S.Y. Novak

On accuracy of multivariate compound Poisson approximation

Eurandom

Technical University of Eindhoven

Report 2000-042

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.
Report 2000-042
On accuracy of multivariate compound Poisson approximation
S.Y. Novak
ISSN: 1389-2355
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 $\text{B}(n, p)$ distribution, where $p = \mathbb{P}(X > u)$. If $p$ is “small” then $\mathcal{L}(N_n(u))$ may be approximated by the Poisson $\Pi(np)$ 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 $\mathbb{Z}_+^m$, where $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$:

$$d_{TV}(X; Y) \equiv d_{TV}(\mathcal{L}(X); \mathcal{L}(Y)) = \sup_{A \subseteq \mathbb{Z}_+^m} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

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

$$d_{TV}(N_n(u); \pi) \leq (1 - e^{-np}) p. \quad (1)$$

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]).

---

1Eurandom, 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 $L(N_n(u))$ 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 represent 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 $\mathcal{F}_{a,b} = \mathcal{F}_{a,b}(u_1, ..., u_m)$ be the $\sigma$-field generated by the events $\{X_i > u_j\}$, $a \leq i \leq b, 1 \leq j \leq m$. Denote

$$\alpha(l) \equiv \alpha(l, \{u_1, ..., u_m\}) = \sup |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|,$$

$$\beta(k) \equiv \beta(l, \{u_1, ..., u_m\}) = \sup E \sup_B |\mathbb{E}(B|\mathcal{F}_{a,b}) - \mathbb{E}(B)|,$$

where the supremum is taken over all $A \in \mathcal{F}_{a,b}$, $B \in \mathcal{F}_{a,b+1,n}$, $j \geq 1$, 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 $\{l_n\} \subset \mathbb{Z}_+$ such that $l_n/n \to 0$ as $n \to \infty$. A vector $Y$ has a multivariate compound Poisson distribution $\Pi(\lambda, \mathcal{L}(Z))$ if

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

2
where $Z, Z_1, \ldots$ are i.i.d. random vectors, $\pi$ is independent of $\{Z_i\}$ and has the Poisson distribution with parameter $\lambda$.

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

$$\limsup_n \frac{n}{m} \mathbb{P}(X > u_m) < \infty.$$  \hspace{1cm} (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

$$\mathcal{L}(\zeta(n)) = \mathcal{L}(N_r | N_r(u_m) > 0),$$  \hspace{1cm} (3)

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

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

and let $\pi$ 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 $N = \sum_{i=1}^r \zeta_i(n)$.

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

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

where $\kappa(l) = 2(1 + 2/m) \{2m^{-1}m^2\alpha^2(l)\}^{1/(2+m)}$ 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 $d_w(N_n; N) \leq (1.65(1 - rp)^{-1/2} + e^{rp})(rp + 2(2rp + nr^{-1}l)p + nr^{-1}\beta(l))$. 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 multinominal 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 \( N_n(u, v) = \sum_{i=1}^{n} I\{u \geq X_i > v\} \) 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

\[
d_{TV}(N^*_n; N^*) \leq (1 - e^{-np})r + (2nr^{-1}l + r')p + nr^{-1} \min\{\beta(l); \kappa(l)\},
\]

(4*)

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

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

\[
\mathbb{P}(N^*_n = (l, \ldots, l)) = \frac{n!}{l! \ldots l! (n-l)!} p_1^{l_1} \ldots p_m^{l_m} (1 - p)^{n-l},
\]

(5)

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

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

\[
d_{TV}(Y_n; Y) \leq (1 - e^{-np})r.
\]

(6)

### 3 Proofs

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

Denote \( \Pi_i = (I\{X > u_1\}, \ldots, I\{X > u_m\}) \), and let

\[
N_{r,j} = \sum_{i=|j+1|}^{(j+1)r/n} \Pi_i \quad (0 \leq j \leq k = \lfloor n/r \rfloor).
\]

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

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

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

\[
N^*_{r,j} = \sum_{i=|j+1|}^{(j+1)r-l} \Pi_i, \quad N^*_n = \sum_{j=0}^{k-1} N^*_{r,j} \quad (0 \leq j < k).
\]
Then \( \mathbb{P} \left( \sum_{j=0}^{k-1} N_{r,j} \neq \sum_{j=0}^{k-1} N_{r,j}^* \right) \leq k \mathbb{P} \left( N_{r,0} \neq N_{r,0}^* \right) \leq kl p \).

Let \( \{ \tilde{N}_{r,j}^* \} \) be independent copies of \( N_{r,0}^* \). Denote
\[
S_i = \sum_{j=0}^{i-1} N_{r,j}^* + \sum_{j=i+1}^{k-1} \tilde{N}_{r,j}^* \quad (0 < i < k).
\]

Notice that \( S_j + \tilde{N}_{r,j}^* = S_{j-1} + N_{r,j-1}^* \). We apply Lindeberg’s device (cf. [15]) in order to replace \( \{ N_{r,j}^* \} \) by \( \{ \tilde{N}_{r,j}^* \} \):
\[
\mathbb{P} \left( \sum_{j=0}^{k-1} N_{r,j}^* \in A \right) - \mathbb{P} \left( \sum_{j=0}^{k-1} \tilde{N}_{r,j}^* \in A \right) = \sum_{j=1}^{k-1} \left\{ \mathbb{P}(S_j + N_{r,j}^* \in A) - \mathbb{P}(S_j + \tilde{N}_{r,j}^* \in A) \right\}.
\]

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

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

The mixing coefficient \( \alpha \) is weaker than \( \beta \). Using Lemma 4 below, we evaluate
\[
\left| \mathbb{P} \left( \sum_{j=0}^{k-1} N_{r,j}^* \in A \right) - \mathbb{P} \left( \sum_{j=0}^{k-1} \tilde{N}_{r,j}^* \in A \right) \right| \text{ in terms of } \alpha(l) \text{. Note that } \mathbb{E} |N_{r,0}^*| = rp.
\]

Inequality (10) with \( b = 1 \) and \( y = rp \) entails the random vectors \( \sum_{i=0}^{j-1} N_{r,i}^* \), \( N_{r,j}^* \) and \( \tilde{N}_{r,j}^* \) can be defined on a common probability space so that \( \mathbb{P} \left( N_{r,j}^* \neq \tilde{N}_{r,j}^* \right) \leq \kappa(l) \) if \( m^{2(m-1)/2} \alpha(l) \leq 1 \). Hence
\[
\left| \mathbb{P} \left( \sum_{j=0}^{k-1} N_{r,j}^* \in A \right) - \mathbb{P} \left( \sum_{j=0}^{k-1} \tilde{N}_{r,j}^* \in A \right) \right| \leq k \min \{ \beta(l); \kappa(l) \}.
\]

Let \( \{ \tilde{N}_{r,j} \} \) be independent copies of \( N_{r,0} \), and set \( \tilde{N}_n = \sum_{j=0}^{k-1} \tilde{N}_{r,j} \). Evidently, \( \mathbb{P} \left( \sum_{j=0}^{k-1} \tilde{N}_{r,j} \neq \sum_{j=0}^{k-1} N_{r,j} \right) \leq kl p \). Combining our estimates, we get
\[
d_{TV} \left( N_n, \tilde{N}_n \right) \leq 2kl p + r'p + k \min \{ \beta(l); \kappa(l) \}.
\]

Denote \( \mu = \sum_{j=0}^{k-1} \mathbb{I} \{ \tilde{N}_{r,j} \neq \tilde{0} \} \), and put
\[
Z_0 = \tilde{0}, Z_j = \zeta_1(n) + \ldots + \zeta_j(n) \quad (j \geq 1).
\]

By Khintchine’s formula (see [12], ch. 2), \( \tilde{N}_n \overset{d}{=} Z_{\mu} \). According to (1), \( d_{TV}(\mu; \pi) \leq (1 - e^{-kq})q \). Using this inequality and an idea from [15], we conclude that
\[
d_{TV} \left( Z_{\mu}, Z_\pi \right) = \frac{1}{2} \sum_i \left| \mathbb{P} (Z_{\mu} = i) - \mathbb{P} (Z_{\pi} = i) \right|.
\]

5
\[
\leq \frac{1}{2} \sum_i \sum_{m=0}^{\infty} \mathbb{P}(Z_m = i) \left| \mathbb{P}(\mu = m) - \mathbb{P}(\pi = m) \right|
\]
\[
= d_{\tau \nu}(\mu, \pi) \leq (1 - e^{-kp}) q \leq (1 - e^{-np}) p .
\]

The result follows. \(\square\)

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

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

\[
n \gg r_n \gg l_n + 1 , \ n r_n^{-1} \alpha_n^{2/(2+m)} \to 0 .
\]

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

If \(N_n \Rightarrow \exists N\) then there exists the limit

\[
\lim \mathbb{P}(N_n(u_m) = 0) := e^{-t} .
\]

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

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

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 \(L(\zeta)\) such that \(\mathbb{E} e^{iv\zeta} = \varphi(v)\). \(\square\)

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

\[
d_{\tau \nu} \left( Y_n, \sum_{j=1}^{\pi} \zeta_j^*(n) \right) \leq (1 - e^{-np}) p .
\]
It is easy to see that
\[
\mathbb{E} \exp \left( iv \sum_{j=1}^{\pi} \zeta_j^* (n) \right) = \exp \left( n \sum_{j=1}^{m} (e^{iuj} - 1) p_j \right) = \mathbb{E} e^{ivY}
\]
for any \( v \in \mathbb{R}^m \). Hence \( \sum_{j=1}^{\pi} \zeta_j^* (n) \overset{d}{=} Y \). \( \square \)

For \( v \in \mathbb{R}^m \), we put \( |v| = \max_{i \leq m} |v_i| \). Let \( (X,Y) \) be a random vector taking values in \( \mathbb{R}^1 \times \mathbb{R}^m \), and let \( \alpha \) be the \( \alpha \)-mixing coefficient corresponding to the \( \sigma \)-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 \( X \), \( \hat{Y} \overset{d}{=} Y \) and \( (y > 0, K \in \mathbb{N}) \)
\[
\mathbb{P} \left( |\hat{Y} - Y| > y \right) \leq 2^{(m+3)/2} K^{-m/2} \alpha + 2 \mathbb{P}(|Y| > Ky).
\] (9)

In particular, if \( \nu = \mathbb{E}^{1/b} |Y|^b < \infty \) and \( b(\nu/y)^b \geq m^{(m-1)/2} \alpha \) then
\[
\mathbb{P} \left( |\hat{Y} - Y| > y \right) \leq 2(1 + 2b/m) \left[ (2^{(m-1)/2} m/b)^{2b} (\nu/y)^b \alpha^{2b} \right]^{1/(2b+m)}.
\] (10)

If \( \nu_{\infty} \equiv \text{ess sup} |Y| < \infty \) then (10) yields
\[
\mathbb{P} \left( |\hat{Y} - Y| > y \right) \leq 2^{(m+3)/2} (\nu_{\infty}/y)^{m/2} \alpha.
\] (11)

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

**Proof** of Lemma 4. Denote \( Y^\prec = Y \mathbb{I}[|Y| \leq Ky] \). Vector \( Y^\prec \) takes values in \([-Ky; Ky]^m \). Splitting \([-Ky; Ky]^m \) into \( 2K \) intervals of length \( y \) induces the partition of \([-Ky; Ky]^m \) into \( N = (2K)^m \) cubes \( H_1, \ldots, H_N \). According to Theorem 2 in [8], one can define \( X, Y^\prec \) and \( \hat{Y}^\prec \) on a common probability space so that \( Y^\prec \) is independent of \( X \), \( \hat{Y}^\prec \overset{d}{=} Y^\prec \) and
\[
\mathbb{P} \left( |\hat{Y}^\prec - Y^\prec| > y \right) = \mathbb{P}(A) \leq \sqrt{8N} \alpha,
\]
where \( A = \{ \hat{Y}^\prec \text{ and } \hat{Y}^\prec \text{ are not elements of the same } H_i \} \).

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

Evidently, \( \hat{Y} \overset{d}{=} Y \). Indeed, \( \mathbb{P}(\hat{Y} = 0) = \mathbb{P}(\hat{Y}^\prec = 0 = Y') = \mathbb{P}(B) \mathbb{P}(Y' = 0) = \mathbb{P}(Y = 0) \), and if \( z \neq 0 \) then
\[
\mathbb{P}(\hat{Y} \in dz) = \mathbb{P}(\hat{Y}^\prec \in dz) + \mathbb{P}(\hat{Y}^\prec = 0, Y' \in dz)
\]
\[ = \mathbb{P}(B_c, Y \in dz) + \mathbb{P}(B)\mathbb{P}(Y \in dz|B) = \mathbb{P}(Y \in dz), \]

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

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

It remains to construct \((X,Y)\) on the base of \((X,Y^c)\). Let \( \{Y_x\} \) be independent random vectors with distributions \( \mathcal{L}(Y_x) = \mathcal{L}(Y|B, X = x) \). Denote \( Y^* = Y^c + \mathbb{1}\{Y^c = 0\}Y_X \). Then \((X,Y^*) \overset{d}{=} (X,Y)\). Indeed,

\[ \mathbb{P}(X \in dx, Y^* = 0) = \mathbb{P}(X \in dx, Y^c = 0 = Y_X) = \mathbb{P}(X \in dx, Y^c = 0)\mathbb{P}(Y_x = 0) = \mathbb{P}(X \in dx, B, Y = 0) = \mathbb{P}(X \in dx, Y = 0). \]

If \( z \neq 0 \) then

\[ \mathbb{P}(X \in dx, Y^* \in dz) = \mathbb{P}(X \in dx, Y^c \in dz) + \mathbb{P}(X \in dx, Y^c = 0, Y_X \in dz) \]

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

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

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

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 \( \frac{2bd}{cm} = \frac{b(\nu/y)^b}{2^{(m+1)/2}\alpha} \), inequality (9) entails (10). The proof is complete. \( \Box \)
References