Poisson approximation

S.Y. Novak
MDX University London

Abstract
We overview results on the topic of Poisson approximation that are missed in existing surveys. The topic of Poisson approximation to the distribution of a sum of integer-valued random variables is presented as well.

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

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

Key words: Poisson approximation, compound Poisson approximation, accuracy of approximation, asymptotic expansions, Poisson process approximation, total variation distance, long head runs, long match patterns.

AMS Subject Classification: 60E15, 60F05, 60G50, 60G51, 60G55, 60G70, 60J75, 62E17, 62E20.

Content
1. Weak convergence to a Poisson law 2
1. Independent random variables 2
1.2 Dependent Bernoulli random variables 3
2. Accuracy of Poisson approximation 6
2.1 Independent Bernoulli random variables 8
2.2 Dependent Bernoulli random variables 13
2.3 Independent integer-valued random variables 14
2.4 Dependent integer-valued random variables 16
2.5 Asymptotic expansions 17
2.6 Sum of a random number of random variables 18
3. Applications 21
3.1 Long head runs 21
3.2 Long match patterns 23
4. Compound Poisson approximation 27
4.1 CP limit theorem 27
4.2 Accuracy of CP approximation 29
4.3 CP approximation to B(n, p) 31
5. Poisson process approximation 32
6. References 38
1 Weak convergence to a Poisson law

Poisson approximation appears natural in situations where one deals with a large number of rare events. The topic has attracted a considerable body of research. It has important applications in insurance, extreme value theory, reliability theory, mathematical biology, etc. (cf. [7, 12, 39, 48, 61]). However, existing surveys are surprisingly sketchy, and miss not only a number of results obtained during the last three decades but even some classical results going back to 1930s.

The paper aims to fill the gap. We present a comprehensive list of results on the topic of Poisson approximation, and formulate a number of open problems. Related results on the topic of compound Poisson approximation are presented as well.

1.1 Weak convergence to a Poisson law

We denote by $\Pi(\lambda)$ a Poisson law with parameter $\lambda$.

The following Poisson limit theorem is due to Gnedenko [36] and Marcinkiewicz [54].

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

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

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

(1)

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

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

Theorem 1 [36, 54] If $\{X_{n,k}\}$ are infinitesimal r.v.s, then

$$\mathcal{L}(S_n) \Rightarrow \Pi(\lambda) \quad (\exists \lambda > 0)$$

(2)

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

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

(3)

$$\sum_{k} \mathbb{P}(X_{n,k} \notin B_\varepsilon) \to 0, \quad \sum_{k} \mathbb{E}X_{n,k} \mathbb{I}\{|X_{n,k}| < \varepsilon\} \to 0,$$

(4)

$$\sum_{k} (\mathbb{E}X_{n,k}^2 \mathbb{I}\{|X_{n,k}| < \varepsilon\} - \mathbb{E}^2X_{n,k} \mathbb{I}\{|X_{n,k}| < \varepsilon\}) \to 0.$$

(5)

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

Let $\mathbb{N}$ denote the set of natural numbers, and let $\mathbb{Z}_+ := \mathbb{N} \cup \{0\}$. Corollary 2 deals with the case of non-negative integer-valued random variables.
Corollary 2 If \{X_{n,1}, ..., X_{n,k_n}\}_{n \geq 1} is a triangle array of independent random variables taking values in \(\mathbb{Z}_+\) such that \(X_{n,i} \overset{d}{=} X_{1,n} (1 \leq i \leq k_n)\), then (2) holds if and only if
\[
k_n \mathbb{P}(X_{1,n} = 1) \to \lambda \quad \text{and} \quad \mathbb{P}(X_{n,1} \geq 2)/\mathbb{P}(X_{n,1} \geq 1) \to 0. \tag{6}
\]

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

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

Example 1.1. Let \{X_{n,1}, ..., X_{n,n}\} be i.i.d. random variables with the distribution
\[
\mathbb{P}(X = 0) = 1 - \lambda/n - 1/n^{1.5}, \quad \mathbb{P}(X = 1) = \lambda/n, \quad \mathbb{P}(X = n) = 1/n^{1.5} \quad (\lambda > 0).
\]
Then (1) and (6) hold, hence \(L(S_n) \Rightarrow \Pi(\lambda)\). Note that \(\mathbb{E}S_n \nrightarrow \lambda\). \(\square\)

The proof of Theorem 1 can be found in [38].

1.2 Dependent Bernoulli random variables

The topic of Poisson approximation to the distribution of a sum of dependent Bernoulli r.v.s has applications in extreme value theory, reliability theory, etc. (cf. [7, 12, 48, 61]).

Let \{X_{n,1}, ..., X_{n,n}\}_{n \geq 1} be a triangle array of 0-1 random variables such that sequence \(X_{n,1}, ..., X_{n,n}\) is stationary for each \(n \in \mathbb{N}\). For instance, in extreme value theory one often has
\[
X_{n,k} = \mathbb{1}\{Y_k > u_n\},
\]
where \(\{Y_i, i \geq 1\}\) is a stationary sequence of random variables and \(\{u_n\}\) is a sequence of “high” levels. The special case where \(\{Y_i, i \geq 1\}\) is a moving average is related to the topic of the Erdős–Rényi partial sums (cf. [61], ch. 2).

Let \(\mathcal{F}_{l,m}(\tau)\) be the \(\sigma\)-field generated by the events \(\{X_{n,i}\}, \ l \leq i \leq m\). Set
\[
\alpha_n(l) = \sup |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|, \quad \varphi(l) = \sup |\mathbb{P}(B|A) - \mathbb{P}(B)|, \quad \beta_n(l) = \sup_{B} \mathbb{E} \sup_{B} |\mathbb{P}(B|\mathcal{F}_{1,m}) - \mathbb{P}(B)|,
\]
where the supremum is taken over \(m \geq 1, A \in \mathcal{F}_{1,m}(\tau), \ B \in \mathcal{F}_{m+i+1,n}\) such that \(\mathbb{P}(A) > 0\). Conditions involving mixing coefficients \(\alpha_n(\cdot), \beta_n(\cdot), \varphi_n(\cdot)\) are slightly weaker than those involving traditional mixing coefficients \(\alpha(\cdot), \beta(\cdot), \varphi(\cdot)\).

Condition \(\Delta\) is said to hold if \(\alpha_n(l_n) \to 0\) for some sequence \(\{l_n\}\) of natural numbers such that \(1 \ll l_n \ll n\).
Class $\mathcal{R}$. If $\Delta$ holds, then there exists a sequence $\{r_n\}$ of natural numbers such that
\[ n \gg r_n \gg l_n \gg 1, \quad nr_n^{-1} \alpha_n^{2/3}(l_n) \to 0 \quad (n \to \infty) \tag{7} \]
(for instance, one can take $r_n = \lceil \sqrt{n \max\{l_n; n\alpha_n(l_n)\}} \rceil$).

We denote by $\mathcal{R}$ the class of all such sequences $\{r_n\}$.

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

Let $\zeta_{r,n}$ be a r.v. with the distribution $L(\zeta_{r,n}) = L(S_r|S_r>0)$. \tag{8}

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

**Theorem 3** Assume condition $\Delta$. If, as $n \to \infty$,
\[ S_n \Rightarrow \pi\lambda \quad (\exists \lambda > 0), \tag{9} \]
then
\[ \zeta_{r,n} \to_P 1 \quad (n \to \infty) \tag{10} \]
for any sequence $\{r=r_n\}$ obeying (7).

If there exists the limit
\[ \lim_{n \to \infty} \mathbb{P}(X_{n,1} = \ldots = X_{n,n} = 0) = e^{-\lambda} \quad (\exists \lambda > 0) \tag{11} \]
and (10) holds for some $\{r=r_n\} \in \mathcal{R}$, then $S_n \Rightarrow \pi\lambda$.

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

Condition (11) is an analogue of (3); it means that $\mathbb{P}(X_{n,1} \neq 0)$ is “properly small”: if $X_{n,1}, \ldots, X_{n,n}$ are i.i.d.r.v.s (see also (11†)).

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

**Remark 2.1.** The following condition $(D')$ has been widely used in extreme value theory (cf. [48, 61]):
\[ \lim_{n \to \infty} n \sum_{i=1}^{r-1} \mathbb{P}(X_{n,i+1} \neq 0, X_{n,1} \neq 0) = 0 \quad (D') \tag{12} \]
for any sequence $\{r=r_n\}$ such that $n \gg r_n \gg 1$. Condition $(D')$ means that there is no asymptotic clustering of extremes. Condition $(D')$ was introduced by Loynes [52].
Closely related is the following condition
\[
\lim_{n \to \infty} \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0 | X_{n,1} \neq 0) = 0. 
\]

If conditions $\Delta$ and (11) hold, then $(D')$ is equivalent to $\tilde{\Delta}$. Indeed, one can check that $\Delta$ and (11) yield
\[
\mathbb{P}(S_r > 0) \sim \frac{\lambda r}{n} \quad (n \to \infty)
\]  
(cf. (16) below). Then
\[
\frac{\lambda r}{n} \sim \mathbb{P}(S_r > 0) \leq r \mathbb{P}(X_{n,1} \neq 0)
\]  
yielding the lower bound in (11).

Denote $p = \mathbb{P}(X_{n,1} \neq 0)$. By Bonferroni’s inequality,
\[
\frac{\lambda r}{n} \sim \mathbb{P}(S_r > 0) \geq r \mathbb{P}(X_{n,1} \neq 0) - \mathbb{P}\left( \bigcup_{1 \leq i < j \leq r} \{X_{n,i} \neq 0, X_{n,j} \neq 0\} \right)
\]  
\[
\geq rp - rp \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0 | X_{n,1} \neq 0).
\]

Therefore,
\[
1 \geq \mathbb{P}(S_r > 0)/rp \geq 1 - \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0 | X_{n,1} \neq 0),
\]
\[
\lambda + o(1) \leq np \leq \frac{(\lambda + o(1)) \left( 1 - \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0 | X_{n,1} \neq 0) \right)}{1 - \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0 | X_{n,1} \neq 0)}.
\]

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

**Remark 2.2.** Condition (10) is weaker than $(D')$: if conditions $\Delta$ and (11) hold, then $(D')$ entails (10). Indeed, $\zeta_{r,n} \geq 1$ by construction. Note that
\[
\mathbb{P}(S_r > 1) = \mathbb{P}\left( \bigcup_{1 \leq i < j \leq r} \{X_{n,i} \neq 0, X_{n,j} \neq 0\} \right)
\]  
\[
\leq r \sum_{i=1}^{r-1} \mathbb{P} (X_{n,i+1} \neq 0, X_{n,1} \neq 0).
\]

Thus, $\mathbb{P}(S_r > 1) = o(r/n)$ if $(D')$ holds. In view of (12), $\mathbb{P}(\zeta_{r,n} > 1) \to 0$ as $n \to \infty$, i.e., (10) holds.
Remark 2.3. If conditions $\Delta$ and $(D')$ hold, then (11) is equivalent to
\[
\lim_{n \to \infty} n \mathbb{P}(X_{n,1} \neq 0) = \lambda. \tag{3}
\]
Indeed, this follows from (12), (13) and (16) (cf. [48], Theorem 3.4.1).

A generalisation of Corollary 2 to the case of stationary $\varphi$-mixing r.v.s has been given by Utev [82], Theorem 10.1, who has shown that conditions (3) and $(D')$ are necessary and sufficient for (9). Sufficient conditions for Poisson convergence without assuming stationarity have been provided by Sevastyanov [75]. A Poisson limit theorem in the case of a two-dimensional random field $\{X_{i,j}\}$ has been given by Banis [8].

Proof of Theorem 3. Let $\{r=r_n\} \in \mathcal{R}$. Condition $\Delta$ and Lemma 2.4.1 from [48] imply that for any $t \in \mathbb{R}$, as $n \to \infty$,
\[
\mathbb{E} \exp(itS_n) = \exp\left(\frac{n}{r} \mathbb{P}(S_r > 0) \mathbb{E}\left\{e^{itS_r} - 1 \mid S_r > 0\right\}\right) + o(1), \tag{15}
\]
\[
\mathbb{P}(S_n = 0) = \mathbb{P}^{n/r}(S_r = 0) + o(1) = \exp\left(-\frac{n}{r} \mathbb{P}(S_r > 0)\right) + o(1) \tag{16}
\]
(cf. (5.10) in [61]).

If (9) holds, then so does (11): $\mathbb{P}(S_n = 0) \to e^{-\lambda}$ as $n \to \infty$. Note that (11) and (16) yield (12). Since
\[
\mathbb{E}e^{itS_n} \to \exp(\lambda(e^{it} - 1)) \quad (\forall t \in \mathbb{R}) \tag{9}
\]
by the assumption, (15) and (12) entail $\mathbb{E}e^{itr_n} \to e^{it}$, i.e., (10) holds.

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

2 Accuracy of Poisson approximation

The problem of evaluating the accuracy of Poisson approximation to the distribution of a sum
\[
S_n = X_1 + \ldots + X_n
\]
of independent 0-1 random variables has attracted a lot of attention among researchers (cf. [12, 61] and references wherein).

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

Historically, the accuracy of Poisson approximation was first studied in terms of the uniform distance (sometimes called the Kolmogorov distance).
The uniform distance \( d_K(X;Y) \equiv d_K(F_X;F_Y) \) between the distributions of random variables \( X \) and \( Y \) with distribution functions (d.f.s) \( F_X \) and \( F_Y \) is defined as
\[
d_K(F_X;F_Y) = \sup_x |F_X(x) - F_Y(x)|.
\]

Many authors evaluated the accuracy of Poisson approximation to \( L(S_n) \) in terms of the total variation distance. Recall that the total variation distance \( d_{TV}(X;Y) \) between the distributions of r.v.s \( X \) and \( Y \) is defined as
\[
d_{TV}(X;Y) = \sup_{A \in \mathcal{A}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|,
\]
where \( \mathcal{A} \) is a Borel \( \sigma \)-field. Evidently, \( d_K(X;Y) \leq d_{TV}(X;Y) \).

The Gini-Kantorovich distance between the distributions of r.v.s \( X \) and \( Y \) with finite first moments (known also as the Kantorovich–Wasserstein distance) is
\[
d_G(X;Y) \equiv d_G(L(X);L(Y)) = \sup_{g \in \mathcal{F}} |\mathbb{E}g(X) - \mathbb{E}g(Y)|,
\]
where \( \mathcal{F} = \{g: |g(x) - g(y)| \leq |x - y|\} \) 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 \( L(X') = L(X) \) and \( L(Y') = L(Y) \) \[32\] \[21\].

The Gini-Kantorovich distance between the distributions of r.v.s \( X \) and \( Y \) with finite first moments (known also as the Kantorovich–Wasserstein distance) is
\[
d_G(X;Y) \equiv d_G(L(X);L(Y)) = \sup_{g \in \mathcal{F}} |\mathbb{E}g(X) - \mathbb{E}g(Y)|,
\]
where \( \mathcal{F} = \{g: |g(x) - g(y)| \leq |x - y|\} \) 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 \( L(X') = L(X) \) and \( L(Y') = L(Y) \) \[83\]. If \( X \) and \( Y \) take values in \( \mathbb{Z}_+ \), then \[67\]
\[
d_G(X;Y) = \sum_{i \geq 1} |\mathbb{P}(X \geq i) - \mathbb{P}(Y \geq i)|.
\]

Distance \( d_G \) was introduced by Kantorovich \[43\] (to be precise, Kantorovich has introduced a class of distances that includes \( d_G \)). We add the name of Gini since Gini \[35\] used \( \mathbb{E}|X - Y| \)-type quantities. Barbour et al. \[12\] called \( d_G \) the “Wasserstein distance” after Dobrushin \[32\] attributed it to Vasershtein \[84\].

If distributions \( P_1 \) and \( P_2 \) have densities \( f_1 \) and \( f_2 \) with respect to a measure \( \mu \), 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
\[
d_H^2 \leq d_{TV} \leq \sqrt{2} d_H.
\]

Certain other distances can be found in \[71\] \[64\]. Below we present estimates of the accuracy of Poisson approximation for \( L(S_n) \) in terms of \( d_K, d_{TV} \) and \( d_G \) distances.
2.1 Independent Bernoulli r.v.s

We denote by $B(n, p)$ the Binomial distribution with parameters $n$ and $p$. Let $\Pi(\lambda)$ denote the Poisson distribution with parameter $\lambda$. A Poisson $\Pi(\lambda)$ random variable is denoted by $\pi_\lambda$.

Let $X_1, X_2, ..., X_n$ be independent Bernoulli $B(p_i)$ r.v.s. Denote

$$\lambda = \mathbb{E}S_n, \quad p_i = \mathbb{P}(X_i = 1) \quad (i \geq 1), \quad \theta = \sum_{i=1}^{n} p_i^2 / \lambda.$$  

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

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

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

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

$$B(n, p) = L(\pi_n(p) | \pi_n(1) = n).$$  

Prohorov [66] has established the existence of an absolute constant $c$ such that

$$d_{inv}(B(n, p); \Pi(np)) \leq cp.$$  

Tsaregradskii [81] has shown that

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

if $X$ and $Y$ are integer-valued r.v.s. Using inequality (22), he has derived the estimate

$$d_K(B(n, p); \Pi(np)) \leq p\pi^2 e^{2p(2-p)}/16(1-p) \quad (p \in (0; 1/2]).$$  

Inequality (23) seems to be the first estimate of the accuracy of Poisson approximation with an explicit constant.

In the case of non-identically distributed Bernoulli $B(p_i)$ random variables Shorgin [78] has proved that

$$d_K(S_n; \pi_\lambda) \leq 1.13 \theta / (1 - \sqrt{\theta})$$  

if $\theta < 1$. Kontoyiannis et al. [46] have shown that

$$d_G^2(S_n; \pi_\lambda) \leq \lambda^{-1} \sum_{i=1}^{n} p_i^3 / (1 - p_i).$$
Many authors worked on the problem of evaluating the total variation distance $d_{TV}(S_n; \pi_\lambda)$ (cf. [12, 61] and references wherein). LeCam [49] presents the following bound:

$$d_{TV}(S_n; \pi_\lambda) \leq \sum_{i=1}^{n} p_i^2. \quad (24)$$

This bound is sharp: according to (2.10) in Deheuvels & Pfeifer [30],

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

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

Note that (24) is a consequence of the property of $d_{TV}$ and the following fact:

$$d_{TV}(B(p); \Pi(p)) = (1-e^{-p})p \leq p^2. \quad (25)$$

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

$$d_{TV}(S_n; \pi_\lambda) \leq d_{TV}(\bar{X}; \bar{\pi}) \leq \sum_{i=1}^{n} d_{TV}(X_i; \pi_{p_i}) \leq \sum_{i=1}^{n} p_i^2. \quad (24')$$

Kerstan [44] has shown that

$$d_{TV}(S_n; \pi_\lambda) \leq 1.05\theta \quad (26)$$

if $p_n^* := \max_{i \leq n} p_i \leq 1/4$. According to Romanowska [68]

$$d_{TV}(B(n, p); \Pi(np)) \leq p/2\sqrt{1-p}. \quad (27)$$

Barbour and Eagleson [9] have derived the popular estimate

$$d_{TV}(S_n; \pi_\lambda) \leq \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^{n} p_i^2. \quad (28)$$

Presman [65] has established an estimate of $d_{TV}(S_n; \pi_\lambda)$ with the constant 0.83 at the leading term. In the case of i.i.d. Bernoulli $B(p)$ r.v.s Presman’s bound becomes

$$d_{TV}(B(n, p); \Pi(np)) \leq 0.83p/(1-p)(1-1/n). \quad (29)$$

Xia [85] has derived an estimate with the constant 0.6844 at the leading term.

Roos [71] (see also Čekanavičius & Roos [25]) has obtained a bound with a correct constant $3/4e$ at the leading term:

$$d_{TV}(S_n; \pi_\lambda) \leq 3\theta/4e(1-\sqrt{\theta})^{3/2}. \quad (30)$$
Roos [71] has shown that if \( \theta \to 0 \) and \( \lambda \to 1 \) as \( n \to \infty \), then

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

Thus, constant \( 3/4e \) cannot be improved.

Denote

\[
p_n^* = \max_{i \leq n} p_i, \quad \varepsilon = \min \left\{ 1; (2\pi[\lambda-p_n^*])^{-1/2} + 2\delta/(1-p_n^*/\lambda) \right\},
\]

\[
\delta = \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} p_i^2, \quad \delta^* = \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} p_i^3.
\]

The following inequality from Novak [61], Theorem 4.12, sharpens the second-order term of the right-hand side of estimate (30):

\[
d_{TV}(S_n; \pi_\lambda) \leq 3\theta/4e + 2\delta^* \varepsilon + 2\delta^2. \tag{31}
\]

An estimate in terms of the Gini-Kantorovich distance is available as well:

\[
d_G(S_n; \pi_\lambda) \leq \left( 1 \wedge \frac{4}{3} \sqrt{2/e\lambda} \right) \sum_{i=1}^{n} p_i^2 \tag{32}
\]

(cf. [61], formula (4.53)).

**Asymptotics of** \( d_{TV}(S_n; \pi_\lambda) \). The asymptotics of \( d_{TV}(S_n; \pi_\lambda) \) in the case of identically distributed Bernoulli \( B(p) \) r.v.s has been established by Prohorov [66]:

\[
d_{TV}(B(n,p); \Pi(np)) = p/\sqrt{2\pi e \left( 1 + O(1/(p + 1/\sqrt{np})) \right)}. \tag{33}
\]

Deheuvels & Pfeifer [29] and Roos [70] have generalised (37) to the case of non-identically distributed \( \{X_i\} \). The following result concerning the asymptotics of \( d_{TV}(S_n; \pi_\lambda) \) uses the notation from [61], ch. 4.

Given a non-negative integer-valued random variable \( Y \), we denote by \( Y^* \) a random variable with the distribution

\[
\mathbb{P}(Y^* = k) = \mathbb{P}(Y = k)(k - \lambda)^2/\lambda \quad (k \in \mathbb{Z}_+). \tag{34}
\]

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

\[
d_{TV}(S_n; \pi_\lambda) = \theta d_{TV}(\pi^*; \pi_\lambda + 1)/2 + O\left( \delta^* \varepsilon + \delta^2 \right). \tag{35}
\]

One can check that

\[
d_{TV}(\pi^*; \pi_\lambda + 1) = \sqrt{2/\pi e \left( 1 + O(1/\sqrt{\lambda}) \right)} \tag{36}
\]
as \( \lambda \to \infty \). Thus,
\[
d_{TV}(S_n; \pi_\lambda) = \theta / \sqrt{2\pi e} \left( 1 + O(\theta + 1/\sqrt{\lambda}) \right)
\]
(37)
if \( \lambda \to \infty \) and \( \theta \to 0 \) as \( n \to \infty \). Deheuvels & Pfeifer [29] present also the asymptotics of \( d_{TV}(S_n; \pi_\lambda) \) in the case \( \lambda \to \text{const} \) as \( n \to \infty \).

**Example 2.1.** Let \( p_i = 1/i, \ i \in \mathbb{N} \). Then \( p_n^* = 1 / n \), \( \lambda = \lambda(n) \to \infty \), \( \theta \to 0 \) as \( n \to \infty \), and (37) entails
\[
d_{TV}(S_n; \pi_\lambda) \sim \theta / \sqrt{2\pi e}.
\]
\[\blacksquare\]

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

Let \( \bar{S}_n \) be a random vector with multinomial distribution \( \mathcal{B}(n,p_1,...,p_m) \):
\[
\mathbb{P}(\bar{S}_n = \bar{l}) = \frac{n!}{l_1!...l_m!(n-l)!}p_1^{l_1}...p_m^{l_m}(1-p)^{n-l},
\]
(38)
where \( l_i \in \mathbb{Z}_+ \) (\( \forall i \)), \( \bar{l} = (l_1,...,l_m) \), \( l = l_1 + ... + l_m \leq n \), \( p = p_1 + ... + p_m \).

Formula (38) describes, in particular, the joint distribution of the increments of the empirical d.f..

Note that
\[
\bar{S}_n \overset{d}{=} \xi_1 + ... + \xi_n,
\]
(39)
where \( \xi, \xi_1, ..., \xi_n \) are i.i.d. random vectors with the distribution
\[
\mathbb{P}(\xi = 0) = 1 - p, \ \mathbb{P}(\xi = e_j) = p_j \quad (1 \leq j \leq m),
\]
where \( e_j \) has the \( j^{th} \) coordinate equal to 1 and the other coordinates equal to 0.

Let
\[
\bar{\pi} = (\pi_1, ..., \pi_m)
\]
be a vector of independent Poisson r.v.s with parameters \( np_1, ..., np_m \), and let \( \pi_n(\cdot) \) denote a Poisson jump process on \([0;1]\) with intensity rate \( n \). Then \( \bar{\pi} \) is a vector of increments of process \( \pi_n(\cdot) \): \( \pi_1 \overset{d}{=} \pi_n(p_1), ..., \pi_m \overset{d}{=} \pi_n(p) - \pi_n(p-p_m) \). Note that
\[
\mathbb{P}(\bar{S}_n = \bar{l}) = \mathbb{P} (\pi_n(p_1)=l_1, ..., \pi_n(p)-\pi_n(p-p_m)=l_n | \pi_n(1)=1) \quad (38)\]
(cf. (20)).

Arenbaev [6] has shown that
\[
d_{TV}(\bar{S}_n; \bar{\pi}) = p / \sqrt{2\pi e} \left( 1 + O(1\wedge 1/\sqrt{np}) \right)
\]
(40)
if \( n \to \infty \) (the term \( 1/\sqrt{np} \) in (40) apparently needs to be replaced with \( p + 1/\sqrt{np} \), cf. (33)). Arenbaev [6], (5)–(9'), has shown also that
\[
d_{TV}(\bar{S}_n; \bar{\pi}) = d_{TV}(\mathcal{B}(n,p); \Pi(np)).
\]
(41)
Using (41) and (31), we deduce
\[ d_{TV}(\bar{S}_n; \bar{\pi}) \leq 3p/4 + 4(1 - e^{-np})p^2. \] (42)

According to Deheuvels & Pfeifer [30],
\[ |d_{TV}(\bar{S}_n; \bar{\pi}) - K_{n,\lambda}| \leq \max\{16p^2; 5np^3\}, \] (43)
where
\[ K_{n,\lambda} = np^2e^{-np}((np)^{\alpha-np}(\alpha-np)/\alpha! - (np)^{\beta-np}(\beta-np)/\beta!)/2, \]
\[ \alpha = np+1/2 + \sqrt{np+1/4}, \beta = np+1/2 - \sqrt{np+1/4}. \]

The case of non-identically distributed random vectors \( \bar{\xi}_1, \ldots, \bar{\xi}_n \) has been treated by Roos [73]. A generalisation of (42) to the case of a stationary sequence of dependent r.v.s is given in [61], Theorem 6.8.

**Shifted Poisson approximation.** Shifted Poisson approximation to \( B(n,p) \) has been considered by a number of authors (see [16, 18, 62] and references therein). The accuracy of shifted Poisson approximation can be sharper than that of pure Poisson approximation. Another advantage of using shifted Poisson approximation is the possibility to derive a more general result (e.g., a uniform in \( p \) estimate of \( d_{TV}(B(n,p); \Pi(np)) \)). We present such a result below.

Set \( [x] := \max\{k \in \mathbb{Z} : k \leq x\}, \{x\} := x - [x], q = 1-p, \) where \( 0 < p < 1, \) and let
\[ P_{n,p} = \mathcal{L}([np^2] + \pi_{npq + np^2}). \]

If random variable \( Y \) has distribution \( \mathcal{L}(Y) = P_{n,p}, \) then
\[ \mathbb{E}Y = np, \text{ var} Y = npq + \{np^2\}. \]

Note that \( \text{var} \pi_{np} - \text{var} S_n = np^2, \text{ var} Y - \text{var} S_n < 1. \) One can expect \( B(n,p) \) be better approximated by \( P_{n,p} \) than by \( \Pi(np) \) if \( np^2 > 1. \) Theorem 5 below confirms that expectation.

**Theorem 5** [62] **As** \( n > 16, \)
\[ \sup_{0 \leq p \leq 1/2} d_{TV}(B(n,p); P_{n,p}) \leq (3 - 4/\sqrt{n})/(\sqrt{n} - 4). \] (44)

Estimate (44) holds if its r.-h.s. is replaced with \( \sqrt{2/n (1+o(1))}. \)

Theorem 5 can be compared with the Berry–Esseen inequality
\[ d_K(B(n,p); N(np, npq)) \leq C/\sqrt{np}, \]
(see, e.g., [77]) as well as with the results by Meshalkin [55] and Pressman [64]. Estimate (14) is uniform in \( p \in [0; 1/2] \). Note that a uniform in \( p \in [0; 1/2] \) Berry–Esseen estimate would be infinite. Inequality (14) has advantages over Meshalkin’s [55] and Pressman’s [64] results as only estimates with explicit constants matter in applications; besides, the structure of the approximating distribution in (14) is simpler and does not involve negative numbers. Bound (14) is preferable to (28) – (31) if \( p > 4e/\sqrt{n} \).

An estimate of the accuracy of shifted Poisson approximation to the distribution of a sum of Bernoulli \( B(p) \) r.v.s in terms of the Gini-Kantorovich distance is given by Barbour & Xia [18].

**Open problem.**

2.1. Improve the constant in (32).

2.2. Generalise Theorem 5 to the case of \( m \)-dependent r.v.s.

### 2.2 Dependent Bernoulli r.v.s

Let \( X_1, ..., X_n \) be (possibly dependent) Bernoulli r.v.s. Set \( p_i = \mathbb{P}(X_i = 1|X_1, ..., X_{i-1}) \). A generalisation of (24) has been given by Serfling [74]:

\[
d_{TV}(S_n; \pi_\lambda) \leq \sum_{i=1}^n (\mathbb{E}p_i)^2 + \sum_{i=1}^n |\mathbb{E}p_i - \mathbb{E}p_i|, \tag{24}
\]

\[
d_K(S_n; \pi_\lambda) \leq \frac{2}{\pi} \sum_{i=1}^n (\mathbb{E}p_i)^2 + \sum_{i=1}^n |\mathbb{E}p_i - \mathbb{E}p_i|. \tag{45}
\]

We present now a generalisation of (28) to the case of dependent Bernoulli r.v.s.

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

The idea of splitting the sample into “strongly dependent” and “almost independent” parts goes back to Bernstein [19] (see also [75]).

Denote

\[
S = \sum_{a \in J} X_a, \quad \lambda = \mathbb{E}S,
\]

and let

\[
\delta_1 = \sum_{a \in J} \sum_{b \in B_a} \mathbb{E}X_a \mathbb{E}X_b, \quad \delta_2 = \sum_{a \in J} \sum_{b \in B_a \setminus \{a\}} \mathbb{E}X_a X_b,
\]

\[
\delta_3 = \sum_{a \in J} \mathbb{E} \left| \mathbb{E}X_a - \mathbb{E} \left\{ X_a \sum_{b \in J \setminus B_a} X_b \right\} \right|.
\]

The following Theorem 6 is due to Arratia et al. [2] and Smith [79].
Theorem 6  There holds
\[ d_{TV}(S; \pi_\lambda) \leq \frac{1-e^{-\lambda}}{\lambda} \left( \delta_1 + \delta_2 \right) + \min\{1; \sqrt{2/e\lambda}\} \delta_3. \]  (46)

In the case of independent random variables one can choose \( B_a = \{a\} \), then (46) coincides with (28).

Theorem 6 has applications to the problem of Poisson approximation to the distribution of the number of long head runs in a sequence of Bernoulli r.v.s, and to the problem of Poisson approximation to the distribution of the number of long match patterns in two sequences (e.g., DNA sequences, see [12, 61] and references therein).

The topic concerning \( \mathcal{L}(S_n) \) in the case of stationary dependent r.v.s \( \{X_i\} \) has applications in extreme value theory [48, 61]. The case where the sequence \( X_1, ..., X_n \) is a moving average is related to the topic concerning the so-called Erdős–Rényi maximum of partial sums ([61], ch. 2).

A generalization of Theorem 6 to the case of compound Poisson approximation has been given by Roos [69].

Open problem.

2.3. Improve the constant before the term \( (\delta_1 + \delta_2) \) in (46).

2.3  Independent integer-valued r.v.s

The topic of Poisson approximation to the distribution of a sum of integer-valued r.v.s has applications in extreme value theory, insurance, reliability theory, etc. (cf. [7, 12, 48, 61]).

For instance, in insurance applications the sum \( S_n = \sum_{i=1}^{n} Y_i \mathbb{I}\{Y_i > y_i\} \) of integer-valued r.v.s allows to account for the total loss from the claims exceeding excesses \( \{y_i\} \). One would be interested if Poisson approximation to \( \mathcal{L}(S_n) \) is applicable.

In extreme value theory one often deals with the number of extreme (rare) events represented by a sum \( S_n = \xi_1 + ... + \xi_n \) of 0-1 r.v.s (indicators of rare events). The r.v.s \( \xi_1, ..., \xi_n \) can be dependent. One way to cope with dependence is to split the sample into blocks, which can be considered almost independent (the so-called Bernstein’s blocks approach [19]). The number of r.v.s in a block is an integer-valued r.v.; thus, the number of rare events is a sum of almost independent integer-valued r.v.s.

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

The problem of evaluating the accuracy of Poisson approximation to the distribution of a sum of independent non-negative integer-valued r.v.s has been considered, e.g., in [10, 11, 61]. LeCam’s inequality [24] and the Barbour-Eagleson inequality (28) has been
generalised to the case of non-negative integer-valued r.v.s by Barbour [10]. Theorem 7 below presents another result of that kind ([61], ch. 4).

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

\[ S_n = X_1 + ... + X_n, \quad \lambda = \mathbb{E} S_n, \]

\( \pi_\lambda \) denotes a Poisson \( \Pi(\lambda) \) r.v..

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

\[ \mathbb{P}(Y^* = m) = (m+1)\mathbb{P}(Y = m+1)/\mathbb{E}Y \quad (m \geq 0). \]  

Distribution (47) differs by a shift from the distribution introduced by Stein [80], p. 171. Note that \( Y^* \overset{d}{=} Y \) if and only if \( \mathcal{L}(Y) \) is Poisson.

**Theorem 7** As \( n \geq 1, \)

\[
\begin{align*}
  d_{TV}(S_n; \pi_\lambda) &\leq \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^{n} d_G(X_i; X_i^*)\mathbb{E}X_i, \quad (48) \\
  d_G(S_n; \pi_\lambda) &\leq \min\left\{ 1; \frac{4}{3} \sqrt{2/e\lambda} \right\} \sum_{i=1}^{n} d_G(X_i; X_i^*)\mathbb{E}X_i. \quad (49)
\end{align*}
\]

In the case of Bernoulli \( B(p_i) \) r.v.s one has \( X_i^* \equiv 0 \), and (48) coincides with (28). In the case of i.i.d.r.v.s (48) becomes

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

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

**Example 2.2.** Let \( \xi, X_1, X_2, ... \) be i.i.d.r.v.s with geometric \( \Gamma_0(p) \) distribution:

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

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

\[ X_i^* \overset{d}{=} X_i + \xi, \quad (50) \]

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

\[ \lambda = n\mathbb{E}\xi = np/(1-p), \quad d_G(X; X^*) = \mathbb{E}\xi. \]
Theorem 1 entails
\[ d_{TV}(S_n; \pi_x) \leq (1 - e^{-np/(1-p)}) p/(1-p), \quad (51) \]
\[ d_{C}(S_n; \pi_x) \leq \min \left\{ 1; \frac{4}{\pi^2} \sqrt{2/cnp} \right\} np^2/(1-p)^2. \quad (52) \]

Inequality (51) is available in [10], p. 758. Estimate (52) is formula (4.53) in [61].

**Shifted Poisson approximation.** A number of authors dealt with shifted Poisson approximation to the distribution of a sum \( S_n \) of integer-valued r.v.s (see [16, 62] and references therein). Let
\[ \sigma^2 = \text{var} S_n, \quad a = \lfloor ES_n - \sigma^2 \rfloor, \quad b = \{ ES_n - \sigma^2 \}, \quad \mu = \sigma^2 + b, \]
where \([x]\) and \( \{x\} = x - [x] \) denote the integer and the fractional parts of \( x \).

Barbour & Chekanavichius [16] have shown that
\[ d_{TV}(S_n; a + \pi_x) \leq (1 \wedge \sigma^{-2}) \left( b + d_n \sum_{i=1}^n \psi_i \right) + \mathbb{P}(S_n < a), \quad (53) \]
where \( d_n = \max_{i \leq n} d_{TV}(S_{n,i}; S_{n,i}+1) \), \( S_{n,i} = S_n - X_i \), \( \psi_i = \sigma_i^2 \mathbb{E}[X_i(X_i-1)] + \|\mathbb{E}[X_i - \sigma_i^2]\| \mathbb{E}(X_i-1)(X_i-2) + \|\mathbb{E}X_i(X_i-1)(X_i-2)\|, \sigma_i^2 = \text{var} X_i. \)

In the case of the Binomial distribution \( B(n, p) \) the r.-h.s. of (53) is \( O(p/\sqrt{np} + 1/np) \); the term \( 1/np \) can be improved to \( O(p \wedge np^2) \) if \( np^2 < 1 \).

### 2.4 Dependent integer-valued r.v.s

Let \( X_1, \ldots, X_n \) be (possibly dependent) non-negative integer-valued r.v.s. Set \( p_i = \mathbb{P}(X_i=1|X_1, \ldots, X_{i-1}). \) A generalisation of (24), (45) has been given by Serfling [74]:
\[ d_{TV}(S_n; \pi_x) \leq \sum_{i=1}^n \left( \mathbb{E}^2 p_i + \mathbb{E}|p_i - \mathbb{E}p_i| + \mathbb{P}(X_i \geq 2) \right), \quad (24') \]
\[ d_{K}(S_n; \pi_x) \leq \sum_{i=1}^n \left( \frac{2}{\pi^2} \mathbb{E}^2 p_i + \mathbb{E}|p_i - \mathbb{E}p_i| + \mathbb{P}(X_i \geq 2) \right). \quad (45') \]

Below we present a generalisation of Theorem 6.

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

Let \( \mathcal{L}(\pi_x) \) denote a Poisson \( \Pi(\lambda) \) r.v.. Set
\[ \delta^*_1 = \sum_{a \in J} \sum_{b \in B_a \setminus \{a\}} \mathbb{E}X_a\mathbb{E}X_b, \quad \delta_4 = \sum_{a \in J} d_{G}(X_a; X^*_a)\mathbb{E}X_a, \]
and let \( \delta_1, \delta_2, \delta_3 \) be defined as in section 2.2. Theorems 8 and 9 are from [61], ch. 4.
Theorem 8 If \( \{X_b, b \in J \setminus B_a\} \) are independent of \( X_a \), then
\[
d_{TV}(S_n; \pi_\lambda) \leq \frac{1-e^{-\lambda}}{\lambda} (\delta_1^* + \delta_2 + \delta_4) .
\] (54)

In Theorem 9 we drop the local dependence condition assumed in Theorem 8.

Theorem 9 Denote \( \delta_5 = \sum_{a \in J} \mathbb{E}X_a(X_a-1)1\{X_a \geq 2\} \). Then
\[
d_{TV}(S_n; \pi_\lambda) \leq \frac{1-e^{-\lambda}}{\lambda} (\delta_1 + \delta_2 + \delta_5) + \min\{1; \sqrt{2/e \lambda}\} \delta_3 .
\] (55)

Open problem.

2.4. Improve the constants in (55).

2.5 Asymptotic expansions

Let \( X_1, \ldots, X_n \) be independent Bernoulli \( \mathcal{B}(p_i) \) r.v.s. Shorgin [78] presents asymptotic expansions of \( \mathbb{P}(S_n \leq x) \) in the case of non-i.i.d. \( \{X_i\} \). Asymptotic expansions to \( \mathbb{P}(S_n \in A), A \subset \mathbb{Z}_+ \), and \( \mathbb{E}h(S_n) \), where function \( h \) obeys certain restrictions, are given by Barbour [10].

The formulation of the full asymptotic expansions is cumbersome and will be omitted.

We present below the first-order asymptotics of \( \mathbb{E}h(S_n) \) for particular classes of functions \( h \).

Of special interest are indicator functions \( h(\cdot) = 1_{\{\cdot \in A\}}, A \subset \mathbb{Z}_+ \). Denote
\[
Q_\lambda(A) = \left[ \mathbb{P}(\pi_\lambda \in A) + \mathbb{P}(\pi_\lambda + 2 \in A) - 2\mathbb{P}(\pi_\lambda + 1 \in A) \right] / 2 ,
\]
where \( \pi_\lambda \) is a Poisson random variable.

Let \( \pi_\lambda^* \) denote a random variable with distribution (34). Then
\[
Q_\lambda(A) = [\mathbb{P}(\pi_\lambda^* \in A) - \mathbb{P}(\pi_\lambda + 1 \in A)] / 2\lambda
\]
(see [61], ch. 4). The following result from [61], ch. 4, sharpens Corollary 2.4 in [10].

Theorem 10 Let \( X_1, \ldots, X_n \) be independent Bernoulli r.v.s, \( \mathcal{L}(X_i) = \mathcal{B}(p_i) \). Then
\[
\left| \mathbb{P}(S_n \in A) - \mathbb{P}(\pi_\lambda \in A) + Q_\lambda(A) \sum_{i=1}^n p_i^2 \right| \leq 2\delta^* \varepsilon + 2\delta^2 ,
\] (56)
where \( \delta = \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n p_i^2 \), \( \delta^* = \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n p_i^3 \).
Asymptotic expansions for \( \mathbb{E}h(S_n) - \mathbb{E}h(\pi) \) in the case of independent 0-1 r.v.s \( \{X_k\} \) and unbounded function \( h \) are given by Barbour et al. [13] and Borisov & Ruzankin [20]. Denote
\[
\Delta h(\cdot) = g(\cdot + 1) - g(\cdot), \; \lambda_k = \sum_{i=1}^{n} p_i^k \; (k \geq 2).
\]

**Theorem 11** [20] *If \( \mathbb{E}|h(\pi)|^4 < \infty \), then*
\[
\left| \mathbb{E}h(S_n) - \mathbb{E}h(\pi) + \lambda_2 \mathbb{E}\Delta^2 h(\pi)/2 \right|
\leq \frac{p^*_n}{(1-p^*_n)^2} \left( \lambda_3 \mathbb{E}|\Delta^3 h(\pi)|/3 + \lambda_2^2 \mathbb{E}|\Delta^4 h(\pi)|/8 \right). \tag{57}
\]

Borisov & Ruzankin [20], Lemma 2, have showed also that
\[
\sup_k \mathbb{P}(S_n = k)/\mathbb{P}(\pi = k) \leq (1-p^*_n)^2.
\]

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

### 2.6 Sum of a random number of random variables

Let \( \nu, X, X_1, X_2, \ldots \) be independent non-negative random variables, where r.v. \( \nu \) takes values in \( \mathbb{Z}_+ \), \( x, x_1, x_2, \ldots \) are i.i.d. random variables.

Set \( S_\nu = X_1 + \ldots + X_\nu \). A natural task is to evaluate the accuracy of Poisson approximation to \( \mathcal{L}(S_\nu) \).

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

Denote \( \bar{\nu} := \mathbb{E}\nu \). Then \( \mathbb{E}S_\nu = p\bar{\nu} \). Using (60) and inequalities
\[
d_{TV}(S_\nu; \pi) \leq \sum_k \mathbb{P}(\nu = k)d_{TV}(S_k; \pi), \tag{58}
\]
\[
d_{TV}(S_k; \pi) \leq d_{TV}(S_k; \pi_k) + d_{TV}(\pi_k; \pi),
\]
Yannaros [88] has shown that
\[
d_{TV}(S_\nu; \pi_{\bar{\nu}}) \leq \min \left\{ \frac{p}{2\sqrt{1-p}}; (1 - \mathbb{E}e^{-p\nu})p \right\} + \min \left\{ p\mathbb{E}|\nu - \bar{\nu}|; \frac{1}{2} \sqrt{p\text{var} \nu / \bar{\nu}} \right\}. \tag{59}
\]

The term \( \min \{ p/2\sqrt{1-p}; (1-\mathbb{E}e^{-p\nu})p \} \) in (59) has been inherited from (27) and (28). The right-hand side of (59) can be sharpened using (31):
\[
d_{TV}(S_\nu; \pi_{\bar{\nu}}) \leq 3p/4e + 2\delta^* \varepsilon + 2\delta^2 + \min \left\{ p\mathbb{E}|\nu - \bar{\nu}|; \frac{1}{2} \sqrt{p\text{var} \nu / \bar{\nu}} \right\}. \tag{59'}
\]
The derivation of the second term in (59) involves the following inequality:

\[ d_{TV}(\pi_\lambda; \pi_\mu) \leq \min\{|\sqrt{\lambda} - \sqrt{\mu}|; |\lambda - \mu|\}. \]  

(60)

The proof of (60) is based on noticing that \( \sum_k \sqrt{\text{IP}(\pi_\lambda = k)\text{IP}(\pi_\mu = k)} = e^{\sqrt{\lambda} - (\lambda + \mu)/2} \) and using the relation (19) between the Hellinger and the total variation distances.

We now consider the situation where r.v. \( \nu \) depends on \( \{X_i\} \).

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

\[ S_n = X_1 + \ldots + X_n \quad (n \geq 1), \]

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

\[ \mu(t) = \max\{n \geq 0 : S_n \leq t\}. \]

Theorems 12–13 below are cited from see [61], ch. 3. They provide estimates of the accuracy of Poisson approximation to the distribution of the number

\[ N_t(x) = \sum_{j=1}^{\mu(t)} \mathbb{I}\{X_j \geq x\} + \mathbb{I}\{t - S_{\mu(t)} \geq x\} \]  

(61)

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

Note that

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

where

\[ M_t = \max\{t - S_{\mu(t)}; \max_{1 \leq i \leq \mu(t)} X_i\} \]  

(62)

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

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

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

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

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

\[ \mathcal{L}(X^<) = \mathcal{L}(X \mid X < x), \mathcal{L}(X^>) = \mathcal{L}(X \mid X \geq x). \]
We set \( p_x = \mathbb{P}(X \geq x) \),

\[
S_0(k) = 0, \quad S_m(k) = \sum_{i=0}^{k} X_i^\geq + \sum_{i=k+1}^{m} X_i^\leq \quad (m \geq 1).
\]

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

\[
\tau_k = \tau_k' - k, \quad \tau_k' = \min\{n : S_n(k) > t-x\},
\]

\[
\lambda_k = \lambda_k(t, x, k) = p_x(t-x-k\mathbb{E}X^\geq)/\mathbb{E}X^\leq.
\]

In Theorems 12–13 we assume the following condition:

there exist constants \( D < \infty \) and \( D_* \in (K_*; K^*) \) such that

\[
\int_x^{\infty} \mathbb{P}(X \geq y)dy \leq D \mathbb{P}(X \geq x) \quad (x \geq D_*).
\]

Condition (63) means the tail of \( \mathcal{L}(X) \) is light (cf. (3.15) in [61]). Inequality (63) holds if function \( g(x) = e^{cx} \mathbb{P}(X \geq x) \) is not increasing as \( x > 1/c \) (\( \exists c > 0 \)). The equality in (63) for all \( x \geq 0 \) may be attained only if \( \mathcal{L}(X) \) is exponential with \( \mathbb{E}X = D \).

**Theorem 12** For any \( k \in \mathbb{Z}_+ \), as \( t \to \infty \),

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

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

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

**Theorem 13** For any \( k \in \mathbb{Z}_+ \), as \( t \to \infty \),

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

One can show that \( N_t(x) \) is “small” when \( x \) is “large”:

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

Theorem 3.7 in [61] presents asymptotic expansions for \( \mathbb{P}(N_\pm(t) = k) \). Note that the asymptotic expansions for \( \mathcal{L}(M_t) \) have been established under a weaker moment assumption (cf. [61], ch. 3).
The number of intervals between consecutive jumps of a Poisson process. Consider a Poisson jump process \( \{ \pi_I(s), s \geq 0 \} \) with parameter \( \lambda > 0 \), and let \( \eta_i \) denote the moment of its \( i \)th jump. Set \( X_i = \eta_i - \eta_{i-1} \). Then \( N_t(x) \) is the number of intervals between consecutive jumps with lengths greater or equal to \( x \). If the points of jumps represent catastrophic/rare events, then \( N_t(x) \) can be interpreted as the number of “long” intervals without catastrophes.

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

\[
\sup_{0 < x < t} | \mathbb{P}(N_t(x) = k) - \mathbb{P}(\pi_{t,x} = k) | = O(t^{-1} \ln t) \tag{64}
\]

(cf. (3.12) in [61]).

Open problems. 2.5. Will asymptotic expansions for \( \mathcal{L}(N_x(t)) \) hold under a weaker moment assumption? 2.6. Generalise the results of Theorems 12–13 to the case of

\[
N_t(x) = \sum_{j=1}^{\mu(t)} Y_j \mathbb{I}\{ X_j \geq x \} + Y_{\mu(t)+1} \mathbb{I}\{ t - S_{\mu(t)} \geq x \},
\]

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

3 Applications

Applications of the theory of Poisson approximation to meteorology, reliability theory and extreme value theory have been discussed in [7, 39, 48, 61]. In this section we present a number of results that are not fully covered in existing surveys.

3.1 Long head runs

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

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

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

Denote

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

Then

\[
W_n(k) = \sum_{i=0}^{n-k} \mathbb{I}\{ A_i \} \quad (n \geq k \geq 1)
\]
is the number of head runs of length $\geq k$ among $\xi_1, ..., \xi_n$ (NLHR).

Set
\[ L_n = \max\{k : \xi_{i+1} = ... = \xi_{i+k} = 1 (\exists i \leq n-k)\}. \] (65)

$L_n$ is the length of the longest head run (LLHR) among $X_1, ..., X_n$. Obviously,
\[ \{L_n < k\} = \{W_n(k) = 0\}. \]

The problem of approximating the distribution of LLHR is a topic of active research; it has applications in reliability theory and psychology (cf. \[7, 61\]).

Let $\{\xi_i, i \geq 1\}$ be i.i.d. Bernoulli $B(p)$ r.v.s, $p \in (0; 1)$, and let $\pi_{\lambda}$ denote the Poisson $\Pi(\lambda)$ r.v.. Theorem 6 with $B_i = [i-k; i+k]$ and
\[ \lambda \equiv \lambda(n, k, p) = p^k(1+(n-k)(1-p)) \]
yields the following

**Corollary 14** As $n \geq k \geq 1$,
\[ d_{TV}(W_n(k); \pi_{\lambda}) \leq (1-e^{-\lambda})(2k+1)p^k. \] (66)

An open question is if the estimate (66) can be improved. Note, for instance, that (66) does not yield (68) even for $j=0$.

There is a close relation between $N_t(x)$ and $W_n(k)$. Let $\eta_0 = 0,$
\[ \eta_i = \min\{k > \eta_{i-1} : \xi_k = 0\}, \ X_i = \eta_i - \eta_{i-1} \ (i \geq 1). \]

Then
\[ W_n(k) = \sum_{j=1}^{\mu(t)} \mathbb{I}\{X_j-1 \geq k\} + \mathbb{I}\{n-\eta_{\mu(n)} \geq k\}. \] (67)

Hence
\[ W_{n-1}(k) = N_n(k+1). \]

Denote $\lambda_k = n(1-p)p^k$. Theorem 13 entails

**Corollary 15** For any $j \in \mathbb{Z}_+$, as $n \to \infty$,
\[ \max_{1 \leq k \leq n} |\mathbb{P}(W_n(k) = j) - \mathbb{P}(\pi_{\lambda_k} = j)| = O\left(n^{-1}\ln n\right). \] (68)
According to Theorem 3.13 in [61], the rate $n^{-1} \ln n$ in (68) cannot be improved.

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

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

Pittel [63] has proved a Poisson limit theorem for NLNR (see also Chryssaphinou et al. [27] concerning the case of a Markov chain).

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

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

Theorem [6] with $B_i = [i-k-1; i+k+1]$ yields the following

**Corollary 16** As $n \geq k \geq 1$,

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

The accuracy of compound Poisson approximation to the distribution of the number of non-decreasing runs of fixed length has been evaluated by Barbour & Chryssaphinou [15], p. 982 (continuous d.f.) and Minakov [59] (discrete d.f.).

**Open problem.**

3.1. Improve the estimates of Corollary [14] and Corollary [16].

3.2. Derive $(44)$-type (i.e., uniform in $k$) estimates of the accuracy of (possibly shifted) Poisson approximation to $\mathcal{L}(W_n(k))$ and $\mathcal{L}(W_n^+(k))$.

**3.2 Long match patterns**

Closely related to the number of long head runs is the number of long match patterns (NLMP) between sequences of independent r.v.s. Information on the distribution of NLMP and the length of the longest match pattern (LLMP) can help recognising “valuable” fragments of DNA sequences (see [2, 3, 58, 60]).

In this section we present results on the accuracy of Poisson approximation to the distribution of NLMP. Theorems [17] [19] and Lemma [21] below have been established by the author (see [61], ch. 4).
Let $X, X_1, ..., X_m, Y, Y_1, ..., Y_n$ be independent non-degenerate random variables taking values in a discrete state space $A$. Denote $(k \in \mathbb{N})$

$$T_{ij} = \mathbb{I}\{X_{i+1} = Y_{j+1}, \ldots, X_{i+k} = Y_{j+k}\},$$

$$\tilde{T}_{ij} = T_{ij}(k) \mathbb{I}\{X_i \neq Y_j\},$$

$$T^*_{ij} = \tilde{T}_{ij} \ (i \geq 1, j \geq 1), \ T^*_{ij} = T_{ij} \ (i = 0 \text{ or } j = 0).$$

Then

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

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

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

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

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

Denote by

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

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

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

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

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

Then

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

Denote

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

and let

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

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

$$p_*^2 < p, \ p^2 \leq q \leq p_*.$$

Taking into account Hölder’s inequality, we conclude that

$$1 \geq c_+ \geq c_* > 1/2.$$  (69)
Note that $c_+ = c_* = 1$ if $\mathcal{L}(X)$ is uniform over a finite alphabet.

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

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

**Theorem 17** If $n \geq k$ and $m \geq k \geq 1$, then

$$d_{TV}(W_{m,n}; \pi_{m,n}) \leq \frac{1-e^{-\lambda}}{\lambda} m'n' (2k+1) (2kq_{2k} + (m' + n' - 1)(p_{2k}^2 + q_{2k}^2)).$$ (71)

Theorem 17 has been derived using Theorem 6 and Lemma 21.

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

**Corollary 18** For any constant $C \in \mathbb{R}$, as $m \to \infty$, $n \to \infty$,

$$\max_{k \geq C + \log mn} \Delta_{m,n}(k) = O\left((m+n)(mn)^{-c_+} (\ln mn) + (mn)^{1-2c_*} (\ln mn)^2\right).$$ (72)

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

$$\max_{1 \leq k \leq m \wedge n} \Delta_{m,n}(k) = O\left((m+n)(mn)^{-c_+} (\ln mn)^{1+c_*} + (mn)^{1-2c_*} (\ln mn)^{1+2c_*}\right).$$ (73)

It is easy to see that the accuracy of estimate (72) depends on the relation between $m$ and $n$. If $\mathcal{L}(X)$ is uniform over a finite alphabet and $(\ln mn)/(m \wedge n) \to 0$, then Corollary 18 implies that

$$\max_{1 \leq k \leq m \wedge n} \left|\mathbb{P}(M^*_{m,n} < k) - e^{-\lambda}\right| = O\left(n^{-1}(\ln n)^2\right)$$ (74)

If $\mathcal{L}(X)$ is uniform over a finite alphabet and $c \leq m/n \leq 1/c$

for some constant $c > 0$, then the right-hand side of (74) becomes $O(n^{-1} \ln n)$. We conjecture that the correct rate of convergence in (74) for the uniform $\mathcal{L}(X)$ is $O(n^{-1} \ln n)$.

The reason why (71) does not yield such a rate is the lack of factor $e^{-\lambda}$ on the right-hand side. Results obtained for LLHR by the method of recurrent inequalities do produce such a factor (cf. Theorem 3.12 in [61]).

In a more general situation one can consider NLMP with say $r$ mismatches allowed. An estimate of the accuracy of Poisson approximation to the distribution of the number of long $r$-interrupted match patterns among $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ (match patterns of length $\geq k$ with $\leq r$ “interruptions”) can be found in [58, 61].
The Zubkov–Mihailov statistic. Let now \( Y_i = X_i \) (\( \forall i \)), \( m = n \). Denote
\[
N^*_n \equiv N^*_n(k) = \sum_{(i,j) \in A(n,k)} T^*_ij,
\]
where
\[
A(n,k) = \{(i,j) : 0 \leq i < j \leq n-k \} \quad (n>k).
\]
\( N^*_n \) is the number of long match patterns in one and the same sequence, \( X_1, ..., X_n \).

Statistic \( N^*_n \) was introduced by Zubkov & Mihailov \[93\] who have shown that \( \mathcal{L}(N^*_n) \) is asymptotically Poisson \( \Pi(\mu) \) if
\[
n^2 p^k (1-p)/2 \to \mu > 0, \quad nk^t p^k \to 0 \quad (\forall t>0).
\]

Note that \( M^*_n = \max\{k \leq n: \max_{(i,j) \in A(n,k)} T_{ij} = 1\} \) is the length of the longest match pattern among \( X_1, ..., X_n \). Obviously,
\[
\{M^*_n<k\} = \{N^*_n = 0\}.
\]

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

**Theorem 19** If \( n > 3k \geq 3 \), then
\[
d_{TV}(N^*_n, \pi^*_n(k)) \leq \frac{1-e^{-\lambda^*}}{\lambda^*} \left((n^*)^3(2k+1)(p^{2k}+q^{k}) + 2(kn^*)^2 q_{2k}\right) + 2kn^* p^k,
\]
where \( \lambda^* \equiv \lambda^*_{n,k} = (n-3k+1) p^k (1+(n-3k)(1-p)/2), \quad n^* = n-k, \quad \mathcal{L}(\pi^*_n(k)) = \Pi(\lambda^*) \).

Theorem 19 has been derived using Theorem 6 and Lemma 21.

Denote
\[
\Delta^*(n,k) = |\mathbb{P}(M^*_n < k) - \exp(-\lambda^*_{n,k})|.
\]

**Corollary 20** As \( n \to \infty \),
\[
\max_{k \geq C+2\log n} \Delta^*(n,k) = O \left(n^{1-2c+\ln n} + n^{2-4c}(\ln n)^2\right), \quad (75)
\]
\[
\max_{1 \leq k < n/3} \Delta^*(n,k) = O \left(n^{1-2c}(\ln n)^{1+c} + n^{2-4c}(\ln n)^{1+2c}\right). \quad (76)
\]

If \( \mathcal{L}(X) \) is uniform over a finite alphabet, then the right-hand side of (75) is \( O(n^{-1} \ln n) \), the right-hand side of (76) is \( O(n^{-1}(\ln n)^2) \).

The key result behind Theorems 17 and 19 is the following
Lemma 21 For all natural $i, j, i', j'$ such that $(i, j) \neq (i', j')$,
\[ P(T_{ij}^* = T_{i'j'}^* = 1) \leq q_{2k}. \tag{77} \]

Denote by
\[ \tau_k = \min\{n : N_n^*(k) \neq 0\} \]
the first instance a match pattern of length $k$ appears in the sequence $\{X_i, i \geq 1\}$. Then
\[ \{\tau_k > n\} = \{M_n^* < k\}. \]

The results on the asymptotics of $\tau_k$ can be derived from the corresponding results on $M_n^*$.

NLMP with a small number of mismatches has been considered by several authors (see [58, 61] and references therein).

A number of authors evaluated the accuracy of compound Poisson approximations to the distribution of NLMP (see [58, 61, 76] and references therein).

Open problems.
3.2. Derive uniform in $k$ estimates of (possibly shifted) Poisson approximation to $L(W_{m,n})$ and $L(N_n^*)$.
3.3. Find the 2nd-order asymptotic expansions for $P(W_{m,n} \in \cdot)$ and $P(N_n^* \in \cdot)$.
3.4. Check if the correct rate of convergence in (73) and (76) in the case of uniform $L(X)$ is $O(n^{-1} \ln n)$.
3.5. Improve the estimate of the rate of convergence in the limit theorem for the length of the longest $r$-interrupted match pattern.

4 Compound Poisson approximation

The topic of compound Poisson (CP) approximation is vast. From a theoretical point of view, the interest to the topic arises in connection with Kolmogorov’s problem concerning the accuracy of approximation to the distribution of a sum of independent r.v.s by infinitely divisible laws (see [5, 50, 64, 66] and references therein).

Recall that the class of infinitely divisible distributions coincides with the class of weak limits of compound Poisson distributions [45].

The topic has applications in extreme value theory, insurance, reliability theory, pattern matching, etc. (cf. [7, 12, 48, 61]). For instance, in (re)insurance applications the sum $S_n = \sum_{i=1}^{n} Y_i 1\{Y_i > x_i\}$ 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 \([19]\). The number of r.v.s in a block is an integer-valued r.v., hence the number of rare events is a sum of almost independent integer-valued r.v.s. In all such situations the block sums are non-zero with small probabilities. More information concerning applications can be found in \([7, 12, 31, 48]\).

This section concentrates on results concerning compound Poisson (CP) approximation that can be derived from the results concerning pure Poisson approximation.

4.1 CP limit theorem

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

\[
\sum_{i=1}^{\tau_\lambda} \zeta_i,
\]

where \( \zeta_0 = 0 \), r.v.s \( \tau_\lambda, \zeta_1, \zeta_2, ... \) are independent, \( \mathcal{L}(\zeta) = \Pi(\lambda), \; \zeta \overset{d}{=} \zeta \; (i \geq 1) \).

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

Typically \( \zeta \neq 0 \) w.p. 1. The requirement \( \zeta \neq 0 \) w.p. 1 may be omitted. Indeed, denote \( p = \mathbb{P}(\zeta \neq 0) \). Then by Khintchin’s formula \([45]\), ch. 2,

\[
\zeta \overset{d}{=} \tau_p \zeta', \tag{78}
\]

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

\[
\Pi(t, \tau_p \zeta') = \Pi(tp, \zeta')
\]

(cf. (6.26) in \([61]\)).

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

\[
S_n = X_{n,1} + ... + X_{n,n}.
\]

Let \( \zeta_{r,n} \) be a r.v. with the distribution \([8]\). The following Theorem \([22]\) generalises Theorem \([3]\) to the case of CP approximation. It states that under certain assumptions weak convergence of the cluster size distribution (see \([30]\) below) is necessary and sufficient for the CP limit theorem for \( S_n \).

In Theorem \([22]\) below we will assume \([11]\) and the following condition:

\[
\limsup_{n \to \infty} n \mathbb{P}(X_{n,1} \neq 0) < \infty. \tag{79}
\]
Note that relation \((11)\) does not imply \((79)\) — for example, consider the case \(X_{n,1} \equiv X\). Denzel & O’Brien [31] present an example of an \(\alpha\)-mixing sequence such that \((11)\) holds though \((79)\) does not.

**Theorem 22** Assume conditions \((11), (79)\) and \(\Delta\). If

\[
\zeta_{r,n} \Rightarrow \zeta \quad (n \to \infty)
\]

for a sequence \(\{r = r_n\} \in \mathcal{R}\), then

\[
S_n \Rightarrow \sum_{i=0}^{\pi(\lambda)} \zeta_i.
\]

The limit in \((81)\) does not depend on the choice of a sequence \(\{r_n\} \in \mathcal{R}\).

If \(S_n\) converges weakly to a random variable \(Y\), then \(\mathcal{L}(Y)\) is compound Poisson \(\mathbf{Pi}(\lambda, \zeta)\), where \(\lambda = -\ln \mathbb{P}(Y=0)\). If \(\lambda > 0\), then \((80)\) holds for some random variable \(\zeta\) and sequence \(\{r=r_n\} \in \mathcal{R}\).

Theorem 22 is effectively Theorem 5.1 from [61].

### 4.2 Accuracy of CP approximation

Let \(\{X_i, i \geq 1\}\) be independent r.v.s that are non-zero with small probabilities (cf. 50, 53, 65, 89). Set \(S_n := X_1 + \ldots + X_n\), and denote

\[
p_i = \mathbb{P}(X_i \neq 0) \quad (i \geq 1).
\]

According to Khintchine’s formula \((78)\),

\[
X_i \overset{d}{=} \tau_i X'_i,
\]

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

\[
S_n \overset{d}{=} \tau_1 X'_1 + \ldots + \tau_n X'_n.
\]

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

\[
\mathcal{L}(Z_n) = \mathbf{Pi}(\lambda, X'_\eta),
\]

where r.v. \(\eta\) is independent of \(X'_1, \ldots, X'_n\), \(\mathbb{P}(\eta = j) = p_j / \lambda \quad (1 \leq j \leq n)\).

A simple estimate of the accuracy of CP approximation to \(\mathcal{L}(S_n)\) follows from the property of \(d_{TV}\) and \((25)\):

\[
d_{TV}(S_n; Z_n) \leq \sum_{i=1}^{n} d_{TV}(\tau_i; \pi_{p_i}) \leq \sum_{i=1}^{n} p_i^2
\]
Zaitsev [89] has derived an estimate of the accuracy of compound Poisson approximation that can be sharper than (24) if $\lambda = p_1 + \ldots + p_n$ is “large”. The following Theorem presents Zaitsev’s result.

**Theorem 23** There exists an absolute constant $C$ such that

$$d_K(S_n; Z_n) \leq Cp_n^*.$$  \hfill (82)

Inequality (82) has been generalised to the multidimensional situation by Zaitsev [90].

We consider now the situation where

$$X'_i \overset{d}{=} X' \ (\forall i).$$

Presman [65] was probably the first to notice that in such a situation an estimate of the accuracy of compound Poisson approximation to $L(S_n)$ follows from the estimate of the accuracy of pure Poisson approximation to $L(\tau_1 + \ldots + \tau_n)$.

Indeed, denote

$$\nu_n = \tau_1 + \ldots + \tau_n, \quad Y = \sum_{i=1}^{\nu_n} X'_i,$$

where Poisson $\Pi(\lambda)$ r.v. $\pi_\lambda$ is independent of $X'_1, X'_2, \ldots$. Then

$$S_n \overset{d}{=} \sum_{i=1}^{\nu_n} X'_i.$$

It is easy to check (see, e.g., [65]) that

$$d_{TV}(S_n; Y) \leq d_{TV}(\nu_n; \pi_\lambda).$$  \hfill (83)

Besides, according to [61], Lemma 5.4,

$$d_G(S_n; Y) \leq d_G(\nu; \pi_\lambda)\mathbb{E}|X'|.$$ \hfill (84)

Presman [65] has evaluated $d_{TV}(\nu_n; \pi_\lambda)$ (and hence $d_{TV}(S_n; Y)$) using (83) and (29).

Michel [53] has applied (83) and the Barbour–Eagleson estimate (28).

An application of (83) and (31) yields

$$d_{TV}(S_n; Y) \leq 3p/4e + 2\delta^*\varepsilon + 2\delta^2,$$ \hfill (85)

where $p = \mathbb{P}(X \neq 0)$. A combination of (32) and (84) entails

$$d_G(S_n; Y) \leq \left(1 \wedge \frac{4}{3}\sqrt{2/\mathbb{E}X'}\right)np^2\mathbb{E}|X'|.$$ \hfill (86)

Further results on the accuracy of compound Poisson approximation can be found in [14, 24, 25, 72, 92, 87].

**Open problem.**

4.1 Evaluate constant $C$ in (82).
4.3 CP approximation to \( B(n, p) \)

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

Let \( X, X_1, \ldots \) be independent Bernoulli \( B(p) \) r.v.s. Presman [64] has shown that

\[
\sup_p d_{TV}(B(n, p); F_{n,p}) = O(n^{-2/3}),
\]

(87)

where the compound Poisson distribution \( F_{n,p} \) is constructed via Poisson distributions (a similar result in terms of \( d_K \) is due to Meshalkin [55]).

We present Presman’s result in Theorem 24 below (see also [5], ch. 4).

Denote

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

Let \( \eta_1, \eta_2, \eta_3 \) be independent r.v.s with distributions

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

(multiplication is superior to division). Set

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

Note that \( Y \) is a CP r.v.. One can check that

\[
\mathbb{E}(X - p)^3 = pq(q - p), \quad \mathbb{E}Y = p, \quad \mathbb{E}(Y - p)^2 = pq, \quad \mathbb{E}(Y - p)^3 = pq(q - p).
\]

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

**Theorem 24** There exists an absolute constant \( C \) such that

\[
d_{TV}(B(n, p); F_{n,p}) \leq C\varepsilon_{n,p} \quad (0 \leq p \leq 1/2),
\]

(88)

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

Bound (87) follows after noticing that \( \sup_{0 \leq p \leq 1/2} \varepsilon_{n,p} = O(n^{-2/3}) \).

**Unit measure (signed measure) approximations.** A number of authors evaluated the accuracy of unit measure (signed measure) approximation to the distribution of a sum of independent Bernoulli r.v.s (see, e.g., [22, 16]). In particular, Borovkov [22] has generalised LeCam’s inequality [24]. Barbour & Chekanavichius [16] present a unit measure approximation to the distribution of a sum of independent integer-valued r.v.s. Note that asymptotic expansion (56) is an example of a unit measure approximation.
Dependent 0-1 r.v.s. Let $X, X_1, \ldots$ be a stationary sequence of 0-1 r.v.s. The following Theorem 25 is an application of (83) in the case of dependent r.v.s.

Let $\pi, \zeta^{(r)}_1, \zeta^{(r)}_2, \ldots$ be independent random variables, where $1 \leq r \leq n$, $\pi_{n,r}$ is a Poisson $\Pi(kq)$ r.v., $\zeta^{(r)}_0 = 0$, \[ L(\zeta^{(r)}_i) = L(S_r|S_r > 0) \quad (i \geq 1), \]
\[ q = \mathbb{P}(S_r \neq 0), \quad k = \lfloor n/r \rfloor. \]

Denote $p = \mathbb{P}(X=1)$,
\[ Y_n = \sum_{i=0}^{\pi_{n,r}} \zeta^{(r)}_i. \]

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

**Theorem 25** If $n>r>l \geq 0$, then
\[ d_{TV}(S_n; Y_n) \leq \kappa_{n,r}rp + (2kl + r')p + nr^{-1}\gamma_n(l), \quad (89) \]
\[ d_{G}(S_n; Y_n) \leq rp \min\left\{ np: \frac{4}{3}\sqrt{2np/e} \right\} + (2kl + r')p + n\gamma_n(l), \quad (90) \]
where $r' = n-rk$, $\kappa_{n,r} = \min\{1-e^{-np}; 3/4e+(1-e^{-np})rp\}$ and $\gamma_n(l) = \min\{4\alpha(l)\sqrt{r}; \beta_n(l)\}$.

Theorem 25 is effectively Theorem 5.2 from [61].

If the random variables $\{X_i\}$ are independent, then (89) with $r=1$, $l=0$ yields (28) and (31).

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

For further results on the accuracy of compound Poisson approximation to a sum of dependent r.v.s, see [26] and references therein.

Open problem.

4.2. Evaluate constant $C$ in (88).

5 Poisson process approximation

Point process counting locations of rare events. Let $\{\xi_i, i \geq 1\}$ be Bernoulli r.v.s (e.g., $\xi_i = \mathbb{I}\{X_i > u_n\}$, where $u_n$ is a “high” level). Then
\[ S_n(\cdot) = \sum_{i=1}^{n} \mathbb{I}\{i/n \in \cdot\} \xi_i \quad (91) \]
is called a "Bernoulli process".

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

\[
N_n(B, u_n) = \sum_{i=1}^{n} \mathbb{1}\{i/n \in B, X_i > u_n\} \quad (B \subset (0; 1]).
\]

(91)

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

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

**Theorem 26** Assume (11), (79) and mixing condition \( \Delta \). If (10) holds, then

\[
N_n(\cdot, u_n) \Rightarrow N(\cdot),
\]

(92)

where \( N(\cdot) \) is a Poisson point process with intensity rate \( \lambda \).

The necessity part of Theorem 26 is given by Theorem 3 if (92) holds, then so does (10). The proof of Theorem 26 follows from the proof of Theorem 7.1 in [61].

Leadbetter et al. [48], Theorem 5.2.1, present a version of Theorem 26 with condition (D') instead of (10). Theorem 26 is a particular case of Theorem 7.2 in [61].

The accuracy of Poisson process approximation to \( \mathcal{L}(S_n(\cdot)) \) has been evaluated by Brown [23] and Kabanov et al. [41], Theorem 3.2: if \( \{\xi_i\} \) are independent, then

\[
d_{TV}(S_n(\cdot); \Xi_n(\cdot)) \leq \sum_{i=1}^{n} p_i^2,
\]

(24')

where \( \Xi_n \) is a Poisson point process with intensity measure \( \lambda(\cdot) = \sum_{i=1}^{n} p_i \mathbb{1}\{i/n \in \cdot\} \), \( p_i = \mathbb{P}(\xi_i = 1) \).

Xia [86] presents an estimate of the accuracy of Poisson process approximation in terms of a \( d_{G} \)-type distance.

In the general case (when the limiting distribution of \( \zeta_{r,n} \) is not degenerate) the limiting distribution of \( N_n(\cdot, u_n) \) is necessarily compound Poisson (Hsing et al. [40], see also [61], ch. 7).

**Excess process.** Let \( X, X_1, X_2, \ldots \) be a stationary sequence of observations. A typical example of a rare event is an exceedance of a high threshold.

If 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 \( N_n(\cdot) \). We give the definition of the excess process below.
Suppose there is a sequence \( \{u_n(\cdot), n \geq 1\} \) of functions on \([0; \infty)\) such that function \( u_n(\cdot) \) is strictly decreasing for all large enough \( n, \; u_n(0) = \infty \),

\[
\limsup_{n \to \infty} n \mathbb{P}(X > u_n(t)) < \infty \quad (0 < t < \infty), \tag{93}
\]

\[
\lim_{n \to \infty} \mathbb{P}(M_n \leq u_n(t)) = e^{-t} \quad (t \geq 0), \tag{94}
\]

where \( M_n = \max\{X_1, \ldots, X_n\} \) is the sample maximum. Conditions (93) and (94) mean that \( u_n(\cdot) \) is a “proper” normalising sequence for the sample maximum.

Set \( N^\varepsilon_n(t) = \sum_{i=1}^{n} \mathbb{I}\{X_i > u_n(t)\} \), where \( t > 0 \). Given \( B \subset [0; \infty) \), we call \( \{N^\varepsilon_n(t), t \in B\} \) the excess process.

Process \( N^\varepsilon_n(\cdot) \) describes variability in the heights of the extremes.

Note that \( N^\varepsilon_n(\cdot) \) is the “tail empirical process” for \( Y_{n,1}, \ldots, Y_{n,n}, \) where

\[
Y_{n,i} = u_{n}^{-1}(X_i): \quad N^\varepsilon_n(t) = \sum_{i=1}^{n} \mathbb{I}\{Y_{n,i} < t\}. \tag{95}
\]

There is considerable amount of research on the topic of tail empirical processes (see, e.g., [28] and references therein).

We present necessary and sufficient conditions for the weak convergence of the excess process to a Poisson process in Theorem 27 below (see [61], ch. 7).

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

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

\[
\tau = (t_1, \ldots, t_k), \quad u_n(\tau) = (u_n(t_1), \ldots, u_n(t_k)).
\]

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

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

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

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

The next condition describes the joint distribution of exceedances of several levels.
We say that condition $C'_\tau$ holds if there exists a sequence $\{r_n\} \in \mathbb{R}(\tau)$ such that for every $1 \leq i < j \leq k$ and every $t_i < t_j$ from $\{t_1, \ldots, t_k\}$

(a) $\mathbb{P}(N_r[u_n(t_{i-1}); u_n(t_i)]) = 1$ \~ $\frac{j}{n} (t_i - t_{i-1})$, $\mathbb{P}(N_r[u_n(t_{i-1}); u_n(t_i)]) = j = o(\frac{j}{n})$ $(j \geq 2)$,

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

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

**Theorem 27** Assume mixing condition condition $\Delta$, and let $\pi(\cdot)$ denote a Poisson process with intensity rate 1. Then

$$N_{n}^\varepsilon(\cdot) \Rightarrow \pi(\cdot)$$

if and only if condition $C'$ holds.

**Example 3.** Let $X, X_1, X_2, \ldots$ be i.i.d.r.v.s with the distribution function (d.f.) $F$. Denote $K^* = \sup\{x: F(x) < 1\}$, and assume that

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

as $x \to K^*$ (Gnedenko’s condition [37]). Set $u_n(t) = F_{c}^{-1}(t/n)$, where

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

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

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

where $\gamma_j(t) \overset{d}{=} \mathbb{I}\{\xi < t\}$ and r.v. $\xi$ has uniform $[0; 1]$ distribution.

The accuracy of approximation $N_{n}^\varepsilon(\cdot) \approx N(\cdot)$ can be evaluated as well (cf. Deheuvels & Pfeifer [30], Kabanov & Liptser [42], Novak [61], ch. 8).

Note that [83] is applicable. Given $T > 0$, let $\pi(np)$ denote a Poisson $\Pi(np)$ r.v., where $p = \mathbb{P}(X > u_n(T))$. Let $\eta, \eta_1, \eta_2, \ldots$ be independent of $\pi(np)$ i.i.d. processes with the distribution

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

$(i \geq 1)$. An application of [83] and [31] yields

$$d_{TV}(N_{n}^\varepsilon(\cdot); \sum_{i=1}^{\pi(np)} \eta_i(\cdot)) \leq 3p/4e + 2(1-e^{-np})p^2 \varepsilon + 2(1-e^{-np})^2 p^2,$$  \hspace{1cm} (97)
where \( \varepsilon = \min\{1; (2\pi(n-1)p)^{-1/2} + 2(1-e^{-np})p/(1-1/n)\} \) (61), Theorem 8.3).

Note that \( \sum_{i=1}^{\pi(np)} \eta_i(\cdot) \) is a Poisson process. If \( F_c \) is a continuous decreasing function, then \( \eta(\cdot) \overset{d}{=} \mathbb{I}\{\xi<\cdot\} \), where \( L(\xi) = U[0; 1] \).

Let \( X, X_1, X_2, \ldots \) be i.i.d.r.v.s, and let \( B \subset [0; \infty) \) be a closed set. According to (6.5) in [30] and (41), the total variation distance between \( \{\sum_{i=1}^{n} \mathbb{I}\{Y_{n,i} < t\}, t \in B\} \) and the approximating Poisson process coincides with \( d_{TV}(B(p); \Pi(p)) \), where \( p = \mathbb{P}(Y_{n,1} \in B) \).

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

\[
\{N_{n}^\varepsilon(\cdot), t \leq T\} \Rightarrow \left\{ \sum_{j=1}^{\pi(T)} \gamma_j(t/T), t \leq T \right\},
\]

where \( \pi(\cdot) \) is Poisson \( \Pi(\cdot) \), \( \{\gamma_j(\cdot)\} \) are independent jump processes.

Process \( \{\sum_{j=1}^{\pi(T)} \gamma_j(t)\} \) can be called \textit{Poisson cluster process} or \textit{compound Poisson process of the second order} (regarding the standard CP process as a “compound Poisson process of the first order”).

Necessary and sufficient conditions for the weak convergence of the excess process to a compound Poisson process or a Poisson cluster processes are presented in [61], ch. 7, 8. For an estimate of the total variation distance between \( N_{n}^\varepsilon(\cdot) \) and the approximating process in the case of weakly dependent r.v.s see [61], Theorem 8.3.

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

\[
N_{n}^*(A) := \sum_{i=1}^{n} \mathbb{I}\{(i/n, u_n^{-1}(X_i)) \in A\}.
\]

If \( \{X_i\} \) are i.d.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]). Theorem 28 below presents Adler’s result.

We will need a multilevel version of Leadbetter’s “declustering” condition \((D')\):

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

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

**Theorem 28** If conditions \( \Delta \) and \((D'_+)\) hold, then \( N_{n}^* \) converges weakly to a pure Poisson point process \( N^* \) on \( (0; 1] \times [0; \infty) \) with the Lebesgue intensity measure.
Example 4. Let $Y, Y_1, Y_2, \ldots$ be a sequence of i.i.d.r.v.s with exponential $\mathbb{E}(1)$ distribution, and set

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

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

Let $u \equiv u_n(t) = \ln[t^{-1}n \ln n], \ t > 0$. Then $\mathbb{P}(X > u_n(t)) \sim t/n$, and condition $(D_+')$ holds. According to Theorem 28, $N_n^* \Rightarrow N^*$, the Poisson point process with the Lebesgue intensity measure (cf. [61], ch. 7).

Adler’s result has been generalised to the case of compound Poisson approximation: necessary and sufficient conditions for the weak convergence of $N_n^*$ to a compound Poisson point process can be found in [61], ch. 7.

Necessary and sufficient conditions for the weak convergence of $N_n^*$ to a Poisson cluster process are given in [61], ch. 8.

An estimate of the accuracy of approximation $N_n^*(\cdot) \approx \sum \pi_j(T) \gamma_j(\cdot)$ in terms of the $d_G(X; Y)$-type distance is given in [17].

Open problem.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[19] Bernstein S.N. (1926) Sur l’extension du theoreme limite du calcul des probabilites aux sommes de quantites dependantes. — Math. Annalen, v. 97, 1–59.
[20] Borisov I.S. and Ruzankin P.S. (2002) Poisson approximation for expectations of unbounded functions of independent random variables. — Ann. Probab., v. 30, No 4, 1657–1680.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[35] Gini C. (1914) Di un misura delle relazioni tra le graduatorie di due caratteri. — In: Appendix to Hancini A. Le Elezioni Generali Politiche del 1913 nel Comune di Roma. Rome: Ludovico Cecelini.

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

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

[38] Gnedenko B.V., Kolmogorov A.N. (1954) Limit distributions for sums of independent random variables. Addison-Wesley: Cambridge, MA.
[39] Haight F.A. (1967) Handbook of the Poisson distribution. New York: Wiley.
[40] Hsing T., Hüsler J. and Leadbetter M.R. (1988) On the exceedance point process for stationary sequence. — Probab. Theory Rel. Fields, v. 78, 97–112.
[41] Kabanov Yu.M., Liptser R.Sh. and Shiryaev A.N. (1983) Weak and strong convergence of the distributions of point processes. — Theory Probab. Appl., v. 28, No 2, 288–319.
[42] Kabanov Yu.M., Liptser R.Sh. (1983) On convergence in variation of the distributions of multivariate point processes. Z. Wahrscheinlichkeitstheor. verw. Geb., v. 63, 475–485.
[43] Kantorovich L.V. (1942) On the translocation of mass.— Doklady USSR Acad. Sci., v. 37, No 7-8, 227–229. Trans: Management Sci. (1958) v. 5, No 1, 1–4.
[44] Kerstan J. (1964) Verallgemeinerung eines satzes von Prochorow und Le Cam. Z. Wahrsch. Verw. Gebiete, v. 2, 173–179.
[45] Khintchin A.Y. (1933) Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete. — Berlin: Springer.
[46] Kontoyiannis I., Harremoes P. and Johnson O.T. (2005) Entropy and the law of small numbers. — IEEE Trans. Inform. Theory, v. 51, No 2, 466–472.
[47] Kozulyaev P.A. (1939) Asymptotic analysis of a fundamental formula of Probability Theory. — Acad. Notes Moscow Univ., v. 15, 179–182.
[48] Leadbetter M.R., Lindgren G. and Rootzen H. (1983) Extremes and Related Properties of Random Sequences and Processes. — New York: Springer Verlag.
[49] LeCam L. (1960) An approximation theorem for the Poison binomial distribution. Pacif. J. Math., v. 19, 3, p. 1181–1197.
[50] LeCam L. (1965) On the distribution of sums of independent random variables. — In: Proc. Intern. Res. Sem. Statist. Lab. Univ. California, pp. 179–202. New York: Springer.
[51] Liapunov A.M. (1901) Nouvelle forme du theor’eme sur la limite des probabilites. — Mem. Acad. Imp. Sci. St.–Peterburg, v. 12, 1–24.
[52] Loynes R.M. (1965) Extreme values in uniformly mixing stationary stochastic processes. — Ann. Math. Statist., v. 36, 993–999.
[53] Michel R. (1987) An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. — ASTIN Bulletin, v. 17, 165–169.
[54] Marcinkiewicz J. (1938) Sur les fonctions independantes II. — Func. Mat., v. 30, 349–364.
[55] Meshalkin L.D. (1960) On the approximation of polynomial distributions by infinitely-divisible laws. — Theory Probab. Appl., v. 5, No 1, 114–124.
[56] Michel R. (1987) An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. — ASTIN Bulletin, v. 17, 165–169.
[57] Mihailov V.G. (2001) Estimate of the accuracy of compound Poisson approximation for the distribution of the number of matching patterns. — Theory Probab. Appl., v. 46, No 4, 667–675.
[58] Mihailov V.G. (2008) Poisson-type limit theorem for the number of pairs of almost matching patterns. — Theory Probab. Appl., v. 53, No 1, 59-71.
[59] Minakov A.A. (2015) Compound Poisson approximation of the distribution of the number of monotone runs of fixed length. — Appl. Discrete Math., v. 28, No 2, 21–29.

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

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

[62] Novak S.Y. (2018) On the accuracy of Poisson approximation. (submitted).

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

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

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

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

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

[68] Romanowska M. (1977) A note on the upper bound for the distribution in total variation between the binomial and the Poisson distribution. — Statist. Neerlandica, v. 31, 127–130.

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

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

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

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

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

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

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

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

[77] Shevtsova I.G. (2011) On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands. — arXiv:1111.6554v1
Shorgin S.Y. (1977) Approximation of a generalized binomial distribution. — Theory Probab. Appl., v. 22, No 4, 846–850.

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

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

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

Utev S.A. (1992) Extremal problems, characterisation, and limit theorems of Probability Theory. — DSc Thesis. Novosibirsk: Novosibirsk Inst. Math., 280 pp.

Vallander S.S. (1973) Calculation of the Wasserstein distance between probability distributions on the line. — Theory Probab. Appl., v. 18, No 4, 824-827.

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

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

Xia A. (2005) Stein’s method and Poisson process approximation. — In: An introduction to Stein’s method (A.D.Barbour and L.H.Y.Chen, eds.) Singapore: World Scientific, 115–181.

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

Yannaros N. (1991) Poisson approximation for random sums of Bernoulli random variables. — Statist. Probab. Lett., v. 11, 161165.

Zaitsev A.Yu. (1983) On the accuracy of approximation of distributions of sums of independent random variables, which are non-zero with a small probability, by accompanying laws. — Theory Probab. Appl., v. 28, No 4, 657–669.

Zaitsev A.Yu. (1988) A multidimensional variant of Kolmogorov’s second uniform limit theorem. Theory Prob. Appl., v. 34, 108–128.

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

Zaitsev A.Yu. (2005) Approximation of a sample by a Poisson point process. — J. Math. Sciences, v. 128, No 1, 2556–2563.

Zubkov A.M. and Mihailov V.G. (1979) On the repetitions of s–tuples in a sequence of independent trials. — Theory Probab. Appl., v. 24, No 2, p. 267–279.