial distribution, see Sason [173] and references therein.
Open problem.
3.1. Evaluate absolute constant C in Zaitsev's inequality (3.8).

3.2. Asymptotic expansions

Construction of asymptotic expansions is based on the following considerations.

Recall (2.1*). Let f_i denote the characteristic functions of X_i'. The characteristic function of S_n can be formally written as

$$\begin{aligned} \mathbb{E}\exp(itS_n) &= \prod_{i=1}^{n}(1+p_i(f_i-1)) = \exp\left(\sum_{i=1}^{n}\ln(1+p_i(f_i-1))\right) \\ &= \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{\infty}(-1)^{j+1}p_i^j(f_i-1)^j/j\right). \end{aligned} \tag{3.16}$$

This leads to the asymptotic expansion

$$\exp\left(\sum_{i=1}^{n} p_i(f_i-1)\right)\left(1-\frac{1}{2}\sum_{i=1}^{n}p_i^2(f_i-1)^2+\cdots\right). \tag{3.17}$$

If we leave more terms in the exponent (3.16), then we arrive at the asymptotic expansion involving a signed compound Poisson measure:

$$\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{(-1)^{j+1}p_i^j(f_i-1)^j}{j}\right)\left(1+\sum_{i=1}^{n}\sum_{j=k+1}^{\infty}\frac{(-1)^{j+1}p_i^j(f_i-1)^j}{j}+\cdots\right). \tag{3.18}$$

This is not the only possible asymptotic expansion. Assume that $0<p_i<1$. Then (3.16) can be written as

$$\exp\left(\sum_{i=1}^{n}(\ln(1-p_i)+\ln(1+p_if_i/(1-p_i)))\right)$$
$$=\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{\infty}(-1)^{j+1}\left(\frac{p_i}{1-p_i}\right)^j(f_i^j-1)/j\right).$$

Leaving a finite number of summands in the exponent, we get yet another possible SCP approximation.

Asymptotic expansions can be traced back to Uspensky [182], see also Herrmann [104]. Herrmann's paper went largely unnoticed; SCP approximations have been rediscovered in 1983 by Kornya [122] and Presman [153].

The following first-order asymptotic expansion in (3.8) is due to Čekanavičius [50]:

$$d_K(\mathcal{L}(S_n);G_1)\le Cp_n^{*\,2}\,, \tag{3.19}$$

where

$$G_1 = \mathcal{L}(\tilde S_n)*\Big(I-\frac{1}{2}\sum_{i=1}^{n}p_i^2(\mathcal{L}(X_i')-I)^{*2}\Big) \tag{3.20}$$
$$= \mathcal{L}(\tilde S_n)-\frac{1}{2}\sum_{i=1}^{n}p_i^2\left(\mathcal{L}(\tilde S_n+X_i'+X_i'')-2\mathcal{L}(\tilde S_n+X_i')+\mathcal{L}(\tilde S_n)\right),$$

$X_i''\stackrel{d}{=}X_i'$, all r.v.s are independent.

An asymptotic expansion in (3.6) has been given by Čekanavičius [54]:

$$d_{TV}(\mathcal{L}(S_n);G_1)\le\frac{8}{3}\Big(\sum_{j=1}^{n}p_j^2\Big)^{3/2}+2\Big(\sum_{j=1}^{n}p_j^2\Big)^2. \tag{3.21}$$

Some other expansions have been presented in [54]. In particular, it was shown that

$$d_{TV}(\mathcal{L}(S_n);G_1')\le 2\Big(\sum_{j=1}^{n}p_j^2\Big)^2, \tag{3.22}$$

where

$$G_1'=\mathcal{L}(\tilde S_n)+\sum_{k=1}^{n}(\mathcal{L}(X_j)-\mathcal{L}(\tilde X_j))*\mathcal{L}(\tilde S_n-\tilde X_j),$$

$\tilde X_j$ denotes an accompanying X_j random variable, see (1.9). The rate of approximation in (3.22) is better than that in (3.21).

Note that the first-order Poisson asymptotic expansion ensures the accuracy of approximation of order $O(p^2)$ in terms of the total variation distance [11, 146] and of order $O(p^2\sqrt{np}\,)$ in terms of the Gini-Kantorovich distance [148].

Next we consider SCP approximations. Given a fixed natural number s, denote

$$\begin{aligned} H_{n,s} &= \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{s} p_i^j(-1)^{j+1}(\mathcal{L}(X_i')-I)^{*j}/j\right), && (3.23)\\ K_{n,s} &= \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{s}(-1)^{j+1}\left(\frac{p_i}{1-p_i}\right)^{j}(\mathcal{L}(X_i')^{*j}-I)/j\right), && (3.24)\\ \tau(i,s) &= (2p_i)^{s+1}/(s+1)(1-2p_i). \end{aligned}$$

Recall that $p_n^* = \max p_i$. If $p_n^* < 1/2$, then (Hipp [106])

$$\begin{aligned} d_{TV}(\mathcal{L}(S_n); H_{n,s}) &\leq \exp\left(\sum_{i=1}^{n}\tau(i,s)\right) - 1, && (3.25)\\ d_{TV}(\mathcal{L}(S_n); K_{n,s}) &\leq \exp\left(\sum_{i=1}^{n}\frac{2}{s+1}\left(\frac{p_i}{1-p_i}\right)^{s+1}\frac{1-p_i}{1-2p_i}\right) - 1. && (3.26) \end{aligned}$$

Estimates (3.25) and (3.26) are of order $O(\sum_{i=1}^{n} p_i^{s+1})$ whenever $\sum_{i=1}^{n} p_i^{s+1} = O(1)$. Therefore, (3.25), (3.26) improve (3.21) even if $s = 2$. For instance, if $p_i = n^{-1/2}$ for all i, then the right-hand side (RHS) of (3.21) is $O(1)$, while the RHS of (3.25), (3.26) are $O(n^{-1/2})$.

Roos [168] has obtained an estimate involving a concentration function.

Let $\tau(i,s)$ be defined as above,

$$\delta = \sum_{i=1}^{n}(\mathrm{e}^{\tau(i,s)-1} - 1),$$

and let P_0, c_i be defined as in (3.7).

Theorem 3.3. [168] If all $\{X_i\}$ are nonnegative, $p_i < 1/2$ and $\delta < 1$, then

$$d_K(\mathcal{L}(S_n); H_{n,s}) \leq \frac{\pi^2}{4(1-\delta)}\sum_{i=1}^{n} c_i(\mathrm{e}^{\tau(i,s)}-1)Q_{P_0}(\mathbb{E}X_i'). \qquad (3.27)$$

Estimate (3.27) has been generalized to the case of distributions that are absolutely continuous with respect to a particular probability measure by Roos [169].

3.3. Dependent random variables

Approximation under mixing conditions.

Let $\{X, X_1, ..., X_n\}$ be a stationary sequence of r.v.s that are non-zero with small probabilities, and let $S_n = X_1 + ... + X_n$.

Given $1 \leq r \leq n$, set $k = [n/r]$, and let $p = \mathbb{P}(X \neq 0)$, $q = \mathbb{P}(S_r \neq 0)$.

Let $\pi_{n,r}, \zeta_1, \zeta_2, \dots$ be independent random variables, $\zeta_0 = 0$, $\mathcal{L}(\pi_{n,r}) = \mathbf{\Pi}(kq)$,

$$\mathcal{L}(\zeta_i) \;=\; \mathcal{L}(S_r | S_r \neq 0) \quad (i \geq 1).$$

Recall that α_n, β_n denote mixing coefficients defined in Section 2.3. Set

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

The distribution of S_n can be approximated by a compound Poisson distribution $\mathcal{L}(Y_n)$.

Theorem 3.4. *If* $n > r > l \geq 0$, *then*

$$d_{TV}(S_n; Y_n) \;\leq\; \kappa_{n,r} rp + (2kl + r')p + nr^{-1}\gamma_n(l), \tag{3.28}$$
$$d_G(S_n; Y_n) \;\leq\; rp \min\Big\{ np \,; \tfrac{4}{3}\sqrt{2np/\mathrm{e}} \Big\} + (2kl + r')p + n\gamma_n(l, \tag{3.29}$$

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

Theorem 3.4 is effectively Theorem 5.2 from [144]. The proof involves Bernstein's blocks method and an application of (2.9).

If $\{X_i\}$ are independent Bernoulli $\mathbf{B}(p)$ r.v.s, then (3.28), (3.29) with $r = 1$, $l = 0$ yield sharp estimates of the accuracy of pure Poisson approximation.

If 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 $\sqrt{mn}$) to get

$$d_{TV}(S_n; Y_n) \leq 4p\lceil \sqrt{mn}\, \rceil. \tag{3.28*}$$

An estimate of the accuracy of Negative Binomial approximation to the distribution of a sum of stationary dependent Bernoullu r.v.s can be found, e.g., in [145].

Locally dependent random variables.

The notion of m-dependent random variables can be generalized to the case of a family $\{X_a\}_{a \in J}$ of r.v.s, where J is an arbitrary index set.

Suppose that for every $a \in J$ there exists "neighborhoods" $\{A_a\}$, $\{B_a\}$ such that $A_a \subset B_a \subset J$, X_a is independent of $\{X_b\}_{b \notin A_a}$, and the family $\{X_b\}_{b \in A_a}$ is independent of $\{X_c\}_{c \notin B_a}$. Then random variables $\{X_a\}_{a \in J}$ are called *locally dependent*.

Let $S = \sum_{a \in J} X_a$, where r.v.s $\{X_a\}$ take values in $\mathbb{Z}_+$. The following estimate of the accuracy of compound Poisson approximation to $\mathcal{L}(S)$ has been given in [14], Theorem 7:

$$d_{TV}(S_n; \nu) \leq 2e^n \sum_{a \in J} \mathbb{P}(X_a \neq 0)\mathbb{P}\Bigg(\sum_{b \in B_a} X_b \neq 0 \Bigg), \tag{3.30}$$

where $\lambda = \sum_{a\in J} \mathbb{E}X_a(Y_a)^{-1}$, $Y_a = \sum_{b\in A_a} X_b$, compound Poisson r.v. ν is defined using measure $\mu(\cdot) = \sum_{a\in J} \mathbb{E}X_a(Y_a)^{-1}\mathbb{1}\{Y_a\in\cdot\}$, $0/0 := 0$.

If r.v.s $\{X_i\}$ are independent and $J=\{1,\ldots,n\}$, then it is natural to choose $A_i = B_i = \{i\}$. In that case $Y_i = X_i$ $(\forall i)$, $\lambda = n$, $\mu(\cdot) = \sum_{i\le n} \mathbb{P}(X_i\in\cdot)$, and (3.30) entails

$$d_{TV}(S_n;\tilde S_n) \le 2e^{\lambda}\sum_{i\le n}\mathbb{P}^2(X_i\ne 0),$$

where $\tilde S_n$ is the sum of accompanying r.v.s.

Let $J = \{1,2,\ldots,n\}$. Suppose that there exist subsets $\mathcal{I}_{1j}$, $\mathcal{I}_{2j}$ of $J\setminus\{j\} = \mathcal{I}_{1j}\cup\mathcal{I}_{2j}$ such that X_j and $\{X_i, i\in\mathcal{I}_{2j}\}$ are independent. According to Barbour et al. [13], Cor. 10.L.1,

$$\begin{aligned} d_{TV}(S_n;\tilde S_n) \;\le\; & \sum_{j=1}^{n}\Bigg(\mathbb{P}^2(X_j\ne 0) \\ & + \sum_{i\in\mathcal{I}_{1j}}\Big(\mathbb{P}(X_i\ne 0\ne X_j) + \mathbb{P}(X_i\ne 0)\mathbb{P}(X_j\ne 0)\Big)\Bigg), \end{aligned} \tag{3.31}$$

where $\tilde S_n$ is the sum of accompanying random variables defined in (1.9).

If r.v.s $X, X_1, ..., X_n$ are identically distributed and $\mathcal{I}_{1j} = \emptyset$, then (3.31) becomes (3.6). Similar estimates have been proved in [35, 36].

Associated random variables.

Let $X_1,\ldots,X_n$ be non-negative integer-valued random variables. Random variables are called *associated* if

$$\mathrm{cov}(f(X_1,\ldots,X_n); g(X_1,\ldots,X_n)) \ge 0$$

for every pair of non-decreasing functions f and g.

Random variables $X_1,\ldots,X_n$ are called *negatively associated* if

$$\mathrm{cov}(f(X_i, i\in A_1); g(X_i, i\in A_2)) \le 0$$

for every pair of disjoint subsets A_1, A_2 of $\{1,2,\ldots,n\}$ and non-decreasing functions f, g.

Let $\mathcal{I}(i)$ be a subset of $\{1,2,\ldots,n\}\setminus\{i\}$. The choice of $\mathcal{I}(i)$ is arbitrary, though $\mathcal{I}(i)$ is supposed to represent the area of "strong dependence" on X_i. Set

$$\hat X_i = \sum_{j\in\mathcal{I}(i)} X_j\,, \quad \lambda_j = \sum_{i=1}^{n}\mathbb{E}X_i\mathbb{1}(X_i+\hat X_i = j)/j \quad (j\ge 1).$$

Denote $\lambda = \sum_{j\ge 1}\lambda_j$,

$$Z = \sum_{j\ge 1} j\pi_{\lambda_j}\,,$$

where $\{\pi_{\lambda_j}\}$ are independent Poisson variables.

Factor $\mathcal{M}(\lambda)$ in Theorem 3.5 obeys (3.13)–(3.15).

Theorem 3.5. [73] If $X_1,\dots,X_n$ are negatively associated r.v.s, then

$$d_{TV}(S_n;Z)\le \mathcal{M}(\lambda)\Bigg(\sum_{i=1}^n\sum_{j\in\mathcal{I}(i)\cup\{i\}}\mathbb{E}X_iX_j-\operatorname{var}S_n\Bigg).$$

If $X_1,X_2,\dots,X_n$ are associated r.v.s, then

$$d_{TV}(S_n;Z)\le \mathcal{M}(\lambda)\Bigg(\operatorname{var}S_n-\sum_{i=1}^n\sum_{j\in\mathcal{I}(i)\cup\{i\}}\mathbb{E}X_iX_j+2\sum_{i=1}^n\sum_{j\in\mathcal{I}(i)}\mathbb{E}X_i\mathbb{E}X_j\Bigg).$$

If $\{X_i\}$ are independent r.v.s, then one can take $\mathcal{I}(i)=\emptyset$ and arrive at (3.12).

Similar results for locally dependent r.v.s. have been proved in [78]. Applications of Theorem 3.5 to the urn model with overflow, extremes and k-runs can be found in [73].

Locally dependent Bernoulli random variables.

Let $\{X_i\}_{i\in\Gamma}$ be locally dependent Bernoulli $\mathbf{B}(p_i)$ random variables, where Γ is a set of indexes. The following result is due to M.Roos [164].

Suppose that for every $i\in\Gamma$ set Γ is split into 4 subsets: $\{i\}$, Γ_i^{vs} ("very strongly" dependent on X_i), Γ_i^{vw} ("very weakly" dependent on $\{X_j\}_{j\in\Gamma_i^{vs}}$), and Γ_i^b (the rest). Denote

$$\begin{aligned}
S &= \sum_{i\in\Gamma}X_i\,,\ S_i=S-X_i\,,\ \hat X_i=\sum_{j\in\Gamma_i^{vs}}X_j\,,\ Z_i=X_i+\hat X_i\,,\\
X_i^b &= \sum_{j\in\Gamma_i^b}X_j\,,\ Y_i=\sum_{j\in\Gamma_i^{vw}}X_j\,,\ S_{i,U}=S_i-\hat X_i\,,\\
D &= \max_{i\in\Gamma}|\Gamma_i^{vs}|,\ \ \varphi=\sum_{i\in\Gamma}(\varphi_{i,1}+\dots+\varphi_{i,l_i}),\ \ l_i=1+|\Gamma_i^{vs}|,
\end{aligned}$$

where $\varphi_{i,j}=\mathbb{E}|\mathbb{E}X_i\mathbb{1}\{Z_i=j\}-\mathbb{E}\{X_i\mathbb{1}\{Z_i=j\}|\sigma(X_l: l\in\Gamma_i^{vw})\}|$.

Let compound Poisson r.v. Z be defined by (1.2), where $\lambda=\lambda_1+\dots+\lambda_{D+1}$,

$$\lambda_j=\sum_{i\in\Gamma}\mathbb{E}X_i\mathbb{1}\{Z_i=j\}/j\quad(1\le j\le D+1),\quad \lambda_j=0\quad(j>D+1).$$

Theorem 3.6. [164] *There holds*

$$d_{TV}(S;Z)\le c_\lambda\varphi+C_\lambda\sum_{i\in\Gamma}\Big(\mathbb{E}^2X_i+\mathbb{E}X_i\mathbb{E}(\hat X_i+X_i^b)+\mathbb{E}X_iX_i^b\Big)\,,\qquad(3.32)$$

where $\max\{c_\lambda;C_\lambda\}\le e^\lambda$.

m-dependent random variables.

Let $\{X,X_1,\dots,X_n\}$ be a stationary sequence of 1-dependent non-negative integer-valued bounded random variables; $\mathcal{L}(X)$ may depend on n.

Theorem 3.5 and (3.31) can be applied to sums of m-dependent r.v.s. Note that by grouping consequent random variables the sum of m-dependent r.v.s can be presented as a sum of 1-dependent r.v.s.

Assume that

$$\mathbb{E}X = o(1),\ \mathbb{E}X(X-1) = o(\mathbb{E}X),\ \mathbb{E}X_1X_2 = o(\mathbb{E}X),\ n\mathbb{E}X \to \infty \tag{3.33}$$

as $n \to \infty$. Set

$$\begin{aligned} G &= \exp\left(n\mathbb{E}X(I_1 - I) + c_n(I_1 - I)^{*2}\right), \\ \tilde{R} &= \mathbb{E}X(X-1)(X-2) + \mathbb{E}X\mathbb{E}X(X-1) + \mathbb{E}^3X \\ &+ \mathbb{E}X_1(X_1-1)X_2 + \mathbb{E}X_1X_2(X_2-1) + \mathbb{E}X\mathbb{E}X_1X_2 + \mathbb{E}X_1X_2X_3\,, \end{aligned}$$

where $c_n = \frac{n}{2}(\mathbb{E}X(X-1) - \mathbb{E}^2X) + (n-1)(\mathbb{E}X_1X_2 - \mathbb{E}^2X)$. Then [150]

$$d_{TV}(\mathcal{L}(S_n); G) = O\Big(\tilde{R}\Big/\mathbb{E}X\sqrt{n\mathbb{E}X}\Big). \tag{3.34}$$

A generalization of (3.34) to the case of non-identically distributed 1-dependent random variables has been given by Čekanavičius & Vellaisamy [67, 71].

Markov Binomial distribution.

Let $\xi_0, \xi_1, \ldots, \xi_n$ be a Markov chain with the initial distribution

$$\mathbb{P}(\xi_0 = 1) = p_0, \quad \mathbb{P}(\xi_0 = 0) = 1 - p_0, \quad (p_0 \in [0,1])$$

and transition probabilities

$$\begin{aligned} \mathbb{P}(\xi_i = 1 \,|\, \xi_{i-1} = 1) &= \beta, & \mathbb{P}(\xi_i = 0 \,|\, \xi_{i-1} = 1) &= 1 - \beta, \\ \mathbb{P}(\xi_i = 1 \,|\, \xi_{i-1} = 0) &= \alpha, & \mathbb{P}(\xi_i = 0 \,|\, \xi_{i-1} = 0) &= 1 - \alpha, \end{aligned}$$

where $\alpha, \beta \in (0,1)$ $(i \in \mathbb{N})$. If $p_0 = \alpha/(1-\beta+\alpha)$, then the chain is stationary.

The distribution of

$$S_n = \xi_1 + \cdots + \xi_n$$

is sometimes called the Markov Binomial (MB) distribution.

MB is a generalization of the Binomial distribution to the case of dependent 0-1 r.v.s. Indeed, if $p_0 = 0$ and $\alpha = \beta = p$, then $\mathcal{L}(S_n)$ is the Binomial $\mathbf{B}(n,p)$ distribution.

One can check that $\mathbb{E}S_n = n\alpha/(1-\beta+\alpha)$,

$$\operatorname{var} S_n = \frac{n\alpha(1-\beta)}{(1-\beta+\alpha)^2} + \frac{2n\alpha(1-\beta)(\beta-\alpha)}{(1-\beta+\alpha)^3} + \frac{2\alpha(1-\beta)(\beta-\alpha)((\beta-\alpha)^n - 1)}{(1-\beta+\alpha)^4}.$$

It is known that a centered normalised Binomial distribution can have either normal, Poisson or degenerate weak limit [123], while in the case of Markov Binomial distribution the class of limit laws for a centered normalised sum has seven different elements (see Dobrushin [76]).

A compound Poisson limit theorem for $\mathcal{L}(S_n)$ can be found in [121, 188, 83]. Hsiau [110] has extended the compound Poisson limit theorem to the case of a stationary Markov chains with more than two states.

Assume that $\operatorname{var} S_n > \mathbb{E}S_n$. Let Y be a Negative Binomial r.v. defined by (1.4), where

$$r = \frac{(\mathbb{E}S_n)^2}{\operatorname{var} S_n - \mathbb{E}S_n}, \quad p = 1 - \frac{\mathbb{E}S_n}{\operatorname{var} S_n}.$$

Below p_0, α, β may depend on n.

Theorem 3.7. [192] If $\operatorname{var} S_n > \mathbb{E}S_n$, then

$$\begin{aligned} d_{TV}(S_n;Y) \quad \le \quad & \frac{|\beta-\alpha|(5+43\max(\alpha,\beta))}{(1-\max(\beta,\alpha))^2}\left(\frac{2\sqrt{5}}{\sqrt{n}}\cdot\frac{\sqrt{1-\beta+\alpha}}{\sqrt{\alpha(1-\beta)\min(1-\alpha,\beta,1/2)}}\right. \\ & \left. + \frac{360}{n}\cdot\frac{(1-\alpha)(1-\beta)^2+\alpha^2\beta}{\alpha(1-\beta)(1-\beta+\alpha)} + \beta^{\lfloor n/4 \rfloor}\right). \end{aligned} \tag{3.35}$$

If $\alpha \equiv \alpha(n) = O(1), \beta \equiv \beta(n) = O(1)$, then the RHS of (3.35) is $O(n^{-1/2})$. The same rate of shifted Poisson approximation without assumption $\operatorname{var} S_n > \mathbb{E}S_n$ has been achieved by Barbour & Lindvall [21].

The case of a non-stationary Markov chain has been investigated in [64, 65]. Denote by G geometric $\Gamma(\beta)$ distribution (i.e., $G = (1-\beta)\sum_{j=0}^{\infty}\beta^j I_{j+1}$). Let

$$\begin{aligned} \gamma_1 &= \frac{(1-\beta)\alpha}{1-\beta+\alpha}, \qquad \gamma_2 = -\frac{(1-\beta)\alpha^2}{(1-\beta+\alpha)^2}\left(\beta + \frac{1-\beta}{1-\beta+\alpha}\right) - \frac{\gamma_1^2}{2}, \\ \gamma_3 &= \frac{\gamma_1^3}{3} + \frac{\gamma_1^2}{(1-\beta)(1-\beta+\alpha)}\left\{\beta^2\alpha + \frac{\beta(1-\beta)(2\alpha-1+\beta)}{1-\beta+\alpha} + \frac{2\alpha(1-\beta)^2}{(1-\beta+\alpha)^2}\right\} \\ & \quad + \frac{\gamma_1^2\alpha}{1-\beta+\alpha}\left(\beta + \frac{1-\beta}{1-\beta+\alpha}\right), \quad H = I + \varkappa_2(G-I), \\ \varkappa_1 &= \gamma_1\left(\frac{\alpha-\beta}{1-\beta+\alpha} - p_0\right), \quad \varkappa_2 = p_0\frac{\beta(1-\beta)}{1-\beta+\alpha}, \\ D_0 &= \exp\left((n-p_0)\gamma_1(G-I)\right), \quad D_{jn} = \exp\left(n\sum_{i=1}^{j}\gamma_i(G-I)^{*i}\right) \quad (1 \le j \le 3). \end{aligned}$$

Set $b = n(\gamma_1 + 4\gamma_2 + 3\gamma_3)$, $\gamma = [b]$, $\tilde{\omega} = \{b\}$,

$$\begin{aligned} \lambda_1 &= n(\gamma_1+4\gamma_2+3\gamma_3) - \tilde{\omega}, \quad \lambda_2 = \tilde{\omega}/6, \quad \lambda_{-1} = -n(2\gamma_2+3\gamma_3) + \tilde{\omega}/3, \\ \tilde{D} &= G^{*\gamma} * \exp\left(\lambda_1(G-I) + \lambda_2(G^{*2}-I) + \lambda_{-1}(I_{-1}-I)/(1-\beta)\right), \\ A_0 &= H * D_0, \quad \tilde{A}_0 = H * D_0 * (I + n\gamma_2(G-I)^{*2}), \\ A_1 &= H * \exp\left(\varkappa_1(G-I)\right) * \tilde{D}, \quad A_2 = H * \exp\left(\varkappa_1(G-I)\right) * D_{2n}, \\ A_3 &= H * \exp\left(\varkappa_1(G-I)\right) * D_{3n}. \end{aligned}$$

Assume that $0 \le \tilde{\omega} < 1$, $\beta \le 1/2$. According to Čekanavičius & Vellaisamy [65],

$$d_{TV}(\mathcal{L}(S_n); A_0) \le C\left(\alpha(\alpha+\beta)(1\wedge 1/\sqrt{n\alpha}) + \min\{\alpha; n\alpha^2\} + \gamma_n\right), \tag{3.36}$$

$$d_{TV}(\mathcal{L}(S_n);\tilde{A}_0) \le C\big(\alpha^2 + \alpha\beta(1\wedge 1/\sqrt{n\alpha}) + \gamma_n\big)\,, \tag{3.37}$$

where c, C are absolute constants, $\gamma_n = (\alpha+\beta)e^{-cn}$. If $\alpha \ge 1/n$, then the RHS of (3.36) and (3.37) are respectively $O(\alpha)$ and $O(\alpha^2)$. If $p_0 = 0$, then A_0 is an analogue of the accompanying distribution.

Theorem 3.8. [176] *If $\beta \le 1/4$, $\alpha \le 1/30$ and $n\alpha \ge 3$, then*

$$d_{TV}(\mathcal{L}(S_n);A_1) \le C \max(n^{-1};(n\alpha)^{-2}).$$

The accuracy of approximation in (3.36) can be improved if one uses a SCP approximation.

Theorem 3.9. [65] *If $\beta \le 1/2$, $\alpha \le 1/30$, $n\alpha \ge 1$, then there exist absolute constants c, C such that*

$$\begin{aligned} d_{TV}(\mathcal{L}(S_n);A_2) &\le C(\beta+\alpha)\left(\min\{\alpha;n^{-1}\} + e^{-cn}\right), \\ d_{TV}(\mathcal{L}(S_n);A_3) &\le C(\beta+\alpha)\left(\min\{\alpha;n^{-1}\} + e^{-cn}\right), \\ d_G(\mathcal{L}(S_n);A_3) &\le C(\beta+\alpha)\left(\min\{\alpha;\sqrt{\alpha/n}\,\} + e^{-cn}\right). \end{aligned}$$

For example, if $\beta \le 1/2$, $\alpha \equiv \alpha(n) \to 0$, $n\alpha \to \infty$, then for all large enough n

$$C_5\alpha \le d_{TV}(\mathcal{L}(S_n);H*D_0) \le C_6\alpha, \quad C_5\alpha\sqrt{n\alpha} \le d_G(\mathcal{L}(S_n);H*D_0) \le C_6\alpha\sqrt{n\alpha}\,.$$

If, in addition, $\beta \equiv \beta(n) = o(1)$, then

$$d_{TV}(\mathcal{L}(S_n);H*D_0) \sim 6\alpha/\sqrt{2\pi e}\,.$$

If β is "small", then the compound Poisson approximation can be simplified. Let

$$w = \frac{\alpha}{1-\beta+\alpha}\,, \quad u = \frac{\alpha(1-\beta)(\beta-\alpha)}{(1-\beta+\alpha)^3} - \frac{w^2}{2}\,.$$

We define SCP G_6 as $\mathcal{L}(\pi_{w-2u} + 2\pi_u)$. If $\beta \le 1/20$ and $\alpha \le 1/30$, then [56]

$$\begin{aligned} d_{TV}(\mathcal{L}(S_n);G_6) \;\le\; & C(\alpha+\beta)^2 \min\{n\alpha;(n\alpha)^{-1/2}\} \\ +\; & C|\alpha-\beta| \min\{1,(n\alpha)^{-1/2}\}. \end{aligned} \tag{3.38}$$

If $\beta = o(\alpha)$, then the RHS of (3.38) is $O(\alpha n^{-1/2})$. Further results can be found in [59, 64].

The case of a symmetric three-state Markov chain has been investigated by Šliogerė & Čekanavičius [177]. Further extensions of the Markov Binomial model were considered in [140, 190, 206]. Large deviations for Markov binomial distribution have been studied by Jensen [112].

Open problems.

3.2. Can the assumptions on α, β in Theorem 3.9 be weakened?

3.3. Generalize Theorem 3.7 and Theorem 3.9 to the case of a Markov chain with more than 3 states.

3.4. Applications

2-run statistic.

Let $\xi_1,\ldots,\xi_n$ be independent and identically distributed Bernoulli $\mathbf{B}(p)$ random variables, where $0<p<1$. Denote

$$X_i = \min(\xi_i;\xi_{i+1}) = \xi_i\xi_{i+1} \qquad (1\le i\le n),$$

where we assume that $\xi_{n+1}=\xi_1$. Then $S_{n,2} := X_1+...+X_n$ is the number of head runs of length 2, i.e., the 2-run statistic.

Barbour & Xia [16] have suggested a two-parameter compound Poisson approximation to the distribution of $S_{n,2}$ with the accuracy $O(p\,n^{-1/2})$.

Let Y be a Negative Binomial $\mathbf{NB}(a/b,b)$ r.v., where

$$a = (1-b)np^2\,, \quad b = \frac{2p-3p^2}{1+2p-3p^2}\,.$$

Negative Binomial approximation to $\mathcal{L}(S_n)$ has been suggested by Gan & Xia [82]:

$$d_{TV}(\mathcal{L}(S_n|S_n\ge 1);\mathcal{L}(Y|Y\ge 1)) \le \frac{32.2p}{\sqrt{(n-1)(1-p)^3}}\,\frac{a\mathbb{P}(Y>1|Y\ge 1)}{(a+b)\mathbb{P}(S_n\ge 1)}\,. \qquad (3.39)$$

If p is "small", then the RHS of (3.39) is asymptotically $16.1p/\sqrt{n(1-p)^3}$.

Sharp estimates of the accuracy of compound Poisson approximation to $\mathcal{L}(S_{n,k})$ has been established by Petrauskienė & Čekanavičius [149], see (4.56), and by Vellaisamy & Čekanavičius [187], see (4.57).

***k*-run statistic.**

Let $\xi_1,\xi_2,\ldots$ be independent Bernoulli $\mathbf{B}(p_i)$ random variables, where $0<p_i<1$. Denote

$$X_i = \xi_i\xi_{i+1}\cdots\xi_{i+k-1}\,, \quad S_{n,k} = X_1+\cdots+X_{n-k+1}\,.$$

where $k\in\mathbb{N}$. Then $S_{n,k}$ is the number of head runs of length k among $X_1,...,X_n$, i.e., $S_{n,k}$ is the k-run statistic. For instance, if $k=1$, then $X_i=\xi_i$ $(\forall i)$, and $S_{n,1}=X_1+\cdots+X_n$. If $k=2$, then $S_{n,2}=\sum_{i=1}^{n-1}\xi_i\xi_{i+1}$ is a 2-run statistic.

For the sake of simplicity we will assume that $\xi_{i+n}=\xi_i$ $(1\le i\le n)$.

The accuracy of Negative Binomial approximation to $S_{n,m}$ has been evaluated by Wang & Xia [191] in the assumption that $\sigma^2>\lambda$, where

$$\lambda = \mathbb{E}S_{n,k}\,, \quad \sigma^2 = \mathrm{var}\,S_{n,k}\,.$$

Let Y_n be a Negative Binomial $\mathbf{NB}(r,p)$ r.v. defined by (1.4), where

$$r = \lambda/(\sigma^2-\lambda), \quad p = 1-\lambda/\sigma^2\,.$$

Let ϑ_l be the lth largest number among $(1-p_{i-1})^2(1-p_i)p_ip_{i+1}\cdots p_{i+k-1}$ $(1\le i\le n)$. Set

$$\phi = \min\left\{2;\ 4.6\left(\sum_{i=4k-1}^{n}\vartheta_i\right)^{-1/2}\right\},\ q_i = \max\{p_j : |j-i| \le 2k-2\}.$$

Theorem 3.10. [191] If $\sigma^2>\lambda$ and $n>4k$, then

$$d_{TV}(S_{n,k};Y_n) \le \frac{4.5(4k-3)(2k-1)\phi}{\lambda}\sum_{i=1}^{n} q_i\mathbb{E}X_i\,. \tag{3.40}$$

If $k=1$, then $S_{n,1} = \xi_1 + ... + \xi_n$, $\lambda = \sum_{i=1}^n p_i$, $\sigma^2 = \sum_{i=1}^n p_i(1-p_i) < \lambda$ meaning Theorem 3.10 is not applicable.

Let $p_i\equiv p\le 1/5$, $k\ge 2$, $D = \exp\left(\sum_{j=1}^{2k-1}\lambda_j(I_j-I)\right)$, where

$$\lambda_j = \begin{cases} np^{k+j-1}(1-p)^2, & j=1,\dots,k-1,\\ np^{k+j-1}j^{-1}(1-p)[2+(2k-j-2)(1-p)], & j=k,\dots,2k-2,\\ np^{3k-2}(2k-1)^{-1} & j=2k-1.\end{cases}$$

Then (Daly [73])

$$d_{TV}(\mathcal{L}(S_{n,k});D) \le \mathcal{M}(\lambda)(2k-1)np^{2k}\,,$$

where $\mathcal{M}(\lambda)$ is defined by (3.13) – (3.15).

Match patterns of length k.

Let $X, X_1, \dots,$ and $Y, Y_1, \dots,$ be two independent sequences of independent random variables taking values in $\mathbb{N}$, $X_i \stackrel{d}{=} X$ $(\forall i)$, $Y_j \stackrel{d}{=} Y$ $(\forall j)$. Then

$$S_{m,n,k} = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}\{(X_i,\dots,X_{i+k-1}) = (Y_j,\dots,Y_{j+k-1})\}$$

is the number of match patterns (NMP) of length k. In particular, if $X\equiv 1$, then

$$S_{1,n,k} = \sum_{j=1}^{n}\mathbb{1}\{Y_j = \cdots = Y_{j+k-1} = 1\}$$

is the number of head runs of length k among $Y_1,\dots,Y_n$.

Set $R=\mathbb{P}(X=Y)$, and suppose that $R<1/2$. Let

$$p_\ell = \mathbb{P}(X=\ell),\ q_\ell = \mathbb{P}(Y=\ell)\ (\forall \ell\in\mathbb{N}),$$

$$r = \max_\ell p_\ell q_\ell, \quad S(R) = (1-R)(1-2R),$$

$$\tilde p = \max_\ell p_\ell\mathbb{1}\{q_\ell>0\}, \quad \tilde q = \max_\ell q_\ell\mathbb{1}\{p_\ell>0\}, \quad \lambda = mn(1-R)R^k\,.$$

Clearly, $\mathbb{E}S_{m,n,k} = mnR^k$.

Information on the distribution of NMP can help recognising "valuable" fragments of DNA sequences (see [138, 147, 174]) and references therein).

$S_{m,n,k}$ can be approximate by a compound Poisson random variable

$$\tilde{Y} = \sum_{i=1}^{\infty} i\pi_{\theta_i},$$

where $\theta_i = \lambda(1-R)R^{i-1}$, $\{\pi_{\theta_i}\}_{i\geq 1}$ are independent Poisson $\mathbf{\Pi}(\theta_i)$ r.v.s. Note that

$$\mathbb{E}\tilde{Y} = mnR^k.$$

A compound Poisson limit theorem has been given by Mikhailov [138]. The following estimate of the accuracy of compound Poisson approximation to $\mathcal{L}(S_{m,n,k})$ is due to Mikhailov [137].

Theorem 3.11. [137] If $2\leq k<\min(n,m)$ and $0<R<1$, then

$$\begin{aligned} d_{TV}(S_{m,n,k};\tilde{Y}) \;\; &\leq \;\; \mathcal{M}(\lambda)\left(2k\lambda(1-R)^{-1}\left(2k(r/R)^k + m\tilde{p}^k + n\tilde{q}^k\right)\right. \\ &+ \;\; \left.\frac{2(4k-3)\lambda^2}{(1-R)^2}\Big(\frac{1}{n}+\frac{1}{m}\Big)\right) + \frac{4\lambda^2}{nmR(1-R)}\,, \end{aligned} \tag{3.41}$$

where $\mathcal{M}(\lambda) \leq (1\wedge 1/\lambda(1-R))\mathrm{e}^{\lambda}$. If $0<R<1/2$, then

$$\mathcal{M}(\lambda) \leq 1 \wedge \left(1/4\lambda S(R) + \ln_+(2\lambda S(R))\right)/\lambda S(R).$$

In the "central zone" where $k = k(m,n)$ obeys $mnR^k \asymp const$ as $m \to \infty$, $n\to\infty$, $m\asymp n$, the RHS of (3.41) is $O(\ln(mn)(1/m+1/n))$.

Mikhailov [137] gives also an estimate of the accuracy of Poisson approximation in the case $k=1$. Note that the distribution of the number of match patterns of length $\geq k$ can be well approximated by the Poisson law (see [144, 147]).

Non-decreasing runs of fixed length.

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with uniform distribution

$$\mathbb{P}(X=k) = 1/N \qquad (k = 0, 1, \ldots, N-1),$$

where $N\geq 3$ is a fixed natural number.

Given $s\in\mathbb{N}$, denote $\eta_i(s) = \mathbb{1}\{X_i \leq X_{i+1} \leq \cdots \leq X_{i+s-1}\}$. Let

$$S_n(s) = \sum_{i=1}^{n} \eta_i(s)\,,$$

and let Z be defined by (1.2):

$$Z = \sum_{j=1}^{\infty} j\pi_{\lambda_j}\,,$$

where $\{\pi_{\lambda_j}\}$ are independent Poisson $\mathbf{\Pi}(\lambda_j)$ variables,

$$\lambda_j = \begin{cases} N^{-s-j-1}n\varkappa_{1,s+j-1}, & j=1,\dots,s-1, \\ j^{-1}N^{-s-j-1}n(2N\varkappa_{2,s+j-1}+(2s-2-j)\varkappa_{1,s+j-1}) & j=s,\dots,2s-2, \\ (2s-1)^{-1}N^{-3s+2}\varkappa_{3,2s-1} & j=2s-1, \\ 0 & j=2s,2s+1,\dots \end{cases}$$

Here

$$\varkappa_{1,k} = \binom{k+N-1}{k+2}\frac{N(k^2+k-1)-k^2-k}{N-2}\,, \quad \varkappa_{2,k} = \binom{k+N}{k+1}\frac{k(N-1)}{N+k}\,,$$

$$\varkappa_{3,k} = \binom{k+N-1}{k}.$$

Then Z is a compound Poisson $\mathbf{\Pi}(\lambda,\zeta)$ r.v. with $\lambda=\sum_{i=1}^{2s-1}\lambda_i$ and multiplicity distribution $\mathcal{L}(\zeta)$ such that the ch.f. $\varphi_\zeta(t)=\sum_{j=1}^{\infty}\lambda_j e^{itj}/\lambda$.

Minakov [139], using Theorem 3.6, has shown that

$$d_{TV}(S_n(s);Z) \le e^{\lambda}\frac{n(6s-5)}{(sN^{-1}+1)^2N^{2s}}\binom{s+N}{s}^2. \tag{3.42}$$

If $n\to\infty$, $s\to\infty$, $s/n\to 0$ and

$$n(s+N)^{N-1}N^{-s-1}/(N-2)! \sim \lambda,$$

then $\mathcal{L}(S_n(s))$ converges to the compound Poisson distribution $\exp(\lambda(N-1)\sum_{j=1}^{\infty}N^{-j}(I_j-I))$.

4. Accuracy of CP approximation: general case

In this section the random variables are no longer assumed to take value 0 with "large" probability. We present estimates of the accuracy of compound Poisson approximation.

4.1. Independent Bernoulli random variables

Though Poisson distribution is a natural proxy to the Binomial one, there exist a compound Poisson approximation that is more accurate.

Let $X, X_1, ..., X_n$ be independent Bernoulli $\mathbf{B}(p)$ r.v.s. Presman [153] has shown that

$$\sup_p d_{TV}(\mathbf{B}(n,p);P_{n,p}) = O(n^{-2/3}), \tag{4.1}$$

where $P_{n,p}$ is a shifted compound Poisson distribution (a similar result in terms of d_K is due to Meshalkin [135]).

The Meshalkin–Presman result is related to the first uniform Kolmogorov's problem, see (7.3). Unlike (7.3), the approximating distribution in (4.1) is given explicitly.

We present Presman's result in Theorem 4.1 below.

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

$$\gamma = \left\lceil 3np^2 - 2np^3 \right\rceil, \ \ \beta = \gamma - 3np^2 + 2np^3, \ \ q = 1-p.$$

Note that $\beta \in [0;1)$.

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 compound Poisson random variable. One can check that

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

matching the first three moments of $X - p$.

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

Theorem 4.1. [153] *There exists an absolute constant C such that*

$$d_{TV}(\mathbf{B}(n,p); \mathbf{\Pi}(np)) \wedge d_{TV}(\mathbf{B}(n,p); P_{n,p}) \le C\varepsilon_{n,p} \qquad (0 \le p \le 1/2), \tag{4.2}$$

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

Bound (4.1) follows after noticing that

$$\sup_{0 \le p \le 1/2} \varepsilon_{n,p} = O(n^{-2/3}),$$

cf. [153] or [6], ch.8. Clearly, it suffices considering only $p \in [0;1/2]$: if $\mathcal{L}(S_n) = \mathbf{B}(n,p)$, then $\mathcal{L}(n - S_n) = \mathbf{B}(n, 1-p)$.

Theorem 4.1 has been derive by the method of characteristic functions.

Presman [153] has shown also that it is impossible to construct a compound Poisson or infinitely divisible distribution approximating $\mathbf{B}(n,p)$ with the accuracy better than $\varepsilon_{n,p}$. Namely, there exists an absolute constant $c > 0$ such that for an arbitrary infinitely divisible distribution P

$$\begin{array}{lcll} d_{TV}(\mathbf{B}(n,p); P) & \ge & c\varepsilon_{n,p} & (0 \le p \le 1/n); \\ d_{TV}(\mathbf{B}(n,p); P) & \ge & c \min\left\{p; (np)^{-2}\right\} & (1/n \le p \le 1/\sqrt{n}\,). \end{array} \tag{4.3}$$

Notice that $\varepsilon_{n,p} = O(n^{-1})$ if $p \ge 1/\sqrt{n}$.

Čekanavičius [55] has extended Presman's result to the case of non-identically distributed 0-1 r.v.s.

Symmetrised Bernoulli random variables.

Consider independent Bernoulli $\mathbf{B}(p_i)$ random variables $\{X_j\}$, where $p_j \in [0,1]$, $q_j = 1 - p_j$ $(j = 1, \dots, n)$.

Let X_j' denote an independent copy of X_j, and set

$$S_n^o = (X_1 - X_1') + \cdots + (X_n - X_n').$$

The characteristic function of S_n^o is equal to

$$\mathbb{E}e^{itS_n^o} = \prod_{j=1}^n |q_j + p_j e^{it}|^2 = \prod_{j=1}^n (q_j^2 + 2p_j q_j \cos t + p_j^2).$$

It is natural to approximate $\mathcal{L}(S_n^o)$ by a symmetrised Poisson distribution.

Let π_{σ^2} and π'_{σ^2} be two independent Poisson random variables with parameter $\sigma^2 = \sum_{j=1}^n p_j q_j$. Set

$$Y = \pi_{\sigma^2} - \pi'_{\sigma^2}.$$

Clearly, Y is a compound Poisson r.v. with the characteristic function

$$\mathbb{E}e^{itY} = \exp\left(2\sigma^2(\cos t - 1)\right).$$

Presman [154] has proved that

$$d_{TV}(S_n^o; Y) \le \min\left(0.7225 \sum_{j=1}^n p_j^2 q_j^2/(\sigma^2 - p_j q_j)^2; 4\sum_{j=1}^n (p_j q_j)^2\right). \tag{4.4}$$

If all $\{p_j\}$ are uniformly bounded away from 0 and 1, then the RHS of (4.4) is $O(1/n)$.

Assume now that $p_i \equiv p \le 1/2$. Presman [153] has proved that

$$C_1 \varepsilon^*_{n,p} \le d_{TV}(S_n^o; Y) \le C_2 \varepsilon^*_{n,p}\,, \tag{4.5}$$

where $\varepsilon^*_{n,p} = \min\{np^2; n^{-1}\}$, cf. (4.2). Thus, in the case of symmetrised r.v.s one can expect the correct rate of the accuracy of compound Poisson approximation be $O(n^{-1})$.

An extension of (4.4) to the case of discrete distributions with non-negative characteristic functions has been given by Čekanavičius [43]. A multivariate version of (4.4) is given in [126].

SCP approximations.

Let $\mathcal{L}(X_i) = \mathbf{B}(p_i)$ $(i = 1, \ldots, n)$. Set

$$\sigma^2 = \operatorname{var} S_n = \sum_{j=1}^n p_j(1 - p_j)\,, \ \lambda_k = \sum_{j=1}^n p_j^k \quad (k \ge 1), \ \lambda = \lambda_1\,, \ \theta = \lambda_2/\lambda.$$

Denote (with some abuse of notation)

$$G_2 = \mathcal{L}(\pi_{\lambda+\lambda_2} + 2\pi_{-\lambda_2/2}), \tag{4.6}$$

where $\pi_{\lambda+\lambda_2}$ and $\pi_{-\lambda_2/2}$ are independent "r.v.s", i.e., G_2 is a convolution of $\mathbf{\Pi}(\lambda+\lambda_2)$ and a compound Poisson unit measure $\mathbf{\Pi}(-\lambda_2/2, 2)$.

Observe that $G_2 = H_{n,2}$ from (3.23), where $\mathcal{L}(X'_i) = I_1$.

Presman [153] approximated the Binomial distribution by G_2 . Kruopis [124] has extended Presman's result to the case of non-identically distributed Bernoulli r.v.s:

$$d_{TV}(\mathbf{B}(n,p); G_2) \le 5\sqrt{e}\,\lambda_3 \min\{1.2\sigma^{-3} + 4.2\lambda_2\sigma^{-6}; 2 + \sigma^2 + 3.4\lambda_2\}. \tag{4.7}$$

Constants in (4.7) have been improved by Barbour & Xia [16] under the additional assumption that $\theta < 1/2$, see also Zacharovas & Hwang [193].

Theorem 4.2. [193] If $\theta<1$, then for any $m\in\mathbb{N}$

$$d_{TV}(\mathcal{L}(S_n);G_2) \le \frac{\lambda_3}{2\lambda^{3/2}}\left(\frac{\sqrt{6}C_2}{(1-\theta)^2}+\frac{\sqrt{3\theta}}{2\sqrt{2}(1-\theta)^{5/2}}\right), \tag{4.8}$$

$$d_G(\mathcal{L}(S_n);G_2) \le \frac{\lambda_3}{\lambda}\left(\frac{\sqrt{2}C_2}{(1-\theta)^{3/2}}+\frac{\sqrt{3\theta}}{4\sqrt{2}(1-\theta)^2}\right), \tag{4.9}$$

$$d_K(\mathcal{L}(S_n);G_2) \le \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{\sqrt{6}C_2}{(1-\theta)^2}+\frac{\sqrt{3\theta}}{2\sqrt{2}(1-\theta)^{5/2}}\right)g_\lambda(m), \tag{4.10}$$

$$|\mathbb{P}(S_n=m)-G_2(\{m\})| \le \frac{\lambda_3}{\lambda^2}\left(\frac{2\sqrt{6}C_2}{(1-\theta)^{5/2}}+\frac{\sqrt{15\theta}}{2\sqrt{2}(1-\theta)^6}\right)g_\lambda(m), \tag{4.11}$$

where $C_2=0.3706$, $g_\lambda(m)=\mathrm{e}^{-(m-\lambda)^2/4(m+\lambda)}$.

In particular,

$$\begin{aligned} &C_3\min\left(p\sqrt{p}/\sqrt{n};np^3\right)\le d_K(\mathbf{B}(n,p);G_2)\\ \le\ & d_{TV}(\mathbf{B}(n,p);G_2)\le C_4\min\left(p\sqrt{p}/\sqrt{n};np^3\right). \end{aligned} \tag{4.12}$$

The upper bound follows from (4.7), the lower bound follows from Theorem 7 in [124]. In some cases (4.12) is sharper than (4.2) – consider, for example, $p=n^{-2/3}$. On the other hand, in some cases (e.g., if $p=1/2$), Presman's bound (4.2) is more accurate.

Does there exist a SCP approximation, which is always better than the best compound Poisson approximation? The answer is affirmative, at least if $\{p_i\}$ are "small". Denote

$$H^*_{n,s}=\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{s}(-1)^{j+1}p_i^j(I_1-I)^{*j}/j\right),$$

cf. (3.23). Assume that $p_i\le 1/4$ ($\forall i$), and let $s\ge 2$ be a fixed integer. According to Theorem 1 in [166],

$$d_{TV}(\mathcal{L}(S_n);H^*_{n,s})\le C(s)\lambda_{s+1}\min(1;\lambda^{-(s+1)/2}). \tag{4.13}$$

In particular,

$$d_{TV}(\mathbf{B}(n,p);H^*_{n,s})\le C(s)\min\left(np^{s+1};p^{(s+1)/2}n^{-(s-1)/2}\right). \tag{4.14}$$

Thus, if $p\le 1/4$, then there exists a SCP measure, which approximates the Binomial distribution with the accuracy $O(n^{-(s-1)/2})$. For example, $H^*_{n,3}$ guarantees the accuracy of approximation as good as in (4.2) if $p=O(1)$: for any $p\le 1/4$ the accuracy of approximation by $H^*_{n,3}$ is $O(n^{-1})$. Moreover, $H^*_{n,3}$ is structurally comparable to Presman's approximation, since both involve three Poisson random variables.

It is easy to check that $H^*_{n,2}$ can be expressed through Hermite polynomials $\{H_m\}$:

$$H^*_{n,2}\{m\} = \mathrm{e}^{-\lambda-\lambda_2/2}\,\frac{\lambda_2^{m/2}}{m!}\,H_m\left(\frac{\lambda+\lambda_2}{\sqrt{\lambda_2}}\right), \tag{4.15}$$

where $H_m(x) = (-1)^m \mathrm{e}^{x^2/2}\frac{\mathrm{d}^m}{\mathrm{d}x^m}\mathrm{e}^{-x^2/2}$. Besides,

$$mH^*_{n,2}\{m\} = (\lambda_1+\lambda_2)H^*_{n,2}\{m-1\} - \lambda_2 H^*_{n,2}\{m-2\} \qquad (m\geq 2) \tag{4.16}$$

[52]. No such simple formulas exist for $P_{n,p}$ and $H^*_{n,s}$ if $s\geq 3$. The approach proposed by Kruopis [125] is to construct asymptotic expansion to $H^*_{n,2}$ rather than to apply a SCP with a longer expansion in the exponent.

Let

$$G^*_3 = H^*_{n,2} * (I+\lambda_3(I_1-I)^{*3}/3).$$

Observe that G^*_3 can be expressed through the first, second or third backward differences of $H^*_{n,2}$. According to Kruopis [124]

$$d_K(\mathcal{L}(S_n);G^*_3) \leq \min(2.3\sqrt{\mathrm{e}}\lambda_4\sigma^{-4} + 7.1\lambda_3^2\sigma^{-6}; 3\sqrt{\mathrm{e}}\lambda_4 + 3.3\mathrm{e}\lambda_3^2). \tag{4.17}$$

If $p_i \leq C < 1$ $(\forall i)$, then

$$d_K(\mathcal{L}(S_n);G^*_3) \leq C\lambda_4 \min(1;\lambda^{-2}), \tag{4.18}$$

that is, the accuracy is the same as in (4.13).

A similar to (4.18) bound in terms of the total variation distance follows from [67], Theorem 3.2, if all $p_i \leq 1/100$.

SCP measure $H^*_{n,s}$ is not the only possible SCP approximation. For instance, SCP

$$\mathcal{L}(\,\pi_{\lambda-\lambda_2/2} - \pi_{-\lambda_2/2})$$

was used in [58], though the rate of the accuracy of approximation was worse than that provided by $H^*_{n,2}$.

A *large deviations* result concerning SCP approximation to the Binomial distribution $\mathbf{B}(n,p)$ has been suggested in [52].

Assume that $n^{-1}\leq p\leq 1/3$, $n\geq 4$, $x\in[np; n(1+p)^2/5]$. Then

$$\frac{\mathrm{I\!P}(S_n=x)}{H^*_{n,2}(\{x\})} = \mathrm{e}^{\Lambda(x)}\left\{1+\frac{\theta_x A(x)}{1-\theta_x A(x)}\right\}, \tag{4.19}$$

where $|\theta_x|\leq 1$,

$$\begin{aligned} A(x) &= 14.4\mathrm{e}^2 y(y-p+\sqrt{y/n}\,), \quad y = x/n, \\ \Lambda(x) &= -n(1-p)\sum_{k=3}^{\infty}\left(\frac{y-p}{1-p}\right)^k \frac{1}{k(k-1)}\left\{1-\sum_{j=0}^{k-2}\binom{j+k-2}{j}(1-p)^{-j}\right\}. \end{aligned}$$

As shown in [52], the "equivalence zone" for $H_{n,2}$ is larger than that for Poisson approximation if $np^2\to\infty$, $p\to 0$.

Open problems.

4.1. Evaluate constant C in Presman's inequality (4.2).

4.2. Independent discrete random variables

A lattice random variable is a linear transform of an integer-valued random variable. In this section we deal with integer-valued r.v.s.

Integer-valued random variables with finite 3rd moments.

Let X be an integer-valued r.v. with a finite 3rd moment, and let $X_i \stackrel{d}{=} X$ $(\forall i)$. Set

$$\mu = \mathbb{E}X, \quad u = \operatorname{var} X / \mathbb{E}X^2 .$$

Denote by $Y_{n,u}$ a compound Poisson r.v. with the distribution $\mathcal{L}(Y_{n,u}) = \mathbf{\Pi}(nu, X)$. Čekanavičius [57] has proved that there exists constant C_X (that depends on $\mathcal{L}(X)$) such that

$$d_{TV}(S_n; [n\mu^3/\mathbb{E}X^2] + Y_{n,u}) \le C_X n^{-1/2} . \tag{4.20}$$

Let $\{X_i\}$ be r.v.s taking values in the set $\mathbb{Z}$ of integer numbers. Denote

$$S_{n,i} = S_n - X_i , \quad e_n = \mathbb{P}\{S_n < 0\}, \quad d_+^{(i)} = d_{TV}(S_{n,i}; S_{n,i}+1),$$

$$a = 2\mathbb{E}S_n - \operatorname{var} S_n, \ b = (\operatorname{var} S_n - \mathbb{E}S_n)/2, \ \tilde{\theta} = |\mathbb{E}S_n - \operatorname{var} S_n| / \mathbb{E}S_n ,$$

$$\hat{\psi}_i = \mathbb{E}|X_i(X_i-1)(X_i-2)| + |\mathbb{E}X_i|\mathbb{E}|X_i(X_i-1)| + 2\mathbb{E}|X_i||\operatorname{var} X_i - \mathbb{E}X_i|.$$

Theorem 4.3. [16] *If* $\frac{2}{3}\operatorname{var} S_n < \mathbb{E}S_n < 2\operatorname{var} S_n$ *, then*

$$d_{TV}(S_n; \pi_a + 2\pi_b) \le \frac{1}{(1-2\tilde{\theta})\mathbb{E}S_n} \left\{ \sum_{i=1}^n \hat{\psi}_i d_+^{(i)} + e_n \right\}. \tag{4.21}$$

Quantity $d_{TV}(S_{n,i}; S_{n,i}+1)$ appears becaused of the method.

According to Barbour & Xia [16], Proposition 4.6,

$$d_{TV}(S_n; S_n+1) \le 2V^{-1/2} , \tag{4.22}$$

where

$$V = \sum_{i=1}^n \min\{1/2; v_i\}, \quad v_i = 1 - d_{TV}(X_i; X_i+1).$$

Estimate (4.22) has been improved by Mattner & Roos [134]:

$$d_{TV}(S_n; S_n+1) \le \sqrt{\frac{2}{\pi}} \Big(\frac{1}{4} + \sum_{j=1}^n v_i \Big)^{-1/2} . \tag{4.23}$$

Set $\varepsilon_{i,n} = \sqrt{2/\pi} \Big/ \Big(1/4 + V - v_i\Big)^{1/2}$. Then

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

According to Lemma 5 in [146], if one approximates S_n by an integer-valued r.v. Y, then

$$d_{TV}(S_n; S_n+1) \le d_{TV}(Y; Y+1) + 2d_{TV}(S_n; Y). \tag{4.24}$$

Bound (4.24) allows for replacing $\varepsilon_{i,n}$ with $\varepsilon_{i,n} \wedge (\varepsilon^* + 2\varepsilon^+)$, where ε^* is an estimate of $d_{TV}(Y; Y+1)$ and ε^+ is an estimate of $d_{TV}(S_n; Y)$, cf. [146, 148].

If $b \geq 0$, then $\mathcal{L}(\pi_a + 2\pi_b)$ is a compound Poisson distribution, otherwise we have a SCP approximation.

If $\{X_i\}$ are Bernoulli $\mathbf{B}(p)$ r.v.s and $2/n < p < 1/3$, then the assumptions of the theorem are satisfied, $a = np + np^2$, $b = -np^2/2$, and SCP measure $\mathcal{L}(\pi_a + 2\pi_b)$ coincides with G_2 from (4.6), $V = np$, $\tilde{\theta} = p$, $\hat{\psi}_i = 2p^3$, and the RHS of (4.21) is bounded by $12\sqrt{2}p\sqrt{p}/\sqrt{n}$, cf. (4.12).

On the other hand, if $X_i \stackrel{d}{=} \pi_1$, then $a = n$, $b = \tilde{\theta} = 0$, $\hat{\psi}_i = 2$, and $d_+^{(i)} \leq 1/\sqrt{2\mathrm{e}(n-1)}$, cf. (2.10) in Čekanavičius [68]. The left-hand side of (4.21) equals zero, while the RHS is not.

Similar approximations involving $\gamma + \pi_a + 2\pi_b$ or $\gamma + \pi_a - \pi_b$ (with possibly negative b) have been suggested in [19].

Integer-valued random variables with finite 4$^{\text{th}}$ moments.

Barbour & Čekanavičius [19] have generalized Presman's estimate (4.2) to the case of non-identically distributed integer-valued r.v.s $\{X_i\}$ with finite 4th moments.

Denote $(i \geq 1)$

$$\mu_i = \mathbb{E}X_i,\ \mu = \mathbb{E}S_n, \sigma_i^2 = \operatorname{var} X_i,\ \sigma^2 = \operatorname{var} S_n, \beta_{3i} = \mathbb{E}(X_i - \mu_i)^3,$$
$$\beta_3 = \mathbb{E}(S_n - \mu)^3.$$

Let

$$\begin{aligned}\lambda_1 &= \sigma^2 - (\beta_3 - \sigma^2 + 2m\delta)/(m-1), \quad \delta = \gamma - \mu + \sigma^2 - (\beta_3 - \sigma^2)/m \\ \lambda_2 &= m\delta/2(m-2), \quad \lambda_m = (\beta_3 - \sigma^2 - 2\delta m/(m-2))/m^2(m-1),\end{aligned}$$

where $m \in \mathbb{Z}\setminus\{0,1,2\}$ and γ are chosen according to the following rules:

a) If $\beta_3 < \sigma^2$, then $\gamma = \lceil \mu - \sigma^2 + m^{-1}(\beta_3 - \sigma^2) \rceil$, $\quad m = -\max\{1; \lceil 8(1 - \beta_3/\sigma^2) \rceil\}$;
b) If $\sigma^2 \leq \beta_3 < \sigma^2 + 3$, then $\gamma = \lceil \mu - \sigma^2 + m^{-1}(\beta_3 - \sigma^2) \rceil + 3$, $\quad m = -2$;
c) If $\beta_3 \geq \sigma^2 + 3$, then $\gamma = \lceil \mu - \sigma^2 + m^{-1}(\beta_3 - \sigma^2) \rceil$, $m = \max\{6; \lceil 8(\beta_3/\sigma^2 - 1) \rceil\}$.

Set

$$Y_m = \gamma + \pi_{\lambda_1} + 2\pi_{\lambda_2} + m\pi_{\lambda_m}\,.$$

The characteristic function of Y_m is

$$\mathbb{E}\exp\left(\mathrm{i}tY_m\right) = \exp\left(\mathrm{i}t\gamma + \lambda_1(\mathrm{e}^{\mathrm{i}t} - 1) + \lambda_2(\mathrm{e}^{2\mathrm{i}t} - 1) + \lambda_m(\mathrm{e}^{\mathrm{i}tm} - 1)\right).$$

The choice of parameters $\lambda_1, \lambda_2, \lambda_m$ ensures matching the first three moments of $\mathcal{L}(S_n)$:

$$\mathbb{E}Y_m = \mu,\ \operatorname{var} Y_m = \sigma^2,\ \mathbb{E}(Y_m - \mathbb{E}Y_m)^3 = \beta_3.$$

If $\mathcal{L}(S_n)$ is Binomial $\mathbf{B}(n,p)$, where $p \leq 1/16$, then

$$\sigma^2 = npq, \quad \beta_3 = npq(1-2p), \quad m = -1, \quad \lambda_1 = npq^2 - \delta, \quad \lambda_2 = \delta/6, \quad \lambda_{-1} = np^2q + \delta/3,$$

and $\mathcal{L}(Y_m)$ coincides with Presman's $P_{n,p}$ from Theorem 4.1.

Denote $S_{n,i} = S_n - X_i$,

$$
\begin{aligned}
d' &= \max_{1\le i\le n} \|\mathcal{L}(S_{n,i}) * (I_1 - I)^{*2}\|, \quad \kappa = \max\{8; \lceil 8|1-\beta_3/\sigma^2|\rceil\}, \\
\psi_i &= |\mu_i - \sigma_i^2 + (\beta_{3i} - \sigma_i^2)/m|\mathbb{E}|(X_i-1)(X_i-2)(X_i-3)| \\
&\quad + |\sigma_i^2 - (\beta_{3i} - \sigma_i^2)/(m-1)|\mathbb{E}|X_i(X_i-1)(X_i-2)| \\
&\quad + |(\beta_{3i} - \sigma_i^2)/(m(m-1))|\mathbb{E}|(X_i+m-1)(X_i+m-2)(X_i+m-3)| \\
&\quad + \mathbb{E}|X_i(X_i-1)(X_i-2)(X_i-3)|.
\end{aligned}
$$

The following result is Theorem 4.3 from [19].

Theorem 4.4. *If* $\sigma^2 \geqslant 24$, *then*

$$
\begin{aligned}
d_{TV}(S_n; Y_m) \quad \le \quad & \frac{d'}{3\sigma^2}\Big\{\sum_{i=1}^n \psi_i + 2\kappa\Big\} + \frac{10}{3}\exp\left(-\frac{5\sigma^2}{48\kappa}\right) \\
& + 42\sigma^{-4} + 14\sigma^{-8}\sum_{i=1}^n \mathbb{E}(X_i - \mu_i)^4. \qquad (4.25)
\end{aligned}
$$

It has been stated in [19] that $d' \le 16/V$ (this can be improved using (4.23)).

If $X_i \stackrel{d}{=} X$ $(\forall i)$ and $\mathcal{L}(X)$ does not depend on n, then the RHS of (4.25) is $O(n^{-1})$.

For the Binomial $\mathbf{B}(n,p)$ distribution we have $\sigma^2 = npq$, $\beta_3 = npq(q-p)$. If $npq \ge 24$ and $p \le 11/16$, then $\psi_i = 30p^2q^2$, $v = np$; $d' \le 16/np$; $\kappa = 8$, and the RHS in Theorem 4.4 is bounded by $C\min(n^{-1}; (np)^{-2})$, cf. (4.2).

Čekanavičius [49, 53] has suggested a SCP measure D_k that matches first k moments of $\mathcal{L}(S_n)$, where $k > 2$. In the case of i.i.d. lattice r.v.s that approximation ensures $d_{TV}(\mathcal{L}(S_n); D_k) = O(n^{-(k-2)/2})$.

Integer-valued r.v.s with a non-negative characteristic function.

Let $X, X', X'', X_1, \ldots, X_n$ be independent and identically distributed symmetric integer-valued r.v.s with a non-negative characteristic function $\widehat{F}(t) = \mathbb{E}e^{itX}$. Assume that $\mathbb{E}|X|^3 < \infty$, and denote

$$\beta_1 = \mathbb{E}|X|, \quad \sigma^2 = \mathbb{E}X^2, \quad \beta_3 = \mathbb{E}|X|^3, \quad q_0 = P(X \neq 0).$$

The distribution of S_n can be approximated by the accompanying compound Poisson law $\mathbf{\Pi}(n, X)$ or by a SCP

$$G_n = \mathbf{\Pi}(2n, X) * \mathbf{\Pi}(-n/2, X' - X'').$$

The Fourier transform of G_n is

$$\exp\left(2n(\widehat{F}(t) - 1)\right) * \exp\left(-n(\widehat{F}(t) - 1)^2/2\right).$$

More generally, let

$$\hat{H}_{n,k} = \exp\left(n\sum_{j=1}^k (-1)^{j+1}(F - I)^{*j}/j\right),$$

where $k \in \mathbb{N}$. Then

$$\hat{H}_{n,1} = \mathcal{L}(\tilde{S}_n) = \mathbf{\Pi}(n, X), \quad \hat{H}_{n,2} = G_n .$$

The following theorem from [43] shows that Hipp's result (3.25) can be improved if $\sigma^2 < \infty$.

Theorem 4.5. [43] *There exist absolute constants C, C_1, C_2 such that*

$$d_{TV}(\mathcal{L}(S_n); \hat{H}_{n,k}) \leq C \min\left(q_0^{-1/4}(n^{-1/2}+\sigma)^{1/2} n^{-k}; n\sigma^{2(k+1)}(1+n\sigma^2) \right). \quad (4.26)$$

In particular,

$$d_{TV}(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) \leq C_1 q_0^{-1/4} n^{-1}(\sigma+n^{-1/2})^{1/2} , \quad (4.27)$$

$$d_{TV}(\mathcal{L}(S_n); G_n) \leq C_2 q_0^{-1/4} n^{-2}(\sigma+n^{-1/2})^{1/2} . \quad (4.28)$$

If $\beta_3/(\sigma^3\sqrt{n}\,) \leq C$ for all n, where C is an absolute constant, then there exist absolute constants C_3, C_4 such that

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) \geq C_3 n^{-1} , \quad d_K(\mathcal{L}(S_n); G_n) \geq C_4 n^{-2} .$$

Non-uniform estimates have been established in [46].

Bound (4.27) can be compared with (4.35).

It was proved in [43] that for any fixed $m \in \mathbb{N}$

$$d_K(\mathcal{L}(S_n); \hat{H}_{n,m}) \leq C_m n^{-2}(1 + \beta_1 n^{-m}) \quad (4.29)$$

if $s > 1$ and $\beta_1 < \infty$.

Negative Binomial approximation.

Recall that Negative Binomial distribution is a particular compound Poisson distribution. Negative Binomial distribution is a natural choice if $\mathbb{E}S_n < \mathrm{var}\, S_n$.

Let $S_n = X_1 + ... + X_n$, where $\{X_i\}$ are independent non-negative integer-valued random variables with finite third moments. Assume that $\mathbb{E}S_n < \mathrm{var}\, S_n$.

Let Y be a Negative Binomial $\mathbf{NB}(r, q)$ r.v. with parameters r, q , where

$$r = \frac{(\mathbb{E}S_n)^2}{\mathrm{var}\, S_n - \mathbb{E}S_n} , \quad q = \frac{\mathbb{E}S_n}{\mathrm{var}\, S_n} . \quad (4.30)$$

Then $\mathbb{E}Y = \mathbb{E}S_n = rq/p$, $\mathrm{var}\, Y = \mathrm{var}\, S_n = rq/p^2$, where $p = 1-q$.

The following estimate has been obtained by Vellaisamy et al. [186]:

$$\begin{aligned} d_{TV}(S_n; Y) \ \leq \ & \frac{2\tau}{rp} \sum_{i=1}^{n} \Bigg(\Big(\frac{3+p}{2} \mathbb{E}X_i + p \Big) \mathbb{E}X_i(X_i-1) \\ & + \frac{p+1}{2} \mathbb{E}X_i(X_i-1)(X_i-2) + q(\mathbb{E}X_i)^3 + p(\mathbb{E}X_i)^2 \Bigg), \quad (4.31) \end{aligned}$$

where $\tau = \max_{1 \leq i \leq n} d_{TV}(S_{n,i}; S_{n,i}+1)$, $S_{n,i} = S_n - X_i$.

Note that (4.31) is not applicable if $\{X_i\}$ are Poisson r.v.s, nor if $\{X_i\}$ are Bernoulli r.v.s.

Example 4.1. Let r.v.s $\{X_i\}$ have geometric distributions

$$\mathbb{P}(X_i = k) = q_i p_i^k \quad (k \geq 0),$$

where $q_i = 1 - p_i$. Assume that $q_i > 1/2$ $(\forall i \geq 1)$. Then

$$\mathbb{E}S_n = \sum_{k=1}^{n} p_k/q_k\,, \quad \operatorname{var} S_n = \sum_{k=1}^{n} p_k/q_k^2\,,$$

and (4.23), (4.31) yield an explicit estimate of $d_{TV}(S_n; Y)$.

In particular, if $p_i = p_o/3$ when i is odd, $p_i = 2p_o/3$ when i is even, where $p_o \in (0;1)$, then $d_{TV}(S_n;Y) = O\left(\sqrt{p_o/n}\right)$. □

Open problems.
4.2. Can an analogue of (4.26) hold under weaker moment assumption?
4.3. Can moment conditions in (4.27) be weakened?

4.3. Discrete non-lattice distributions

Any discrete distribution can include zero in its support after a proper shift. W.l.o.g., we will assume that in this section.

The following result is due to Čekanavičius & Wang [62].

Assume that $X_1, \ldots, X_n$ are independent r.v.s taking on values $x_0 = 0, x_1, x_2, \ldots, x_N$, where $\{x_j\}$ and N are fixed numbers.

Denote $p_{ki} = \mathbb{P}(X_k = x_i)$ $(k = 1, \ldots, n,\ i = 0, 1, \ldots N)$. Set

$$G_4 = \exp\left(\sum_{k=1}^{n} (\mathcal{L}(X_k) - I) - \frac{1}{2} \sum_{k=1}^{n} (\mathcal{L}(X_k) - I)^{*2}\right).$$

Theorem 4.6. [62] *Suppose that there exists an absolute constant $\tilde{C} \in (0;1)$ such that $p_{k0} \geq \tilde{C} > 0$ $(k = 1, \ldots, n)$. Then there exists constant C_N depending only on N such that*

$$\begin{aligned} d_K(\mathcal{L}(S_n); G_4) \quad \leq \quad & C_N \frac{\max_{1 \leq j \leq n} |x_j|}{\min_{1 \leq j \leq n} |x_j|} \left(\Big(\sum_{k=1}^{n} (1 - p_{k0}) \Big)^{-1/2} \ln n + 1 \right) \\ & \times \sum_{j=1}^{N} \left(\sum_{k=1}^{n} p_{kj}^3 \right) \left(\sum_{m=1}^{n} p_{mj} \right)^{-3/2} + \mathrm{e}^{-n}\,. \end{aligned} \tag{4.32}$$

If N is fixed, $\max_j |x_j| / \min_j |x_j| < C$, $0 < C_9 < p_{kj} < C_{10} < 1$, then the RHS of (4.32) is $O(n^{-1/2})$. Similar results have been given in [60, 66].

4.4. Special classes of distributions

In this section we present estimates of the accuracy of compound Poisson approximation to the distribution of the sum S_n of i.i.d.r.v.s when $\mathcal{L}(X)$ belongs to a particular class of distributions. Denote

$\mathcal{F}$	the set of all distributions
$\mathcal{F}_s$	the set of *symmetric* distributions
$\mathcal{F}_+ \subset \mathcal{F}_s$	the set of distributions with *non-negative* characteristic functions
$\mathcal{F}_\alpha \subset \mathcal{F}_s$	the set of distributions such that the ch.f. φ obeys $\varphi(t) \geq -1+\alpha$ $(\forall t)$
$\mathcal{F}_{0,\beta}$	the set of zero-mean distributions such that $\mathbb{E}\lvert X\rvert^\beta < \infty$

Symmetric random variables.

If $X, X_1, \ldots, X_n$ are i.i.d. *symmetric* random variables, then estimate (4.48) can be improved. Zaitsev [194] (upper bound) and Studnev [178] (lower bound) have shown that there exist absolute constants $0<c<C<\infty$ such that

$$cn^{-1/2} \leq \sup_{\mathcal{L}(X)\in\mathcal{F}_s} d_K(S_n; \tilde{S}_n) \leq Cn^{-1/2}, \tag{4.33}$$

where $\tilde{S}_n = \tilde{X}_1 + ... + \tilde{X}_n$ is the sum of accompanying independent random variables.

Denote $P = \mathcal{L}(X)$. In terms of convolutions of distributions (4.33) states that

$$cn^{-1/2} \leq \sup_{P\in\mathcal{F}_s} d_K\Big(P^{*n}; \exp(n(P-I))\Big) \leq Cn^{-1/2}.$$

Moreover [194],

$$\sup_{P\in\mathcal{F}_s} d_K\Big(P^{*n}; \exp(\tfrac{n}{2}(P^{*2}-I))\Big) \leq Cn^{-1/2}.$$

The derivation of (4.33) relies on a result of Arak [3] for r.v.s with a non-negative ch.f.. Note that no moment assumption is needed.

According to Prokhorov [159],

$$cn^{-1} \leq d_K(\mathbf{B}(n, 1/2); \mathcal{D}) \leq Cn^{-1}, \tag{4.34}$$

where $\mathcal{D}$ denotes the set of infinitely divisible distributions, c, C are absolute constants.

A similar result holds for a sum S_n^* of i.i.d. symmetrised Bernoulli random variables $X_i^* = X_i - \hat{X}_i$, where $\mathcal{L}(X_i) = \mathcal{L}(\hat{X}_i) = \mathbf{B}(p)$ $(\forall i)$, $\hat{X}_i$ is an independent copy of X_i:

$$C_1 \min\{np^2; n^{-1}\} \leq d_{TV}(S_n^*; Y) \leq C_2 \min\{np^2; n^{-1}\}, \tag{4.5*}$$

(Presman [153]), where C_1, C_2 are absolute constants, $Y = \pi_{npq} - \hat{\pi}_{npq}$, $\hat{\pi}_{npq}$ is and independent copy of a Poisson $\mathbf{\Pi}(npq)$ r.v. π_{npq} (see also Theorem 4.2.1 in [6]).

Zaitsev [204] has conjectured that for any distribution $\mathcal{L}(X)$ there exist a constant C_X such that

$$d_K(\mathcal{L}(S_n); \mathcal{D}) \leq C_X n^{-1}.$$

Random variables with a non-negative characteristic function

Let $X, X_1, \ldots, X_n$ be i.i.d.r.v.s. Denote $P = \mathcal{L}(X)$.

Arak [2, 3] has obtained a sharper estimate in the case of r.v.s with a non-negative characteristic function: if $\mathcal{L}(X) \in \mathcal{F}_+$, then there exist absolute constants $0 < c < C < \infty$ such that

$$cn^{-1} \le \sup_{\mathcal{L}(X)\in\mathcal{F}_+} d_K(S_n; \tilde{S}_n) \le Cn^{-1}. \tag{4.35}$$

Čekanavičius [42] has suggested the following asymptotic expansion in (4.35):

$$\sup_{P\in\mathcal{F}_+} d_K(P^{*n}; \exp(n(P-I)) * (I - n(P-I)^{*2}/2)) \le Cn^{-2}. \tag{4.36}$$

Zaitsev [199] (see also [6], Theorem 5.1) has shown that the upper bound in (4.35) holds for a more general class of distributions: if $\mathcal{L}(X) \in \mathcal{F}_\alpha$ $(\exists \alpha \in (0;2))$, then there exist an absolute constant $C < \infty$ such that

$$d_K(S_n; \tilde{S}_n) \le C\big(n^{-1} + \exp\big(-n\alpha + C\ln^3 n\big)\big) = O(n^{-1}). \tag{4.37}$$

A non-uniform analogue of (4.37) has been proved by Zaitsev [201].

Similar estimates hold for distributions with a symmetric component. Recall that any distribution $P := \mathcal{L}(X)$ admits representation

$$X \stackrel{d}{=} (1-\tau)\xi + \tau\eta, \tag{3.10*}$$

where $\mathcal{L}(\tau) = \mathbf{B}(p)$, $p \in [0;1]$, random variables τ, ξ, η are independent. Equivalently,

$$\mathcal{L}(X) = (1-p)U + pV, \tag{3.10*}$$

where $U = \mathcal{L}(\xi)$, $V = \mathcal{L}(\eta)$.

Let

$$G_5 = \mathcal{L}(\tilde{S}_n) - n\mathcal{L}(\tilde{S}_{n-1}) * (\mathcal{L}(X) - I)^{*2}/2.$$

Čekanavičius [42] presents asymptotic expansions for $\mathcal{L}(S_n)$ in the assumption that $U \in \mathcal{F}_+$, $V \in \mathcal{F}$:

$$\sup_{U\in\mathcal{F}_+}\sup_{V\in\mathcal{F}} d_K(S_n; \tilde{S}_n) \le C(n^{-1}+p), \ \sup_{U\in\mathcal{F}_+}\sup_{V\in\mathcal{F}} d_K(\mathcal{L}(S_n); G_5) \le C(n^{-1}+p)^2. \tag{4.38}$$

The RHSs of estimates (4.38) are small if p is small.

If $p = 0$, then the upper bound in (4.35) follows from the first estimate in (4.38); the second estimate in (4.38) becomes $O(n^{-2})$.

Let $p = 0$, so that $\mathcal{L}(X) = U$, $\mathcal{L}(S_n) = U^{*n}$. Set

$$G_6 = \exp(n(U-I)) * (I - n(U-I)^{*2}/2 + n(U-I)^{*3}/3 + n^2(U-I)^{*4}/8).$$

Then [42]

$$\sup_{U\in\mathcal{F}_+} d_K(\mathcal{L}(S_n); G_6) \le Cn^{-3}. \tag{4.39}$$

Let $X, X_1, ..., X_n$ be i.i.d.r.v.s. The following SCP approximations to $\mathcal{L}(S_n)$ has been suggested by Čekanavičius [51]. Set

$$\begin{aligned} G_7 &= \exp\big((1-p)(U-I) + p(V-I) - p^2(V-I)^{*2}/2\big), \\ A_3 &= -n(1-p)p(U-I) * (V-I) + np^3(V-I)^{*3}/3. \end{aligned}$$

Note that $G_7 = \exp(P-I-p^2(V-I)^{*2}/2)$.

Theorem 4.7. [51] *If* $0 \le p \le C_0 < 1$*, then there exists an absolute constant* C *such that*

$$\sup_{P \in \mathcal{F}_+} \sup_{V \in \mathcal{F}} d_K(\mathcal{L}(S_n); G_7^{*n}) \le C\Big(1/n + \sqrt{p/n}\,\Big), \tag{4.40}$$

$$\sup_{P \in \mathcal{F}_+} \sup_{V \in \mathcal{F}} d_K(\mathcal{L}(S_n); G_7^{*n} * (I + A_3)) \le C/n. \tag{4.41}$$

Inequality (4.40) is a generalization of the upper bound in (4.35). Estimates (4.40) and (4.41) demonstrate that approximations of order $O(n^{-1/2})$ and $O(n^{-1})$ can be achieved if (3.10*) holds for a particular $U \in \mathcal{F}_+$ and $p \le C_0 < 1$.

Proofs of a number of results concerning compound Poisson approximations to the distribution of a sum of symmetric r.v.s. can be found in [6] and [68], Sec. 2.7.

Distributions obeying certain moment assumptions.

Let $X, X_1, \ldots, X_n$ be i.i.d.r.v.s such that $\mathcal{L}(X) \in \mathcal{F}_{0,\beta}$ $(\exists \beta \in (1;2])$. Zaitsev [204] has proved that there exists a constant C_X such that

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) \le C_X n^{-\alpha}\,, \tag{4.42}$$

where $\alpha = \min\{1/2; \beta - 1\}$. If $\beta \in (3/2; 2)$ and $\mathbb{E}X^2 = \infty$, then

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) = o(n^{-1/2}),$$

where $\tilde{S}_n$ is the sum of accompanying r.v.s. defined by (1.9).

Zaitsev's estimate (4.42) can be improved if $\mathbb{E}X = 0$, $\mathbb{E}|X|^{1+\beta} < \infty$ $(\exists \beta \in (0;1])$, and $\mathcal{L}(X)$ satisfies Cramér's condition

$$\limsup_{|t| \to \infty} |\widehat{F}_X(t)| < 1. \tag{4.43}$$

Namely, in such case there exists C_X such that

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) \le C_X n^{-\beta}\,, \quad d_K(\mathcal{L}(S_n); G_8) \le C_X n^{-2\beta}\,, \tag{4.44}$$

where

$$G_8 = \mathbf{\Pi}(n, X) * (I - n(\mathcal{L}(X) - I)^{*2}/2)$$

[57]. In particular,

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n, X)) \le C_X n^{-1} \tag{4.45}$$

if $\mathbb{E}X=0$, $\mathbb{E}X^2<\infty$ (Studnev [178]).

Under the additional assumption that $\mathbb{E}X^4<\infty$ Studnev [178] has shown that

$$\sup_x |\mathbb{P}(S_n/\sigma\sqrt{n}<x) - F_n(x) - x(3-x^2)\varphi(x)/8n| = o(n^{-1})\,, \tag{4.46}$$

where $\sigma^2=\mathbb{E}X^2$, F_n is the d.f. of $\mathbf{\Pi}(n,X/\sigma\sqrt{n})$ and φ is the density of the standard normal distribution. If $\mathbb{E}X=0$, $\mathbb{E}X^2<\infty$, but instead of (4.43) one assumes that $\mathcal{L}(X)$ is non-lattice, then the RHS of (4.45) is $o(n^{-1/2})$.

Further results on the topic can be found in [205, 94].

Open problem.

4.4. Recall (3.10^*). Let $\mathcal{L}(\tilde{X})$ be the accompanying $\mathcal{L}(X)$ distribution defined by (1.6). For any $m\in\mathbb{N}$ we set

$$B_m(U,V) = \sum_{j=0}^{m} \binom{n}{j} \mathcal{L}(\tilde{X})^{*(n-j)} * (\mathcal{L}(X)-\mathcal{L}(\tilde{X}))^{*j}\,.$$

For instance, $B_0(U,V)=\mathcal{L}(\tilde{S}_n)$,

$$B_1(U,V) = \mathcal{L}(\tilde{S}_n) + n\mathcal{L}(\tilde{S}_{n-1}) * (\mathcal{L}(X)-\mathcal{L}(\tilde{X})).$$

It is known [42] that

$$\sup_{U\in\mathcal{F}_+}\sup_{V\in\mathcal{F}} d_K(\mathcal{L}(S_n);B_m(U,V)) \le C_m(n^{-1}+p)^{m+1}\,. \tag{4.47}$$

Will (4.47) hold if assumption $U\in\mathcal{F}_+$ is replaced with $U\in\mathcal{F}_s$?

4.5. Shifted compound Poisson approximation

Let $X,X_1,\dots,X_n$ be independent and identically distributed r.v.s. Denote by $\tilde{X}_{1,a},\dots,\tilde{X}_{n,a}$ accompanying $X_1+a,\dots,X_n+a$ random variables, and let

$$\tilde{S}_{n,a} = \tilde{X}_{1,a} + \dots + \tilde{X}_{n,a}\,.$$

Clearly, $\tilde{S}_{n,a}$ is a compound Poisson $\mathbf{\Pi}(n,X+a)$ random variable.

Le Cam [129] has shown that

$$\sup_{\mathcal{L}(X)}\inf_{a\in\mathbb{R}} d_K(S_n+na;\tilde{S}_{n,a}) \le 132n^{-1/3}\,. \tag{4.48}$$

A detailed procedure of finding a suitable shift has been described in [111]. According to Ibragimov & Presman [111], constant 132 in (4.48) can be replaced with 8.

Presman ([6], ch. VIII.4) has shown that

$$\sup_{\mathcal{L}(X)}\inf_{a\in\mathbb{R}} d_K(S_n+na;\tilde{S}_{n,a}) \ge cn^{-1/3}\,, \tag{4.49}$$

where $c>0$ is an absolute constant; the bound holds if the class of distributions is reduced to the family of Bernoulli random variables.

Note that the rate of the accuracy of shifted Poisson approximation to the Binomial distribution is $n^{-1/2}$ (see Theorem 6 in [147]); according to (4.1), the rate of the accuracy of shifted compound Poisson approximation to the Binomial distribution is $n^{-2/3}$.

Zaitsev [204] has conjectured that for every $\mathcal{L}(X)$ there exist a constant C_X such that

$$\inf_{a\in\mathbb{R}} d_K(\mathcal{L}(S_n+na);\tilde{S}_{n,a}) \le C_X n^{-1/2}\,. \tag{4.50}$$

A first-order asymptotic expansion.

Let $X, X_1, \ldots, X_n$ be i.i.d.r.v.s. Set

$$B_{n,a}(X) = \mathcal{L}(\tilde{S}_{n,a}) * \Big(I - \tfrac{n}{2}(\mathcal{L}(X+a)-I)^{*2}\Big). \tag{4.51}$$

Then [47]

$$\sup_{\mathcal{L}(X)\in\mathcal{F}} \inf_{a\in\mathbb{R}} d_K(\mathcal{L}(S_n+na);B_{n,a}(X)) \le C_5 n^{-2/5}\,. \tag{4.52}$$

Čekanavičius [42] has proved that

$$\sup_{\mathcal{L}(X)\in\mathcal{F}_+} d_K(\mathcal{L}(S_n);B_{n,0}(X)) \le C n^{-2}\,. \tag{4.53}$$

If $0\le p\le C_0<1$, then [51]

$$d_K(\mathcal{L}(S_n);H^{*n}) \le C\Big(1/n + \sqrt{p/n}\Big). \tag{4.54}$$

Open problem.
4.5. Improve the accuracy of approximation in (4.52).

4.6. Other results

Arak's method has been applied in order to derive an asymptotic expansion with the accuracy $O(n^{-1+\varepsilon})$ for any fixed $0<\varepsilon\le 1/3$, see Čekanavičius [60]. However, only the existence of such asymptotic expansion has been established.

Chen & Roos [38] have evaluated the accuracy of compound Poisson approximation to $\mathbb{E}f(S_n)$, where f is an unbounded function. Borisov [31] has proved that

$$\mathbb{E}f(S_n) \le \mathbb{E}f(\tilde{S}_n)$$

for a class of functions f, where $\tilde{S}_n$ is a sum of accompanying $X, X_1, \ldots$ r.v.s. Besides, he showed that

$$\mathbb{E}f(S_n) \le \mathbb{E}f(\tilde{S}_n)/\mathbb{P}(X=0)$$

for a non-negative measurable function f if $X, X_1, \ldots$ are i.i.r.v.s, see also [30], §5.

Let $X, X_1, ..., X_n$ be i.i.d.r.v.s such that $\mathbb{E}X = 0 < \mathbb{E}X^2 < \infty$, $\mathbb{E}|X|^k < \infty$ $(\exists k \geq 3)$. Assume that $\mathcal{L}(X)$ does not depend on n and satisfies Cramér's condition (4.43).

Let $\eta, \pi_{\alpha_1}, ..., \pi_{\alpha_k}$ be independent random variables, where η is a standard normal random variable, π_{α_j} is a Poisson r.v. with parameter α_j $(j \geq 1)$. There exist $\alpha_1 > 0, \ldots, \alpha_k > 0, \beta_1 \in \mathbb{R}, \ldots \beta_k \in \mathbb{R}$ such that

$$d_K(S_n/\sqrt{n\mathbb{E}X^2}\,; \eta + \beta_1\pi_{\alpha_1} + \ldots + \beta_k\pi_{\alpha_k}) = O(n^{-(k-1)/2}) \qquad (4.55)$$

as $n \to \infty$. An explicit algorithm for choosing $\{\alpha_i, \beta_i\}$ has been described in Čekanavičius [49].

Open problem.

4.6. Let $X, X_1, ..., X_n$ be i.i.d.r.v.s. Denote $P_X = \mathcal{L}(X)$. Will SCP measure

$$G_9 = \exp\left(n(P_X - I) - \frac{n}{2}(P_X - I)^{*2}\right)$$

approximate $\mathcal{L}(S_n)$ with the rate $o(n^{-1})$?

4.7. Applications

2-run statistic.

Let $\xi_1, \xi_2, \ldots$ be independent Bernoulli $\mathbf{B}(p)$ random variables, where $0 < p < 1$. Denote $S_{n,2} = \sum_{i=1}^{n-1} \xi_i\xi_{i+1}$,

$$G_+ = \exp\left(np^2(I_1 - I) + \gamma_2(I_1 - I)^{*2} + \gamma_3(I_1 - I)^{*3}\right),$$

where

$$\gamma_2 = np^3(1 - \tfrac{3}{2}p) - p^3(1-p)\,,\ \gamma_3 = np^4(1 - 4p + \tfrac{10}{3}p^2) - 2p^4(1-p)(1-2p)\,.$$

Then $S_{n,2}$ is a 2-run statistic, G_+ is the distribution of $\pi_{\lambda_1} + 2\pi_{\lambda_2} + 3\pi_{\lambda_3}$, where π_{λ_1}, π_{λ_2}, π_{λ_3} are independent Poisson random variables with parameters $\lambda_1 = np^2 - 2\gamma_2 + 3\gamma_3$, $\lambda_2 = \gamma_2 - 3\gamma_3$, $\lambda_3 = \gamma_3$.

A sharp estimate of the accuracy of compound Poisson approximation to $\mathcal{L}(S_{n,2})$ has been established by Petrauskienė & Čekanavičius [149].

Theorem 4.8. [149] *Assume that* $p \leq 1/5$. *There exists an absolute constant* C *such that*

$$d_{TV}(\mathcal{L}(S_{n,2}); G_+) \leq C \min\left(np^5; p/n\right) \qquad (n \geq 3). \qquad (4.56)$$

Further reading on the topic includes [40, 41, 67, 187, 70, 71].

(k_1, k_2)**-run statistic**.

k-run statistic is not the only one explicitly related to a sequence of independent Bernoulli random variables.

Given two natural numbers k_1, k_2, a (k_1, k_2)-run is a pattern consisting of at least k_1 consecutive failures followed by at least k_2 consecutive successes.

Let $\{\xi_i\}$ be independent Bernoulli $\mathbf{B}(p_i)$ random variables $(0<p_i<1)$. Set $m=k_1+k_2$,

$$X_j = (1-\xi_{j-m+1})\cdots(1-\xi_{j-k_2})\xi_{j-k_2+1}\cdots\xi_j \qquad (j\geq m).$$

Denote

$$S_n(k_1,k_2) = X_m + X_{m+1} + \ldots + X_n\,.$$

If $k_1 = 1$, then $S_n(1,k)$ is the number of head runs of length $\geq k$ among $\xi_1,\ldots,\xi_n$.

Approximations to $\mathcal{L}(S_n(k_1,k_2))$ have been suggested in [67, 180, 184, 187]. We present below an analogue of (4.1) established by Vellaisamy & Čekanavičius [187].

Let $\mathcal{S} = \{p\colon m(1-p)^{k_1}p^{k_2} \leq 0.01\}$. Assume that $p\in\mathcal{S}$. Then there exist a compound Poisson distribution Y and a constant C_m such that

$$d_{TV}(\mathcal{L}(S_n(k_1,k_2));Y) \leq C_m n^{-2/3} \qquad (n>C_m). \tag{4.57}$$

An urn model with overflow.

Suppose that n balls are distributed into m urns, and each ball is equally likely to be assigned to any urn. Each urn can hold at most k balls, where $k\geq 2$ is a fixed number. If a ball is assigned to an urn that is already full, that ball is placed in an additional "overflow urn" of unlimited capacity.

Let W be the number of balls allocated to the overflow urn. A compound Poisson approximation to W has been suggested in [35, 73]. Set

$$\begin{aligned}\lambda_i &= \binom{n}{i+k}\left(\frac{1}{m}\right)^{i+k-1}\left(1-\frac{1}{m}\right)^{n-i-k} \qquad (i=1,\ldots,n-k),\\ \lambda &= \lambda_1+\ldots+\lambda_{n-k}\,,\quad Z_n = \sum_{i=1}^{n-k} i\pi_{\lambda_i}\,.\end{aligned}$$

Daly [73] has shown that

$$\begin{aligned}d_{TV}(W;Z_n) &\leq \mathcal{M}(\lambda)\Bigg\{m^2\left(\sum_{i=k}^{n}(i-k)\binom{n}{i}\left(\frac{1}{m}\right)^i\left(1-\frac{1}{m}\right)^{n-i}\right)^2\\ &- m(m-1)\sum_{i=k+1}^{n}\sum_{j=k+1}^{n-i}\frac{(i-k)(j-k)n!}{i!j!(n-i-j)!}\left(\frac{1}{m}\right)^{i+j}\left(1-\frac{2}{m}\right)^{n-i-j}\Bigg\},\end{aligned}$$

where factor $\mathcal{M}(\lambda)$ obeys (3.13)–(3.15).

Other applications.

An overview of compound Poisson approximation results obtained via Stein's method can be found in [17, 18].

Compound Poisson approximation to the distribution of the number of k-out-of-n isolated vertices of a rectangular lattice on a torus has been presented in [164].

Compound Poisson approximation to the number of overlapping and non-overlapping occurrences of word patterns has been suggested in [87]. Compound Poisson approximation to the number of overlapping sequences has been studied in [40].

A Negative Binomial approximation to the number of parasites has been suggested in [23].

A review of compound Poisson approximations to the number of dependent claims has been given by Genest et al. [84].

5. Multivariate compound Poisson approximation

A compound Poisson random vector is defined by (1.1), where $\{\zeta_i\}$ are independent random vectors. We denote by $\mathbf{\Pi}(\lambda,\zeta) \equiv \mathbf{\Pi}(\lambda,\mathcal{L}(\zeta))$ the multivariate compound Poisson distribution with intensity λ and compounding (multiplicity) distribution $\mathcal{L}(\zeta)$:

$$\mathbf{\Pi}(\lambda,\zeta) = \mathcal{L}\left(\sum\nolimits_{i=0}^{\pi_\lambda} \zeta_i\right),$$

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

5.1. Multivariate compound Poisson limit theorem

This section presents a multivariate compound Poisson limit theorem.

Let $\{X,X_1,...,X_n\}$, where $X_i = (X_i^{(1)},...,X_i^{(k)})$, be a sequence of k-dimensional random vectors that are non-zero with "small" probabilities. Set

$$S_n = X_1 + ... + X_n\,.$$

Example 5.1. Let $\{\xi_i\}$ be a sequence of random variables. Denote

$$N_n(x) = \sum_{i=1}^n 1\!\!\mathrm{I}\{\xi_i > x\}, \quad N_n[a,b) = \sum_{i=1}^n 1\!\!\mathrm{I}\{a \geq \xi_i > b\} \qquad (a > b).$$

Given a set $x_1 > ... > x_k$ of numbers ("levels"), set

$$S_n = (N_n(x_1), N_n[x_1;x_2), ..., N_n[x_{k-1};x_k))\,. \tag{5.1}$$

Then $S_n = X_1 + ... + X_n$, where $X_i = (1\!\!\mathrm{I}\{\xi_i > x_1\}, 1\!\!\mathrm{I}\{x_1 \geq \xi_i > x_2\}, ..., 1\!\!\mathrm{I}\{x_{k-1} \geq \xi_i > x_k\})$.

Random vector (5.1) plays a role in extreme value theory when one deals with a joint distribution of exceedances of several level (cf. [144], ch. 6). □

Random vector S_n can be approximated by a compound Poisson random vector. Indeed, Theorem 2.1 clearly holds if $\{X_i\}$ are i.i.d. random vectors.

Let $\{X,X_1,...,X_n\}$ be a stationary sequence of random vectors. The argument of the proof of Theorem 2.3 remains valid. Hence Theorem 2.3 holds if $\{X_i\}$ are random vectors.

Theorem 2.3* *Assume that*

$$\limsup_{n\to\infty} n\mathbb{P}(X \neq \bar{0}) < \infty, \tag{5.2}$$

and there exists the limit

$$\lim_{n\to\infty} \mathbb{P}(S_n = \bar{0}) := e^{-\lambda} \qquad (\exists \lambda > 0). \tag{5.3}$$

If $\mathcal{L}(S_r | S_r \neq \bar{0}) \Rightarrow \mathcal{L}(\zeta)$ *as* $n \to \infty$ *for a random vector* ζ *and a sequence* $\{r = r_n\} \in \mathcal{R}$*, then*

$$\mathcal{L}(S_n) \Rightarrow \mathbf{\Pi}(\lambda, \zeta). \tag{2.15*}$$

The limiting distribution $\mathbf{\Pi}(\lambda, \zeta)$ *in (2.15*) does not depend on the choice of a sequence* $\{r_n\}$*.*

If $\mathcal{L}(S_n)$ *converges weakly to a random vector* S*, then there exists* $\lambda \geq 0$ *and a random vector* ζ *such that* $\mathcal{L}(S) = \mathbf{\Pi}(\lambda, \zeta)$*, where* $\lambda = -\ln \mathbb{P}(S = \bar{0})$*, and (5.3) holds. If* $\lambda > 0$*, then there exist a sequence* $\{r = r_n\} \in \mathcal{R}$ *such that (2.14) holds.*

Theorem 2.3* is essentially Theorem 6.6 from [144].

Random vector ζ in Theorem 2.3* may have dependent components. The following theorem presents a necessary and sufficient condition for a weak convergence of S_n to a vector with *independent* compound Poisson components.

Given $t_0 = 0 < t_1 < \dots < t_k < \infty$, denote $\bar{t} = (t_1, \dots, t_k)$. Set

$$S_n^{(j)} = X_1^{(j)} + \dots + X_n^{(j)}, \quad p_j = (t_j - t_{j-1})/t_k \qquad (1 \leq j \leq k).$$

Condition $(C_{\bar{t}})$.

We say that condition $(C_{\bar{t}})$ *holds if there exists a random variable* ζ *taking values in* $\mathbb{N}$ *and a sequence* $\{r = r_n\}$ *such that* $n \gg r \gg 1$*,*

(a) for every $1 \leq i \leq k$*,* $\ell \geq 1$*,*

$$\mathbb{P}(S_r^{(i)} = \ell) \sim \frac{r}{n} \mathbb{P}(\zeta = \ell)(t_i - t_{i-1}) \qquad (n \to \infty),$$

(b) for every $1 \leq i < j \leq k$

$$\mathbb{P}(S_r^{(i)} > 0, S_r^{(j)} > 0) = o(r/n) \qquad (n \to \infty).$$

Condition $(C_{\bar{t}})$ is necessary and sufficient for the weak convergence of $\mathcal{L}(S_n)$ to a vector with *independent* compound Poisson components.

Note that conditions (a) and (5.2) yield

$$\mathbb{P}(S_r^{(i)} > 0) \sim (t_i - t_{i-1}) r/n \qquad (n \to \infty) \tag{5.4}$$

$(1 \leq i \leq m,\ \ell \geq 1)$. Hence (a) means

$$\mathbb{P}(S_r^{(i)} = \ell | S_r^{(i)} > 0) \sim \mathbb{P}(\zeta = \ell) \qquad (1 \leq i \leq k,\ \ell \geq 1) \tag{a^*}$$

as $n \to \infty$. If condition Δ holds, then (5.4) is equivalent to

$$\lim_{n\to\infty} \mathbb{P}(S_n^{(i)} = 0) = e^{-t_i - t_{i-1}}.$$

Thus, instead of assuming (5.3), one could have added (5.4) as item (c) of condition $(C_{\bar t})$ (cf. [143]).

Condition (b) means components of a random vector ζ_r with the distribution $\mathcal{L}(\zeta_r) = \mathcal{L}(S_r | S_r \neq 0)$ are asymptotically independent.

Let $\{\pi(s), s \geq 0\}$ be a Poisson process with intensity rate 1, and let $\eta, \eta_1, \eta_2, ...$ be a sequence of i.i.d.r.v.s taking values in $\mathbb{N}$. Denote

$$Q(t) = \sum_{j=1}^{\pi(t)} \eta_j .$$

Then $\{Q(t), t \geq 0\}$ is a compound Poisson jump process. Equivalently,

$$\tilde{Q}(B) := \int_B Q(dt)$$

is a compound Poisson point process with the Lebesgue intensity measure and multiplicity distribution $\mathcal{L}(\eta)$.

Denote

$$S_{\bar t} = \{Q(t_1), Q(t_2) - Q(t_1), ..., Q(t_k) - Q(t_{k-1})\}.$$

Clearly, $S_{\bar t}$ is a random vector with *independent* compound Poisson components.

The ch.f. of $S_{\bar t}$ is

$$\mathbb{E}\exp(i\bar{s}S_{\bar t}) = \exp\Big(t_k\Big(\sum_{j=1}^{k} p_j \varphi_\eta(s_j) - 1\Big)\Big) \qquad (\forall \bar{s} = (s_1, ..., s_k) \in \mathbb{R}^k),$$

where φ_η is a ch.f. of $\mathcal{L}(\eta)$.

Theorem 5.1. *Assume conditions Δ and (5.2), and suppose that for a vector $\bar t = (t_1, ..., t_k)$, where $t_0 = 0 < t_1 < ... < t_k < \infty$, there exist the limits*

$$\lim_{n\to\infty} \mathbb{P}(S_n^{(j)} = 0) = e^{-t_j + t_{j-1}} \qquad (\forall j \in \{1, ..., k\}). \tag{5.5}$$

Weak convergence

$$S_n \Rightarrow S_{\bar t} \tag{5.6}$$

holds if and only if condition $(C_{\bar t})$ holds.

Theorem 5.1 is essentially Theorem 6.3 from [144].

Example 5.2. deals with sample extremes. We rewrite the sample $X_1, \ldots, X_n$ in the non-increasing order:

$$X_{1:n} \geq ... \geq X_{n:n} .$$

Then $X_{1:n}, ..., X_{n:n}$ are called the "order statistics". In particular, $M_n = X_{1:n}$ is the sample maximum, $X_{l:n}$ is the l^{th} sample maximum.

Denote by $N_n(x) = \sum_{i=1}^{n} \mathbb{1}\{X_i > x\}$ the number of exceedances over the threshold x. It is easy to see that

$$\{X_{l:n} \le x\} = \{N_n(x) < l\} \qquad (1 \le l \le n). \tag{5.7}$$

Let $\{u_n(\cdot)\}$ be a non-decreasing normalising sequence such that

$$\limsup_{n\to\infty} n\mathbb{P}(X > u_n(t)) < \infty, \ \lim_{n\to\infty} \mathbb{P}(M_n \le u_n(t)) = e^{-t} \qquad (\forall t > 0). \tag{5.8}$$

Assume condition Δ. The following result can be deduced from Theorem 5.1 for the joint limiting distribution of $X_{1:n}$ and $X_{l:n}$: if $0 < s < t$, then

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}(X_{1:n} \le u_n(s), \ X_{l:n} \le u_n(t)) \qquad (5.9)\\ &= e^{-t}\left\{1 + \sum_{j=1}^{l-1} (t-s)^j \, \mathbb{P}\left(\sum_{i=1}^{j} \zeta_i < k\right) \Big/ j!\right\} \qquad (l \ge 2). \end{aligned}$$

In particular,

$$\lim_{n\to\infty} \mathbb{P}(X_{1:n} \le u_n(s), X_{2:n} \le u_n(t)) = e^{-t}\left(1 + (t-s)\mathbb{P}(\zeta = 1)\right).$$

Similarly, if $0 < q < s < t$, then Theorem 5.1 yields

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}(X_{1:n} \le u_n(q), X_{2:n} \le u_n(s), X_{3:n} \le u_n(t)) \qquad (5.10)\\ &\quad = e^{-t}\big\{1 + (t-q)\mathbb{P}(\zeta = 1) + (t-s)^2\mathbb{P}^2(\zeta = 1)/2\\ &\quad + (t-s)(s-q)\mathbb{P}^2(\zeta = 1) + (t-s)\mathbb{P}(\zeta = 2)\big\}. \end{aligned}$$

Formulas (5.9), (5.10) demonstrate the impact of the asymptotic clustering of extremes on the limiting distribution of upper order statistics. □

Remark. Condition $(C_{\bar{t}})$ stipulates the "regular" way of asymptotic clustering of extremes. Waiving it makes the situation more complicated (cf. formula (6.10) in [144]).

Example 5.3. Let $\{X_i, i \ge 1\}$ be a strictly stationary α–mixing sequence. Hsing [107] has shown that $\lim_{n\to\infty} \mathbb{P}(X_{1:n} \le u_n(s), X_{l:n} \le u_n(t))$, if exists, is necessarily expressed via a compound Poisson distribution, cf. (5.11).

Necessary and sufficient conditions for the convergence of $\mathbb{P}(X_{1:n} \le u_n(s),$ $X_{l:n} \le u_n(t))$: if (5.8) holds, then the probability

$$\mathbb{P}(X_{1:n} \le u_n(s), X_{l:n} \le u_n(t)) = \mathbb{P}(N_n(u_n(s)) = 0, N_n(u_n(t)) < l)$$

converges for every $t > s > 0$ if and only if there exist functions $f_i(\cdot)$ and a sequence $\{r = r_n\}$ such that $n \gg r \gg 1$ and

$$\lim_{n\to\infty} \mathbb{P}(N_r(u_n(s)) = 0, \ N_r(u_n(t)) = i \,|\, N_r(u_n(t)) > 0) = f_i(s/t)$$

for each $t > s > 0$ and $i \in \{1, \ldots, l-1\}$. The limit is expressed via a compound Poisson distribution:

$$\lim_{n\to\infty} \mathbb{P}\left(N_n(u_n(s))=0,\, N_n(u_n(t))<l\right) = \mathbb{P}\left(\sum_{i=1}^{\pi(t)} \zeta_i^* < l\right), \tag{5.11}$$

where $\{\zeta_i^*\}$ are i.i.d.r.v.s, $\mathbb{P}(\zeta^* = i) = f_i(s/t)$ (Novak [143]). In particular, the limiting cluster size distribution depends on the ratio s/t.

Sufficient conditions for the weak convergence of the random vector $\{N_n(u_n(s)),$ $N_n(u_n(t))\}$ can be found in Novak [143], Proposition 6 (see also [144], p. 107).
□

5.2. Accuracy of multivariate CP approximation: rare events

Estimates of the accuracy of univariate compound Poisson approximation to $\mathcal{L}(S_n)$ have been given in section 3. Definitions of metrics, accompanying r.v.s, an exponent of a measure, etc., remain valid in the multivariate case.

We present below results concerning the accuracy of multivariate compound Poisson approximation to the distribution of the sum $S_n = X_1 + \ldots + X_n$ of random vectors $X_1, \ldots, X_n$.

Let $\{X_i\}$ be independent random vectors that are non-zero with small probabilities. Recall that $\tilde{S}_n$ denote the sum of accompanying random vectors, see (1.9). Set

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

Khintchine's formula (2.1*) holds for random vectors:

$$X_i \stackrel{d}{=} \tau_i X_i',$$

where τ_i and X_i' are independent r.v.s,

$$\mathcal{L}(X_i') = \mathcal{L}(X_i | X_i \neq 0), \;\; \mathcal{L}(\tau_i) = \mathbf{B}(p_i).$$

Therefore, (2.9) remains valid: if $\{X_i'\}$ are identically distributed, then

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

Besides, (3.6) entails

$$d_{TV}(S_n; \tilde{S}_n) \leq \sum_{i=1}^{n} p_i^2\,.$$

A number of univariate results have been generalized to the multi-dimensional case. In particular, inequality (3.8) has been generalized by Zaitsev [199] to the case of independent random vectors taking values in $\mathbb{R}^k$ that are zero with large probabilities: there exists constant $C(k)$ such that

$$d_K(S_n; \tilde{S}_n) \leq C(k) \max_{1\leq i\leq n} p_i\,. \tag{5.12}$$

Related results can be found in [95, 96].

Bound (3.9) holds in the multivariate case as well. Estimates (3.25) and (3.26) hold in the multivariate case (cf. [68], p. 28). Multivariate versions of (4.7) for independent random vectors have been given by Roos [166, 167].

Independent 0-1 random vectors.

Let $\{X_i\}$ be independent k-dimensional random vectors. Set $\bar{0}=(0,\ldots,0)$,

$$p_{j,r} = \mathbb{P}(X_j = \bar{e}_r), \quad p_j = \mathbb{P}(X_j \neq \bar{0}) = \sum_{r=1}^{k} p_{j,r}, \quad \lambda_r = \sum_{i=1}^{n} p_{i,r} > 0,$$

$$\tilde{p}_0 = \sum_{r=1}^{k} \max_{1\le i\le n} p_{i,r}\,.$$

Tecall that I denotes the distribution concentrated at $\bar{0}$, $I_{\bar{e}_j}$ is the distribution concentrated at $\bar{e}_j$. Denote

$$F_n = \prod_{j=1}^{*n}\left((1-p_j)I + \sum_{r=1}^{k} p_{j,r}(I_{\bar{e}_r}-I)\right), \quad G_* = \exp\left(\sum_{r=1}^{k} \lambda_r (I_{\bar{e}_r}-I)\right).$$

If $\{X_i\}$ are i.i.d., then $F_n = \mathcal{L}(S_n)$ is a multinomial distribution. Results on the accuracy of Poisson approximation to the multinomial distribution can be found, e.g., in [144]. Set

$$A_j = \sum_{r=1}^{k} p_{j,r}(I_{\bar{e}_r}-I); \quad H_{n,s} = \exp\left(\sum_{m=1}^{s} \frac{(-1)^{m+1}}{m} \sum_{j=1}^{n} A_j^{*m}\right) \qquad (s\in\mathbb{N}).$$

If $\tilde{p}_0 \le 1/4$, then (Roos [166]) there exists constant C_s such that

$$d_{TV}(\mathcal{L}(S_n); H_{n,s}) \le C_s \sum_{j=1}^{n}\left(\min\left\{\sum_{r=1}^{k}\frac{p_{j,r}^2}{\lambda_r}; p_j^2\right\}\right)^{(s+1)/2}. \tag{5.13}$$

If k,s are fixed and $p_{i,j} \asymp C$, then the rate of accuracy in (5.13) is $O(n^{-(s-1)/2})$.

Further SCP approximations to the sum of 0-1 random vectors can be found in [167]. A one dimensional version of (5.13) is (4.13).

The next theorem estimates the accuracy of approximation $F_n \approx G_*$.

Denote $q_{i,r} = p_{i,r}/p_i$ $(i \in \{1,\ldots,n\},\ r\in\{1,\ldots,k\})$. Clearly, $\sum_{r=1}^{k} q_{i,r} = 1$ $(\forall i)$. Set

$$\lambda = \sum_{j=1}^{n} p_j = \sum_{r=1}^{k} \lambda_r\,,$$

$$\alpha = \sum_{j=1}^{n} g(2p_j) p_j^2 \sum_{r=1}^{k} q_{j,r} \min\{2^{-3/2} q_{j,r}/\lambda_r; 2\},$$

$$\beta = \sum_{j=1}^{n} p_j^2 \sum_{r=1}^{k} q_{j,r} \min\{q_{j,r}/\lambda_r; 1\},$$

where $g(z) = 2e^z(e^{-z} - 1 - z)/z^2 \quad (z > 0)$.

Theorem 5.2. [170] *The following estimate holds:*

$$d_{TV}(F_n; G_*) \le 7.8\beta. \tag{5.14}$$

If $\alpha \le 2^{-3/2}$, *then*

$$d_{TV}(F_n; G_*) \le \alpha/(1 - 2\sqrt{2}\alpha). \tag{5.15}$$

If $k = 1$ and $q_{j,1} \equiv 1$ $(\forall j)$, then (5.14) becomes an estimate of the accuracy of univariate Poisson approximation to the distribution of a sum of independent Bernoulli r.v.s.

Dependent 0-1 random vectors.

We now consider the case of weakly dependent 0-1 random vectors.

Let $X, X_1, \ldots, X_n$ be a strictly stationary sequence of k-dimensional 0-1 random vectors such that not more than one coordinate of a vector may equal 1. Set $X_i = (X_i^{(1)}, ..., X_i^{(k)})$,

$$S_n = X_1 + ... + X_n \,.$$

Denote $\bar{e}_j = (0, \ldots, 1, \ldots, 0)$, i.e., vector $\bar{e}_j$ has the j^{th} coordinate equal 1, the other coordinates equal zero. Assume that

$$\mathbb{P}(X = \bar{0}) = 1 - p, \ \mathbb{P}(X = \bar{e}_j) = p_j \qquad (1 \le j \le k), \tag{5.16}$$

where $\bar{0} = (0, \ldots, 0)$, $p = \mathbb{P}(X \ne \bar{0})$.

If $\{X_i\}$ are independent, then $\mathcal{L}(S_n)$ is multinomial $\mathbf{B}(n, p_1, ..., p_k)$. An estimate of the accuracy of Poisson approximation to $\mathbf{B}(n, p_1, ..., p_k)$ can be found in [147].

Given $r \in \{1, ..., n\}$, let $\zeta, \zeta_1, \zeta_2, \ldots$ be independent random vectors with the common distribution

$$\mathcal{L}(\zeta) = \mathcal{L}(S_r | S_r \ne \bar{0}).$$

Denote

$$q = \mathbb{P}(S_r \ne 0), \ k = [n/r], \ r' = n - rk, \ \lambda = kq.$$

We approximate $\mathcal{L}(S_n)$ by the multivariate compound Poisson distribution $\mathbf{\Pi}(kq, \zeta)$.

Theorem 5.3. *If* $n > r > l \ge 0$ *and* $\mathcal{L}(Y) = \mathbf{\Pi}(kq, \zeta)$, *then*

$$d_{TV}(S_n; Y) \le C_{n,r} rp + (r' + 2nr^{-1}l)p + nr^{-1} \min\{\beta(l); \kappa(l)\}, \tag{5.17}$$

where $C_{n,r} = \min\{3/4e + (1 - e^{-np})rp; 1 - e^{-np}\}$ *and* $\kappa(l) = 1$ *if* $m2^{(m-1)/2}\alpha(l) > 1$, $\kappa(l) = 2(1 + 2/m)\left(2^{m-1}m^2\alpha^2(l)\right)^{1/(2+m)}$ *if* $m2^{(m-1)/2}\alpha(l) \le 1$.

Theorem 5.3 is effectively Theorem 6.8 from [144].

If $\{\xi_i\}$ are i.i.d. $\mathbf{B}(p)$ random variables, then (5.17) with $l=0$ and $r=1$ becomes an estimate of the accuracy of Poisson approximation to $\mathcal{L}(S_n)$ with a correct constant $3/4e$ at the leading term:

$$d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)) \le 3p/4\mathrm{e} + (1-\mathrm{e}^{-np})p^2.$$

The next corollary applies (5.17) to the case of m-dependent random vectors.

Corollary 5.4. *If vectors $\{\xi_i\}$ are m-dependent and $m<r<n$, then*

$$d_{TV}(S_n;Y) \le C_{n,r}rp + (r'+2nm/r)p. \tag{5.18}$$

If we choose $r \asymp \sqrt{n}$, then the right-hand side of (5.18) is $O(p\sqrt{n}\,)$.

Open problem.

5.1. The term $(2nr^{-1}l+r')p$ appears in (5.17) because of the method (Bernstein's blocks approach). An open question is if it can be removed.

5.3. Accuracy of multivariate CP approximation: general case

Let $X, X_1, \dots, X_n$ be independent and identically distributed random vectors taking values in $\mathbb{R}^k$. Denote by $\tilde{X}_a, \tilde{X}_{1,a}, \dots, \tilde{X}_{n,a}$ accompanying $X_1+a, \dots, X_n+a$ independent random vectors, and let

$$\tilde{S}_{n,a} = \tilde{X}_{1,a} + \dots + \tilde{X}_{n,a}\,.$$

Recall that $\tilde{S}_{n,a}$ is a compound Poisson random vector.

Estimate (4.48) of the accuracy of compound Poisson approximation has been generalized to the multivariate case by Presman [152]: there exists an absolute constants C such that

$$\sup_{\mathcal{L}(X)} \inf_a d_K(S_n;\tilde{S}_{n,a}-na) \le Cn^{-1/3}\,. \tag{5.19}$$

At a moment (5.19) is the best available estimate of the accuracy of multivariate compound Poisson approximation without extra assumptions on $\mathcal{L}(X)$.

Bentkus et al. [25] state that the rate of approximation to $\mathcal{L}(S_n)$ by $\mathcal{L}(\tilde{S}_n)$ is $O(n^{-1})$ if $\mathbb{E}\|X\|^{8/3}<\infty$. Here norm is understood as a square root of a scalar product of X with itself. Namely, assume that $\mathcal{L}(X)$ is not concentrated on a hyperspace in $\mathbb{R}^k$, $\mathbb{E}X=0$, $\mathbb{E}\|X\|^{8/3}<\infty$. Then for any $a\in\mathbb{R}^k$, as $n\to\infty$,

$$\sup_x |\mathbb{P}(\|S_n-a\|^2 < x) - \mathbb{P}(\|\tilde{S}_n-a\|^2 < x)| = O((1+\|a\|^4)n^{-1})\,. \tag{5.20}$$

Asymptotic expansions.

The next result presents a SCP approximation in (5.19).

Recall Khintchine's formula (3.10):

$$X \stackrel{d}{=} \tau X_A + (1-\tau)X_{A^c}\,,$$

where X^A, X^{A^c}, τ are independent r.v.s, $\mathcal{L}(\tau)=\mathbf{B}(p)$, $p=\mathbb{P}(X\in A)$,

$$\mathcal{L}(X_A) = \mathcal{L}(X|X\in A),\, \mathcal{L}(X_{A^c}) = \mathcal{L}(X|X\in A^c)$$

([115], ch. 2). One may choose a bounded set A and take $a=\mathbb{E}X_A$ in (5.21). Denote

$$P = \mathcal{L}(X), \quad P_a = \mathcal{L}(X+a), \quad V=\mathcal{L}(X_A).$$

Theorem 5.5. [48] For any $n\in\mathbb{N}$ and any k-dimensional distribution P there exists constant $C_k(P_a)$ such that

$$\inf_a d_K(\mathcal{L}(S_n+na); \exp\left(n(P_a-I) - n^2(V-I)^{*2}/2\right)) \le C_k(P_a)n^{-1/2}\,. \tag{5.21}$$

Further results concerning asymptotic expansions can be found in [152, 48].

Symmetric random vectors.

Let $\{X_i\}$ be i.i.d. symmetric random vectors taking values in $\mathbb{R}^k$.

Multivariate analogues of (4.33) and (4.37) for independent random vectors have been established by Zaitsev [198, 200]:

$$d_K(S_n,\tilde S_n) \le C_k n^{-1/2}\,. \tag{5.22}$$

If $\mathcal{L}(X)$ has a non-negative characteristic function, then

$$d_K(S_n,\tilde S_n) \le C_k n^{-1}\,. \tag{5.23}$$

Infinite-dimensional versions of (5.12), (5.22), (5.23) can be found in Götze & Zaitsev [97].

Čekanavičius [44] investigated the case of mixtures of distributions with a dominant symmetric part.

Let P be a *symmetric* distribution, and let V be the distribution of an arbitrary k-dimensional random vector. Consider the situation where

$$\mathcal{L}(X) = (1-p)\sum_{j=1}^{s} q_j P^{*j} + pV$$

for some p, q_i, $s\in\mathbb{N}$ such that $0<p<1/2$, $q_i\in[0,1]$ $(i\ge 1)$, $q_1+\cdots+q_s=1$. Denote

$$H_n = \left((1-p)\sum_{j=1}^{s} q_j P^{*j} + pV\right)^{*n},$$

$$D_n = \exp\left(n\sum_{j=1}^{s} q_j(P-I) + np(V-I) - \frac{np^2}{2}(V-I)^{*2}\right),$$

Čekanavičius [44] has shown that for any $s\in\mathbb{N}$ there exists an absolute constant $C(s,k)$ such that

$$d_K(H_n;D_n) \le C(s,k)\left(p^{1/2}n^{-1/4} + s^3 n^{-1/2}\Big(\sum_{j=1}^{s} jq_j\Big)^{-1/2}\right). \tag{5.24}$$

If $s=1$, then the RHS of (5.24) is $O(p^{1/2}n^{-1/4}+n^{-1/2})$.

Suppose now that

$$X \stackrel{d}{=} \sum_{j=0}^{\xi} \eta_j ,$$

where r.v. ξ takes values in $\mathbb{N}$, $0 \le \xi \le s \in \mathbb{N}$, $\eta_0 = 0$, $\eta, \eta_1, \eta_2 \ldots$ are i.i.d. random vectors with a *non-negative* characteristic function, $\{\eta_j\}$ and $\xi \ge 0$ are independent. Set $\mu = \mathbb{E}\xi$. It is shown in [44] that

$$d_K(\mathcal{L}(S_n); \mathbf{\Pi}(n\mu, \eta)) \le C_k s^3 (n\mu)^{-1} . \tag{5.25}$$

Estimate (5.25) demonstrates that additional information about $\mathcal{L}(X)$ helps improvinging the accuracy of compound Poisson approximation. For example, let $\mathcal{L}(X) = 0.2I + 0.3P + 0.5P^{*5}$. Then $\mu = \mathbb{E}\xi = 2.8$. It follows from (5.25) that

$$\sup_{P \in \mathcal{F}_+(k)} d_K(\mathcal{L}(S_n); \exp(2.8n(P-I))) \le C_k n^{-1} .$$

Here $\mathcal{F}_+(k)$ denotes the class of k-dimensional distributions with non-negative characteristic functions.

Symmetric integer-valued random vectors.

Similarly to (5.16) we denote

$$\bar{0} = (0, \ldots, 0), \ \ \bar{e}_j = (0, \ldots, 1, \ldots, 0) \quad (1 \le j \le k),$$

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

Let $\{X_i\}$ be independent integer-valued random vectors with distributions concentrated on coordinate axes of $\mathbb{R}^k$:

$$\sum_{r=1}^{k} \sum_{m=-\infty}^{\infty} \mathbb{P}(X_j = m\bar{e}_r) = 1 \quad (\forall j).$$

Set $(r=1, \ldots, k, \ j=1, \ldots, n)$

$$p_{j,r} = \mathbb{P}(X_j \in \bar{e}_r \mathbb{Z} \setminus \{\bar{0}\}), \quad p_j = \sum_{r=1}^{k} p_{j,r} \, , \quad p_{j,0} = 1 - p_j \, .$$

Denote

$$F_r\{m\bar{e}_r\} = \mathbb{P}(X_j = m\bar{e}_r)/p_{j,r} \qquad (m \in \mathbb{Z} \setminus \{0\}).$$

We assume that F_r does not depend on j. Such distribution F_r always exist in the case of identically distributed random vectors (but not in the general case). Then

$$\mathcal{L}(S_n) = \prod_{j=1}^{*n} \left((1-p_j)I + \sum_{r=1}^{k} p_{j,r} F_r \right), \quad \mathcal{L}(\tilde{S}_n) = \exp\left(\sum_{j=1}^{n} \sum_{r=1}^{k} p_{j,r}(F_r - I) \right).$$

Let σ_r^2 denote the variance of F_r, and let

$$g(z) = 2\mathrm{e}^z(\mathrm{e}^{-z}-1-z)/z^2\,, \quad \lambda_{n,r} = \sum_{j=1}^n p_{j,r}\,,$$

$$\alpha_0 = \sum_{i=1}^n g(2(1-p_{i,0}))\min\Big\{2^{-3/2}\sum_{r=1}^k p_{i,r}^2/\lambda_{n,r}; p_i^2\Big\}.$$

Theorem 5.6. [126] *Suppose that* F_r *is a symmetric distribution,* $\sigma_r^2 < \infty$ $(r=1,\dots,k)$, $2\alpha_0\mathrm{e}<1$. *Then*

$$d_{TV}(S_n;\tilde{S}_n) \le \frac{8}{(1-2\alpha_0\mathrm{e})^{3/2}}\sum_{i=1}^k(1+\sigma_i)\sum_{r=1}^k \lambda_{0,r}^{-2}\sum_{j=1}^n p_{j,r}^2\,. \tag{5.26}$$

If $\{\sigma_r, p_{j,r}\}$ are bounded away from 0, then the RHS of (5.26) is $O(n^{-1})$, i.e., the accuracy is comparable with that of (4.37).

If $k=1$, σ_1^2 is fixed and $p_{j,1}\equiv p$ $(\forall j)$, then (5.26) is comparable to (4.27).

Infinite-dimensional spaces.

Very few results are known for a sum $S_n = X_1+\ldots+X_n$ of random elements $X_1,\ldots,X_n$ taking values in a general measurable space.

Bakštys & Paulauskas [7, 8] dealt with random elements $X_1,\ldots,X_n$ taking values in a separable Banach space $\mathcal{B}$.

Denote by $\tilde{X}_a, \tilde{X}_{1,a},\ldots,\tilde{X}_{n,a}$ accompanying $X{+}a, X_1{+}a,\ldots,X_n{+}a$ independent random elements, and let

$$\tilde{S}_{n,a} = \tilde{X}_{1,a}+\ldots+\tilde{X}_{n,a}\,, \quad \tilde{S}_n = \tilde{X}_{1,0}+\ldots+\tilde{X}_{n,0}\,.$$

Let $\mathcal{U}$ be the set of all convex Borel sets. Suppose that for any $\epsilon>0$ there exists a finite-dimensional subspace $\mathcal{V}_\epsilon$ such that $\mathbb{P}(X\in\mathcal{V}_\epsilon)\le\epsilon$. Then [8]

$$\lim_{n\to\infty}\inf_a\sup_{A\in\mathcal{U}}|\mathbb{P}(S_n\in A)-\mathbb{P}(\tilde{S}_{n,a}-na\in A)| = 0. \tag{5.27}$$

If X is symmetric random element taking values in a Hilbert space, then

$$\lim_{n\to\infty}\inf_a\sup_{A\in\mathcal{V}}|\mathbb{P}(S_n\in A)-\mathbb{P}(\tilde{S}_{n,a}-na\in A)| = 0, \tag{5.28}$$

where $\mathcal{V}$ is a set of all open balls in that Hilbert space (Bakštys [9]).

Let $X, X', X_1,\ldots,X_n$ be a sequence of i.i.d. random elements taking values in a real separable Hilbert space H with scalar product $(\cdot,\cdot)$ and norm $\|\cdot\|_H$. Nagaev [142] has derived estimates of the accuracy of approximation $\mathcal{L}(S_n)\approx\mathcal{L}(\tilde{S}_n)$.

Namely, for any $x\in H$ and constant $C>0$ set

$$B(x;C) = \mathbb{E}^{1/2}(X-X',x)^2\,\mathbb{1}\{\|X\|_H\wedge\|X'\|_H\le C\}, \quad B(x) = B(x,\infty).$$

Let $\sigma_1^2(C)\ge\sigma_2^2(C)\ge\ldots$ be the eigenvalues of the quadratic form $\{B(\cdot;C)\times B(\cdot;C)\}$.

Denote $V(a;r)=\{x\in H: \|x-a\|_H\le r\}$, and let

$$\sigma^2(C)=\prod_{j=1}^{\infty}\sigma_j^2(C),\ \ \Lambda_l(C)=\prod_{j=1}^{l}\sigma_j^2(C),\ \ \sigma^2=\mathbb{E}\|X\|_H^2 .$$

Theorem 5.7. [142] If $\mathbb{E}X=0$ and $\sigma^2<\infty$, then for any $a\in H$ and $C\in(0;\infty)$

$$\sup_r \|\mathbb{P}(S_n\in V(a;r))-\mathbb{P}(\tilde S_n\in V(a;r))\| \le \frac{\sigma^4+B^2(a)/n}{\Lambda_5^{2/5}(C)\sqrt{n}}+\frac{C\sigma(C)}{\Lambda_3^{1/3}(C)\sqrt{n}} . \quad (5.29)$$

If $0<\sigma(C)<\infty$, then the RHS of (5.29) is $O(n^{-1/2})$.
Open problem.
5.2. Evaluate constant C_k in (5.21).

6. Compound Poisson process approximation

Let r.v.s $X_1,X_2,...$ represent rare events (i.e., $\{X_j\}$ are non-zero with "small" probability). Poisson process approximation to corresponding empirical point processes of exceedances has been studied by many authors (see, e.g., [31, 144] and references therein). However, Poisson process approximation is applicable only if the limiting cluster size distribution is degenerate.

If the limiting cluster size distribution is not degenerate, then the limiting distribution of the number of exceedances is typically compound Poisson; the limiting distribution of an empirical point processes of exceedances can be more complex than compound Poisson (cf. [144], ch. 8).

This section presents results concerning compound Poisson process approximation.

Compound Poisson process is a process with independent compound Poisson increments. Namely, a point process $S(\cdot)$ is called a compound Poisson process with intensity measure Q and multiplicity distribution $\mathcal{L}(\zeta)$ if it has independent increments (i.e., for arbitrary disjoint measurable sets $A_1,...,A_k$ r.v.s $S(A_1),...,S(A_k)$ are independent) and for any measurable set A random variable $S(A)$ is compound Poisson $\mathbf{\Pi}(Q(A),\mathcal{L}(\zeta))$.

If $Q=\lambda m$, where m is the Lebesgue measure, then we say that $S(\cdot)$ is a compound Poisson process with intensity rate λ and compounding (multiplicity) distribution $\mathcal{L}(\zeta)$.

6.1. Empirical processes

Let r.v.s $X,X_1,X_2,...,X_n$ represent rare events (i.e., they are non-zero with "small" probability). Define the point process

$$S_n(\cdot)=\sum_{i=1}^{n}X_i\,\mathbb{I}\{i/n\in\cdot\}. \quad (6.1)$$

A particular case is the jump process

$$S_{n,t} = \sum_{i=1}^{[nt]} X_i \qquad (0<t\le 1)$$

known also as a random broken line. Note that $S_{n,t} = S_n((0;t])$,

$$S_n \equiv X_1 + ... + X_n = S_n((0;1]).$$

If $\{X_i\}$ are Bernoulli r.v.s, then point process (6.1) counts *locations* of rare events.

Example 6.1. A typical example is a process of exceedances of a "high" threshold. Let $\{\xi_i, i\ge 1\}$ be a stationary sequence of random variables, and let $\{u_n\}$ be a sequence of levels. Set $X_i = \mathbb{I}\{\xi_i > u_n\}$. Process $S_n(\cdot) = N_n(\cdot, u_n)$, where

$$N_n(B, u_n) = \sum_{i=1}^{n} X_i \mathbb{I}\{\, i/n \in B\} \qquad (B \subset (0;1]), \tag{6.2}$$

counts locations of exceedances of level u_n. □

A natural approximation to $S_n(\cdot)$ is a compound Poisson process.

Let $\{X, X_1, ..., X_n\} \equiv \{X_{n,0}, X_{n,1}, ..., X_{n,n}\}$, $n \ge 1$, be a triangle array of dependent r.v.s, strictly stationary in each row. In applications r.v.s $\{X_i\}$ are typically non-negative; they usually represent rare events. Therefore, we assume that $X_i \ge 0$ $(\forall i)$.

If a sequence $\{r = r_n\}$ of natural numbers obeys $n \gg r_n \ge 1$, we denote by $\zeta_r \equiv \zeta_{r,n}$ a r.v. with the distribution

$$\mathcal{L}(\zeta_{r,n}) = \mathcal{L}(S_r \,|\, S_r > 0). \tag{6.3}$$

The next theorem presents necessary and sufficient conditions for the weak convergence of $S_n(\cdot)$ to a compound Poisson point process. It is essentially Theorem 7.1 from [144].

Theorem 6.1. *Assume mixing condition Δ, and suppose that (2.13) holds. If, as $n \to \infty$,*

$$\mathbb{P}(S_n = 0) \to e^{-\lambda} \qquad (\exists \lambda > 0), \tag{2.11*}$$

$$\mathcal{L}(S_r | S_r \neq 0) \Rightarrow \mathcal{L}(\zeta) \tag{2.14*}$$

for a sequence $\{r = r_n\}$ obeying (2.10), then

$$S_n(\cdot) \Rightarrow S(\cdot), \tag{6.4}$$

where $S(\cdot)$ is a compound Poisson point process with intensity rate λ and multiplicity distribution $\mathcal{L}(\zeta)$.

If $S_n(\cdot)$ converges weakly to a point process $S(\cdot)$, then $S(\cdot)$ is a compound Poisson process on $(0;1]$ with intensity rate λ given by (2.11). If $\lambda > 0$, then (2.14*) is valid for some r.v. ζ and sequence $\{r_n\}$ that obeys (2.10).*

Condition (2.13) can be relaxed to allow for $np \to \infty$ at a certain "slow" rate.

Example 6.2. Let $\{X_i, i \geq 0\}$ be a *regenerative* process, i.e., there exist integer-valued r.v.s $0 < \xi_0 < \xi_1 < \ldots$ such that the "cycles"

$$\{X_i,\, 0 \leq i < \xi_0\},\ \{X_i,\, \xi_0 \leq i < \xi_1\}, \ldots$$

are i.i.d.. We define r.v.s $Y, Y_1, Y_2, \ldots$ as follows:

$$Y = \max_{0 \leq i < \xi_0} X_i\,,\ Y_1 = \max_{\xi_0 \leq i < \xi_1} X_i\,, \ldots$$

Denote

$$T_j = \sum_{i=\xi_{j-1}}^{\xi_j - 1} \mathbb{I}\{X_i > u_n\} \qquad (j \in \mathbb{N}),$$

where $\{u_n\}$ is a sequence of levels. Suppose that ξ_0 is aperiodic, $\mu := \mathbb{E}\xi_0 < \infty$ and $\mathbb{P}(Y > \max_{1 \leq j \leq k} Y_i) \to 0$ as $k \to \infty$.

Process $N_n(\cdot, u_n)$ converges weakly to a non-degenerate point process N if and only if there exist $\lambda > 0$ and a distribution P such that

$$n\mathbb{P}(Y > u_n)/\mu \to \lambda \text{ and } \mathcal{L}(T_1 | T_1 > 0) \Rightarrow P$$

as $n \to \infty$; necessarily N a compound Poisson point process with intensity rate λ and multiplicity distribution P (Rootzén [171]). □

Concerning random broken line $\{S_{n,t}, 0 \leq t \leq 1\}$, Borisov & Borovkov [27] use a Poisson component in order to improve the rate of approximation in the Donsker-Prokhorov invariance principle.

6.2. Excess process

Let $X, X_1, X_2, \ldots, X_n$ be a stationary sequence of r.v.s. When one is interested in the joint distribution of exceedances of several levels among $X_1, \ldots, X_n$, a natural tool is the excess process $N_n^\varepsilon(\cdot)$. This section presents necessary and sufficient conditions for the weak convergence of the excess process to a compound Poisson process.

Given a sequence $\{u_n(\cdot), n \geq 1\}$ of monotone functions on $[0; \infty)$, denote

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

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

Process $N_n^\varepsilon(\cdot)$ describes variability in the *heights* of observations $X_1, X_2, \ldots, X_n$.

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

$$N_n^\varepsilon(t) = \sum_{i=1}^{n} \mathbb{I}\{Y_{n,i} < t\}. \tag{6.5}$$

There is a considerable amount of research on the topic of tail empirical processes (see, e.g., [72, 133] and references therein).

We present necessary and sufficient conditions for the weak convergence of the excess process to a compound Poisson process in Theorem 6.2 below.

Suppose that function $u_n(\cdot)$ is strictly decreasing for all large enough n, $u_n(0)=\infty$,

$$\limsup_{n\to\infty} n\mathbb{P}(X>u_n(t))<\infty \qquad (0<t<\infty), \tag{6.6}$$

$$\lim_{n\to\infty} \mathbb{P}(N_n(u_n(t))=0)=\mathrm{e}^{-t} \qquad (t>0). \tag{6.7}$$

Condition (6.7) means $u_n(\cdot)$ is a "proper" normalising sequence.

Given $t_0=0<t_1<...<t_k<\infty$, denote $\bar t=(t_1,...,t_k)$. Recall condition $C_{\bar t}$.

Definition. Condition (C) holds if condition $C_{\bar t}$ is valid for every $0<t_1<...<t_k<\infty$, $k\in\mathbb{N}$.

Let $\{\pi_s, s\geq 0\}$ be a Poisson process with intensity rate 1, and let $\zeta_1,\zeta_2,...$ be a sequence of i.i.d. copies of ζ. Denote

$$Q_\zeta(t)=\sum_{j=1}^{\pi_t}\zeta_j\,. \tag{6.8}$$

Then $\{Q_\zeta(t), t\geq 0\}$ is a compound Poisson jump process. Equivalently,

$$\tilde Q_\zeta(B):=\int_B Q_\zeta(dt)$$

is a compound Poisson point process with the Lebesgue intensity measure and multiplicity distribution $\mathcal{L}(\zeta)$. We do not distinguish between Q_ζ and $\tilde Q_\zeta$ in the sequel.

Theorem 6.2. *Assume mixing condition condition Δ, (6.6), (6.7), and let $\pi_\zeta(\cdot)$ denote a compound Poisson process with intensity rate 1 and multiplicity distribution $\mathcal{L}(\zeta)$. Then*

$$N_n^\varepsilon(\cdot)\Rightarrow Q_\zeta(\cdot) \tag{6.9}$$

as $n\to\infty$ if and only if condition (C) holds.

Theorem 6.2 is Theorem 7.2 from [144].

General situation.

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

$$\{N_n^\varepsilon(t),\, t\leq T\}\Rightarrow\left\{\sum_{j=1}^{\pi_T}\gamma_j(t/T),\, t\leq T\right\} \tag{6.10}$$

as $n\to\infty$, where π_T is a Poisson r.v., $\{\gamma_j(\cdot)\}$ are independent jump processes.

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

Results concerning approximation (6.10) can be found in [144], ch. 8. The accuracy of approximation to the distribution of an excess process can be evaluated in terms of the total variation distance (cf. [144], Theorem 8.3).

6.3. General point processes of exceedances

Both (6.2) and (6.5) are one–dimensional processes of exceedances. Below we deal with a general point process of exceedances N_n^*, which counts locations of extremes (rare events) as well as their heights.

For any Borel set $A \subset (0;1] \times [0;\infty)$ denote

$$N_n^*(A) := \sum_{i=1}^{n} \mathbb{I}\{\, \big(i/n, u_n^{-1}(X_i)\big) \in A \,\}. \tag{6.11}$$

If $\{X_i\}$ are i.i.d.r.v.s, or if $\{X_i, i \geq 1\}$ is a strictly stationary sequence obeying certain mixing conditions, then $N_n^*(\cdot)$ converges weakly to a pure Poisson point process (Adler [1], see also [147]).

The following theorem presents necessary and sufficient conditions for the weak convergence of point process $N_n^*(\cdot)$ to a compound Poisson point process.

Denote by $N^*(\cdot)$ a compound Poisson point process on $(0;1] \times [0;\infty)$ with the Lebesgue intensity measure and multiplicity distribution $\mathcal{L}(\zeta)$. Note that

$$Q_\zeta(t) \stackrel{d}{=} N^*((0;1] \times [0;t)).$$

Theorem 6.3. *Assume conditions* Δ, *(6.6), (6.7). Then*

$$N_n^* \Rightarrow N^* \qquad (n \to \infty) \tag{6.12}$$

if and only if condition (C) *holds.*

Theorem 6.3 is Theorem 7.4 from [144].

Example 6.3. Let $\{\xi_i\}, \{\alpha_i\}$ be independent sequences of i.i.d. r.v.s, $\mathbb{P}(\xi_i \leq x) = F(x)$ and $\alpha_i \in \mathbf{B}(\theta)$, where $\theta \in (0;1)$. Put $X_1 = \xi_1$, and let

$$X_i = \alpha_i \xi_i + (1-\alpha_i) X_{i-1} \qquad\qquad (i \geq 2). \tag{6.13}$$

Then $\{X_i, i \geq 1\}$ is a stationary sequence of r.v.s with the marginal d.f. F, the cluster sizes have the geometric distribution with mean $1/\theta$, and the extremal index equals θ.

Notice that sequence $\{X_i, i \geq 1\}$ is φ–mixing and

$$\varphi(k) \leq (1-\theta)^k \qquad\qquad (k \geq 1).$$

Furthermore,

$$\mathbb{P}(\max_{i \leq n} X_i \leq u) = F(u)\mathbb{E}(1-p)^\nu = F(u)(1-\theta p)^{n-1},$$

where $\nu = \sum_{i=2}^{n} \alpha_i$ is a Binomial $\mathbf{B}(n-1, \theta)$ r.v..

Denote $K^* = \sup\{x\colon F(x) < 1\}$, and assume that

$$\mathbb{P}(X \geq x)/\mathbb{P}(X > x) \to 1 \tag{6.14}$$

as $x \to K^*$ (Gnedenko's condition [90]). Then there exists a sequence $\{u_n\}$ such that $n\mathbb{P}(X > u_n) \to 1$ (cf. Theorem 1.7.13 in [127]). Put

$$u_n(t) = u_{[\theta n/t]} \qquad (t > 0).$$

Then

$$\mathbb{P}(X > u_n(t)) \sim t/n\theta, \ \mathbb{P}(N_r(u_n(t)) > 0) \sim tr/n, \tag{6.15}$$

and $\{u_n(\cdot)\}$ obeys (6.7).

In order to check condition (C), we need to check items (a), (b) of condition $(C_{\bar{t}})$. Let $0 < s < t < v < \infty$. Condition (b) follows from (6.15) and estimate

$$\begin{aligned}&\mathbb{P}(N_r[u_n(t); u_n(v)) > 0, N_r[u_n(s); u_n(t)) > 0)\\ &\leq r^2 \mathbb{P}(u_n(v) < \xi \leq u_n(t)) \mathbb{P}(u_n(t) < \xi \leq u_n(s)) = O((r/n)^2).\end{aligned}$$

Random variables $\{X_i, ..., X_{i+m}\}$ form a cluster of size m if $\alpha_i = 1, \alpha_{i+1} = ... = \alpha_{i+m-1} = 0, \alpha_{i+m} = 1$. Denote

$$W = \mathbb{1}_1 + \sum_{i=2}^{r} \alpha_i \mathbb{1}_i\,,$$

where $\mathbb{1}_i = \mathbb{1}\{\xi_i \in (u_n(t); u_n(s)]\}$. Asymptotically, only one cluster among $X_1, ..., X_r$ may hit $(u_n(t); u_n(s)]$. Therefore,

$$\begin{aligned}&\mathbb{P}(N_r[u_n(s); u_n(t)) = j) \sim \mathbb{P}(N_r[u_n(s); u_n(t)) = j, N_r(u_n(s)) = 0)\\ &= \mathbb{P}(N_r[u_n(s); u_n(t)) = j, N_r(u_n(s)) = 0, W = 1) + O((r/n)^2)\\ &\sim r\theta^2 (1-\theta)^{j-1} \mathbb{P}(u_n(t) < \xi \leq u_n(s)) \sim (t-s)\mathbb{P}(\zeta = j) r\theta/n,\end{aligned} \tag{6.16}$$

where $\mathcal{L}(\zeta) = \mathbf{\Gamma}(1-\theta)$. Thus, condition (a) holds, and Theorem 6.3 entails

$$N_n^* \Rightarrow N^*$$

as $n \to \infty$, where N^* is a compound Poisson point process with the Lebesgue intensity measure and multiplicity distribution $\mathbf{\Gamma}(1-\theta)$. □

Results concerning weak convergence of point process $N_n^*(\cdot)$ to a Poisson cluster process can be found in [144], ch. 8. An estimate of the accuracy of approximation to $\mathcal{L}(N_n^*(\cdot))$ in terms of a d_G-type distance has been established in [20].

Open problem.

6.1. Improve the estimate of the accuracy of approximation $N_n^* \approx N^*$ presented in [20].

7. Kolmogorov's problem

Let $\{X_1, \ldots, X_n\}_{n\geq 1}$ be independent infinitesimal random variables, $S_n = X_1 + ... + X_n$. It is well-known [116] that if the limiting distribution of S_n exists, then it is infinitely divisible.

The notion of the infinitely divisible distribution was introduced by de Finetti in 1925. A well-known result due to Khintchine [117] states that the class $\mathcal{D}$ of infinitely divisible distributions coincides with the class of weak limits of compound Poisson distributions. Thus, the topics of compound Poisson and infinitely divisible approximations are closely related.

This section is devoted to Kolmogorov's problem.

7.1. Kolmogorov's first problem

In early 1950s Kolmogorov has raised the problem of evaluating the accuracy of infinitely divisible approximation to $\mathcal{L}(S_n)$.

Kolmogorov's first problem is concerned with i.i.d.r.v.s, while Kolmogorov's second problem deals with independent but not necessarily identically distributed random variables. The problem is called "uniform" since the estimate of the accuracy of approximation established by Kolmogorov is *uniform* over the class $\mathcal{F}$ of all probability distributions.

Prokhorov [155, 157] (see also [160]) has proved that for any distribution $\mathcal{L}(X)$ there exists a sequence of infinitely divisible distributions that are "close" to $\mathcal{L}(S_n)$, hence

$$d_K(\mathcal{L}(S_n);\mathcal{D}) \equiv \inf_{P\in\mathcal{D}} d_K(\mathcal{L}(S_n);P) \to 0 \qquad (n\to\infty). \tag{7.1}$$

If $\mathcal{L}(X)$ has an absolute continuous component or is a discrete distribution, then d_K in (7.1) can be replaced with d_{TV}.

Kolmogorov [120] has derived an estimate of the accuracy of approximation that is uniform over $\mathcal{F}$ (the so-called first Kolmogorov's theorem): there exists an absolute constant C such that

$$\sup_{\mathcal{L}(X)\in\mathcal{F}} d_K(\mathcal{L}(S_n);\mathcal{D}) \leq Cn^{-1/3}\,. \tag{7.2}$$

Observe the extreme generality of estimate (7.2) — there are no moment or structural assumptions.

Many authors worked on deriving upper and lower bounds to $d_K(\mathcal{L}(S_n);\mathcal{D})$ (see, e.g., [6, 135, 205] and references therein). It took over 25 years of research by various mathematicians before the correct rate of the accuracy of approximation in (7.2) has been established by Arak [4, 5] (a comprehensive history of the problem can be found in the monograph by Arak & Zaitsev [6]).

Arak's [4, 5] theorem states that there exist absolute constants $0<C_1<C_2<\infty$ such that

$$C_1 n^{-2/3} \leq \sup_{F\in\mathcal{F}} d_K(\mathcal{L}(S_n);\mathcal{D}) \leq C_2 n^{-2/3}\,. \tag{7.3}$$

The lower bound in (7.3) sets a limit to the rate of the accuracy of compound Poisson approximation (as well as a limit to the rate of the accuracy of approximation by any other infinitely divisible distribution).

There is no multidimensional analogue of Arak's result (7.3).

Arak has proved that the rate $n^{-2/3}$ in (7.3) can be achieved using shifted compound Poisson approximation. Therefore, (7.3) demonstrates universality of compound Poisson approximation.

The main drawback of (7.3) is that only existence of an approximating compound Poisson distribution has been established.

Relations (4.2), (4.25), (4.33), (4.37) can be viewed as solutions of Kolmogorov's first problem for special classes of distributions.

Zaitsev [202] has shown that d_K in (7.3) cannot in general be replaced by the total variation distance.

Let $X, X_1, \ldots, X_n$ be i.i.d. integer-valued r.v.s. Studnev [178] reports that Gusak has shown that

$$d_K(\mathcal{L}(S_n); \mathcal{D}) = O(n^{-1}). \tag{7.4}$$

We are not aware if Gusak's result (7.4) has been published.

Zaitsev [204] has conjectured that for any distribution $\mathcal{L}(X)$ there exist a constant C_X such that

$$\inf_{a\in\mathbb{R}} d_K(\mathcal{L}(S_n{+}na); \mathbf{\Pi}(n, X{+}a)) \leq C_X n^{-1/2}\,. \tag{7.5}$$

It is shown in [6] that

$$Cn^{-1} \leq \sup_{\mathcal{L}(X)\in\mathcal{F}_+} d_K(\mathcal{L}(S_n); \mathcal{D}) \leq \sup_{\mathcal{L}(X)\in\mathcal{F}_+} d_K(\mathcal{L}(S_n); \mathcal{L}(\tilde{S}_n)).$$

If (7.5) is true, then the accompanying compound Poisson distribution ensures the best possible rate of infinitely divisible approximation in the class $\mathcal{F}_+$ of distributions with non-negative characteristic functions.

Open problems.

7.1 Derive a multidimensional analogue of Arak's inequality (7.3).

7.2. Is it true that

$$\sup_{\mathcal{L}(X)\in\mathcal{F}_s} d_K(\mathcal{L}(S_n); \mathcal{D}) \leq Cn^{-1}\,, \tag{7.6}$$

where C is an absolute constant?

7.2. Kolmogorov's second problem

Kolmogorov's second problem deals with independent but not necessarily identically distributed random variables. In general, the problem has no solution.

Let $X_1, \ldots, X_n$ be independent random variables. Denote by d_L the Lévy distance, and let

$$d_L(P; \mathcal{D}) = \inf_{D\in\mathcal{D}} d_L(P; D).$$

Recall that by (3.10)

$$\mathcal{L}(X_i) = (1-p_i)U_i + p_iV_i \qquad (0 \le p_i \le 1),$$

where distribution U_i may be chosen concentrated on a finite interval of length say T. Set

$$a_i = \int_{\mathbb{R}} xU_i(\mathrm{d}x), \quad p_n^* = \max\{p_1, \ldots, p_n\}.$$

Denote by $\tilde{X}_{1,a_1}, ..., \tilde{X}_{n,a_n}$ accompanying $X_1+a_1, ..., X_n+a_n$ independent random variables. Let

$$\tilde{S} = \tilde{X}_{1,a_1} + ... + \tilde{X}_{n,a_n} .$$

If $a_i = 0$ $(\forall i)$, then $\tilde{S} = \tilde{S}_n$, cf. (1.9).

According to Zaitsev & Arak [196] (see also [6]), there exists an absolute constant C such that

$$d_L(\mathcal{L}(S_n); \mathcal{D}) \le d_L(S_n; \tilde{S}) \le C(p_n^* + T\ln(1/T)). \tag{7.7}$$

Zaitsev & Arak [196] have proved that estimate (7.7) is of correct order. A multivariate version of this result has been derived by Zaitsev [199].

The following result is Theorem 4 from Arak & Zaitsev [6], p. 5.

Theorem 7.1. *If $\varepsilon > 0$ and $d_L(\mathcal{L}(X_i); I_{\beta_i}) \le \varepsilon$ for some β_i $(i = 1, \ldots, n)$, then there exist $a_1, a_2, \ldots, a_n$ and an absolute constant $0 << C < \infty$ such that*

$$d_L(S_n; \tilde{S}) \le C\varepsilon(|\ln\varepsilon| + 1).$$

For any $\delta \in (0,1]$ there exist i.i.d. random variables $X, X_1, ..., X_n$ and $n \in \mathbb{N}$ such that $d_L(\mathcal{L}(X); I) \le \delta$ and

$$d_L(S_n; \mathcal{D}) \ge c\delta(|\ln\delta| + 1),$$

where $c > 0$ is an absolute constant.

An infinite-dimensional version of (7.7) has been established by Götze & Zaitsev [98], see also [96].

Acknowledgments

The authors are grateful to the referee for helpful remarks.

References

[1] Adler R.J. (1978) Weak convergence results for extremal processes generated by dependent random variables. — *Ann. Probab.*, **6**(4), 660–667. MR0494408

[2] Arak T.V. (1980) On the approximation of n-fold convolutions of distributions having nonnegative characteristic functions by infinitely divisible distributions. — *Trudy Tallin. Politekhn. Inst.*, **482**, 81–85 (Russian). MR0572558

[3] Arak T.V. (1980) On the approximation by the accompanying laws of n-fold convolutions of distributions with non-negative characteristic functions. — *Teor. Veroyatn. Primen.*, **25**(2), 225–246 (Russian). Transl.: *Theory Probab. Appl.*, **25**(2), 221–243. MR0572558

[4] Arak T.V. (1981) On the convergence rate in Kolmogorov's uniform limit theorem I, II. — *Teor. Veroyatn. Primen.*, **26**(2), 225–245; (3), 449–463 (Russian). Transl.: *Theory Probab. Appl.*, **26**(2), 219–239; (3), 437–451. MR0627854

[5] Arak T.V. (1982) An improvement of the lower bound for the rate of convergence in Kolmogorov's uniform limit theorem. — Teor. Veroyatn. Primen., v. 27, 767–772. MR0681467

[6] Arak T.V. and Zaitsev A.Yu. (1986) Uniform limit theorems for sums of independent random variables. *Trudy Mat. Inst. Steklov (Monograph)*, **174** (Russian). Transl.: *Proc. Steklov Inst. Math.*, **174**, 3–214. MR0871856

[7] Bakstys G. and Paulauskas V.J. (1985) Approximation of distributions of sums of Banach-valued random elements by infinitely divisible laws I. — *Litovsk. Mat. Sb.*, **26**(3) 403–414 (Russian). Transl.: *Lith. Math. J.*, **26**(3), 207-215. MR0867223

[8] Bakstys G. and Paulauskas V.J. (1987) Approximation of distributions of sums of Banach-valued random elements by infinitely divisible laws II. — *Litovsk. Mat. Sb.*, **27**(2) 224–235 (Russian). Transl.: *Lith. Math. J.*, **27**(2), 106–113. MR0916798

[9] Bakstys G. (1989) Approximation with accompanying laws. — *Litovsk. Mat. Sb.*, **29**(3), 423–428 (Russian). Transl.: *Lith. Math. J.*, **29**(3), 211–215. MR1045548

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

[11] Barbour A.D. (1987) Asymptotic expansions in the Poisson limit theorem. — *Ann. Probab.*, **15**(2), 748–766. MR0885141

[12] Barbour A.D. and Jensen J.L. (1989) Local and tail approximations near the Poisson limit. — *Scand. J. Statist.*, **16**(1), 75–87. MR1003970

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

[14] Barbour A.D., Chen L.H.Y. and Loh W.-L. (1992) Compound Poisson approximation for non-negative random variables via Stein's method. — *Ann. Probab.*, **20**(4), 1843–1866. MR1188044

[15] Barbour A.D. and Utev S.A. (1999) Compound Poisson approximation in total variation. — *Stochastic Process. Appl.*, **82**(1), 89–125. MR1695071

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

[17] Barbour A.D., Chryssaphinou O. (2001) Compound Poisson approximation: a user's guide. — *Appl. Probab.*, **11**(3), 964–1002.

[18] Barbour A.D., Chryssaphinou O. and Vaggelatou E. (2001) Applications of compound Poisson approximation. — In: Charalambides C.A., Koutras M.V. and Balakrishnan N. (Eds.) *Probability and Statistical Models with Applications*, 41—62. Chapman & Hall.

[19] Barbour A.D. and Čekanavičius V. (2002) Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.*, **30**(2), 509–545.

[20] Barbour A.D., Novak S.Y. and Xia A. (2002) Compound Poisson approximation for the distribution of extremes. *Adv. Appl. Probab.*, **34**(1), 223–240. MR1865030

[21] Barbour A.D. and Lindvall T. (2005) Translated Poisson approximation for Markov chains. — *J. Theoret. Probab.*, **19**(3), 609–630. MR2280512

[22] Barbour A.D., Čekanavičius V. and Xia A. (2007) On Stein's method and perturbations. — *ALEA, Lat. Am. J. Probab. Math. Stat.*, **3**, 31–53. MR2324747

[23] Barbour A.D., Gan H.L. and Xia A. (2015) Stein factors for negative Binomial approximation in Wasserstein distance. — *Bernoulli*, **21**(2), 1002–1013.

[24] Barbour A. D., Johnson O., Kontoyiannis I. and Madiman M. (2010) Compound Poisson approximation via information functionals. — *Electron. J. Probab.*, **15**, 1344–1369. MR3338654

[25] Bentkus V., Götze F. and Zaitsev A.Yu. (1997) Approximation of quadratic forms of independent random vectors by accompanying laws. — Theory Probab. Appl., v. 42, No 2, 308–335.

[26] Bernstein S.N. (1926) Sur l'extensiori du theoreme limite du calcul des probabilites aux sommes de quantites dependantes. — *Math. Annalen*, **97**, 1–59. MR1474712

[27] Borisov I.S., Borovkov A.A. (1986) Second-order approximation in the Donsker-Prokhorov invariance principle. — Theory Probab. Appl., v. 31, No 2, 225–245.

[28] Borisov I.S. (1993) Strong Poisson and mixed approximations of sums of independent random variables in Banach spaces. — Siberian Adv.Math. v.3, no 2, 1–13. MR1233608

[29] Borisov I.S. (1996) Poisson approximation of the partial sum process in Banach spaces. — *Sibirsk. Mat. Zh.*, **37**(4), 723–731 (Russian). Transl.: *Sib. Math. J.*, **37**(4), 627–634. MR1643350

[30] Borisov I.S. and Ruzankin P.S. (2002) Poisson approximation for expectations of unbounded functions of independent random variables. — *Ann. Probab.*, **30**(4), 1657–1680. MR1944003

[31] Borisov I.S. (2003) Moment inequalities connected with accompanying

Poisson laws in Abelian groups. — Ind. J. MMS, v. 44, 2771–2786. MR2003788

[32] Borisov I.S. (2003) A note on Dobrushin's theorem and couplings in Poisson approximation in Abelian groups. — *Teor. Veroyatn. Primen.*, **48**(3), 576–583 (Russian). Transl.: *Theory Probab. Appl.*, **48**(3), 521–528. MR2141351

[33] Borovkov K.A. and Pfeifer D. (1996) On improvements of the order of approximation in the Poisson limit theorem. — *J. Appl. Probab.*, **96**(33), 146–155. MR1371962

[34] Borovkov K.A. and Novak S.Y. (2010) On limiting cluster size distributions for processes of exceedances for stationary sequences. — Statist. Probab. Letters, v. 80, 1814–1818. MR2734246

[35] Boutsikas M.V. and Koutras M.V. (2002) On the number of overflown urns and excess balls in an allocation model with limited urn capacity. — *J. Statist. Plan. Inf.*, **104**(2), 259–286. MR1906011

[36] Boutsikas M.V. and Vaggelatou E. (2016) A new method for obtaining sharp compound Poisson approximation error estimates for sums of locally dependent random variables. — *Bernoulli*, **16**(2), 301–330. MR2668903

[37] Brown T.C. and Xia A. (2001) Stein's method and birth-death processes. *Ann. Probab.*, **29**(3), 1373–1403 MR1872746

[38] Chen L.H.Y. and Roos M. (1995) Compound Poisson approximation for unbounded functions on a group with application to large deviations. — *Probab. Theory Rel. Fields*, **103**, 515–528. MR1360203

[39] Chen L.H.Y. and Röllin A. (2013) Approximating dependent rare events. — *Bernoulli*, **19**(4), 1243–1267. MR3102550

[40] Chryssaphinou O. and Vaggelatou E. (2001) Compound Poisson approximation for long increasing sequences. — *J. Appl. Probab.*, **38**(2), 449–463.

[41] Chryssaphinou O. and Vaggelatou E. (2002) Compound Poisson approximation for a multiple runs in a Markov chain. — *Ann. Inst. Stat. Math.*, **54**(2), 411–424. MR1910182

[42] Čekanavičius V. (1989) Approximation with accompanying distributions and asymptotic expansions I. — *Litovsk. Mat. Sb.*, **29**(1), 171–178 (Russian). Transl.: *Lith. Math. J.*, **29**(1), 75–80.

[43] Čekanavičius V. (1990) An approximation of integer-valued measures by generalized Poisson measures. — *In: Probabability Theory and Mathematical Statistics 2. Proceedings of the Fifth Vilnius Conference (1989)*, B.Grigelionis et al. (Eds.), 228–237. MR1153814

[44] Čekanavičius V. (1991) An approximation of mixtures of distributions. — *Litovsk. Mat. Sb.*, **31**(2), 351–368 (Russian). Transl.: *Lith. Math. J.*, **31**(2), 243–257. MR1161374

[45] Čekanavičius V. (1993) On the approximation by convolution of the generalized Poisson measure and the Gaussian distribution I. — *Lith. Math. J.*, **32**(3), 265–274.

[46] Čekanavičius V. (1993) Non-uniform theorems for discrete measures. — *Lith. Math. J.*, **33**(2), 114–126.

[47] Čekanavičius V. (1995) On asymptotic expansions in the first uniform

Kolmogorov theorem. — *Statist. Probab. Lett.*, **25**(2), 145–151.
[48] Čekanavičius V. (1996) On multivariate Le Cam theorem and compound Poisson measures. — *Statist. Probab. Lett.*, **28**(1), 33–39. MR1394416
[49] Čekanavičius V. (1997) Asymptotic expansions in the exponent: a compound Poisson approach. — *Adv. Appl. Probab.*, **29**(2), 374–387.
[50] Čekanavičius V. (1997) Approximation of the generalized Poisson Binomial distribution: asymptotic expansions. — *Lith. Math. J.*, **37**(1), 1–12.
[51] Čekanavičius V. (1997) Asymptotic expansions for compound Poisson measures. — *Lith. Math. J.*, **37**(4), 426–447.
[52] Čekanavičius V. and Vaitkus P. (1997) Large deviations for signed compound Poisson approximations. — *Statistics and Decisions*, **15**(4), 379–396. MR1450935
[53] Čekanavičius V. (1998) On signed normal-Poisson approximations. — *Probab. Theory Rel. Fields*, **111**, 563–583. MR1641834
[54] Čekanavičius V. (1998) Estimates in total variation for convolutions of compound distributions. — *J. Lond. Math. Soc.*, **58**(3), 748–760.
[55] Čekanavičius V. (1999) Remarks on infinitely divisible approximations to the Binomial law. — *In: Probability Theory and Mathematical Statistics. Proceedings of the 7'th Vilnius Conference*, B.Grigelionis et al. (Eds.), VSP/TEV, 135–146. MR1678161
[56] Čekanavičius V. and Mikalauskas M. (1999) Signed Poisson approximations for Markov chains. — *Stochastic Process. Appl.*, **82**(2), 205–227.
[57] Čekanavičius V. (1999) On compound Poisson approximations under moment restrictions. — *Teor. Veroyatn. Primen.*, **44**(1), 74–86 (Russian). Transl.: *Theory Probab. Appl.* **44**(1), 18–28. MR1700006
[58] Čekanavičius V. and Kruopis J. (2000) Signed Poisson approximation: a possible alternative to normal and Poisson laws. — *Bernoulli*, **6**(4), 591–606.
[59] Čekanavičius V. and Mikalauskas M. (2001) Local theorems for the Markov Binomial distribution. — *Lith. Math. J.*, **41**(3), 219–231.
[60] Čekanavičius V. (2003) Infinitely divisible approximations for discrete nonlattice variables. — *Adv. Appl. Probab.*, **35**(4), 982–1006.
[61] Čekanavičius V. (2002) On multivariate compound distributions. — *Teor. Veroyatn. Primen.*, **47**(3), 583–594 (Russian). Transl.: — *Theory Probab. Appl.* **47**(3), 493–505, 2003. MR1777684
[62] Čekanavičius V. and Wang Y.H. (2003) Compound Poisson approximation for sums of discrete nonlattice random variables. — *Adv. Appl. Probab.*, **35**(1), 228–250.
[63] Čekanavičius V. and Roos B. (2006) An expansion in the exponent for compound Binomial approximations. — *Lith. Math. J.*, **46**(1), 54–91. MR1975512
[64] Čekanavičius V. and Roos B. (2009) Poisson type approximations for the Markov Binomial distribution. — *Stochastic Process. Appl.*, **119**(1), 190–207. MR2485024
[65] Čekanavičius V. and Vellaisamy P. (2010) Compound Poisson and signed compound Poisson approximations to the Markov Binomial law. —

Bernoulli, **16**(4), 1114–1136.

[66] Čekanavičius V. and Elijio A. (2014) Smoothing effect of compound Poisson approximations to the distributions of weighted sums. — *Lith. Math. J.*, **54**(1), 35–47. MR2759171

[67] Čekanavičius V. and Vellaisamy P. (2015) Discrete approximations for sums of m-dependent random variables. — *ALEA, Lat. Am. J. Probab. Math. Stat.*, **12**, 765–792. MR3446037

[68] Čekanavičius V. (2016) *Approximation methods in Probability Theory*. — Springer: Universitext.

[69] Čekanavičius V. and Vellaisamy P. (2018) On closeness of two discrete weighted sums. — *Mod. Stoch. Theory Appl.*, **5**(2), 207–224. MR3813092

[70] Čekanavičius V. and Vellaisamy P. (2019) On large deviations for sums of discrete m-dependent random variables. — *Stochastics*, **91**(8), 1092–1108.

[71] Čekanavičius V. and Vellaisamy P. (2020) Lower bounds for discrete approximations to sums of m-dependent random variables. — *Probab. Math. Stat.*. MR4023509

[72] Dabrowski A., Ivanoff G., Kulik R. (2009) Some notes on Poisson limits for empirical point processes. — *Canadian J. Statist.*, **37**(3), 347–360.

[73] Daly F. (2013) Compound Poisson approximation with association or negative association via Stein's method. — *Electron. Commun. Probab.*, **18**(30), 1–12.

[74] Dall' Aglio G. (1956) Sugli estremi dei momenti delle funzioni di ripartizione dopia. — *Ann. Scuola Normale Superiore Di Pisa, ser. Cl. Sci.*, **3**(1), 33–74.

[75] Denzel G.E. and O'Brien G.L. (1975) Limit theorems for extreme values of chain–dependent processes. *Ann. Probab.*, **3**(5), 773–779. MR3056067

[76] Dobrushin R.L. (1953) Limit theorems for a Markov chain of two states. — *Izv. Akad. Nauk. USSR Ser. mat.*, **17**, 291–330. (Russian) Transl. *Select. Transl. Math. Stat. and Probab.*, **1**, 97–134, 1961. Providence, R.I.: Inst. Math. Stat., Amer. Math. Soc.. MR0058150

[77] Dobrushin R.L. (1970) Prescribing a system of random variables by conditional distributions. — *Teor. Veroyatn. Primen.*, **15**(3), 469–497 (Russian). Transl. *Theory Probab. Appl.*, **15**(3), 458–486. MR0298716

[78] Eichelsbacher P. and Roos M. (1999) Compound Poisson approximation for dissociated random variables via Stein's method. — *Combinatorics, Probability and Computing*, **8**, 335–346. MR1723647

[79] Elijio A. and Čekanavičius V. (2015) Compound Poisson approximation to weighted sums of symmetric discrete variables. — *Ann. Inst. Stat. Math.*, **67**, No 1, 195–210. MR3297864

[80] Erhardsson T. (2000) Compound Poisson approximation for counts of rare patterns in Markov chains and extreme sojourns in birth-death chains. — Ann. Appl. Probab., v. 10, 573–591. MR1768222

[81] Erhardsson T. (2005) Stein's method for Poisson and compound Poisson approximation. — *In: An Introduction to Stein's Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., v. 4, 61–113.* Singapore: Singapore Univ. Press. MR2235449

[82] Gan H.L. and Xia A. (2015) Stein's method for conditional compound Poisson approximation. — *Statist. Probab. Lett.*, **100**, 19–26. MR3324070

[83] Gani J. (1982) On the probability generating function of the sum of Markov-Bernoulli random variables. — *J. Appl. Probab.* **19A**, 321–326. MR0633201

[84] Genest C., Marceau E. and Mesfioui M. (2003) Compound Poisson approximations for individual models with dependent risks. — *Insurance Math. Econom.*, **32**(1), 73–91. MR3600678

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

[86] Gerber H.U. (1984) Error bounds for the compound Poisson approximation. *Insurance Math. Econom.*, **3**(3), 191–194. MR0752200

[87] Geske M.X., Godbole A.P., Schaffner A.A., Skolnick A.M., Wallstrom G.L. (1995) Compound Poisson approximations for word patterns under Markovian hypotheses. — *J. Appl. Probab.*, **32**(4), 877–892. MR1363330

[88] Gibbs A.L. and Su F. (2002) On choosing and bounding probability metrics. — *Intern. Statist. Rev.*, **70**(3), 419–435.

[89] Gini C. (1914) Di una 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.

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

[91] Gnedenko B.V. (1943) Sur la distribution du terme maximum d'une série aléatoire. — *Ann. Math.*, **44**(3), 423–453. MR0008655

[92] Gnedenko B.V. (1944) Limit theorems for sums of independent random variables. — Uspehi Mat. Nauk, **10**, 115–165. (Russian) Tranls: Proc. Amer. Math. Soc., v. 45. MR0012385

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

[94] Götze F. and Zaitsev A.Yu. (2004) Approximation of convolutions by accompanying laws without centering. — *Zap. Nauchn. Sem. POMI*, **320**, 44–53. Transl.: *J. Math. Sci.*, **137**(1), 4510–4515, 2006. MR2115864

[95] Götze F. and Zaitsev A.Yu. (2017) Rare events and Poisson point processes. — *Zap. Nauchn. Sem. POMI*, **466**, 109–119. Transl.: (2020) *J. Math. Sci.*, **244**(1), 771–778. MR3760046

[96] Götze F., Zaitsev A.Yu and Zaporozhets D. (2019) An improved multivariate version of Kolmogorov's second uniform limit theorem. — *Zap. Nauchn. Sem. POMI*, **486**, 71–85. Transl.: *J. Math. Sci.*, **258**, 782–792. MR4060230

[97] Götze F. and Zaitsev A.Yu. (2021) Convergence to infinite-dimensional compound Poisson distributions on convex polyhedra. — *Zap. Nauchn. Sem. POMI*, **501**, 118–125.

[98] Götze F. and Zaitsev A.Yu. (2022) On alternative approximating distributions in the multivariate version of Kolmogorov's second uniform limit the-

orem. — *Theory Probab. Appl.*, v. 67, No 1, 3–22 (in Russian). MR0993957

[99] Grigelionis B. (1962) On the asymptotic expansion of the remainder term in case of convergence to the Poisson law. (Russian) — *Litovsk. Mat. Sb.*, **2**(1), 35–48. MR0150814

[100] Grigelionis B. (1999) Asymptotic expansions in the compound Poisson limit theorem. — *Acta Applicand. Math.*, **58**, 125–134. MR1734745

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

[102] van Harn K. (1978) Classifying infinitely divisible distributions by functional equations. — Mathematical Centre Tracts, No 103 (Mathematical Centre, Amsterdam). MR0526929

[103] Harremoës P. and Ruzankin P.S. (2004) Rate of convergence to Poisson law in terms of information divergence. — *IEEE Trans. Inform Theory*, **50**(9), 2145–2149. MR2097199

[104] Herrmann H. (1965) Variationsabstand zwischen der Verteilung einer Summe unabhängiger nichtnegativer ganzzahliger Zufallsgrössen und Poissonschen Verteilungen. — *Math. Nachr.*, **29**(5), 265–289. MR0190979

[105] Hipp C. (1985) Approximation of aggregate claims distributions by compound Poisson distributions. — *Insurance Math. Econom.*, **4**(4), 227–232. Correction: **6**, p. 165, 1987. MR0810720

[106] Hipp C. (1986) Improved approximations for the aggregate claims distribution in the individual model. — *ASTIN Bull.*, **16**(2), 89–100.

[107] Hsing T. (1988) On the extreme order statistics for a stationary sequence. — Stochastic Processes Appl., v. 29, 155–169. MR0952827

[108] Hsing T. (1991) Estimating the parameters of rare events. — Stochastic Process. Appl., v. 37, 117–139. MR1091698

[109] Hsing T. (1993) On some estimates based on sample behavior near high level excursions. — Probab. Theory Related Fields, v. 95, 331–356. MR1213195

[110] Hsiau S.R. (1997). Compound Poisson limit theorems for Markov chains. *J. Appl. Probab.*, **34**(1), 24–34. MR1429051

[111] Ibragimov I. A. and Presman E. L. (1973) On the rate of approach of the distributions of sums of idependent random variables to accompanying distributions. — *Teor. Veroyatn. Primen.*, **18**(4), 753–766 (Russian). Transl.: *Theory Probab. Appl.*, **18**(4), 713–727. MR0331475

[112] Jensen J.L. (2013) On a saddlepoint approximation to the Markov Binomial distribution. *Braz. J. Probab. Stat.*, **27**(2), 150–161. MR3028801

[113] Johnson N.L., Kemp A.W. and Kotz S. (2005) *Univariate discrete distributions.* — Wiley. MR2163227

[114] Kantorovich L.V. (1942) On the translocation of masses. — *Doklady USSR Acad. Sci.*, **37**(7-8), 227–229. (Russian) Transl. *Management Sci.*, **5**(1), 1–4, 1958, *J. Math. Sci.*, **133**(4), 1381–1382. MR0096552

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

[116] Khintchine A.Y. (1937) Zur Theorie der unbeschränkt teilbaren

Verteilungsgesetze. — Rec. Math. [Mat. Sbornik] N.S., v. 2(44), No 1, 79–119.

[117] Khintchine A.Y. (1938) *Limit laws for sums of random variables.* — Moscow: ONTI. (in Russian)

[118] Kobus M. (1995) Generalized Poisson distributions as limits of sums for arrays of dependent random vectors. — *J. Multivariate Anal.*, **52**(2), 199–244. MR1323331

[119] Kolmogorov A.N. (1956) Two uniform limit theorems for sums of independent random variables. — *Teor. Veroyatn. Primen.*, **1**(4), 426–436 (Russian). Transl.: *Theory Probab. Appl.*, **1**(4), 384–394. MR2014969

[120] Kolmogorov A.N. (1963) On approximation of the distribution of a sum of independent random variables by infinitely divisible distributions. — *Trudy MMO*, **12**, 437–451. MR0169264

[121] Koopman B.O. (1950) A generalization of Poisson's distribution for Markoff chains. — *Proc. Nat. Acad. Sci. USA*, **36**, 202–207. MR0033467

[122] Kornya P. (1983) Distribution of aggregate claims in the individual risk theory model. — *Society of Actuaries: Transactions*, **35**, 823–858.

[123] Kozulyaev P.A. (1939) Asymptotic analysis of a fundamental formula of Probability Theory. — *Uch.. Zap. Moskov. Gos. Univ.*, **15**, 179–182. (Russian)

[124] Kruopis J. (1986) Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures. *Litovsk. Mat. Sb.*, **26**(1), 53–69 (Russian). Transl.: *Lith. Math. J.*, **26**(1), 37–49. MR0847204

[125] Kruopis J. (1986) Approximations for distributions of sums of lattice random variables I. — *Litovsk. Mat. Sb.*, **26**(3), 455–467 (Russian). Transl.: *Lith. Math. J.*, **26**(3), 234–244. MR0867228

[126] Kruopis J. and Čekanavičius V. (2014) Compound Poisson approximations for symmetric vectors. — *J. Multivariate Anal.*, **123**, 30–42. MR3130419

[127] Leadbetter M.R., Lindgren G. and Rootzen H. (1983) *Extremes and related properties of random sequences and processes.* — New York: Springer. MR0691492

[128] Le Cam L. (1960) An approximation theorem for the Poison Binomial distribution. — *Pacif. J. Math.*, **10**(4), 1181–1197. MR0142174

[129] Le Cam L. (1965) On the distribution of sums of independent random variables. — In: *Bernoulli, Bayes, Laplace (Anniversary volume).* Springer: Berlin, 179–202. MR0199871

[130] Z.-W.Liao, Y.Ma, A.Xia (2020) On Stein's factors for Poisson approximation in Wasserstein distance with non-linear transportation costs. — arXiv:2003.13976v1

[131] Liu Q. and Xia A. (2020) On moderate deviations in Poisson approximation. — *arXiv:1906.10016v2*. MR4148068

[132] Logunov P.L. (1990) Estimates for the convergence rate to the Poisson distribution for random sums of independent indicators. — *Teor. Veroyatn. Primen.*, **35**(3), 580–582 (Russian). Transl.: *Theory Prob. Appl.*, **35**(3), 587–590, 1990. MR1091218

[133] Major P. (1990) A note on the approximation of the uniform empirical

process. — Ann. Probab., v. 18, No 1, 129–139. MR1043940

[134] Mattner L. and Roos B. (2007) A shorter proof of Kanter's Bessel function concentration bound. — *Probab. Theory Rel. Fields*, **139**, 191–205. MR2322695

[135] Meshalkin L.D. (1960) On the approximation of polynomial distributions by infinitely-divisible laws. — *Teor. Veroyatn. Primen.*, **5**(1), 114–124 (Russian). Transl.: *Theory Probab. Appl.*, **5**(1), 106–114. MR0132586

[136] Michel R. (1987) An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. — *ASTIN Bulletin*, **17**(2), 165–169.

[137] Mikhailov V.G. (2001) Estimate of the accuracy of the compound Poisson approximation for the distribution of the number of matching patterns. — *Teor. Veroyatn. Primen.*, **46**(4), 713–723 (Russian). Transl.: *Theory Probab. Appl.*, **46**(4), 667–675. MR1971829

[138] Mikhailov V.G. (2008) A Poisson-type limit theorem for the number of pairs of matching sequences. — *Teor. Veroyatn. Primen.*, **53**(1), 59–71 (Russian). Transl.: *Theory Probab. Appl.*, **53**(1), 106–116. MR2760565

[139] Minakov A.A. (2015) Compound Poisson approximation of the number distribution for monotone strings of fixed length in a random sequence. (Russian) — *Prikl. Diskr. Mat.*, **28**(2), 21–29.

[140] Minkova L. and Omey E. (2014) A new Markov Binomial distribution. — *Comm. Statist. Theory Methods*, **43**(13), 2674–2688. MR3223703

[141] de Moivre A. (1712) De mensura sortis. — Philosophical Transactions, **27**, 213–264. Transl: Hald A. (1984) A. de Moivre: 'De Mensura Sortis' or 'On the measurement of chance'. — International Statistical Review, v. 52, No 3, 229–262. MR0867173

[142] Nagaev S.V. (1998) Concentration functions and the accuracy of approximation by infinitely divisible laws in a Hilbert space. (Russian) — *Dokl. Akad. Nauk*, **359**(4), 461–463. MR1668408

[143] Novak S.Y. (1998) On the limiting distribution of extremes. — *Siberian Adv. Math.*, v. 8, No 2, 70–95. MR1650530

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

[145] Novak S.Y. and Xia A. (2012) On exceedances of high levels. — Stochastic Processes Appl., v. 122, 582–599. MR2868931

[146] Novak S.Y. (2019) On the accuracy of Poisson approximation. — *Extremes*, **22**, 729–748. MR4031855

[147] Novak S.Y. (2019) Poisson approximation. — *Probability Surveys*, **16**, 228–276; (2021) v. 18, 272–275.

[148] Novak S.Y. (2021) Poisson approximation in terms of the Gini–Kantorovich distance. — Extremes, v. 24, No 1, 67–84. MR3992498

[149] Petrauskienė J. and Čekanavičius V. (2011) Compound Poisson approximations for sums of one-dependent random variables I. — *Lith. Math. J.*, **50**(3), 323–336. MR2719567

[150] Petrauskienė J. and Čekanavičius V. (2011) Compound Poisson approximations for sums of one-dependent random variables II. — *Lith. Math.*

J., **51**(1), 51–65. MR2784377

[151] Pinelis I. (2020) Monotonicity properties of the Poisson approximation to the binomial distribution, with applications to statistical testing. — Statistics and Probability Letters. MR4143570

[152] Presman E.L. (1973) On a multidimensional version of the Kolmogorov uniform limit theorem. — *Teor. Veroyatn. Primen.*, **18**(2), 396–402 (Russian). Transl.: *Theory Probab. Appl.*, **18**(2), 378–384. MR0317384

[153] Presman E.L. (1983) Approximation of Binomial distributions by infinitely divisible ones. — *Teor. Veroyatn. Primen.*, **28**(2), 372–382 (Russian). Transl.: *Theory Probab. Appl.*, **28**(2), 393–403. MR0700218

[154] Presman E.L. (1985) Approximation in variation of the distribution of a sum of independent Bernoulli variables with a Poisson law. — *Teor. Veroyatn. Primen.*, **30**(2), 391–396 (Russian). Transl.: *Theory Probab. Appl.*, **30**(2), 417–422. MR0792634

[155] Prokhorov Yu.V. (1952) *Limit theorems for sums of independent random variables.* (Russian) PhD/DSc Thesis. — Moscow: Steklov Institute of Mathematics, 42 pp.

[156] Prokhorov Yu.V. (1953) Asymptotic behavior of the Binomial distribution. — *Uspehi Matem. Nauk*, **8**(55), 135–142 (Russian). Transl.: *Select. Transl. Math. Statist. and Probability*, **1**, 87–95, 1961. Providence, R.I.: Inst. Math. Statist. & Amer. Math. Soc.. MR0056861

[157] Prokhorov Yu.V. (1955) On sums of identically distributed random variables. — *Dokl. Akad. Nauk SSSR (N.S.)*, **105**(4), 645–647 (Russian). MR0077008

[158] Prokhorov Yu.V. (1956) Convergence of random processes and limit theorems in probability theory. — *Teor. Veroyatn. Primen.*, **1**(2), 177–238 (Russian). Transl.: *Theory Probab. Appl.*, **1**(2), 157–214. MR0084896

[159] Prokhorov Yu.V. (1960) On a uniform limit theorem by A.N.Kolmogorov. — *Theory Probab. Appl.*, **5**(1), 103–113. MR0132585

[160] Prokhorov Yu.V. (2012) *Selected works*, v. 1. — Moscow: Torus Press, 775 pp. (Russian)

[161] Rachev S.T. (1984) The Monge–Kantorovich mass transference problem and its stochastic applications. *Teor. Veroyatn. Primen.*, **29**(4), 625–653 (Russian). Transl.: *Theory Probab. Appl.*, **29**(4), 647–676, 1985. MR0773434

[162] Rachev S.T. and Rüschendorf L. (1990) Approximation of sums by compound Poisson distributions with respect to stop-loss distances. *Adv. Appl. Probab.*, **22**(2), 350–374. MR1053235

[163] Robert C.Y. (2009) Inference for the limiting cluster size distribution of extreme values. Ann. Stattist., v. 37, No 1, 271–310. MR2488352

[164] Roos M. (1994) Stein's method for compound Poisson approximation: the local approach. — *Ann. Appl. Probab.*, **4**(4), 1177–1187. MR1304780

[165] Roos B. (2001) Sharp constants in the Poisson approximation. *Statist. Probab. Lett.*, **52**, 155–168. MR1841404

[166] Roos B. (2002) Kerstan's method in the multivariate Poisson approximation: an expansion in the exponent. *Teor. Veroyatn. Primen.*, **43**(2), 397–

402. Reprinted in *Theory Probab. Appl.*, **47**(2), 358–363, 2003. MR2003208

[167] Roos B. (2003) Poisson approximation via the convolution with Kornya-Presman signed measures. — *Teor. Veroyatn. Primen.*, **48**(3), 628–632. Reprinted in *Theor. Probab. Appl.*, **48**(3), 555–560, 2004. MR2141358

[168] Roos B. (2005) On Hipp's compound Poisson approximations via concentration functions. — *Bernoulli*, **11**(3), 533–557. MR2147774

[169] Roos B. (2007) On variational bounds in the compound Poisson approximation of the individual risk model. — *Insurance Math. Econom.*, **40**(3), 403–414. MR2310979

[170] Roos B. (2017) Refined total variation bounds in the multivariate and compound Poisson approximation. *ALEA, Lat. Am. J. Probab. Math. Stat.*, **14**, 337–360. MR3647296

[171] Rootzén H. (1988) Maxima and exceedances of stationary Markov processes. — Adv. Appl. Probab., v. 20, 371–390. MR0938151

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

[173] Sason I. (2013) Improved lower bounds on the total variation distance for the Poisson approximation. — *Statist. Probab. Lett.*, **83**, 2422–2431. MR3093834

[174] Schbath S. (2000) An overview on the distribution of word counts in Markov chains. — *J. Comput. Biology*, **7**(1-2), 193–201.

[175] Šiaulys J. and Čekanavičius V. (1988) Approximation of distributions of integer-valued additive functions by discrete charges I. — *Litovsk. Mat. Sb.*, **28**(4), 795–810 (Russian). Transl.: *Lith. Math. J.*, **28**(4), 392–401. MR0987874

[176] Šliogerė J. and Čekanavičius V. (2015) Two limit theorems for Markov binomial distribution. — *Lith. Math. J.*, **55**(3), 451–463. MR3379037

[177] Šliogerė J. and Čekanavičius V. (2016) Approximation of symmetric three-rtate Markov chain by compound Poisson law. — *Lith. Math. J.*, **56**(3), 417–438. MR3530227

[178] Studnev Yu.P. (1960) An approximation to the distribution of sums by infinitely divisible laws. — *Teor. Veroyatn. Primen.*, **5**(4), 465–469 (Russian). Transl.: *Theory Probab. Appl.*, **5**(4), 421–426.

[179] Tsybakov A.B. (2009) *Introduction to nonparametric estimation.* — Springer. MR2724359

[180] Upadhye N.S. and Kumar A.N. (2018) Pseudo-Binomial approximation to (k_1, k_2)-runs. — *Statis. Probab. Lett.*, **141**, 19–30.

[181] Upadhye N.S. and Vellaisamy P. (2013) Improved bounds for approximations to compound distributions. — *Statist. Probab. Lett.*, **83**, 467–473. MR3006978

[182] Uspensky J.V. (1931) On Ch.Jordan's series for probability. — *Ann. Math.*, **32**(2), 306–312. MR1502999

[183] Vallander S.S. (1973) Calculation of the Wasserstein distance between probability distributions on the line. — *Teorya Veroyatn. Primen.*, **18**(4), 824–827 (Russian). Transl.: *Theory Probab. Appl.*, **18**(4), 784–786, 824–827. MR0328982

[184] Vellaisamy P. (2004) Poisson approximation for $(k_1; k_2)$ events via the Stein-Chen method. — *J. Appl. Probab.*. **41**(4), 1081–1092.

[185] Vellaisamy P., Upadhye N.S. (2009) Compound negative Binomial approximations for sums of random variables. — *Probab. Math. Stat.*, **29**(2), 205–226. MR2122802

[186] Vellaisamy P., Upadhye N.S. and Čekanavičius V. (2012) On negative Binomial approximation. — *Theory Probab. Appl.*, **57**(1), 97–109. Reprinted: *Teorya Veroyatn. Primen.*, **57**(1), 141–154. MR3201640

[187] Vellaisamy P. and Čekanavičius V. (2018) Infinitely divisible approximations for sums of m-dependent random variables. — *J. Theoret. Probab.*, **31**(4), 2432–2445. MR3866620

[188] Wang Y.H. (1981) On the limit of the Markov Binomial distribution. — *J. Appl. Probab.*, **18**(4), 937–942. MR0633240

[189] Wang Y.H. (1991) A compound Poisson convergence theorem. — *Ann. Probab.*, **19**(1), 452–455. MR1085347

[190] Wang Y.H., Chang H.-F. and Chen S.-Y. (2003) Convergence theorems for the lengths of consecutive successes of Markov Bernoulli sequences. — *J. Appl. Probab.*, **14**(3), 741–749. MR1993264

[191] Wang X. and Xia A. (2008) On Negative Binomial Approximation to k-Runs. — *J. Appl. Probab.*, **45**(2), 456–471. MR2426844

[192] Xia A. and Zhang M. (2009) On approximation of Markov Binomial distributions. — *Bernoulli*, **15**(4), 1335–1350. MR2597595

[193] Zacharovas V. and Hwang H.-K. (2010) A Charlier-Parseval approach to Poisson approximation and its applications. — *Lith. Math. J.*, **50**(1), 88–119. MR2607681

[194] Zaitsev A.Yu. (1981) Some properties of n-fold convolutions of distributions. — *Teor. Veroyatn. Primen.*, **26**(1), 152–156 (Russian). Transl.: *Theory Probab. Appl.*, **26**(1), 148–152. MR0605644

[195] 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. — *Teor. Veroyatn. Primen.*, **28**(4), 625–636 (Russian). Transl.: *Theory Probab. Appl.*, **28**(4), 657–669. MR0726889

[196] Zaitsev A.Yu. and Arak T. (1983) On the rate of convergence in the second Kolmogorov's uniform limit theorem. — *Teor. Veroyatn. Primen.*, **28**(2), 333–353. Transl.: *Theory Probab. Appl.*, **28**(2), 351–374. MR0700213

[197] Zaitsev A.Yu. (1987) On the uniform approximation of distributions of sums of independent random variables. — *Teor. Veroyatn. Primen.*, **32**(1), 45–52 (Russian). Transl.: *Theory Probab. Appl.*, **32**(1), 40–47. MR0890929

[198] Zaitsev A.Yu. (1988) Estimates for the closeness of successive convolutions of multidimensional symmetric distributions. — *Probab. Theory Rel. Fields*, **79**, 175–200. MR0958287

[199] Zaitsev A.Yu. (1989) Multidimensional version of the second uniform limit theorem of Kolmogorov. — *Teor. Veroyatn. Primen.*, **34**(1), 128–151 (Russian). Transl.: *Theory Prob. Appl.*, **34**(1), 108–128. MR0993957

[200] Zaitsev A.Yu.(1989) Approximation of convolutions of multi-dimensional

symmetric distributions by accompanying laws. — *Zap. Nauchn. Sem. LOMI V. A. Steklova AN SSSR*, **177**, 55–72 (Russian). Transl.: *J. Soviet Mathematics*, **61**(1), 1859–1872. MR1053124

[201] Zaitsev A.Y. (1990) Certain class of nonuniform estimates in multidimensional limit theorems. — *Zap. Nauchn. Sem. POMI*, **184**, 92–100 (Russian). Transl.: *J. Math. Sci.*, **68**, 459–468. MR1098691

[202] 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 distance in variation. — *Teor. Veroyatn. Primen.*, **36**(2), 356–361 (Russian). Transl.: *Theory Probab. Appl.*, **36**(2), 419–425. MR1119511

[203] Zaitsev A.Yu. (1992) Approximation of convolutions of probability distributions by infinitely divisible laws under weakened moment restrictions. — *Zap. Nauchn. Sem. POMI*, **194**, 79–90 (Russian). Transl.: *J. Math. Sci.*, **75**(5), 1922–1930. MR1175738

[204] Zaitsev A.Yu. (1996) Approximation of convolutions by accompanying laws under the existence of moments of low orders. — *Zap. Nauchn. Sem. POMI*, **228**, 135–141 (Russian). Transl.: *J. Math. Sci.*, **93**(3), 336–340. MR1449852

[205] Zaitsev A.Yu. (2003) Approximation of a sample by a Poisson point process. — *Zap. Nauchn. Sem. POMI*, **298**, 111–125 (Russian). Transl.: *J. Math. Sci.*, **128**(1), 2556–2563. MR2038866

[206] Zhang M. (2011) Approximation for counts of head runs. — *Sci. China Math.*, **54**(2), 311–324. MR2771207