

On Error Bounds for the Approximation of Random Sums

Topic 4: Other

Roos, Bero

Fachbereich Mathematik, SPST, Universität Hamburg,

Bundesstr. 55, 20146 Hamburg, Germany.

Tel.: +49 40 42838 4925, Fax: +49 40 42838 5260,

E-mail: roos@math.uni-hamburg.de

Pfeifer, Dietmar

Fachbereich Mathematik, Universität Oldenburg, 26111 Oldenburg, Germany.

Tel.: +49 441 798 3229, Fax: +49 441 798 3004,

E-mail: pfeifer@mathematik.uni-oldenburg.de

Abstract

We present bounds for the total variation distance between the distributions of random sums with different random summation indices. The summands are assumed to be nonzero with small probability.

Key words and phrases. Excess of loss reinsurance, finite signed measures, random sums, total variation distance, upper bounds.

1 Introduction

In this paper, we consider the approximation of the distribution of a random sum with independent and identically distributed summands, which are nonzero with small probability. Such random sums appear, for example, in the excess of loss (XL) reinsurance (see, for example, Heilmann (1988, 6.2.2, p. 212)): We start with an arbitrary collective model with claims Y_1, Y_2, \dots , random claim number M and aggregate claim $\sum_{j=1}^M Y_j$. Clearly, the Y_j are non-negative, independent and identically distributed random variables, which are also independent of M . In the excess of loss reinsurance, each claim Y_j ($j \in \{1, \dots, M\}$) is divided between the ceding company and the reinsurer, i.e. the reinsurer has to pay the excess $X_j = \max\{Y_j - t, 0\}$ over an agreed retention (priority) $t > 0$, whereas the ceding company has to pay the remaining amount $Y_j - X_j$. The reinsurer's aggregate claim is now given by $S_M = \sum_{j=1}^M X_j$, whereby we may assume that t is large enough such that $p = P(X_1 > 0) = P(Y_1 > t)$ is small.

In what follows, we consider a somewhat more general situation. In fact, the summands can also assume negative values. Let X, X_1, X_2, \dots be independent and identically distributed random variables in \mathbf{R} . Set $p = 1 - q = P(X \neq 0) \in [0, 1]$ and $S_n = \sum_{j=1}^n X_j$ for $n \in \mathbf{Z}_+ = \{0, 1, 2, \dots\}$. Let M and N be random variables in \mathbf{Z}_+ independent of the X_j .

One can find several contributions on the normal approximation of the distribution $\mathcal{L}(S_M)$ of S_M (e.g. see Gut (1988), Kruglov and Korolev (1990), and Gnedenko and Korolev (1996) and the references therein). But also the (compound) Poisson approximation (e.g. see Logunov (1990), Yannaros (1991), Barbour et al. (1992, equation (3.8), p. 39) and Vellaisamy and Chaudhuri (1996)) was already considered. Here, the summation index M has to be replaced with an independent Poisson $\text{Po}(t)$ distributed random variable, say N . Intuitively, we expect a good approximation when we choose a Poisson distribution with the same mean as M , i.e. $t = E(M)$, provided $E(M)$ is finite. A further improvement of the approximation can be expected when also the variances of M and N coincide. But this would mean that $\text{Var}(M) = E(M)$, which is rarely the case. Therefore, in this context, it is better to allow the random variable N to have another distribution than the Poisson. This is the main subject of this paper.

In order to make the approximation results mathematically precise, we use the total variation distance

$$d_{\text{TV}}(Q_1, Q_2) = \sup_A |Q_1(A) - Q_2(A)| \quad (1)$$

between probability distributions Q_1 and Q_2 on the line, where the supremum is over all Borel measurable sets $A \subseteq \mathbf{R}$. If Z_i , ($i = 1, 2$) are random variables with distributions Q_i , then we set $d_{\text{TV}}(Z_1, Z_2) = d_{\text{TV}}(Q_1, Q_2)$. In what follows, we shall discuss upper bounds for the distance

$$d_\tau := d_{\text{TV}}(S_M, S_N),$$

where N has an arbitrary distribution on \mathbf{Z}_+ . The following lemma shows that it is reasonable to expect upper bounds for d_τ , which are small when p is small or when the distributions of M and N are close. In what follows, let $\stackrel{\text{d}}{=}$ denote equality in distribution.

Lemma 1 $S_M \stackrel{\text{d}}{=} S_N$ if and only if $M \stackrel{\text{d}}{=} N$ or $p = 0$.

Proof. The if part is obvious. The only if part can be show as follows: Let us assume that $S_M \stackrel{\text{d}}{=} S_N$ and that $p \neq 0$. If, for the characteristic function $\varphi_X(t) = \text{E}(e^{itX})$ of X , $|\varphi_X(t)| = 1$ for all $t \in \mathbf{R}$, then $X =: c \neq 0$ is a.s. constant (see Lukacs (1970, Theorem 2.1.4, p. 18)), so that, in this case, $Mc \stackrel{\text{d}}{=} S_M \stackrel{\text{d}}{=} S_N \stackrel{\text{d}}{=} Nc$, giving $M \stackrel{\text{d}}{=} N$. Let us now assume that $t_0 \in \mathbf{R}$ exists such that $|\varphi_X(t_0)| < 1$. Set $B = \{z \in \mathbf{C} \mid |z| < 1\}$. Since φ_X is continuous, the set $A := B \cap \varphi_X(\mathbf{R})$ has infinitely many elements and contains limit points. Let denote ψ_M and ψ_N the probability generating functions of M and N , respectively. Because of $A \subseteq \{z \in \mathbf{C} \mid \psi_M(z) = \psi_N(z), |z| < 1\}$, we obtain $\psi_M = \psi_N$ on B and therefore $M \stackrel{\text{d}}{=} N$ also in this case. \square

It should be mentioned that there is a ad hoc bound, which can easily be shown with the help of coupling arguments originally due to Doeblin (1938) (cf. Lindvall (1992, Theorem 5.2, p. 19)), giving

$$d_\tau \leq d_{\text{TV}}(M, N). \quad (2)$$

But, since the bound in (2) is independent of p , it is useless if p is small. In this case, there are better estimates.

2 The case $E(M) = E(N) < \infty$

We have shown the following theorem.

Theorem 1 *If $E(M) = E(N) < \infty$, then, for all $m \in \mathbf{Z}_+$,*

$$d_\tau \leq p^2 \sum_{n=0}^{\infty} (n-m)(n-m-1) |a_n|, \quad (3)$$

where $a_n = P(M = n) - P(N = n)$, ($n \in \mathbf{Z}_+$).

This result is in fact a considerable improvement of bounds given by Logunov (1990, p. 588) and Vellaisamy and Chaudhuri (1996, proof of their Lemma 2.1). In fact, their bounds are comparable with our bound, where $m = 0$. It is not difficult to show that, in order to minimize the bound in (3), we must set $m = \lfloor \mu \rfloor$, where μ is the expectation of a random variable Z with

$$P(Z = n) = \frac{|a_n|}{\sum_{k=0}^{\infty} |a_k|}, \quad (n \in \mathbf{Z}_+)$$

and $\lfloor x \rfloor$ denotes the largest integer $\leq x$, ($x \in \mathbf{R}$).

The bound in Theorem 1 can be simplified:

Corollary 1 *If $E(M) = E(N) < \infty$, then, for all $t \in \mathbf{R}$,*

$$d_\tau \leq p^2 \sum_{n=0}^{\infty} (n-t)^2 |a_n|. \quad (4)$$

Proof. In the case $t \in (-\infty, 1/2]$, the assertion follows from (3) with $m = 0$, since, for $n \in \mathbf{Z}_+$, $n(n-1) \leq (n-t)^2$. Let us now consider the case $t \in (1/2, \infty)$. Let $\alpha = t + 1/2 - \lfloor t + 1/2 \rfloor \in [0, 1]$ and $m = \lfloor t + 1/2 \rfloor - 1 \in \mathbf{Z}_+$. By using (3) we obtain

$$\begin{aligned} d_\tau &\leq p^2 \sum_{n=0}^{\infty} \left[(1-\alpha)(n-m)(n-m-1) + \alpha(n-m-1)(n-m-2) \right] |a_n| \\ &\leq p^2 \sum_{n=0}^{\infty} \left(n - \left(m + \frac{1}{2} + \alpha \right) \right)^2 |a_n|. \end{aligned}$$

Since $t = m + 1/2 + \alpha$, the proof is completed. \square

It is well-known that (4) will be minimized by using $t = E(Z)$. This yields

$$d_\tau \leq 2p^2 d_{TV}(M, N) \text{Var}(Z),$$

where we used the well-known fact that $d_{\text{TV}}(M, N) = 2^{-1} \sum_{n=0}^{\infty} |a_n|$. In particular, we learn that, here, d_τ is of order $O(p^2)$ as $p \rightarrow 0$ when $d_{\text{TV}}(M, N)\text{Var}(Z)$ is bounded. But what can be said when this term is large or infinite? The following theorem shows a further inequality, which can be useful in this case.

Theorem 2 *If $E(M) = E(N) < \infty$, then, for all $m \in \mathbf{Z}_+$,*

$$d_\tau \leq \frac{p}{\sqrt{2}q} \sum_{n=0}^{\infty} \left(m - n + n \ln \left(\frac{n + 1/2}{m + 1/2} \right) \right) |a_n|. \quad (5)$$

Clearly, the bound in this theorem exhibits only the order $O(p)$, but the accompanying factor can be very small. For example, letting $m = \lfloor \mu \rfloor$, we obtain, from (5),

$$d_\tau \leq \frac{\sqrt{2}p}{q} d_{\text{TV}}(M, N) E \left[Z \ln \left(\frac{Z + 1/2}{\lfloor \mu \rfloor + 1/2} \right) \right]. \quad (6)$$

In order to compare the order of the bounds in (4) and (5), we derive from (5) a further bound without a log-factor.

Corollary 2 *If $E(M) = E(N) < \infty$, then*

$$d_\tau \leq \frac{\sqrt{2}p}{q(\lfloor \mu \rfloor + 1/2)} d_{\text{TV}}(M, N) E(Z - \lfloor \mu \rfloor)^2. \quad (7)$$

Proof. The assertion can be shown with the help of the simple inequality

$$E \left(\lfloor \mu \rfloor - Z + Z \ln \left(\frac{Z + 1/2}{\lfloor \mu \rfloor + 1/2} \right) \right) \leq \frac{E(Z - \lfloor \mu \rfloor)^2}{\lfloor \mu \rfloor + 1/2}. \quad \square$$

Letting $m = \lfloor \mu \rfloor$ in (4), we see that (7) contains an additional factor

$$\frac{1}{\sqrt{2}pq(\lfloor \mu \rfloor + 1/2)}, \quad (8)$$

which leads to a better upper bound in the case of $pq\mu$ being large.

Example 1 If $M = m$, ($m \in \mathbf{Z}_+$) is almost surely constant and $E(N) = m$, then we obtain from Corollary 1 and Theorem 2 that

$$\begin{aligned} d_\tau &\leq p^2 \min \left\{ \frac{1}{\sqrt{2}pq} E \left(N \ln \frac{N + 1/2}{m + 1/2} \right), \text{Var}(N) \right\} \\ &\leq \text{Var}(N) p^2 \min \left\{ \frac{1}{\sqrt{2}pq(m + 1/2)}, 1 \right\}. \end{aligned}$$

For the second inequality, we used the simple fact that $\ln(1 + x) \leq x$ for $x > -1$. If additionally N is Poisson $\text{Po}(m)$ distributed, then $\text{Var}(N) = m$. In this case, we obtain

an upper bound, which cannot be improved much for p being sufficiently small, since as shown by Barbour and Hall (1984), here, d_τ and $\min\{p, mp^2\}$ have the same order, provided that the X_j are Bernoulli distributed with success probability p . In fact they proved the inequalities $\frac{1}{32} \min\{p, mp^2\} \leq d_\tau \leq \min\{p, mp^2\}$. Note that their upper bound also holds for general X_j (see Le Cam (1965, p. 187) or Michel (1987, p. 167)).

3 The case $E(M^j) = E(N^j) < \infty$ for $j = 1, 2$

Above, we mentioned that the precision of approximation should increase when we use a random variable N having the same expectation and the same variance as M . As in the previous section, we shall present two bounds for d_τ . For $j \in \mathbf{Z}_+$ and $x \in \mathbf{R}$, set $\binom{x}{j} = \prod_{i=1}^j [(x - i + 1)/i]$.

Theorem 3 *If, for all $j \in \{1, 2\}$, $E(M^j) = E(N^j) < \infty$, then, for arbitrary $m \in \mathbf{Z}_+$,*

$$d_\tau \leq 4p^3 \sum_{n=0}^{\infty} |a_n| \left| \binom{n-m}{3} \right|, \quad (9)$$

$$d_\tau \leq \frac{2\sqrt{6}}{\sqrt{m+1}} \left(\frac{p}{q}\right)^{3/2} \sum_{n=0}^{\infty} \frac{|a_n| |m-n|^3}{(\sqrt{m+1} + \sqrt{n+1})^2}. \quad (10)$$

We see that the first bound has the order $O(p^3)$ as $p \rightarrow 0$, when $d_{\text{TV}}(M, N)E(|Z - \lfloor \mu \rfloor|^3)$ is bounded, whereas the second bound exhibits only the order $O(p^{3/2})$. However, comparing the bounds, we see that (10) contains a extra factor having at least the order $(pq(\lfloor \mu \rfloor + 1))^{-3/2}$, which coincides, up to constants, with the factor (8) of the previous section to the power $3/2$. Therefore, (10) is better than (9), if $pq\mu$ is large.

Example 2 In the important case $s := \text{Var}(M) > E(M) =: t > 0$, we can use a negative binomial $\text{NB}(q, \alpha)$ distributed N with parameters $q \in (0, 1)$ and $\alpha \in (0, \infty)$ and with the probabilities

$$\text{NB}(q, \alpha)(\{m\}) = \binom{\alpha + m - 1}{m} (1 - q)^\alpha q^m, \quad (m \in \mathbf{Z}_+).$$

With this definition, we have $E(N) = \alpha/\beta$ and $\text{Var}(N) = \alpha(\beta^{-1} + \beta^{-2})$, where $\beta = (1 - q)q^{-1}$. Therefore, equality of the first two moments of M and N can be achieved by setting $q := (s - t)/s$ and $\alpha := t^2/(s - t)$.

4 Idea of the proofs of Theorems 1–3

The theorems given above are easy consequences of a more general result on the approximation with finite signed measures. Let us write

$$P^X = q\varepsilon_0 + pQ, \quad Q = P(X \in \cdot | X \neq 0),$$

where ε_x is the Dirac measure at point $x \in \mathbf{R}$. We use the finite signed measures

$$R_{k,m} := P^{S_N} + \sum_{j=0}^k \beta_{j,m} p^j (Q - \varepsilon_0)^{*j} * P^{S_m}, \quad (k, m \in \mathbf{Z}_+ \text{ fixed}),$$

where we assume that

$$\sum_{n=0}^{\infty} |a_n| n^k < \infty, \quad (11)$$

such that

$$\beta_{j,m} = \sum_{n=0}^{\infty} a_n \binom{n-m}{j} \quad (12)$$

is absolutely convergent for all $j \in \{0, \dots, k\}$. Here, for a finite signed measure G on \mathbf{R} , let G^{*n} , ($n \in \mathbf{N} = \{1, 2, \dots\}$) denote the n -fold convolution of G and set $G^{*0} = \varepsilon_0$.

One of the most important properties of these signed measures is that $R_{0,m} = P^{S_N}$ for all $m \in \mathbf{Z}_+$ and that, more generally, $R_{k,m} = P^{S_N}$ for $m \in \mathbf{Z}_+$ and $k \in \mathbf{N}$, if $E(M^j) = E(N^j)$ for all $j \in \{1, \dots, k\}$. This means, that if $E(M) = E(N) < \infty$, then $R_{1,m} = P^{S_N}$ for all $m \in \mathbf{Z}_+$. Further, if $E(M^j) = E(N^j) < \infty$ for $j \in \{1, 2\}$, then $R_{2,m} = P^{S_N}$ for all $m \in \mathbf{Z}_+$. Therefore, under these assumptions, bounds for the approximation error between $\mathcal{L}(S_M)$ and $R_{k,m}$ are also bounds for d_τ .

For the main result in this section, we use the following notation. For $m, n \in \mathbf{Z}_+$, set $m \vee n = \max\{m, n\}$. For $x \in \mathbf{R}$ and $n \in \mathbf{Z}_+$, let $(x)_n = \prod_{j=0}^{n-1} (x + j)$ denote the Pochhammer symbol. Let the total variation distance between two finite signed measures be defined as it was given in (1) for probability measures. As indicated above, whenever we use the finite signed measure $R_{k,m}$ for $k, m \in \mathbf{Z}_+$, we assume that $\sum_{n=0}^{\infty} |a_n| n^k < \infty$. Theorems 1–3 are immediate consequences of the following theorem.

Theorem 4 (Cf. Roos and Pfeifer (2003)). *Let $k, m \in \mathbf{Z}_+$, $c_0 = 1$, and $c_j = \frac{1}{j - \sqrt{j(j-1)}} \in [1, 2]$, ($j \in \mathbf{N}$). Then*

$$d_{\text{TV}}(\mathcal{L}(S_M), R_{k,m}) \leq 2^k p^{k+1} \sum_{n=0}^{\infty} |a_n| \left| \binom{n-m}{k+1} \right|. \quad (13)$$

If additionally $k \geq 2$, then

$$d_{\text{TV}}(S_M, S_N) \leq \frac{1}{2} \sqrt{\frac{p}{q}} \sum_{n=0}^{\infty} c_{m \vee n} |a_n| |\sqrt{m} - \sqrt{n}|, \quad (14)$$

$$d_{\text{TV}}(\mathcal{L}(S_M), R_{1,m}) \leq \frac{p}{\sqrt{2}q} \sum_{n=0}^{\infty} |a_n| \left(m - n + n \ln \left(\frac{n + 1/2}{m + 1/2} \right) \right), \quad (15)$$

$$d_{\text{TV}}(\mathcal{L}(S_M), R_{k,m}) \leq \frac{2\sqrt{(k+1)!}}{\sqrt{(m+1)_{k-1}}} \left(\frac{p}{q} \right)^{(k+1)/2} \sum_{n=0}^{\infty} \frac{|a_n| |m-n|^{k+1}}{(\sqrt{m+1} + \sqrt{n+1})^2}. \quad (16)$$

Clearly, using (13) with $k = 0$ and (14), it is also possible to give bounds for the approximation error d_τ in the case, where $E(M)$ and $E(N)$ are not necessarily equal. Further, note that the constant $1/2$ in (14) is best possible. From Theorem 4, we see that the accuracy of approximation by $R_{k,m}$, that is to say the convergence rate for $p \rightarrow 0$, increases with k . But, with the help of (13), we see that, under certain assumptions, the distance in fact converges to zero as $k \rightarrow \infty$:

Corollary 3 *If $p < 1/2$ and if all moments of Z exist such that $\lim_{k \rightarrow \infty} E \frac{(2pZ)^{k+1}}{(k+1)!} = 0$, then, for all $m \in \mathbf{Z}_+$,*

$$\lim_{k \rightarrow \infty} d_{\text{TV}}(\mathcal{L}(S_M), R_{k,m}) = 0.$$

In particular, letting $m = 0$, we have

$$d_{\text{TV}}(\mathcal{L}(S_M), R_{k,0}) \leq d_{\text{TV}}(M, N) E \left[\frac{(2pZ)^{k+1}}{(k+1)!} \right].$$

Proof. Using (13), we obtain

$$d_{\text{TV}}(\mathcal{L}(S_M), R_{k,m}) \leq d_{\text{TV}}(M, N) \left[(2p)^{k+1} \binom{m+k}{k+1} + E \left(\frac{(2pZ)^{k+1}}{(k+1)!} \mathbf{1}_{[m+1, \infty)}(Z) \right) \right],$$

where, for a set A , $\mathbf{1}_A(Z) = 1$ when $Z \in A$ and $\mathbf{1}_A(Z) = 0$ otherwise. The assumptions show that the right-hand side tends to zero as $k \rightarrow 0$. This yields the first assertion. The second assertion is clear. \square

Note that the second assumption of Corollary 3 is valid when we assume that the moment generating function $E(e^{tZ})$ of Z is finite for $t = 2p$.

5 Consequence of Theorem 4: Asymptotic result

Using Theorem 4, it is possible to show several asymptotic results. In what follows, we present the simplest one of them under the condition that the summands X_j assume only

two values, i.e. zero and $x \neq 0$, say. With other words, $Q = \varepsilon_x$.

Theorem 5 *If $E(M) = E(N) < \infty$ and if $Q = \varepsilon_x$ with $x \neq 0$, then, for all $m \in \mathbf{Z}_+$,*

$$|d_\tau - p^2|\text{Var}(M) - \text{Var}(N)|| \leq p^3 \left[4 \sum_{n=0}^{\infty} |a_n| \binom{n-m}{3} + 2m(1+2p)^{m-1} |\text{Var}(M) - \text{Var}(N)| \right].$$

Proof. Under the above assumptions, we have $\beta_{2,m} = (\text{Var}(M) - \text{Var}(N))/2$. Therefore

$$|d_\tau - p^2|\text{Var}(M) - \text{Var}(N)|| \leq |d_\tau - d_{\text{TV}}(R_{2,m}, R_{1,m})| + |d_{\text{TV}}(R_{2,m}, R_{1,m}) - 2p^2|\beta_{2,m}||.$$

It is easy to show that the second term on the right-hand side is bounded by

$$2mp^3(1+2p)^{m-1}|\text{Var}(M) - \text{Var}(N)|.$$

Since $d_\tau = d_{\text{TV}}(\mathcal{L}(S_M), R_{1,m})$, the first term is bounded by $d_{\text{TV}}(\mathcal{L}(S_M), R_{2,m})$. Using (13), the assertion is proved. \square

Corollary 4 *If $E(M) = E(N) < \infty$, $Q = \varepsilon_x$ with $x \neq 0$, and $E(Z^3) < \infty$, then*

$$d_\tau = p^2|\text{Var}(M) - \text{Var}(N)| + O(p^3), \quad p \rightarrow 0.$$

References

- [1] BARBOUR, A.D. AND HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Camb. Philos. Soc.* **95**, 473–480.
- [2] BARBOUR, A.D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. Clarendon Press, Oxford.
- [3] DOEBLIN, W. (1938). *Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états*. *Rev. Math. Union Interbalkanique* **2**, 77–105.
- [4] GNEDENKO, B.V. AND KOROLEV, V.YU. (1996). *Random Summation. Limit Theorems and Applications*. CRC Press, Boca Raton.
- [5] GUT, A. (1988). *Stopped Random Walks. Limit Theorems and Applications*. Springer-Verlag, New York.
- [6] HEILMANN, W.-R. (1988). *Fundamentals of Risk Theory*. Verlag Versicherungswirtschaft, Karlsruhe.
- [7] KRUGLOV, V.M. AND KOROLEV, V.YU. (1990). *Limit Theorems for Random Sums*. (Russian). Izdatel'stvo Moskovskogo Universiteta, Moskva.
- [8] LE CAM, L. (1965). *On the distribution of sums of independent random variables*. In: *Bernoulli, Bayes, Laplace*. Eds. J. Neyman and L. Le Cam, Springer-Verlag, New York, 179–202.
- [9] LINDVALL, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.

- [10] LOGUNOV, P.L. (1990). *Estimates for the convergence rate to the Poisson distribution for random sums of independent indicators*. Theory Probab. Appl. **35**, 587–590.
- [11] LUKACS, E. (1970). *Characteristic Functions*. Second ed., Griffin, London.
- [12] MICHEL, R. (1987). *An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio*. ASTIN Bulletin **17**, 165–169.
- [13] ROOS, B. AND PFEIFER, D. (2003). *On the distance between the distributions of random sums*. J. Appl. Probab. **40**(1), 87–106.
- [14] VELLAISAMY, P. AND CHAUDHURI, B. (1996). *Poisson and compound Poisson approximations for random sums of random variables*. J. Appl. Probab. **33**, 127–137 .
- [15] YANNAROS, N. (1991). *Poisson approximation for random sums of Bernoulli random variables*. Statist. Probab. Lett. **11**, 161–165.