I N S T I T U T D E S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N P A P E R

0034

ON THE STOP-LOSS AND TOTAL VARIATION DISTANCES BETWEEN RANDOM SUMS

M. DENUIT and S. VAN BELLEGEM

http://www.stat.ucl.ac.be

ON THE STOP-LOSS AND TOTAL VARIATION DISTANCES BETWEEN RANDOM SUMS

Michel Denuit & Sébastien Van Bellegem
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
denuit@stat.ucl.ac.be & vanbellegem@stat.ucl.ac.be

December 2000

Abstract

The purpose of this work is to provide upper bounds on the stop-loss and total variation distances between random sums. The main theoretical argument consists in defining discrete analogs of the classical ideal metrics considered by Rachev and Rüschendorf (1990). An application in risk theory enhance the relevance of the approach proposed in this paper.

Key words and phrases: probability metrics, stop-loss distances, total variation distances, random sums, s-convex orderings, risk theory

1 Introduction and Motivation

In this paper, we are interested in the stop-loss and total variation probability metrics that can be found in Rachev and Rüschendorf (1990). The latter map a couple of distribution functions to the extended half-positive real line $\overline{\mathbb{R}}^+$ and possess the metric properties of symmetry, triangle inequality and a suitable analog of the identification property. The use of metrics in many problems of applied probability allows one to decide whether the proposed stochastic model provides a satisfactory approximation to the real model, and if so, within what limit. Recently, metrics have received considerable interest in probability theory and mathematical statistics. For a thorough presentation of probability metrics, we refer the interested reader e.g. to Rachev (1991).

The metrics now used in applied probability number in the dozens. Let us recall the definition of several classical distances that we will consider in this paper. First, the total-variation distance d_{TV} between two non-negative random variables X and Y with respective distribution functions F_X and F_Y is given by

$$d_{TV}(X,Y) = \int_{t\in\mathbb{R}^+} |dF_X(t) - dF_Y(t)|;$$

the latter definition slightly differs from the standard one (by a constant factor, in fact) for reasons which will later become obvious to the reader. It is worth mentioning that we write by abuse of notation $d_{TV}(X,Y)$ whereas $d_{TV}(F_X,F_Y)$ would be more appropriate since d_{TV} does not depend on the particular versions of the random variables X and Y (i.e., d_{TV}, as all the other metrics considered in this paper, is a simple metric). Another popular metric is the Kolmogorov (or uniform) metric d_K, defined as

$$d_K(X,Y) = \sup_{t\in\mathbb{R}^+} |F_X(t) - F_Y(t)|.$$

The definition of d_K is based on the famous Kolmogorov-Smirnov test statistic. The Wasserstein distance, also known as the Dudley or Kantorovitch distance, is defined as

$$d_W(X,Y) = \int_{t\in\mathbb{R}^+} |F_X(t) - F_Y(t)|dt.$$

In actuarial sciences, Gerber (1981) defined the stop-loss distance as

$$d_{s\ell}(X,Y) = \sup_{t\in\mathbb{R}^+} |\mathbb{E}(X-t)_+ - \mathbb{E}(Y-t)_+|,$$

where $x_+ = \max\{x,0\}$. The metric $d_{s\ell}$ has been used e.g. by Gerber (1981, 1984), and De Pril and Dhaene (1992). In addition to the distances recalled above, we also consider the classes of the stop-loss and total-variation distances that can be found in Rachev and Rüschendorf (1990).

The problem examined in the present paper can be described as follows. Let d be a probability distance. Given the sequences $\{X_1,X_2,X_3,\dots\}$ and $\{Y_1,Y_2,Y_3,\dots\}$ of independent and identically distributed non-negative random variables, and two counting random variables N and M independent of the X_i's and of the Y_i's, respectively, let us define the compound sums

$$S_N = \sum_{k=1}^{N} X_k, \;\; S_M = \sum_{k=1}^{M} X_k \text{ and } T_M = \sum_{k=1}^{M} Y_k \tag{1.1}$$

with the convention that the empty sum equals 0. Considering non-negative integer-valued summands, Vellaisamy and Chaudhuri (1996) have shown in their Lemma 3.1 that

$$d_{TV}(S_N, S_M) \leq d_{TV}(N, M); \tag{1.2}$$

in the special case of Bernoulli summands, (1.2) has been obtained by Finkelstein, Tucker and Veeh (1990, Lemma 4). Vellaisamy and Chaudhuri (1996) also proved that

$$d_{TV}(S_N, T_M) \leq d_{TV}(N, M) + \min\{\mathbb{E}M, \mathbb{E}N\} d_{TV}(X_1, Y_1); \tag{1.3}$$

see their Theorem 3.1. One of the aim of this paper is to generalize the inequalities (1.2) and (1.3) to d_K, d_W and $d_{s\ell}$, as well as to the whole classes of stop-loss distances and total-variation distances. Moreover, we do no more assume that the summands X_i and Y_i are integer-valued but we deal with the general situation of non-negative random variables.

Our main theoretical vehicle are discrete distances, suitable to evaluate the closeness of counting random variables; these discrete counterparts of the stop-loss and total variation distances have been defined in Denuit, Lefèvre and Utev (2000) and have proved their usefulness in various problems encountered in actuarial risk theory.

In the building of appropriate stochastic models, it is often interesting to be able to compare random variables relating to the different models. To this end, one usually resorts on stochastic orderings, i.e. on binary relations defined on sets of distribution functions. These relations aim to mathematically translate intuitive ideas as "being larger", or "being more variable", for instance. For a comprehensive background on stochastic orderings including a variety of applications, the reader is referred e.g. to Shaked and Shanthikumar (1994). In this paper, we will be interested in classes of integral stochastic orderings; such orderings are defined with the help of classes of measurable functions using the operator expectation. Specifically, considering two random variables X and Y, X is said to be smaller than Y in the integral stochastic order $\prec_{\mathcal{F}}$ generated by the class $\mathcal{F}$ of measurable functions if, $\mathbb{E}\phi(X) \leqslant \mathbb{E}\phi(Y)$ for all functions $\phi \in \mathcal{F}$, provided that the expectations exist. A general study of these orderings is made in Műller (1997). Taking for $\mathcal{F}$ the class of the non-decreasing functions yields the stochastic dominance $\preceq_{st}$. Similarly, if $\mathcal{F}$ is the class of the (non-decreasing) convex functions then $\preceq_{\mathcal{F}}$ is the (increasing) convex order $\preceq_{(i)cx}$.

Traditionally, these orderings are defined for random variables X and Y which take on values in $\mathbb{R}$ or $\mathbb{R}^+$ (or an entire interval). Orderings proper to discrete random variables X and Y have received much less attention. The basic situation is when all outcomes lie in the set $\mathbb{N}$ of the non-negative integers. In this framework, the classes $\mathcal{F}$ represent classes of functions defined only on $\mathbb{N}$ and, roughly, as in the continuous case but with the operator of forward difference substituted for the operator of derivative. We refer the interested reader e.g. to Fishburn and Lavalle (1995), Lefèvre and Utev (1996), Denuit and Lefèvre (1997) and Denuit, Lefèvre and Utev (1999) for more details.

A crucial point, which has often been ignored in the literature, is the close link existing between probability metrics and stochastic orderings. In that respect, we expand here on the pionneering work of Lefèvre and Utev (1998), who were the first authors to systematically exploit the relationships between metrics and orderings generated by classes of measurable functions.

Henceforth, we confine our analysis to non-negative random variables. Given a random variable X, F_X denotes the cumulative distribution function (cdf, in short) of X, i.e. $F_X(t) = \mathbb{P}[X \leqslant t]$, $t \in \mathbb{R}$, and $\overline{F}_X \equiv 1 - F_X$ denotes the decumulative distribution function (ddf, in short) or survival function of X, i.e. $\overline{F}_X(t) = \mathbb{P}[X > t]$, $t \in \mathbb{R}$. The probability density function (pdf, in short) of X with respect to some dominating measure (Lebesgue measure on the half-positive real line $\mathbb{R}^+$ or counting measure on the non-negative integers $\mathbb{N}$) is denoted as f_X. We use capital letters X, Y, Z, ... to represent continuous random variables and capital letters M, N, O, ... to represent integer-valued random variables.

To end this introduction, let us say a few words about the organization of the present paper. Denuit, Lefèvre and Shaked (1998) showed how to generalize the standard stochastic dominance, convex and increasing convex orders to general classes of integral stochastic orderings called s-convex and s-increasing convex orders; we briefly recall this construction in Section 2. A similar generalization is possible in the spirit of Rachev and Rüschendorf (1990) for d_{TV}, d_K, d_W and $d_{s\ell}$; this will be done in Section 3. Under appropriate ordering condition, distances between random variables reduce to differences of moments. In both Sections 2 and 3, we stress the similarities and complementarities between discrete and continuous orderings/distances. In Section 4, upper bounds for various distances between random sums are derived extending (1.2) and (1.3). The key technical argument is the discrete version of the probabilistic distances introduced in Section 3. Finally, Section 5 gives some applications in risk theory which enhance the relevance of the approach developed in the present paper.

2 Stochastic orderings of convex-type

2.1 Comparing continuous random variables

2.1.1 Continuous iterated right-tails

Let X be a random variable valued in $\mathbb{R}^+$ with pdf f_X. To X is associated the sequence $\{\overline{F}_X^{[k]},\ k = 1, 2, 3, \ldots\}$ of its iterated right-tails obtained by repeated integration of the pdf over regions $(t, +\infty)$. Specifically, the $\overline{F}_X^{[k]}$'s are defined recursively as follows:

$$\overline{F}_X^{[0]} \equiv f_X \text{ and } \overline{F}_X^{[k+1]}(t) = \int_{x=t}^{+\infty} \overline{F}_X^{[k]}(x)dx, \;\; t \in \mathbb{R}^+. \tag{2.1}$$

Note in passing that $\overline{F}_X^{[1]}$ coincides with the ddf $\overline{F}_X$ of X; $\overline{F}_X^{[k]}$ is referred to as the k-th iterated right-tail of X. Iterated right-tail functions are closely related to the stop-loss transforms or partial moments. Specifically, it can be shown by induction and successive application of Fubini's theorem that

$$(s-1)!\overline{F}_X^{[s]}(t) = \mathbb{E}(X-t)_+^{s-1}, \;\; x \in \mathbb{R}^+ \tag{2.2}$$

with the convention that $(x-t)_+^0 = 1$ if $x > t$ and 0 otherwise. The name "stop-loss" transform comes from the fact that $\overline{F}_X^{[2]}(t) = \mathbb{E}(X-t)_+$ is called the stop-loss premium with deductible t in risk theory.

2.1.2 Continuous orderings of convex-type

Many comparisons of random variables X and Y are based on the iterated right-tails, as those introduced by Denuit *et al.* (1998) whose definition is recalled next. Given two non-negative random variables X and Y with finite moments of degree $s-1$, X is said to precede Y in the s-convex sense, denoted as $X \preceq_s^{\mathbb{R}^+} Y$, if $\mathbb{E}X^k = \mathbb{E}Y^k$ for $k = 1, 2, \ldots, s-1$ and $\overline{F}_X^{[s]}(t) \leqslant \overline{F}_Y^{[s]}(t)$ for all $t \in \mathbb{R}^+$. As particular cases of the stochastic orderings $\preceq_s^{\mathbb{R}^+}$, $\preceq_1^{\mathbb{R}^+}$ is the usual stochastic dominance $\preceq_{st}$, while $\preceq_2^{\mathbb{R}^+}$ coincides with the well-known convex order $\preceq_{cx}$.

2.2 Comparing discrete random variables

2.2.1 Discrete iterated right-tails

Let us define for arithmetic random variables the analogs of the continuous iterated right tails defined in (2.1). Let N be a random variable valued in $\mathbb{N}$. To N is associated the sequence $\{\overline{F}_N^{[k]},\ k = 1, 2, \ldots\}$ of its iterated right-tails obtained by repeated summation of the pdf over regions $\{j, j+1, j+2, \ldots\}$. Specifically, the $\overline{F}_N^{[k]}$'s are defined recursively as follows:

$$\overline{F}_N^{[0]} \equiv f_N \text{ and } \overline{F}_N^{[k+1]}(j) = \sum_{i=j}^{+\infty} \overline{F}_N^{[k]}(i), \qquad j \in \mathbb{N}. \tag{2.3}$$

It is interesting to compare (2.1) and (2.3). It is worth mentioning that we do not formally adopt a different notation for the discrete iterated right-tails $\overline{F}_N^{[k]}$ and the continuous ones $\overline{F}_X^{[k]}$; the context will make clear which integrated right-tails are used.

Of course, the reader familiar with measure theory has rapidly discovered that a general definition of iterated right-tails calling upon dominating measures could easily be developed. However, we will not follow this route here and confine ourselves to the standard continuous and discrete cases, which appear to be sufficient for the applications we have in mind.

Lefèvre and Utev (1996) proved that the $\overline{F}_N^{[s]}$'s correspond to some factorial moments: specifically, for any $j, k \in \mathbb{N}$, the equalities

$$\overline{F}_N^{[s]}(j) = E\binom{N-j+s-1}{s-1} = \frac{1}{(s-1)!}\mathbb{E}\Big[(N-j+s-1)\ldots(N-j+1)\Big], \tag{2.4}$$

hold true, with the convention that $\binom{\epsilon_1}{\epsilon_2} = 0$ if $\epsilon_1 < \epsilon_2$. Again, compare (2.2) to (2.4).

2.2.2 Discrete orderings of convex-type

Denuit and Lefèvre (1997) defined the s-convex stochastic orderings among integer-valued random variables as follows. Let N and M be two random variables valued in $\mathbb{N}$, N is said to precede M in the s-convex sense, written as $N \preceq_s^{\mathbb{N}} M$, if $EN^k = EM^k$ for $k = 1, \dots, s-1$, and $\overline{F}_N^{[s]}(j) \leqslant \overline{F}_M^{[s]}(j)$ for all $j \in \mathbb{N}$.

The orderings $\preceq_s^{\mathbb{N}}$ and $\preceq_s^{\mathbb{R}^+}$ are not equivalent for $s \geqslant 3$, whereas they coincide for $s = 1, 2$, as shown by Denuit, Lefèvre and Utev (1999). More specifically, given two integer-valued random variables N and M, we have that

$$N \preceq_1^{\mathbb{N}} M \Leftrightarrow N \preceq_1^{\mathbb{R}^+} M \Leftrightarrow N \preceq_{st} M$$

$$N \preceq_2^{\mathbb{N}} M \Leftrightarrow N \preceq_2^{\mathbb{R}^+} M \Leftrightarrow N \preceq_{cx} M$$

and

$$N \preceq_s^{\mathbb{N}} M \Rightarrow N \preceq_s^{\mathbb{R}^+} M$$

with a generally false reciprocal. From the three latter relations, we see that a ranking in the $\preceq_s^{\mathbb{N}}$-sense for counting random variables is more informative as soon as $s \geq 3$.

3 Metrics of convex type

3.1 Distances between continuous random variables

Rachev and Rüschendorf (1990) introduced useful distances based on iterated right-tails; these authors apply them to measure the closeness between the individual and the collective models in risk theory. We resort here on a slight modification of their definition.

Definition 3.1. *(i) Given two random variables X and Y valued in $\mathbb{R}^+$, the stop-loss distance of degree s between X and Y, denoted as $d_s^{\mathbb{R}^+}(X,Y)$, is defined by*

$$d_s^{\mathbb{R}^+}(X,Y) = \sup_{t \in \mathbb{R}^+} \left| \overline{F}_X^{[s]}(t) - \overline{F}_Y^{[s]}(t) \right|$$

where $\overline{F}_X^{[s]}$ and $\overline{F}_Y^{[s]}$ are defined according to (2.1).

(ii) Given two random variables X and Y valued in $\mathbb{R}^+$, the total variation distance of degree s between X and Y, denoted as $D_s^{\mathbb{R}^+}(X,Y)$, is defined as

$$D_s^{\mathbb{R}^+}(X,Y) = \int_{t \in \mathbb{R}^+} \left| \overline{F}_X^{[s]}(t) - \overline{F}_Y^{[s]}(t) \right| dt$$

where $\overline{F}_X^{[s]}$ and $\overline{F}_Y^{[s]}$ are defined according to (2.1).

The main difference between our Definition 3.1 and Rachev and Rüschendorf (1990)'s study is that we define $d_s^{\mathbb{R}^+}$ by a supremum over $t \in \mathbb{R}^+$ rather than over $t \in \mathbb{R}$. This modification enables us to measure the distance between random variables with different moments. For instance, the distance between two Negative Exponential distributions is easily computed with our definition and reduce up to a constant factor to the difference of the $(s-1)$th moments for $s \geqslant 2$ whereas this distance is infinite for $s \geqslant 2$ in Rachev and Rüschendorf (1990) framework. Considering the introduction, we have $d_1^{\mathbb{R}^+}(X,Y) = d_K(X,Y)$ and $d_2^{\mathbb{R}^+}(X,Y) = d_{s\ell}(X,Y)$.

The metric $D_s^{\mathbb{R}^+}$ is closely related to the Zolotarev metric ζ_s (denoted as $d_{s,1}$ in Rachev and Rüschendorf (1990)); it can be shown that they are equivalent provided X and Y have identical first $s-1$ moments. Nevertheless, $D_s^{\mathbb{R}^+}$ can be used to measure the distance between two random variables with different first moments, whereas ζ_s is infinite in such a case. Considering the introduction, we have $D_0^{\mathbb{R}^+}(X,Y) = d_{TV}(X,Y)$ and $D_1^{\mathbb{R}^+}(X,Y) = d_W(X,Y)$.

3.2 Distances between discrete random variables

The discrete analogs of $d_s^{\mathbb{R}^+}$ and $D_s^{\mathbb{R}^+}$ have been introduced by Denuit, Lefèvre and Utev (2000). As it was the case for stochastic orderings, the construction of the discrete versions of the stop-loss and total variation metrics is similar to their continuous counterparts, substituting the discrete iterated right-tails for the continuous ones.

Definition 3.2. *(i) Given the random variables N and M valued in $\mathbb{N}$, the stop-loss distance of degree s between N and M, denoted as $d_s^{\mathbb{N}}(N,M)$, is defined as*

$$d_s^{\mathbb{N}}(N,M) = \sup_{j\in\mathbb{N}} \left| \overline{F}_N^{[s]}(j) - \overline{F}_M^{[s]}(j) \right|$$

where $\overline{F}_N^{[s]}$ and $\overline{F}_M^{[s]}$ are defined according to (2.3).

(ii) Given the random variables N and M valued in $\mathbb{N}$, the total variation distance of degree s between N and M, denoted as $D_s^{\mathbb{N}}(N,M)$, is defined as

$$D_s^{\mathbb{N}}(N,M) = \sum_{j=0}^{\infty} \left| \overline{F}_N^{[s]}(j) - \overline{F}_M^{[s]}(j) \right|$$

where $\overline{F}_N^{[s]}$ and $\overline{F}_M^{[s]}$ are defined according to (2.3).

Given two integer-valued random variables N and M, we can obviously resort on $d_s^{\mathbb{R}^+}$ and $d_s^{\mathbb{N}}$ to assert their respective distance. As a matter of fact, if s is equal to 1 or 2, it is equivalent to compare N and M with the help of the discrete or the continuous stop-loss distances. A similar result holds for total variation distances. More precisely, we have the following property.

Property 3.3. *Given two random variables N and M valued in $\mathbb{N}$, we have $d_s^{\mathbb{N}}(N,M) = d_s^{\mathbb{R}^+}(N,M)$ for $s = 1$ or 2. Similarly, $D_s^{\mathbb{N}}(N,M) = D_s^{\mathbb{R}^+}(N,M)$ for $s = 0$ or 1.*

Proof. The equality $d_1^{\mathbb{N}}(N,M) = d_1^{\mathbb{R}^+}(N,M)$ is obvious. Let us now turn to $s = 2$. Since

$$\mathbb{E}\begin{pmatrix} N-j+1 \\ 1 \end{pmatrix} = \mathbb{E}\max\{N-j+1, 1\} = 1 + \mathbb{E}(N-j)_+$$

the stop-loss distance is equivalently given by

$$d_2^{\mathbb{N}}(N,M) = \sup_{j\in\mathbb{N}} |\mathbb{E}(N-j)_+ - \mathbb{E}(M-j)_+|.$$

Now, for $k \leqslant t < k+1$, it is easily seen that the formula

$$\mathbb{E}(N-t)_+ = \mathbb{E}(N-k)_+ + (t-k)\overline{F}_N(k)$$

holds true. Therefore, denoting as $[t]$ the integer part of the real number t, we have

$$\begin{aligned} d_2^{\mathbb{R}^+}(N,M) &= \sup_{t\in\mathbb{R}^+} |\mathbb{E}(N-t)_+ - \mathbb{E}(M-t)_+| \\ &= \sup_{t\in\mathbb{R}^+} \Big|\mathbb{E}(N-[t])_+ - \mathbb{E}(M-[t])_+ + (t-[t])\Big\{\overline{F}_N([t]) - \overline{F}_M([t])\Big\}\Big| \\ &= d_2^{\mathbb{N}}(M,N), \end{aligned}$$

as announced. By the same argument, a similar result holds when we consider the total variation distance of degre $s = 0$ or 1. □

Considering Property 3.3, we see that the discrete and continuous versions of the standard distances d_{TV}, d_K, d_W and $d_{s\ell}$ coincide. However, this is no more necessarily true for higher degrees stop-loss and total variation metrics, as pointed out in Denuit *et al.* (2000). More precisely, given two random variables N and M valued in $\mathbb{N}$, we only have

$$d_s^{\mathbb{R}^+}(N,M) \leqslant d_s^{\mathbb{N}}(N,M) \text{ for } s \geqslant 3. \tag{3.1}$$

As we will see in the following result, this inequality may be strict in some general cases. In other words, considering arithmetic random variables as continuous ones underestimates the distance existing between them. Intuitively speaking, the arithmetic random variables N and M look more similar if we consider the set of all the non-negative random variables (because they have the discreteness property in common) than if we restrict ourselves to random variables valued in $\mathbb{N}$. Let us now provide a sufficient condition to have a strict inequality in (3.1).

Property 3.4. *Let N and M be two random variables valued in $\mathbb{N}$ and let an integer $s \geqslant 3$ be given. If the stochastic inequality $N \preceq_p^{\mathbb{N}} M$ strictly holds for some $p < s-1$ (i.e. N and M are not identically distributed) then we have a strict inequality in (3.1).*

Proof. First, we note that the $\preceq_p^{\mathbb{N}}$ ranking implies that

$$\overline{F}_N^{[k]}(t) \leqslant \overline{F}_M^{[k]}(t) \text{ for all } t \in \mathbb{R}^+ \text{ and any } k \geqslant p+1.$$

By differentiation, we see that the supremum defining $d_s^{\mathbb{R}^+}$ is attained for $t = 0$ so that

$$d_s^{\mathbb{R}^+}(N,M) = \frac{EM^{s-1} - EN^{s-1}}{(s-1)!}.$$

Secondly, the $\preceq_p^{\mathbb{N}}$ ranking also yields

$$\overline{F}_N^{[k]}(j) \leqslant \overline{F}_M^{[k]}(j) \text{ for all } j \in \mathbb{N} \text{ and } k \geqslant p+1.$$

Therefore, the supremum defining $d_s^{\mathbb{N}}$ is attained for $j = 0$ and

$$d_s^{\mathbb{N}}(N, M) = \frac{1}{(s-1)!}\left\{EM^{s-1} - EN^{s-1} + \sum_{k=p}^{s-2} \alpha_k \left(EM^k - EN^k\right)\right\}$$

where α_k are the coefficients of the Newton expansion which are non-negative. It finally follows that $d_s^{\mathbb{N}}(N, M) > d_s^{\mathbb{R}^+}(N, M)$, as announced. □

The condition given in Property 3.4 is general enough. One may take for instance for N and M two Poisson distributed random variables with $\mathbb{E}N < \mathbb{E}M$. In this case, the condition is fulfilled with $p = 1$.

A similar result holds for the total variation distance. In this case, we have for any two random variables N and M valued in $\mathbb{N}$ that

$$D_s^{\mathbb{R}^+}(N, M) \leqslant D_s^{\mathbb{N}}(N, M) \text{ for } s \geqslant 2. \tag{3.2}$$

The counterpart of Property 3.4 for total variation distances is stated next. Since the proof follows the same lines than the reasoning developped in Property 3.4, we omitted it for brievety.

Property 3.5. *Let N and M be two random variables valued in $\mathbb{N}$ and let an integer $s \geqslant 2$ be given. If the stochastic inequality $N \preceq_p^{\mathbb{N}} M$ strictly holds for some $p < s$ (i.e. N and M are not identically distributed) then we have a strict inequality in (3.2).*

Let us now show that the use of $d_s^{\mathbb{R}^+}$ in lieu of $d_s^{\mathbb{N}}$ may yield different conclusions. Indeed, there exists some random variables N, M and O valued in $\mathbb{N}$ such that $d_s^{\mathbb{N}}(N, M) < d_s^{\mathbb{N}}(N, O)$ but $d_s^{\mathbb{R}^+}(N, M) > d_s^{\mathbb{R}^+}(N, O)$. In such a case, we conclude on the basis of $d_s^{\mathbb{N}}$ that M is closer to N than O but resorting on $d_s^{\mathbb{R}^+}$ yields the opposite conclusion.

For simplicity, we will show an example for $s = 3$. Let N be Poisson distributed with parameter θ_1, O conform to the Poisson distribution with parameter θ_2 and M obey to a mixture of Poisson laws with random parameter Θ such that $E\Theta = \theta_1$ and $\theta_2^2 + \theta_2 - \theta_1^2 < E\Theta^2 < \theta_2 + \theta_2^2 - \theta_1^2 + 3(\theta_2 - \theta_1)$ (take for example $\theta_1 = 1$ and $\theta_2 = 2$). For these variables, it is easy to see that

$$\begin{aligned}
d_3^{\mathbb{R}^+}(N, O) &= \frac{1}{2}\left(\theta_2^2 - \theta_1^2 + \theta_2 - \theta_1\right) \\
&< d_3^{\mathbb{R}^+}(N, M) = d_3^{\mathbb{N}}(N, M) = \frac{1}{2}\left(E\Theta^2 - \theta_1\right) \\
&< d_3^{\mathbb{N}}(N, O) = \frac{1}{2}\left(\theta_2^2 - \theta_1^2 + \theta_2 - \theta_1\right) + 3(\theta_2 - \theta_1).
\end{aligned}$$

4 Analysis of random sums

The metrics $D_s^{\mathbb{R}^+}$ and $d_s^{\mathbb{R}^+}$ are both ideal; as a consequence, the inequalities $d_s^{\mathbb{R}^+}(S_n, T_n) \leqslant n d_s^{\mathbb{R}^+}(X_1, Y_1)$ and $D_s^{\mathbb{R}^+}(S_n, T_n) \leqslant n D_s^{\mathbb{R}^+}(X_1, Y_1)$ both hold true. Ideal probabilistic distances are appropriate to deal with various approximation problems involving sums of random variables (see e.g. Chapter 14 of Rachev (1991)). A straight extension of this result consists in replacing the deterministic number of terms n by a couting random variable N independent of the X_i's and of the Y_i's. This results in

$$d_s^{\mathbb{R}^+}(S_N,T_N)\leqslant \mathbb{E}N d_s^{\mathbb{R}^+}(X_1,Y_1) \text{ and } D_s^{\mathbb{R}^+}(S_N,T_N)\leqslant \mathbb{E}N D_s^{\mathbb{R}^+}(X_1,Y_1).$$

However, the derivation of an upper bound on $d_s^{\mathbb{R}^+}(S_N,T_M)$ seems to be an open problem. This is due to the lack of the discrete counterpart $d_s^{\mathbb{N}}$ of $d_s^{\mathbb{R}^+}$. Partial solutions have been proposed in the literature for d_{TV}, e.g. by Finkelstein, Tucker and Veeh (1990) who considered in their Lemma 4 compound sums of Bernoulli random variables and by Vellaisamy and Chaudhuri (1996) who considered in their Lemma 3.1 integer-valued summands. In this section, we aim to bound the distances $D_s^{\mathbb{R}^+}$ and $d_s^{\mathbb{R}^+}$ between two compound sums S_N and T_M.

Before stating our main result, we need the following technical lemma.

Lemma 4.1. *Let Δ denote the classical forward difference operator, i.e. given a function $\phi:\mathbb{N}\to\mathbb{R}$, $\Delta\phi(n)=\phi(n+1)-\phi(n)$, $n\in\mathbb{N}$. Further, let Δ^s denote the sth iterated of Δ (with the convention that Δ^0 is the identity). We have that $\Delta^s\mathbb{E}[S_n^s]=s!(\mathbb{E}X_1)^s$.*

Proof. The announced result is true for $s=1$ since $\Delta\mathbb{E}[S_n]=\mathbb{E}[S_{n+1}-S_n]=\mathbb{E}X_1$. Let us now proceed by induction: we assume that $\Delta^k\mathbb{E}[S_n^k]=k!(\mathbb{E}X_1)^k$ is valid for any $k\leqslant s-1$, and we aim to establish the desired result for s. Let us start from

$$\begin{aligned}\Delta^s\mathbb{E}S_n^s &= \mathbb{E}\Delta^s S_n^s=\mathbb{E}\Delta^s\Big(S_n^{s-1}S_n\Big)\\ &= \mathbb{E}\sum_{j=0}^{s}\begin{pmatrix} s\\ j\end{pmatrix}\Delta^j S_n^{s-1}\Delta^{s-j}S_{n+j}\end{aligned}$$

which follows from the Leibniz formula for forward differences. Now, with the aid of the Newton formula, it is not difficult to check that

$$\Delta^{s-j}S_{n+j}=\sum_{\ell=0}^{s-j}\begin{pmatrix} s-j\\ \ell\end{pmatrix}(-1)^{s-j-\ell}\Big(X_{n+j+1}+\ldots+X_{n+j+\ell}\Big)$$

so that $\Delta^j S_n^{s-1}$ and $\Delta^{s-j}S_{n+j}$ are independent random variables. This yields

$$\begin{aligned}\Delta^s\mathbb{E}S_n^s &= \sum_{j=0}^{s}\begin{pmatrix} s\\ j\end{pmatrix}\Delta^j\mathbb{E}S_n^{s-1}\Delta^{s-j}\Big\{(n+j)\mathbb{E}X_1\Big\}\\ &= s\Delta^{s-1}\mathbb{E}S_n^{s-1}\mathbb{E}X_1=s\mathbb{E}X_1(s-1)!(\mathbb{E}X_1)^{s-1},\end{aligned}$$

which ends the proof. □

With the help of Lemma 4.1, we are now in a position to bound the distance between the compound sums S_M and S_N.

Proposition 4.2. *(i) $D_s^{\mathbb{R}^+}(S_N,S_M)\leqslant(\mathbb{E}X_1)^s D_s^{\mathbb{N}}(N,M)$;*

(ii) $d_s^{\mathbb{R}^+}(S_N,S_M)\leqslant(\mathbb{E}X_1)^{s-1}d_s^{\mathbb{N}}(N,M)$.

Proof. (i) We start from

$$\begin{aligned}
\overline{F}^{[s]}_{S_N}(t)-\overline{F}^{[s]}_{S_M}(t) &= \sum_{n=0}^{+\infty}\Big\{\mathbb{P}[N>n-1]-\mathbb{P}[M>n-1]\Big\}\overline{F}^{[s]}_{S_n}(t)\\
&\quad -\sum_{n=0}^{+\infty}\Big\{\mathbb{P}[N>n]-\mathbb{P}[M>n]\Big\}\overline{F}^{[s]}_{S_n}(t)\\
&= \sum_{n=0}^{+\infty}\Big\{\overline{F}^{[1]}_{N}(n)-\overline{F}^{[1]}_{M}(n)\Big\}\Delta\overline{F}^{[s]}_{S_n}(t).
\end{aligned}$$

Iterating this reasoning yields

$$\overline{F}^{[s]}_{S_N}(t)-\overline{F}^{[s]}_{S_M}(t)=\sum_{n=0}^{+\infty}\Big\{\overline{F}^{[s]}_{N}(n)-\overline{F}^{[s]}_{M}(n)\Big\}\Delta^s\overline{F}^{[s]}_{S_n}(t).$$

Now,

$$\begin{aligned}
D_s^{\mathbb{R}^+}(S_N,S_M) &= \int_{t\in\mathbb{R}^+}\left|\sum_{n=0}^{+\infty}\Big\{\overline{F}^{[s]}_{N}(n)-\overline{F}^{[s]}_{M}(n)\Big\}\Delta^s\overline{F}^{[s]}_{S_n}(t)\right|dt\\
&\leqslant \sum_{n=0}^{+\infty}\left|\overline{F}^{[s]}_{N}(n)-\overline{F}^{[s]}_{M}(n)\right|\int_{t\in\mathbb{R}^+}\Delta^s\overline{F}^{[s]}_{S_n}(t)dt\\
&= \frac{1}{s!}D_s^{\mathbb{N}}(N,M)\Delta^s\mathbb{E}S_n^s=(\mathbb{E}X_1)^sD_s^{\mathbb{N}}(N,M),
\end{aligned}$$

where the last equality follows from Lemma 4.1.

(ii) Proceeding as in the proof of (i), we have

$$\begin{aligned}
d_s^{\mathbb{R}^+}(S_N,S_M) &= \sup_{t\in\mathbb{R}^+}\left|\sum_{n=0}^{+\infty}\Big\{\overline{F}^{[s]}_{N}(n)-\overline{F}^{[s]}_{M}(n)\Big\}\Delta^s\overline{F}^{[s]}_{S_n}(t)\right|\\
&\leqslant \sup_{t\in\mathbb{R}^+}\sup_{j\in\mathbb{N}}\left|\overline{F}^{[s]}_{N}(j)-\overline{F}^{[s]}_{M}(j)\right|\sum_{n=0}^{+\infty}\Delta^s\overline{F}^{[s]}_{S_n}(t)\\
&= d_s^{\mathbb{N}}(N,M)\sup_{t\in\mathbb{R}^+}\Delta^{s-1}\overline{F}^{[s]}_{S_n}(t),
\end{aligned}$$

Now,

$$\frac{d}{dt}\Delta^{s-1}\overline{F}^{[s]}_{S_n}(t)=-\Delta^{s-1}\overline{F}^{[s-1]}_{S_n}(t)\leqslant 0$$

so that the supremum is attained for $t=0$. Therefore,

$$\begin{aligned}
d_s^{\mathbb{R}^+}(S_N,S_M) &\leqslant d_s^{\mathbb{N}}(N,M)\Delta^{s-1}\overline{F}^{[s]}_{S_n}(0)\\
&= \frac{1}{(s-1)!}d_s^{\mathbb{N}}(N,M)\Delta^{s-1}\mathbb{E}S_n^{s-1}=(\mathbb{E}X_1)^{s-1}d_s^{\mathbb{N}}(N,M),
\end{aligned}$$

where the last equality follows from Lemma 4.1. □

It is worth mentioning that in the special case $s = 0$, Proposition 4.2(i) reduces to (1.2). Let us now turn to the general case and consider the distance between S_N and T_M.

Proposition 4.3. *(i)* $D_s^{\mathbb{R}^+}(S_N, T_M)$

$$\leqslant \min\Big\{(\mathbb{E}X_1)^s D_s^{\mathbb{N}}(N, M) + \mathbb{E}M D_s^{\mathbb{R}^+}(X_1, Y_1), (\mathbb{E}Y_1)^s D_s^{\mathbb{N}}(N, M) + \mathbb{E}N D_s^{\mathbb{R}^+}(X_1, Y_1)\Big\}$$

(ii) $d_s^{\mathbb{R}^+}(S_N, T_M)$

$$\leqslant \min\Big\{(\mathbb{E}X_1)^{s-1} d_s^{\mathbb{N}}(N, M) + \mathbb{E}M d_s^{\mathbb{R}^+}(X_1, Y_1), (\mathbb{E}Y_1)^{s-1} d_s^{\mathbb{N}}(N, M) + \mathbb{E}N d_s^{\mathbb{R}^+}(X_1, Y_1)\Big\}.$$

Proof. The results simply follow from the triangle inequality. Let us prove (i); the reasoning yielding (ii) is similar. Let write

$$\begin{aligned} D_s^{\mathbb{R}^+}(S_N, T_M) &\leqslant D_s^{\mathbb{R}^+}(S_N, S_M) + D_s^{\mathbb{R}^+}(S_M, T_M) \\ &\leqslant (\mathbb{E}X_1)^s D_s^{\mathbb{N}}(N, M) + \mathbb{E}M D_s^{\mathbb{R}^+}(X_1, Y_1); \end{aligned}$$

similarly, one gets

$$D_s^{\mathbb{R}^+}(S_N, T_M) \leqslant (\mathbb{E}Y_1)^s D_s^{\mathbb{N}}(N, M) + \mathbb{E}N D_s^{\mathbb{R}^+}(X_1, Y_1).$$

This ends the proof. □

It is worth mentioning that in the special case $s = 0$, Proposition 4.3(i) reduces to (1.3).

5 Application to the individual and collective models of risk theory

Consider an insurance portfolio consisting of n policies. This portfolio is observed during a given reference period (one year, say). To each contract is associated an indicator random variable I_i for the event "Policy no. i produces at least one claim during the reference period", $i = 1, 2, \ldots, n$; I_i is a Bernoulli random variable with mean q_i. The number of policies producing claims is denoted by $I_\bullet$ and is given by $I_\bullet = I_1 + I_2 + \ldots + I_n$.

Risk i produces a claim amount X_i during a given reference period (one year, say); the individual risks $X_1, X_2, \ldots, X_n$ are assumed to be mutually independent. The aggregate claims of the portfolio during the reference period is denoted by

$$S_{ind} = X_1 + X_2 + \ldots + X_n.$$

It is common to represent each X_i as

$$X_i = I_i V_i = \sum_{j=1}^{I_i} V_i = \begin{cases} V_i \text{ if } I_i = 1 \\ 0 \text{ if } I_i = 0, \end{cases}$$

$i = 1, 2, \ldots, n$, where V_i is the total claim amount produced by risk i.

The individual model is of primary importance for ratemaking on an individual basis. However, it is not convenient when calculations at the portfolio level have to be performed.

Therefore, it is customize to resort on an approximation of the individual model. A number of approximations for S_{ind} are provided in the actuarial literature; see e.g. Kaas and Gerber (1994) and the references therein.

The classical approximation can be seen as a two-stage procedure. We first approximate X_i by $\widetilde{X}_i$ defined as

$$\widetilde{X}_i = \sum_{k=1}^{N_i} V_{i;k},$$

where N_i is Poisson distributed with mean λ_i and is independent of the $V_{i;k}$'s, $k = 1, 2, \ldots$, the latter being independent copies of V_i. Then, we compute the cdf of S_{coll} given by

$$S_{coll} = \sum_{i=1}^{n} \widetilde{X}_i = \sum_{i=1}^{n} \sum_{k=1}^{N_i} V_{i;k} =_d \sum_{i=1}^{N} Y_i$$

where N is Poisson distributed with mean $\lambda = \sum_{i=1}^{n} \lambda_i$ and the Y_i's are independent and distributed according to the mixture

$$\frac{1}{\lambda} \sum_{i=1}^{n} \lambda_i \mathbb{P}[X_i \leq x | X_i > 0], \;\; x \in \mathbb{R}^+.$$

There are two standard choices for the λ_i's. The first one is $\lambda_i = q_i$ yielding $\mathbb{E}I_i = \mathbb{E}N_i$ for $i = 1, 2, \ldots, n$ and hence $\mathbb{E}I_\bullet = \mathbb{E}N_\bullet$. The mean number of policies producing claims is identical in the individual and the collective models. With this choice, the pure premium for the whole portfolio remains the same, i.e. $\mathbb{E}S_{ind} = \mathbb{E}S_{coll}$ but the variance is inflated, i.e. $\mathrm{Var}[S_{ind}] \leq \mathrm{Var}[S_{coll}]$. The variance being traditionally considered as a risk measure, switching from S_{ind} to S_{coll} appears as a prudent strategy; it can be shown that $S_{ind} \preceq_{cx} S_{coll}$.

The second standard choice is $\lambda_i = -\ln(1-q_i)$ which ensures that $\mathbb{P}[N_i = 0] = \mathbb{P}[I_i = 0] = \mathbb{P}[X_i = 0]$ holds for $i = 1, 2, \ldots, n$. In this case, the no-claim probabilities are equivalent in the collective and individual models for each policy. Similarly, at the portfolio level, $\mathbb{P}[N_\bullet = 0] = \mathbb{P}[I_\bullet = 0]$. It can be shown that now $S_{ind} \preceq_{st} S_{coll}$.

For these two particular cases, let us bound the distances $D_s^{\mathbb{R}^+}(S_{ind}, S_{coll})$ and $d_s^{\mathbb{R}^+}(S_{ind}, S_{coll})$ for various values of s.

Proposition 5.1. *(i) if $\lambda_i = -\ln(1 - q_i)$, $i = 1, 2, \ldots, n$, then*

$$\begin{aligned} D_0^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq 2 \sum_{i=1}^{n} \Big\{ 1 - (1 - q_i)(1 + \lambda_i) \Big\} \\ d_1^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq \sum_{i=1}^{n} \Big\{ 1 - (1 - q_i)(1 + \lambda_i) \Big\} \end{aligned}$$

and for any $s \geq 1$,

$$D_s^{\mathbb{R}^+}(S_{ind}, S_{coll}) = d_{s+1}^{\mathbb{R}^+}(S_{ind}, S_{coll}) = \frac{\mathbb{E}\{S_{coll}\}^s - \mathbb{E}\{S_{ind}\}^s}{s!}.$$

(ii) If $\lambda_i = q_i$ for $i = 1, 2, \ldots, n$, then

$$\begin{aligned}
D_0^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq 2\sum_{i=1}^{n} q_i\Big(1-\exp(-q_i)\Big)\\
d_1^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq \sum_{i=1}^{n} \max\Big\{q_i - 1 + \exp(-q_i), 1-\exp(-q_i)(1+q_i)\Big\}\\
D_1^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq 2\sum_{i=1}^{n} \mathbb{E}V_i\Big(q_i - 1 + \exp(-q_i)\Big)\\
d_2^{\mathbb{R}^+}(S_{ind}, S_{coll}) &\leq \sum_{i=1}^{n} \mathbb{E}V_i\Big(q_i - 1 + \exp(-q_i)\Big)
\end{aligned}$$

and for any $s \geq 2$,

$$D_s^{\mathbb{R}^+}(S_{ind}, S_{coll}) = d_{s+1}^{\mathbb{R}^+}(S_{ind}, S_{coll}) = \frac{\mathbb{E}\{S_{coll}\}^s - \mathbb{E}\{S_{ind}\}^s}{s!}.$$

Proof. (i) starting from

$$D_0^{\mathbb{R}^+}(S_{ind}, S_{coll}) \leq \sum_{i=1}^{n} D_0^{\mathbb{R}^+}(X_i, \widetilde{X}_i) \leq \sum_{i=1}^{n} D_0^{\mathbb{N}}(I_i, N_i)$$

a direct computation gives

$$D_0^{\mathbb{N}}(I_i, N_i) = 2\Big\{1 - (1-q_i)(1+\lambda_i)\Big\}$$

which gives the announced result. The same reasoning yields the inequality for $d_1^{\mathbb{R}^+}(S_{ind}, S_{coll})$. Since $S_{ind} \preceq_{st} S_{coll}$, the extrema defining $d_s^{\mathbb{R}^+}(S_{ind}, S_{coll})$ is attained for $t = 0$ for $s \geq 2$ and the integrand in the definition of $D_s^{\mathbb{R}^+}(S_{ind}, S_{coll})$ has always the same sign for $s \geq 1$, whence the announced result follows.

(ii) The proof follows the same lines as that for (i) and is thus omitted. □

Note that some of the results contained in Proposition 5.1 are known; specifically, the bounds for $d_1^{\mathbb{R}^+}(S_{ind}, S_{coll})$ and $d_2^{\mathbb{R}^+}(S_{ind}, S_{coll})$ can be found in De Pril and Dhaene (1992), where they have been obtained via a direct approach. Nevertheless, the others are new to the best of our knowledge, and the unified approach proposed here to derive these bounds is entirely original. To end with, let us mention that some of the results of Proposition 5.1 are related to those derived recently by Vellaisamy and Chaudhury (1999).

Acknowledgements

The second author gratefully acknowledges the financial support of the FSR from the Université catholique de Louvain, as well as the Belgian Fonds National de la Recherche Scientifique (FNRS).

References

Denuit, M., and Lefèvre, Cl. (1997). Some new classes of stochastic order relations among arithmetic random variables, with applications in actuarial sciences. *Insurance: Mathematics and Economics* **20**, 197-214.

Denuit, M., Lefèvre, Cl., and Shaked, M. (1998). The s-convex orders among real random variables, with applications. *Mathematical Inequalities and Their Applications* **1**, 585-613.

Denuit, M., Lefèvre, Cl., and Utev, S. (1999) Stochastic orderings of convex/concave - type on an arbitrary grid. *Mathematics of Operations Research* **24**, 835-846.

Denuit, M., Lefèvre, Cl., and Utev, S. (2000). Measuring the impact of dependence on the number of claimants in the individual model. Mimeo.

De Pril, N., and Dhaene, J. (1992). Error bounds for compound Poisson approximation of the the individual risk model. *ASTIN Bulletin* **22**, 135-148.

Fishburn, P.C., and Lavalle, I.H. (1995). Stochastic dominance on unidimensional grid. *Mathematics of Operations Research* **20**, 513-525. Erratum in Vol. 21, page 252.

Finkelstein, M., Tucker, H.G., and Veeh, J.A. (1990). The limit distribution of the number of rare mutants. *Journal of Applied Probability* **27**, 239-250.

Kaas, R., and Gerber, H.U. (1994). Some alternatives for the individual model. *Insurance: Mathematics and Economics* **15**, 127-132.

Gerber, H.U. (1981). *An Introduction to Mathematical Risk Theory.* Huebner Foundation Monograph.

Gerber, H.U. (1984). Error bounds for the compound Poisson approximation. *Insurance: Mathematics and Economics* **3**, 191-194.

Lefèvre, Cl., and Utev, S. (1996). Comparing sums of exchangeable Bernoulli random variables, *Journal of Applied Probability* **33**, 285-310.

Lefèvre, Cl., and Utev, S. (1998). On order-preserving properties of probability metrics. *Journal of Theoretical Probability* **11**, 907-920.

Müller, A. (1997). Stochastic orderings generated by integrals : a unified study. *Advances in Applied Probability* **29**, 414-428. 907-920.

Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models.* John Wiley, New York.

Rachev, S.T., and Rüschendorf, L. (1990). Approximation of sums by compound Poisson distributions with respect to stop-loss distances. *Advances in Applied Probability* **22**, 350-374.

Shaked, M., and Shanthikumar, J.G. (1994). *Stochastic Orders and their Applications.* Academic Press. New York.

Vellaisamy, P., and Chaudhuri, B. (1996). Poisson and compound Poisson approximation for random sums of random variables. *Journal of Applied Probability* **33**, 127-137.

Vellaisamy, P., and Chaudhuri, B. (1999). On compound Poisson approximation for sums of random variables. *Statistics and Probability Letters* **41**, 179-189.