On accuracy of multivariate compound Poisson approximation

Report 2000-042

On accuracy of multivariate compound Poisson approximation

S.Y.Novak

ISSN: 1389-2355

S.Y.Novak[1]

On accuracy of multivariate compound Poisson approximation

Abstract

We present multivariate generalisations of some classical results on accuracy of Poisson approximation for the distribution of a sum of 0–1 random variables.

1 Introduction

Let $X, X_1, X_2, \ldots$ be a stationary sequence of dependent random variables (r.v.s). The key object in Extreme Value Theory is the number of exceedances

$$N_n(u) = \sum_{i=1}^{n} \mathbb{I}\{X_i > u\}.$$

Investigation of $N_n(u)$ is motivated by applications in finance, insurance, network modelling, meteorology, etc. (cf. [11, 19]).

In the independent case, $N_n(u)$ has binomial $\mathbf{B}(n,p)$ distribution, where $p = \mathbb{P}(X > u)$. If p is "small" then $\mathcal{L}(N_n(u))$ may be approximated by the Poisson $\mathbf{\Pi}(np)$ distribution. Accuracy of Poisson approximation for a binomial distribution has been investigated by famous authors (see, e.g., [17, 14, 10, 3] and references in [6]). The case of a sum of dependent 0–1 random variables was the subject of [9, 2, 3] (see also references in [3]).

The natural measure of closeness of discrete distributions is the total variation distance (TVD). Recall the definition of the TVD between the distributions of random vectors X and Y taking values in $\mathbf{Z}_+^m$, where $\mathbf{Z}_+ = \mathbb{N} \cup \{0\}$:

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

Let π be a Poisson random variable with the parameter np. According to Barbour and Eagleson [2],

$$d_{TV}(N_n(u);\pi) \le \left(1 - e^{-np}\right) p. \tag{1}$$

This is probably the best universal estimate of the TVD between binomial and Poisson distributions; it improves the results of Prokhorov [17] and LeCam [14]. Sharper bounds are available under extra restrictions (see [10, 20]).

[1]Eurandom, PO Box 513, Eindhoven 5600 MB, Netherlands.

Dependence can cause clustering of extremes, and the Poisson approximation may no longer be valid. It is known that under a mild mixing condition, the limiting distribution of $N_n(u)$ is compound Poisson.

Accuracy of compound Poisson approximation for $\mathcal{L}(N_n(u))$ has been evaluated in [1, 15, 18], among others. The feature of the estimate given in [15] is that it coincides with (1) in the particular case of independent r.v.s.

A natural problem is to investigate the distribution of the vector

$$N_n = (N_n(u_1), ..., N_n(u_m))$$

of the numbers of exceedances given a set of distinct levels $u_1, ..., u_m$. The problem has applications in insurance and finance. For instance, a stationary sequence $\{X_i\}$ of (dependent) random variables can represents claims to an insurance company. Let $N(u_i)$ denote the number of claims exceeding a level u_i. It can be of interest to approximate the probability that the number of claims exceeding u_i equals n_i, $1 \leq i \leq m$. This question can be easily addressed if the distribution of the vector N_n has been approximated.

We show that under natural conditions, the limiting distribution of N_n is necessarily compound Poisson. We evaluate accuracy of multivariate compound Poisson approximation for the distribution of N_n. In particular, we improve the corresponding results of Barbour et al. [4] and Novak [15]. In the case of independent trials, our result yields an estimate of accuracy of multivariate Poisson approximation for a multinomial distribution.

2 Results

We may assume $u_1 > ... > u_m$. Let $\mathcal{F}_{a,b} \equiv \mathcal{F}_{a,b}(u_1, ..., u_m)$ be the σ-field generated by the events $\{X_i > u_j\}$, $a \leq i \leq b, 1 \leq j \leq m$. Denote

$$\begin{aligned} \alpha(l) &\equiv \alpha(l, \{u_1, ..., u_m\}) = \sup |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|, \\ \beta(k) &\equiv \beta(l, \{u_1, ..., u_m\}) = \sup \mathbb{E} \sup_B |\mathbb{P}(B|\mathcal{F}_{1,j}) - \mathbb{P}(B)|, \end{aligned}$$

where the supremum is taken over all $A \in \mathcal{F}_{1,j}$, $B \in \mathcal{F}_{j+l+1,n}$, $j \geq 1$, such that $\mathbb{P}(A) > 0$.

Condition $\Delta_m \equiv \Delta_m\{u_1, ..., u_m\}$ is said to hold if

$$\alpha_n \equiv \alpha(l_n, \{u_1, ..., u_m\}) \to 0$$

for some sequence $\{l_n\} \subset \mathbf{Z}_+$ such that $l_n/n \to 0$ as $n \to \infty$. A vector Y has a multivariate compound Poisson distribution $\mathbf{\Pi}(\lambda, \mathcal{L}(Z))$ if

$$Y = \sum_{i=1}^{\pi} Z_i ,$$

where $Z, Z_1, \ldots$ are i.i.d. random vectors, π is independent of $\{Z_i\}$ and has the Poisson distribution with parameter λ.

Theorem 1 *Assume condition Δ_m, and suppose that $u_m \equiv u_m(n)$ obeys*

$$\limsup n\mathbb{P}(X > u_m) < \infty . \tag{2}$$

If N_n converges weakly to a random vector Y then Y has a multivariate compound Poisson distribution.

Let $\zeta(n), \zeta_1(n), \zeta_2(n), \ldots$ be independent random vectors with the common distribution

$$\mathcal{L}(\zeta(n)) = \mathcal{L}(N_r | N_r(u_m) > 0) , \tag{3}$$

where $r \in \{1, \ldots, n\}$. The proof of Theorem 1 shows that $Y \stackrel{d}{=} \Pi(\lambda, \mathcal{L}(Z))$, where $\lambda = -\lim_{n\to\infty} \ln \mathbb{P}(N_n(u_m) = 0)$ and $\mathcal{L}(\zeta)$ is the weak limit of $\mathcal{L}(\zeta(n))$ for an appropriate sequence $r = r_n$.

Denote

$$p = \mathbb{P}(X > u_m) , \; q = \mathbb{P}(N_r(u_m) > 0) , \; k = [n/r] , \; r' = n - rk ,$$

and let π be a Poisson random variable with parameter kq.

In Theorem 2 below we approximate the distribution of N_n by the multivariate compound Poisson distribution $\mathcal{L}(N)$, where $N = \sum_{i=1}^{\pi} \zeta_i(n)$.

Theorem 2 *If $n > r > l \geq 0$ then*

$$d_{TV}(N_n; N) \leq (1 - e^{-np})rp + (2nr^{-1}l + r')p + nr^{-1} \min\{\beta(l); \kappa(l)\} , \tag{4}$$

where $\kappa(l) = 2(1+2/m)\left\{2^{m-1}m^2\alpha^2(l)\right\}^{1/(2+m)}$ if $m2^{(m-1)/2}\alpha(l) \leq 1$, otherwise $\kappa(l) = 1$.

Barbour et al. [4] evaluated accuracy of compound Poisson approximation for general empirical point processes of exceedances in terms of a weaker Wasserstein–type distance d_W. Concerning the approximation $\mathcal{L}(N_n) \approx \mathcal{L}(N)$, Theorem 3.1 in [4] yields $d_W(N_n; N) \leq \left(1.65(1 - rp)^{-1/2} + e^{rp}\right) rp + 2(2rp + nr^{-1}l)p + nr^{-1}\beta(l)$. In the case $m = 1$ (the 1–dimensional situation), (4) improves a result from [15] (cf. also [1]). If $m = 1$ and the random variables $\{X_i\}$ are independent then (4) with $l = 0, r = 1$ yields (1).

As a consequence of Theorem 2, we derive an estimate of accuracy of multivariate Poisson approximation for a multinomial distribution.

Let $i = (i_1, \ldots, i_m)$, where $i_1 \leq \ldots \leq i_m$. Denote $i^* = (i_1, i_2 - i_1, \ldots, i_m - i_{m-1})$,

$$N_n^* = (N_n(u_1), N_n(u_1, u_2), \ldots, N_n(u_{m-1}, u_m)) ,$$

where $N_n(u,v) = \sum_{i=1}^n \mathbb{I}\{u \geq X_i > v\}$ as $u > v$. Evidently, the distribution of N_n determines that of N_n^* and vice versa.

The statement of Theorem 2 can be reformulated as follows: if $n > r > l \geq 0$ then

$$d_{TV}(N_n^*; N^*) \leq (1 - e^{-np})rp + (2nr^{-1}l + r')p + nr^{-1}\min\{\beta(l); \kappa(l)\}, \qquad (4^*)$$

where $N^* = \sum_{i=1}^{\pi} \zeta_i^*(n)$, random vectors $\zeta^*(n), \zeta_1^*(n), ...$ are independent and have the common distribution $\mathbb{P}(\zeta^*(n) = i^*) = \mathbb{P}(\zeta(n) = i)$.

If the random variables $\{X_i\}$ are independent and $r = 1$ then N_n^* has the multinomial distribution $\mathbf{B}(n, p_1, ..., p_m)$ with parameters $p_1 = \mathbb{P}(X > u_1)$, $p_2 = \mathbb{P}(u_1 \geq X > u_2)$, ..., $p_m = \mathbb{P}(u_{m-1} \geq X > u_m)$:

$$\mathbb{P}(N_n^* = (l_1, .., l_m)) = \frac{n!}{l_1!...l_m!(n-l)!} p_1^{l_1} ... p_m^{l_m} (1-p)^{n-l}, \qquad (5)$$

where $l = l_1 + ... + l_m \leq n$, $p = p_1 + ... + p_m$. Theorem 2 yields an estimate of accuracy of multivariate Poisson approximation for the multinomial distribution $\mathbf{B}(n, p_1, ..., p_m)$.

Corollary 3 *Let $\pi_1, ..., \pi_m$ be independent Poisson random variables with parameters $np_1, ..., np_m$. Denote $Y = (\pi_1, ..., \pi_m)$. If $\mathcal{L}(Y_n) = \mathbf{B}(n, p_1, ..., p_m)$ then*

$$d_{TV}(Y_n; Y) \leq \left(1 - e^{-np}\right) p. \qquad (6)$$

3 Proofs

Proof of Theorem 2 incorporates some ideas from [15] and results of Berbee [5] and Bradley [8].

Denote $\mathbb{I}_i = (\mathbb{I}\{X > u_1\}, ..., \mathbb{I}\{X > u_m\})$, and let

$$N_{r,j} = \sum_{i=jr+1}^{(j+1)r \wedge n} \mathbb{I}_i \qquad (0 \leq j \leq k = [n/r]).$$

Evidently, $N_n = \sum_{j=0}^k N_{r,j}$. Notice that the last block $N_{r,k}$ may be omitted:

$$d_{TV}\left(N_n; \sum_{j=0}^{k-1} N_{r,j}\right) \leq \mathbb{P}(N_{r,k} \neq \bar{0}) \leq r'p.$$

Following Bernstein's "blocks" approach, we subtract a subblock of length l from each block $X_{jr+1}, ..., X_{(j+1)r}$ of length r. Denote

$$N_{r,j}^* = \sum_{i=jr+1}^{(j+1)r-l} \mathbb{I}_i, \; N_n^* = \sum_{j=0}^{k-1} N_{r,j}^* \qquad (0 \leq j < k).$$

Then $\mathbb{P}\left(\sum_{j=0}^{k-1} N_{r,j} \neq \sum_{j=0}^{k-1} N^*_{r,j}\right) \leq k\mathbb{P}\left(N_{r,0} \neq N^*_{r,0}\right) \leq klp$.

Let $\{\hat{N}^*_{r,j}\}$ be independent copies of $N^*_{r,0}$. Denote

$$S_i = \sum_{j=0}^{i-1} N^*_{r,j} + \sum_{j=i+1}^{k-1} \hat{N}^*_{r,j} \qquad (0 < i < k).$$

Notice that $S_j + \hat{N}^*_{r,j} = S_{j-1} + N^*_{r,j-1}$. We apply Lindeberg's device (cf. [15]) in order to replace $\{N^*_{r,i}\}$ by $\{\hat{N}^*_{r,i}\}$:

$$\mathbb{P}\left(\sum_{j=0}^{k-1} N^*_{r,j} \in A\right) - \mathbb{P}\left(\sum_{j=0}^{k-1} \hat{N}^*_{r,j} \in A\right) = \sum_{j=1}^{k-1}\left\{\mathbb{P}(S_j + N^*_{r,j} \in A) - \mathbb{P}(S_j + \hat{N}^*_{r,j} \in A)\right\}.$$

According to Berbee's lemma ([5], ch. 4), the random vectors $\sum_{l=0}^{j-1} N^*_{r,l}$, $N^*_{r,j}$ and $\hat{N}^*_{r,j}$ can be defined on a common probability space so that $\mathbb{P}\left(N^*_{r,j} \neq \hat{N}^*_{r,j}\right) \leq \beta(l)$. Therefore,

$$\left|\mathbb{P}\left(\sum_{j=0}^{k-1} N^*_{r,j} \in A\right) - \mathbb{P}\left(\sum_{j=0}^{k-1} \hat{N}^*_{r,j} \in A\right)\right| \leq k\beta(l).$$

The mixing coefficient α is weaker than β. Using Lemma 4 below, we evaluate $\left|\mathbb{P}\left(\sum_{j=0}^{k-1} N^*_{r,j} \in A\right) - \mathbb{P}\left(\sum_{j=0}^{k-1} \hat{N}^*_{r,j} \in A\right)\right|$ in terms of $\alpha(l)$. Note that $\mathbb{E}|N^*_{r,0}| = rp$. Inequality (10) with $b = 1$ and $y = rp$ entails the random vectors $\sum_{l=0}^{j-1} N^*_{r,l}$, $N^*_{r,j}$ and $\hat{N}^*_{r,j}$ can be defined on a common probability space so that $\mathbb{P}\left(N^*_{r,j} \neq \hat{N}^*_{r,j}\right) = \mathbb{P}\left(|N^*_{r,j} - \hat{N}^*_{r,j}| \geq 1\right) \leq \kappa(l)$ if $m2^{(m-1)/2}\alpha(l) \leq 1$. Hence

$$\left|\mathbb{P}\left(\sum_{j=0}^{k-1} N^*_{r,j} \in A\right) - \mathbb{P}\left(\sum_{j=0}^{k-1} \hat{N}^*_{r,j} \in A\right)\right| \leq k\min\{\beta(l); \kappa(l)\}.$$

Let $\{\hat{N}_{r,j}\}$ be independent copies of $N_{r,0}$, and set $\hat{N}_n = \sum_{j=0}^{k-1} \hat{N}_{r,j}$. Evidently, $\mathbb{P}\left(\sum_{j=0}^{k-1} \hat{N}_{r,j} \neq \sum_{j=0}^{k-1} \hat{N}^*_{r,j}\right) \leq klp$. Combining our estimates, we get

$$d_{TV}\left(N_n; \hat{N}_n\right) \leq 2klp + r'p + k\min\{\beta(l); \kappa(l)\}.$$

Denote $\mu = \sum_{j=0}^{k-1} \mathbb{I}\{\hat{N}_{r,j} \neq \bar{0}\}$, and put

$$Z_0 = \bar{0},\ Z_j = \zeta_1(n) + \dots + \zeta_j(n) \qquad (j \geq 1).$$

By Khintchin's formula (see [12], ch. 2), $\hat{N}_n \stackrel{d}{=} Z_\mu$. According to (1), $d_{TV}(\mu; \pi) \leq \left(1 - e^{-kq}\right) q$. Using this inequality and an idea from [15], we conclude that

$$d_{TV}\left(Z_\mu, Z_\pi\right) = \frac{1}{2}\sum_{\bar{i}}\left|\mathbb{P}\left(Z_\mu = \bar{i}\right) - \mathbb{P}\left(Z_\pi = \bar{i}\right)\right|$$

$$\begin{aligned} &\leq \frac{1}{2}\sum_{\bar{i}}\sum_{m=0}^{\infty} \mathbb{P}(Z_m = \bar{i})\left|\mathbb{P}(\mu = m) - \mathbb{P}(\pi = m)\right| \\ &= d_{TV}(\mu, \pi) \leq (1 - e^{-kq})q \leq (1 - e^{-np})rp . \end{aligned}$$

The result follows. □

The proof of Theorem 2 shows that the term $(1 - e^{-np})rp$ in the right–hand side of (4) may be replaced by any other estimate of $d_{TV}(\mu, \pi)$ (cf. [10, 20]).

Proof of Theorem 1. Let $\{r = r_n\}$ be a sequence of natural numbers such that

$$n \gg r_n \gg l_n + 1\,, \; nr_n^{-1}\alpha_n^{2/(2+m)} \to 0\,. \tag{7}$$

Such a sequence exists: one can put $r_n = \max\left\{\left[n\alpha_n^{1/(2+m)}\right]; \left[\sqrt{n(l_n+1)}\right]\right\}$ (note that $rp \to 0$ because of (2)).

If $N_n \Rightarrow \exists N$ then there exists the limit

$$\lim \mathbb{P}(N_n(u_m) = 0) := e^{-t}\,. \tag{8}$$

If $t = 0$ then $N_n(u_m) \to 0$, and the assertion of Theorem 1 trivially holds. Evidently, $t < \infty$ (otherwise $1 + o(1) = \mathbb{P}(N_n(u_m) \geq 1) \leq \mathbb{E}N_n(u_m) = rp \to 0$). Thus, $t \in (0; \infty)$.

It is known (cf. [13, 16]) that (8) with $t \in (0; \infty)$ is equivalent to $\mathbb{P}(N_r(u_m) > 0) \sim tr/n$. Therefore, if $N_n \Rightarrow \exists N$ then Theorem 2 implies

$$\mathbb{E}e^{ivN_n} = \exp\left(t\left(\varphi_{\zeta(n)}(v) - 1\right)\right) + o(1) \to \mathbb{E}e^{ivN} \qquad (\forall v \in \mathbb{R}^m)$$

as $n \to \infty$, where $\varphi_{\zeta(n)}$ is the characteristic function of $\zeta(n)$. Hence there exists the limit $\lim_{n\to\infty} \varphi_{\zeta(n)}(v) := \varphi(v)$. As a limit of a sequence of characteristic functions, it is a characteristic function itself. Therefore,

$$\mathbb{E}e^{ivN} = \exp\left(t(\varphi(v) - 1)\right)\,.$$

This is a characteristic function of a compound Poisson random vector with intensity t and multiplicity distribution $\mathcal{L}(\zeta)$ such that $\mathbb{E}e^{iv\zeta} = \varphi(v)$. □

Proof of Corollary 3. Let $r = 1$ and $l = 0$. Then $\zeta^*(n)$ takes values $(1, 0, ..., 0)$,..., $(0, ..., 0, 1)$ with probabilities $p_1/p, ..., p_m/p$ and $\mathcal{L}(\pi) = \mathbf{\Pi}(np)$. By Theorem 2,

$$d_{TV}\left(Y_n; \sum_{j=1}^{\pi} \zeta_j^*(n)\right) \leq \left(1 - e^{-np}\right)p\,.$$

It is easy to see that

$$\mathbb{E}\exp\left(iv\sum_{j=1}^{\pi}\zeta_j^*(n)\right) = \exp\left(n\sum_{j=1}^{m}\left(e^{iv_j}-1\right)p_j\right) = \mathbb{E}e^{ivY}$$

for any $v \in \mathbb{R}^m$. Hence $\sum_{j=1}^{\pi}\zeta_j^*(n) \stackrel{d}{=} Y$. □

For $v \in \mathbb{R}^m$, we put $|v| = \max_{i\le m}|v_i|$. Let (X,Y) be a random vector taking values in $\mathbb{R}^l \times \mathbb{R}^m$, and let α be the α–mixing coefficient corresponding to the σ–fields $\sigma(X)$ and $\sigma(Y)$.

Lemma 4 *One can define random vectors X, Y and $\hat{Y}$ on a common probability space in such a way that $\hat{Y}$ is independent of X, $\hat{Y} \stackrel{d}{=} Y$ and $(y>0, K\in\mathbb{N})$*

$$\mathbb{P}\left(|\hat{Y}-Y|>y\right) \le 2^{(m+3)/2}K^{m/2}\alpha + 2\mathbb{P}(|Y|>Ky). \tag{9}$$

In particular, if $\nu = \mathbb{E}^{1/b}|Y|^b < \infty$ and $b(\nu/y)^b \ge m2^{(m-1)/2}\alpha$ then

$$\mathbb{P}\left(|\hat{Y}-Y|>y\right) \le 2(1+2b/m)\left[(2^{(m-1)/2}m/b)^{2b}(\nu/y)^{bm}\alpha^{2b}\right]^{1/(2b+m)}. \tag{10}$$

If $\nu_\infty \equiv \operatorname{ess\,sup}|Y| < \infty$ then (10) yields

$$\mathbb{P}\left(|\hat{Y}-Y|>y\right) \le 2^{(m+3)/2}(\nu_\infty/y)^{m/2}\alpha. \tag{11}$$

In the case $m=1$, (10) improves the result of Theorem 3 in [8].

Proof of Lemma 4. Denote $Y^< = Y\mathbb{I}\{|Y|\le Ky\}$. Vector $Y^<$ takes values in $[-Ky;Ky]^m$. Splitting $[-Ky;Ky]$ into $2K$ intervals of length y induces the partition of $[-Ky;Ky]^m$ into $N=(2K)^m$ cubes $H_1,\dots,H_N$. According to Theorem 2 in [8], one can define $X, Y^<$ and $\hat{Y}^<$ on a common probability space so that $\hat{Y}^<$ is independent of X, $\hat{Y}^< \stackrel{d}{=} Y^<$ and

$$\mathbb{P}\left(|\hat{Y}^< - Y^<|>y\right) = \mathbb{P}(A) \le \sqrt{8N}\alpha,$$

where $A = \{\hat{Y}^< \text{ and } \hat{Y}^< \text{ are not elements of the same } H_i\}$.

Now we construct a vector $\hat{Y}$ on the base of $\hat{Y}^<$ such that $\hat{Y} \stackrel{d}{=} Y$. We put $\hat{Y} = \hat{Y}^< + \mathbb{I}\{\hat{Y}^< = 0\}Y'$, where Y' is independent of all other random vectors, $\mathcal{L}(Y') = \mathcal{L}(Y|B)$ and $B = \{Y^< = 0\} = \{Y=0 \text{ or } |Y|>Ky\}$.

Evidently, $\hat{Y} \stackrel{d}{=} Y$. Indeed, $\mathbb{P}(\hat{Y}=0) = \mathbb{P}(\hat{Y}^< = 0 = Y') = \mathbb{P}(B)\mathbb{P}(Y'=0) = \mathbb{P}(Y=0)$, and if $z \ne 0$ then

$$\mathbb{P}(\hat{Y}\in dz) = \mathbb{P}(\hat{Y}^< \in dz) + \mathbb{P}(\hat{Y}^< = 0, Y' \in dz)$$

$$= \mathbb{P}(B_c, Y \in dz) + \mathbb{P}(B)\mathbb{P}(Y \in dz|B) = \mathbb{P}(Y \in dz),$$

where $B_c = \{0 < |Y| < Ky\}$ is the complement to B. It is easy to see that $\mathbb{P}(\hat{Y} \neq \hat{Y}^<) = \mathbb{P}(\hat{Y}^< = 0 \neq Y') = \mathbb{P}(B)\mathbb{P}(Y \neq 0|B) = \mathbb{P}(|Y| > Ky)$. Hence

$$\mathbb{P}\left(|\hat{Y} - Y^<| > y\right) \leq \sqrt{8N}\alpha + \mathbb{P}(\hat{Y} \neq \hat{Y}^<) \leq \sqrt{8N}\alpha + \mathbb{P}(|Y| > Ky).$$

It remains to construct (X, Y) on the base of $(X, Y^<)$. Let $\{Y_x\}$ be independent random vectors with distributions $\mathcal{L}(Y_x) = \mathcal{L}(Y|B, X = x)$. Denote $Y^* = Y^< + \mathbb{I}\{Y^< = 0\}Y_X$. Then $(X, Y^*) \stackrel{d}{=} (X, Y)$. Indeed,

$$\begin{aligned} \mathbb{P}(X \in dx, Y^* = 0) &= \mathbb{P}(X \in dx, Y^< = 0 = Y_X) = \mathbb{P}(X \in dx, Y^< = 0)\mathbb{P}(Y_x = 0) \\ &= \mathbb{P}(X \in dx, B, Y = 0) = \mathbb{P}(X \in dx, Y = 0). \end{aligned}$$

If $z \neq 0$ then

$$\mathbb{P}(X \in dx, Y^* \in dz) = \mathbb{P}(X \in dx, Y^< \in dz) + \mathbb{P}(X \in dx, Y^< = 0, Y_X \in dz)$$

$$= \mathbb{P}(X \in dx, B_c, Y \in dz) + \mathbb{P}(X \in dx, B)\mathbb{P}(Y_x \in dz) = \mathbb{P}(X \in dx, Y \in dz).$$

Note that $\mathbb{P}(Y^* \neq Y^<) = \mathbb{P}(Y^< = 0 \neq Y_X) = \mathbb{P}(|Y| > Ky)$. Therefore,

$$\mathbb{P}\left(|\hat{Y} - Y| > y\right) \leq \mathbb{P}\left(|\hat{Y} - Y^<| > y\right) + \mathbb{P}(|Y| > Ky).$$

Combining our estimates, we get (9).

Using Chebyshev's inequality, we deduce

$$\mathbb{P}\left(|\hat{Y} - Y| > y\right) \leq cK^{m/2} + dK^{-b},$$

where $c = 2^{(m+3)/2}\alpha$ and $d = 2(\nu/y)^b$. The function $f(x) = cx^{m/2} + dx^{-b}$ takes its minimum in $x \geq 1$ at $x_o = \max\{(2bd/cm)^{2/(m+2b)}; 1\}$. Since $\frac{2bd}{cm} = \frac{b(\nu/y)^b}{2^{(m-1)/2}m\alpha}$, inequality (9) entails (10). The proof is complete. □

References

[1] Barbour A.D., Chen L.H.Y. and Loh W.-L. (1992) Compound Poisson approximation for nonnegative random variables via Stein's method. — Ann. Probab., v. 20, No 4, 1843–1866.

[2] 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.

[3] Barbour A.D., Holst L. and Janson S. (1992) *Poisson Approximation.* Oxford: Clarendon Press, 277 pp.

[4] Barbour A.D., Novak S.Y. and Xia A. (1999) Compound Poisson approximation for the distribution of extremes. — Technical University of Eindhoven: Eurandom research report No 99-040.

[5] Berbee H.C.P. (1979) *Random walks with stationary increments and renewal theory.* Amsterdam: Mathematisch Centrum Tract 112.

[6] Borisov I.S. (1993) Strong Poisson and mixed approximations of sums of independent random variables in Banach spaces. — Siberian Adv. Math., v. 3, No 2, 1–13.

[7] Bosq D. (1996) *Nonparametric statistics for stochastic processes.* — New York: Springer Verlag, 169 pp.

[8] Bradley R. (1983) Approximation theorems for strongly mixing random variables. — Michigan Math. J., v. 30, 69-81.

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

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

[11] Embrechts P., Klüppelberg C. and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance.* — Berlin: Springer Verlag.

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

[13] Leadbetter M.R. (1974) On extreme values in stationary sequences. — Z. Wahrsch. Ver. Geb., v. 28, 289–303.

[14] LeCam L. (1965) On the distribution of sums of independent random variables. — In: Proc. Internat. Res. Sem. Statist. Lab. Univ. California, 179–202. New York: Springer Verlag.

[15] Novak S.Y. (1998) On the limiting distribution of extremes. — Siberian Adv. Math., v. 8, No 2, 70–95.

[16] O'Brien G.L. (1974) Limit theorems for the maximum term of a stationary process. — Ann. Probab., v. 2, No 3, 540–545.

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

[18] Raab M. (1997) *On the number of exceedances in Gaussian and related sequences.* — PhD thesis. Stockholm: Royal Institute of Technology.

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

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