On the accuracy of multivariate compound Poisson approximation

S.Y.Novak

Statist. Probab. Letters, 2003, v. 62, No 1, 35-43.
Abridged version

Abstract

We present multivariate generalizations of some classical results on the accuracy of Poisson approximation for the distribution of a sum of 0–1 random variables. A multivariate generalization of Bradley's theorem [7] is established as well.

Keywords: compound Poisson approximation, dependent random variables.

1 Introduction

Let $X, X_1, X_2, ...$ 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. [10, 18]).

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. The accuracy of Poisson approximation for a binomial distribution has been investigated by famous authors (see, e.g., [16, 13, 9, 3] and references in [6]). The case of a sum of dependent 0–1 random variables was the subject of [8, 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 $\mathbf{Z}_+ = \mathbb{N} \cup \{0\}$:

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

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 [16] and LeCam [13]. Sharper bounds are available under extra restrictions (see [9, 19]).

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.

The accuracy of compound Poisson approximation for $\mathcal{L}(N_n(u))$ has been evaluated in [1, 14, 17], among others. The feature of the estimate given in [14] 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, $1 \le i \le m$. 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 the 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 [14]. In the case of independent trials, Theorem 2 yields an estimate of the accuracy of multivariate Poisson approximation for a multinomial distribution. The results allow evident reformulation in terms of random vectors with 0–1 components, but we prefer the present notation in order to keep in touch with applications to Extreme Value Theory.

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 $\{X_i > u_j\}$, $a \le i \le b, 1 \le j \le m$. Denote

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

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

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

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

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 $Z, Z_1, ...$ 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 $r \in \{1, ..., n\}$. The proof of Theorem 1 shows that $Y \stackrel{d}{=} \mathbf{\Pi}(\lambda, \mathcal{L}(\zeta))$, 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 the 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 [14] (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 the accuracy of multivariate Poisson approximation for a multinomial distribution.

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

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

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 $p_1=\mathbb{P}(X>u_1)$, $p_2=\mathbb{P}(u_1\geq X>u_2)$, ..., $p_m=\mathbb{P}(u_{m-1}\geq X>u_m)$:

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

where $l=l_1+...+l_m\leq n$, $p=p_1+...+p_m$. Theorem 2 yields an estimate of the 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*

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

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. (2002) Compound Poisson approximation for the distribution of extremes. — Adv. Appl. Probab., v. 34, No 1, 223–240.

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

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

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

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

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

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

[13] 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.

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

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

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

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

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

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