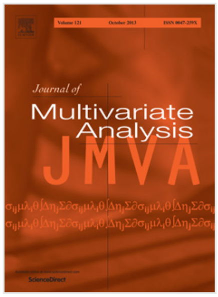
Accepted Manuscript



Extreme negative dependence and risk aggregation

Bin Wang, Ruodu Wang

PII: S0047-259X(15)00011-1
DOI: http://dx.doi.org/10.1016/j.jmva.2015.01.006
Reference: YJMVA 3867

To appear in: *Journal of Multivariate Analysis*

Received date: 5 October 2014

Please cite this article as: B. Wang, R. Wang, Extreme negative dependence and risk aggregation, *Journal of Multivariate Analysis* (2015), http://dx.doi.org/10.1016/j.jmva.2015.01.006

Extreme Negative Dependence and Risk Aggregation

Bin Wang* and Ruodu Wang†

December 30, 2014

Abstract

We introduce the concept of an extremely negatively dependent (END) sequence of random variables with a given common marginal distribution. An END sequence has a partial sum which, subtracted by its mean, does not diverge as the number of random variables goes to infinity. We show that an END sequence always exists for any given marginal distributions with a finite mean and we provide a probabilistic construction. Through such a construction, the partial sum of identically distributed but dependent random variables is controlled by a random variable that depends only on the marginal distribution of the sequence. We provide some properties and examples of our construction. The new concept and derived results are used to obtain asymptotic bounds for risk aggregation with dependence uncertainty.

Key-words: negative dependence; variance reduction; sums of random variables; central limit theorem; risk aggregation.

Mathematics Subject Classification (2010): 60F05, 60E15

1 Introduction

For a given univariate distribution (function) F with finite mean μ, let $X_1, X_2, \dots$ be any sequence of random variables from the distribution F and denote the partial sum $S_n = X_1 + \cdots + X_n$ for $n \in \mathbb{N}$. The distribution of S_n varies under different assumptions of dependency (joint distribution) among the sequence $(X_i, i \in \mathbb{N})$. For example, if we assume that the variance of F is finite (i.e. $(X_i,\ i \in \mathbb{N})$ is square integrable), then it is well-known that

(a) if $X_1, X_2, \dots$ are independent, S_n has a variance of order n, and $(S_n - n\mu)/\sqrt{n}$ converges weakly to a normal distribution (Central Limit Theorem);

*Department of Mathematics, Beijing Technology and Business University, Beijing 100048, China.

†Corresponding author. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L3G1, Canada. (email: wang@uwaterloo.ca).

(b) if $X_1, X_2, \ldots$ are *comonotonic* (when $X_1, X_2, \ldots$ are identically distributed, this means $X_1 = \cdots = X_n$ a.s.), S_n has a variance of order n^2 and S_n/n is always distributed as F.

However, the following question remains: among all possible dependencies, is there one dependency which gives the following **(c1)** or **(c2)**?

(c1) $S_n, n \in \mathbb{N}$ have variance bounded by a constant. Equivalently, S_n has a variance of order $O(1)$ as $n \to \infty$;

(c2) $(S_n - n\mu)/k_n$ converges a.s. for any $k_n \to \infty$ as $n \to \infty$. It is easy to see that this limit has to be zero.

The research on questions of the above type is closely related to the following general question:

(A) for a fixed n, what are the possible distributions of the random variable S_n without knowing the dependence structure of $(X_1, \ldots, X_n)$?

Theoretically, S_n here can be replaced by any functional of $(X_1, \ldots, X_n)$. In this paper we focus on S_n for it is the most typical functional studied in the literature, and it has self-evident interpretations in applied fields. Question **(A)** is a typical question concerning uncertain dependence structures of random vectors. It involves optimization over functional spaces with non-linear constraints. One particular problem related to Question **(A)** is a special case of the multi-dimensional Kantorovich problem (see Ambroso and Gigli, 2013): for a cost function $c : \mathbb{R}^n \to \mathbb{R}$, minimize

$$\int_{\mathbb{R}^n} c(x_1, \ldots, x_n) \mathrm{d}H(x_1, \ldots, x_n) \tag{1.1}$$

over the set of probability measures H on $\mathbb{R}^n$, whose margins are F. Often c is chosen as a function of $x_1 + \cdots + x_n$ or $x_1 \times \cdots \times x_n$; in such cases (1.1) is reduced to a one-dimensional optimization problem over all possible distributions in Question **(A)**. The question is naturally associated with research on copula theory, optimal mass transportation, Monte-Carlo (MC) and Quasi-MC (QMC) simulation, and quantitative risk management. The interested reader is referred to Nelsen (2006) (for copula theory), Villani (2009) (for mass transportation), Glasserman (2006) (for (Q)MC simulation) and McNeil et al. (2005) (for quantitative risk management). Moreover, in Rüschendorf (2013) (Parts I and II), these links as well as recent research developments are extensively discussed with a perspective of financial risk analysis.

As a special case, when $c(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)^2$, (1.1) is equivalent to the minimization of the variance of S_n:

$$\min\{\mathrm{Var}(S_n) : X_i \sim F,\ i = 1, \ldots, n\}, \tag{1.2}$$

where we see a clear connection to **(c1)**-**(c2)**. When $n = 2$, (1.1) for supermodular functions c are well studied already in Tchen (1980). When $n \geqslant 3$, even for the variance problem (1.2), analytical solutions are unknown for general marginal distributions. See Rüschendorf and Uckelmann (2002), Wang and Wang (2011) and Bernard et al. (2014) for recent research on explicit solutions to (1.2) for $n \geqslant 3$ under particular assumptions, and Embrechts et al. (2013) and the references therein for numerical calculations. It is obvious that the questions **(c1)**-**(c2)**, in an asymptotic manner, are directly linked to (1.2).

Questions **(c1)**-**(c2)** are also relevant to the study of *risk aggregation with dependence uncertainty*. We refer the interested reader to Embrechts et al. (2014) for a review on recent developments in this field, with an emphasis on financial risk management. The aggregate position S_n represents the total risk or loss random variable in a given period, where $X_1, \dots, X_n$ are individual risk random variables. Assume we know the marginal distributions of $X_1, \dots, X_n$ but the joint distribution of $(X_1, \dots, X_n)$ is unknown. This assumption is not uncommon in risk management where interdependency modeling relies very heavily on data and computational resources. A regulator or manager may for instance be interested in a particular risk measure ρ of S_n. However, without information on the dependence structure, $\rho(S_n)$ cannot be calculated. It is then important to identify the extreme cases: the largest and smallest possible values of $\rho(S_n)$, and this relates to question **(A)** and, in many cases, to **(c1)**-**(c2)** if n is large. To obtain extreme values of $\rho(S_n)$ for finite n, a strong condition of *complete mixability* is usually imposed in the literature, and explicit values are only available for some specific choices of marginal distributions; see for example Wang and Wang (2011), Wang et al. (2013), and Embrechts et al. (2013). On the other hand, there is limited research on the asymptotic behavior of $\rho(S_n)$ as $n \to \infty$. In this paper, we use the concept of END to derive asymptotic estimates for the popular risk measures VaR and ES of S_n as $n \to \infty$ for any marginal distribution F. As a consequence, our results based on END lead to the asymptotic equivalence between worst-case VaR and ES, shown recently by Puccetti and Rüschendorf (2014) and Puccetti et al. (2013) under different assumptions on F. As an improvement, our result does not require any non-trivial conditions on F, and gives the convergence rate of this asympotic equivalence.

In this paper, we answer questions **(c1)**-**(c2)**. Contrary to the positive dependence in **(b)**, we use the term *extreme negative dependence* (END) for a dependence scenario which gives **(c2)**. We show that there is always an END that yields **(c2)** if F has finite mean, and the same dependency also gives **(c1)** if we further assume that the third moment of F is finite. Within our framework **(c1)** is stronger since it at least requires a finite variance and **(c2)** always has a positive answer, although **(c1)** and **(c2)** are not comparable for a general sequence. Moreover, we show that there exists a dependency among random variables $X_1, X_2, \dots$ such that $|S_n - n\mu|$

is controlled by a single random variable Z, the distribution of which depends on F but not on n.

The rest of the paper is organized as follows. In Section 2, we study the sum of END random variables, and show that the sum is controlled by a random variable with distribution derived from F. As an application of END results, asymptotic bounds for expected convex functions and risk measures of the aggregate risk are studied in Section 3. Some final remarks are put in Section 4. In this paper, we assume that all random variables that we discuss in this paper are defined on a common general atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In such a probability space, we can generate independent random vectors with any distribution.

2 Extreme Negative Dependence

2.1 Main results

Throughout the paper, we denote $S_n = X_1+\cdots+X_n$ where $X_1, \ldots, X_n$ are random variables with distribution F, if not specified otherwise, and we assume that the mean μ of F is finite. We also define the generalized inverse function of any distribution function F by $F^{-1}(t) = \inf\{x : F(x) \geqslant t\}$ for $t \in (0,1]$ and its left endpoint $F^{-1}(0) = \inf\{x : F(x) > 0\}$. U$[0,1]$ represents the standard uniform distribution. For a real number x, we denote by $\underline{x} = x - \lfloor x \rfloor \in [0,1)$ the fractional part of x (and $\lfloor x \rfloor$ is the integer part of x).

First, we give a formal definition of *extreme negative dependence*. Recall the two questions given in the introduction:

(c1) $S_n,\ n \in \mathbb{N}$ have variance bounded by a constant;

(c2) $(S_n - n\mu)/k_n \to 0$ a.s. for any $k_n \to \infty$ as $n \to \infty$.

Definition 2.1. Suppose that $(X_i,\ i \in \mathbb{N})$ is a sequence of random variables with common distribution F. We say that $(X_i,\ i \in \mathbb{N})$ is *extremely negatively dependent* (END), if **(c2)** holds. Moreover, we say that $(X_i,\ i \in \mathbb{N})$ is *strongly extremely negatively dependent* (SEND), if **(c1)-(c2)** hold and

$$\sup_{n\in\mathbb{N}} \mathrm{Var}(S_n) \leqslant \sup_{n\in\mathbb{N}} \mathrm{Var}(Y_1 + \cdots + Y_n)$$

for any sequence of random variables $(Y_i,\ i \in \mathbb{N})$ with common distribution F.

The SEND structure can be treated as the most negative correlation between random variables in a sequence, and hence serves as a potential candidate in many variance minimization problems. Also note that any finite number of random variables in a sequence does not affect the property of END but they do affect the property of SEND.

Remark 2.1. The criterion of minimizing $\sup_{n\in N} \mathrm{Var}(S_n)$ in the definition of an SEND sequence can be replaced by another optimization criterion, such as $\sup_{n\in\mathbb{N}} \mathbb{E}[g(S_n)]$ or $\limsup_{n\to\infty} \mathbb{E}[g(S_n)]$ for a convex function g. The reason why we choose the variance as the criterion is that it gives a comparison with the classic Central Limit Theorem, and also meets the interests of variance reduction in applied fields.

In this section we will show that an END sequence can always be constructed if F has finite mean. More specifically, we will show that there exists a sequence of random variables $X_1, X_2, \dots$ with common distribution F, such that $|S_n - n\mu|$ is controlled by a random variable Z that does not depend on n. It turns out that such a random variable Z has a distribution $\hat{F}$ derived directly from F.

The idea behind out construction is that we try to find a sequence of random variables $X_1, X_2, \dots$ such that each of the member compensates the sum S_n. For each random variable X_i, we consider two possibilities: X_i is "large" and X_i is "small". We design a dependence such that the number of "large" X_i's and the number of "small" X_i's are balanced in a specific way. Moreover, the "large" part and the "small" part are counter-monotonic so that they compensate each other. We first introduce some notation.

Let

$$H(t) = \int_0^t (F^{-1}(s) - \mu)\mathrm{d}s, \quad s \in [0,1],$$

and denote $\nu^- = F(\mu-)$ and $\nu^+ = F(\mu)$. It is obvious that if F does not have a probability mass at μ, $\nu^- = \nu^+$. It is easy to see that the function H is bounded, strictly decreasing on $[0,\nu^-]$, strictly increasing on $[\nu^+,1]$, $H(0) = H(1) = 0$, and the minimum value of $H(t)$ is attained at $c := H(\nu^-) = H(\nu^+) < 0$. Moreover, H is a convex function and hence is almost everywhere (a.e.) differentiable on $[0,1]$. For each $s \in [c,0]$, let

$$A(s) = \inf\{t \in [0,1] : H(t) = s\} \quad \text{and} \quad B(s) = \sup\{t \in [0,1] : H(t) = s\}.$$

It is obvious that $A(s) \in [0,\nu^-]$, $B(s) \in [\nu^+,1]$, $A(c) = \nu^-$, $B(c) = \nu^+$ and $H(A(s)) = H(B(s)) = s$, i.e. A and B are the inverse functions of H on the two intervals $[0,\nu^-]$ and $[\nu^+,1]$, respectively. Note that since H has an a.e. non-zero derivative, A and B are a.e. differentiable on $[c,0]$. Let

$$K(s) = \begin{cases} 1 & s > 0, \\ B(s) - A(s) & c \leqslant s \leqslant 0, \\ 0 & s < c. \end{cases}$$

$K(s)$ is right-continuous, increasing, $K(c-) = 0$, and $K(0) = 1$, hence it is a distribution function on $[c,0]$ with probability mass $\nu^+ - \nu^-$ at c and K is continuous on $(c,0]$. Note that H, A, B, and K all depend on F. Later, we will see that B leads to the "large" values of X_i and A leads to the "small" values of X_i. Moreover, define

$$u(s) = \frac{\mu - F^{-1}(A(s))}{F^{-1}(B(s)) - F^{-1}(A(s))}, \quad s \in (c, 0);$$

and in addition we let $u(c) = 1$. It is easy to see that $u(s) \in [0, 1]$ for $s \in [c, 0)$.

The way that the "large" values of X_i and the "small" values of X_i are balanced is via the distribution function K and the weighting function u, as indicated by Lemma 2.1 below.

Lemma 2.1. *Suppose F is a distribution with mean μ. Let Y be a random variable with distribution K and U be a* U$[0,1]$ *random variable, independent of Y. Let*

$$X = A(Y)\mathrm{I}_{\{U \geqslant u(Y)\}} + B(Y)\mathrm{I}_{\{U < u(Y)\}},$$

Then $F^{-1}(X) \sim F$.

Proof. Note that $A(Y) < \nu^-$ and $B(Y) > \nu^+$ whenever $Y \neq c$. Hence, the possible values of X are divided into three subsets: $\{X \in [0, \nu^-)\} = \{U \geqslant u(Y)\} \cap \{Y \neq c\}$, $\{X = \nu^+\} = \{Y = c\}$ a.s. and $\{X \in (\nu^+, 1]\} = \{U < u(Y)\} \cap \{Y \neq c\}$. For $t \in [0, \nu^-)$,

$$\begin{aligned}
\mathbb{P}(X \leqslant t) &= \mathbb{P}(A(Y) \leqslant t, U \geqslant u(Y), Y \neq c) \\
&= \mathbb{P}(Y \geqslant H(t), U \geqslant u(Y)) \\
&= \int_{H(t)}^{0} (1 - u(y))\mathrm{d}K(y) \\
&= \int_{H(t)}^{0} \frac{F^{-1}(B(y)) - \mu}{F^{-1}(B(y)) - F^{-1}(A(y))}\mathrm{d}(B(y) - A(y)).
\end{aligned}$$

Since B and A are the inverse functions of H, we have a.e.

$$\mathrm{d}B(y) = \frac{1}{H'(B(y))}\mathrm{d}y = \frac{1}{F^{-1}(B(y)) - \mu}\mathrm{d}y,$$

and

$$\mathrm{d}A(y) = \frac{1}{H'(A(y))}\mathrm{d}y = \frac{1}{F^{-1}(A(y)) - \mu}\mathrm{d}y.$$

Thus

$$\begin{aligned}
\mathbb{P}(X \leqslant t) &= \int_{H(t)}^{0} \frac{F^{-1}(B(y)) - \mu}{F^{-1}(B(y)) - F^{-1}(A(y))} \frac{F^{-1}(A(y)) - F^{-1}((B(y))}{(F^{-1}(B(y)) - \mu)(F^{-1}(A(y)) - \mu)}\mathrm{d}y \\
&= \int_{H(t)}^{0} \frac{1}{\mu - F^{-1}(A(y))}\mathrm{d}y \\
&= \int_{t}^{0} \frac{1}{\mu - F^{-1}(s)}(F^{-1}(s) - \mu)\mathrm{d}s \\
&= t.
\end{aligned}$$

Similarly, we can show that $\mathbb{P}(X > t) = 1 - t$ for $t \in (\nu^+, 1]$. Hence, there exists a random variable $U \sim$ U$[0, 1]$ such that $X = U$ when $U \in [0, \nu^-) \cup (\nu^+, 1]$. It is also easy to see that, when $\nu^- \neq \nu^+$ and $U \in [\nu^-, \nu^+]$, we have $X = \nu^+$ and $F^{-1}(X) = F^{-1}(\nu^+) = \mu = F^{-1}(U)$ a.s. In conclusion, $F^{-1}(X) = F^{-1}(U)$ a.s. and thus $F^{-1}(X) \sim F$. $\square$

Now we are ready to present our main result.

Theorem 2.2. *Suppose F is a distribution with mean μ, then there exist $X_i \sim F$, $i \in \mathbb{N}$ and $Z \sim \tilde{F}$, such that for each $n \in \mathbb{N}$,*

$$|S_n - n\mu| \leqslant Z, \tag{2.1}$$

where $\tilde{F}$ is the distribution of $F^{-1}(B(Y)) - F^{-1}(A(Y))$, $Y \sim K$.

Proof. We prove this theorem by construction. Let Y be a random variable with distribution K and U be a $\mathrm{U}[0,1]$ random variable independent of Y. For $k \in \mathbb{N}$, define

$$Y_k = A(Y)\mathrm{I}_{\{\underline{U+ku(Y)} \geqslant u(Y)\}} + B(Y)\mathrm{I}_{\{\underline{U+ku(Y)} < u(Y)\}}, \tag{2.2}$$

(recall that $\underline{x} \in [0,1)$ is the fractional part of a real number x) and

$$X_k = F^{-1}(Y_k). \tag{2.3}$$

It is easy to see that $\underline{U + ku(Y)}$ is $\mathrm{U}[0,1]$ distributed and is independent of Y. Hence, by Lemma 2.1 we know that $X_k \sim F$, $k \in \mathbb{N}$.

An intuition of this construction is as follows. Denote $W_1 = F^{-1}(B(Y))$ and $W_2 = F^{-1}(A(Y))$. As we can see, there are two possibilities for the random variable X_k: it is either W_1 (roughly speaking, representing large values of X_k) or W_2 (representing small values of X_k). Note that $u(Y)W_1 + (1 - u(Y))W_2 = \mu$. By constructing random variables X_k, $k \in \mathbb{N}$ in this specific way, we aim to let W_1 and W_2 compensate each other, leading to an S_n that is close to its mean. In the following we complete the proof.

Denote $C_k = \{\underline{U + ku(Y)} < u(Y)\}$ for $k \in \mathbb{N}$. It is easy to see that

$$\mathrm{I}_{C_k} = \#\{N \in \mathbb{N} : N \in (U + (k-1)u(Y), U + ku(Y)]\}.$$

Thus, for $n \in \mathbb{N}$,

$$\sum_{i=1}^{n} \mathrm{I}_{C_i} = \#\{N \in \mathbb{N} : N \in (U, U + nu(Y)]\} = \#(\mathbb{N} \cap (U, U + nu(Y)]).$$

It follows that

$$\lfloor nu(Y) \rfloor \leqslant \sum_{i=1}^{n} \mathrm{I}_{C_i} \leqslant \lfloor nu(Y) \rfloor + 1.$$

We have that, when $\sum_{i=1}^n \mathrm{I}_{C_i} = \lfloor nu(Y) \rfloor$, or equivalently $U < 1 - \underline{nu(Y)}$,

$$\begin{aligned} S_n &= W_1 \sum_{i=1}^n \mathrm{I}_{C_i} + W_2 \left(n - \sum_{i=1}^n \mathrm{I}_{C_i} \right) \\ &= \lfloor nu(Y) \rfloor W_1 + (n - \lfloor nu(Y) \rfloor) W_2 \\ &= n \left(u(Y) W_1 + (1 - u(Y)) W_2 \right) - \underline{nu(Y)} \left(W_1 - W_2 \right) \\ &= n\mu - \underline{nu(Y)} \left(W_1 - W_2 \right), \end{aligned} \tag{2.4}$$

and when $\sum_{i=1}^n \mathrm{I}_{C_i} = \lfloor nu(Y) \rfloor + 1$, or equivalently $U \geqslant 1 - \underline{nu(Y)}$, that

$$\begin{aligned} S_n &= n\mu - \underline{nu(Y)} \left(W_1 - W_2 \right) + W_1 - W_2 \\ &= n\mu + \left(1 - \underline{nu(Y)} \right) \left(W_1 - W_2 \right). \end{aligned} \tag{2.5}$$

By (2.4)-(2.5), we have

$$S_n - n\mu = \left(W_1 - W_2 \right) \left(\mathrm{I}_{\{U \geqslant 1 - \underline{nu(Y)}\}} - \underline{nu(Y)} \right). \tag{2.6}$$

Thus, we obtain $|S_n - n\mu| \leqslant W_1 - W_2$, and by definition $W_1 - W_2 = F^{-1}(B(Y)) - F^{-1}(A(Y)) \sim \tilde{F}$. □

Remark 2.2. If F does not have a probability mass at μ, X in Lemma 2.1 is U$[0, 1]$ distributed, and Y is a continuous random variable on $[c, 0]$. $\hat{F}$ can be interpreted as the distribution of the large values of X_i (represented by W_1) minus the small values of X_i (represented by W_2), controlling the largest possible departure of S_n from its mean.

Remark 2.3. From the proof of Theorem 2.2, we can see that for $n > m$, $S_n - S_m = \sum_{i=m}^n X_i$ also satisfies $|S_n - S_m - (n - m)\mu| \leqslant Z$. In the above proof, the σ-field of $(X_i,\ i \in \mathbb{N})$ is generated by two independent random variables U and Y.

Using Theorem 2.2, we have the following immediate corollary. It gives general bounds for the sum S_n and the existence of an END sequence.

Corollary 2.3. *Suppose F is a distribution with mean μ.*

(a) There exists an END sequence of F-distributed random variables.

(b) If the support of F is contained in $[a, b]$, $a, b \in \mathbb{R}$, then there exist $X_i \sim F$, $i \in \mathbb{N}$ such that for each $n \in \mathbb{N}$,

$$|S_n - n\mu| \leqslant b - a. \tag{2.7}$$

Remark 2.4. The sequence of probability measures generated by the sequence $S_n - n\mu, n \in \mathbb{N}$ in Corollary 2.3 is *tight* (see for example, Bilingsley (1999), Chapter 1).

One may wonder about the relationship between F and $\tilde{F}$. The following lemma gives a link between the moments of both distribution functions.

Lemma 2.4. *If F has finite k-th moment, $k > 1$, then $\tilde{F}$ has finite $(k-1)$-st moment.*

Proof. Without loss of generality, we assume $\mu = 0$. We use the notation W_1 and W_2 as in the proof of Theorem 2.2. Note that by definition, $W_1 \geqslant 0$, $W_2 \leqslant 0$, and $\mathbb{E}[\min\{W_1, |W_2|\} = 0$ if and only if $W_1 = 0 = W_2$ (the lemma holds trivially in this case). In the following we assume $\mathbb{E}[\min\{W_1, |W_2|\}] > 0$.

By (2.6), setting $n = 1$, we have

$$\begin{aligned}
&\mathbb{E}[|X_1 - \mu|^k] \\
&= \mathbb{E}[|S_1 - \mu|^k] \\
&= \mathbb{E}[(W_1 - W_2)^k |\mathrm{I}_{\{U \geqslant 1-u(Y)\}} - u(Y)|^k] \\
&= \mathbb{E}[(W_1 - W_2)^k \mathbb{E}[|\mathrm{I}_{\{U \geqslant 1-u(Y)\}} - u(Y)|^k | Y]] \\
&= \mathbb{E}\left[(W_1 - W_2)^k u(Y)(1-u(Y)) \left((1-u(Y))^{(k-1)} + u(Y)^{(k-1)}\right)\right] \\
&= \mathbb{E}\left[(W_1 - W_2)^k \frac{-W_2}{W_1 - W_2} \frac{W_1}{W_1 - W_2} \left(\left(\frac{W_1}{W_1 - W_2}\right)^{(k-1)} + \left(\frac{-W_2}{W_1 - W_2}\right)^{(k-1)}\right) \mathrm{I}_{\{W_2 \neq W_1\}}\right] \\
&= \mathbb{E}\left[\frac{-W_2 W_1}{W_1 - W_2}(W_1^{k-1} + (-W_2)^{k-1}) \mathrm{I}_{\{W_2 \neq W_1\}}\right] \\
&\geqslant \mathbb{E}\left[\frac{-W_2 W_1}{W_1 - W_2}(\max\{W_1, |W_2|\})^{k-1} \mathrm{I}_{\{W_1 > 0\}}\right] \\
&\geqslant \mathbb{E}\left[\frac{\max\{W_1, |W_2|\} \min\{W_1, |W_2|\}}{2\max\{W_1, |W_2|\})}(\max\{W_1, |W_2|\})^{k-1} \mathrm{I}_{\{W_1 > 0\}}\right] \\
&= \mathbb{E}\left[\frac{1}{2}\min\{W_1, |W_2|\}(\max\{W_1, |W_2|\})^{k-1} \mathrm{I}_{\{W_1 > 0\}}\right] \\
&= \mathbb{E}\left[\frac{1}{2}\min\{W_1, |W_2|\}(\max\{W_1, |W_2|\})^{k-1}\right]. \qquad (2.8)
\end{aligned}$$

Since $\mathbb{E}[|X_1 - \mu|^k]$ is finite, $\mathbb{E}\left[\min\{W_1, |W_2|\}(\max\{W_1, |W_2|\})^{k-1}\right]$ is finite. Note that W_1 and $|W_2|$ are comonotonic by definition, hence

$$\mathbb{E}\left[\min\{W_1, |W_2|\}(\max\{W_1, |W_2|\})^{k-1}\right] \geqslant \mathbb{E}[\min\{W_1, |W_2|\}]\mathbb{E}[\max\{W_1, |W_2|\})^{k-1}]. \qquad (2.9)$$

Recall that we assume $\mathbb{E}[\min\{W_1, |W_2|\}] > 0$, and hence $\mathbb{E}[(\max\{W_1, |W_2|\})^{k-1}] < \infty$ follows from (2.8)-(2.9). Finally, by definition, $W_1 - W_2$ has distribution $\tilde{F}$, thus $\tilde{F}$ has finite $(k-1)$-st moment. □

Remark 2.5. From (2.8), we can see that if the distribution functions of W_1 and $|W_2|$ are asymptotically equivalent (i.e. $\mathbb{P}(W_1 > x)/\mathbb{P}(|W_2| > x) = O(1)$ and $\mathbb{P}(|W_2| > x)/\mathbb{P}(W_1 > x) = O(1)$ as $x \to \infty$), then the finiteness of the k-th moment of F actually implies the finiteness of the k-th moment of $\tilde{F}$. When one of W_1 and W_2 is bounded but the other one is unbounded, only the finiteness of the $(k-1)$-st moment of $\tilde{F}$ is guaranteed. The relation (2.8) is sharp in the sense that the two inequalities used in (2.8) are sharp inequalities which at most reduce the quantity by three fourths.

Proposition 2.5. *Suppose F is a distribution with mean μ, and F has finite m-th moment, $m > 1$. Then there exist $X_i \sim F$, $i \in \mathbb{N}$ such that uniformly in $n \in \mathbb{N}$, as $k \to \infty$,*

$$\mathbb{P}(|S_n - n\mu| > k) = o(k^{-m+1}). \tag{2.10}$$

In particular, as $n \to \infty$, for all $\varepsilon > 0$,

$$\mathbb{P}(|S_n - n\mu| < n^{\varepsilon}) = 1 - o(n^{-(m-1)\varepsilon}). \tag{2.11}$$

Proof. The finiteness of the $(m-1)$-st moment of $\tilde{F}$ guarantees that $x^{m-1}(1-\tilde{F}(x)) \to 0$ as $x \to \infty$. Hence, by Theorem 2.2, we have that $\mathbb{P}(|S_n - n\mu| > k) \leqslant 1 - \tilde{F}(k) = o(k^{-m+1})$. $\square$

To seek for a possible SEND sequence, we present a link between the variances of F and $\tilde{F}$.

Proposition 2.6. *Suppose $\tilde{F}$ has finite variance. Then there exist $X_i \sim F$, $i \in \mathbb{N}$ and $Z \sim \tilde{F}$ such that for $n \in \mathbb{N}$,*

$$\mathrm{Var}(S_n) \leqslant \frac{1}{4}\mathbb{E}[Z^2].$$

In particular, for such $X_i \sim F$, $i \in \mathbb{N}$, we have that

(a) $\mathrm{Var}(S_n) \leqslant (b-a)^2/4$ *if F is supported on $[a, b]$, $a, b \in \mathbb{R}$;*

(b) $\mathrm{Var}(S_n) \leqslant C$ *for some constant C that does not depend on n if F has finite third moment, and*

(c) the sequence $X_i \sim F$, $i \in \mathbb{N}$ is SEND if $\mathrm{Var}(X_1) = \mathbb{E}[Z^2]/4$.

Proof. We use the notation W_1 and W_2 as in the proof of Lemma 2.4, and let $Z = W_1 - W_2$. By (2.6),

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathbb{E}[(S_n - n\mu)^2] \\
&= \mathbb{E}[(W_1 - W_2)^2(\mathrm{I}_{\{U \geqslant 1 - \underline{nu(Y)}\}} - \underline{nu(Y)})^2] \\
&= \mathbb{E}[(W_1 - W_2)^2\mathbb{E}[(\mathrm{I}_{\{U \geqslant 1 - \underline{nu(Y)}\}} - \underline{nu(Y)})^2|Y]] \\
&= \mathbb{E}[(W_1 - W_2)^2\underline{nu(Y)}(1 - \underline{nu(Y)})] \\
&\leqslant \frac{1}{4}\mathbb{E}[(W_1 - W_2)^2].
\end{aligned}$$

The results follow from this:

(a) This can be seen from the fact that $0 \leqslant Z = W_2 - W_1 \leqslant |b-a|$.

(b) By Lemma 2.4, when F has finite third moment, $Z \sim \tilde{F}$ has finite second moment. Thus, $\mathrm{Var}(S_n) \leqslant \mathbb{E}[Z^2]/4 < C$.

(c) For any sequence $Y_i \sim F,\ i \in \mathbb{N}$,

$$\sup_{n\in\mathbb{N}} \mathrm{Var}(Y_1 + \cdots + Y_n) \geqslant \mathrm{Var}(Y_1) = \frac{1}{4}\mathbb{E}[Z^2] \geqslant \sup_{n\in\mathbb{N}} \mathrm{Var}(S_n).$$

Hence, $X_i \sim F,\ i \in \mathbb{N}$ are SEND.

□

Yet, when $\mathrm{Var}(X_1) < \mathbb{E}[Z^2]/4$, it remains unclear to find an SEND sequence. From the examples in the next section, we would say that the bound $\mathbb{E}[Z^2]/4$ already gives good estimates of the smallest variance of S_n in general.

Remark 2.6. Finding sequences of random variables with small total variance (such as the END sequence) is a classical question in variance reduction and simulation. It is especially important in Monte-Carlo (MC) and Quasi Monte-Carlo (QMC) simulation (for instance, see Glasserman (2006) for (Q)MC methods and their applications in finance), where typically a dependence structure is chosen to generate a random sample such that the error $|S_n/n - \mu|$ is approximately $a/\sqrt{n}$ with a small value of a. QMC techniques, such as low-discrepancy methods, aim for an error of order $O(n^{-(1-\varepsilon)})$, $\varepsilon > 0$, by choosing (usually deterministic) discretization points. In our paper, we give a dependence structure which generates a random sample with an asymptotic error of order $O(1/n)$ which significantly improves the convergence rate. However, any kind of extremal dependence would inevitably lead to almost deterministic dependence structure (for instance comonotoninicty and countermonotonicity). This also applies to the constructive END sequence used in Theorem 2.2. Due to the degeneracy, a direct application of our construction does not lead to a meaningful simulation; see Remark 2.3 and examples in Section 2.2. Regardless, the concept of END is useful for variance minimization with given margins, and there might exist other END sequences suitable for the purpose of QMC simulation. The details of possible new random sample generation techniques, as well as the setup for high-dimensionality, need further research.

We conclude this section by a final remark on the variance of S_n under three representative dependencies. As long as the third moment of F is finite,

- if $X_i,\ i \in \mathbb{N}$ are independent, $\mathrm{Var}(S_n) = O(n)$, and $(S_n - n\mu)/\sqrt{n} \xrightarrow{d}$Normal;
- if $X_i,\ i \in \mathbb{N}$ are comonotonic, $\mathrm{Var}(S_n) = O(n^2)$ and $S_n/n \overset{\text{a.s.}}{=} X_1 \sim F$;
- if $X_i,\ i \in \mathbb{N}$ are END, $\mathrm{Var}(S_n) = O(1)$ and $(S_n - n\mu)/n^{\varepsilon} \xrightarrow{\text{a.s.},\ L_2} 0$ for any $\varepsilon > 0$.

2.2 Examples

In this section we give some examples of END sequences and the corresponding $\tilde{F}$. These examples show that some of the bounds given in Section 2.1 are sharp in the most general sense.

Example 2.1. Suppose F is a Bernoulli distribution on $\{0, 1\}$ with parameter $p \in (0, 1)$:

$$F(x) = (1-p)\mathrm{I}_{\{x \geqslant 0\}} + p\mathrm{I}_{\{x \geqslant 1\}}.$$

Then

$$H(s) = -ps\mathrm{I}_{\{0 \leqslant s \leqslant 1-p\}} - (1-p)(1-s)\mathrm{I}_{\{1-p < s \leqslant 1\}}, \quad s \in [0, 1],$$

$H(s)$ attains its minimum at $H(1-p) = -p(1-p)$, and

$$A(t) = -\frac{t}{p}, \quad B(t) = 1 + \frac{t}{1-p}, \quad t \in [-p(1-p), 0].$$

Therefore, $F^{-1}(B(t)) = 1$, $F^{-1}(A(t)) = 0$ for all $t \in (-p(1-p), 0)$. This leads to $F^{-1}(B(Y)) - F^{-1}(A(Y)) = 1$ a.s. Thus, $\tilde{F}$ is a degenerate distribution at 1, and there exists a sequence of $X_1, X_2, \ldots$ with common distribution F such that for all $n \in \mathbb{N}$,

$$|S_n - np| \leqslant 1. \tag{2.12}$$

One can calculate that $u(t) = p$ for all $t \in (-p(1-p), 0)$. Therefore, an END sequence can be constructed via (2.2) and (2.3) as

$$X_k = \mathrm{I}_{\{\underline{U+kp} < p\}}, \quad k \in \mathbb{N},$$

for some $U \sim \mathrm{U}[0, 1]$. Moreover, we can show that the above sequence is SEND. Let $m_n = \lfloor np \rfloor$. By (2.12), S_n takes value in $\{m_n, m_n + 1\}$ and $\mathbb{E}[S_n] = np$. Therefore,

$$\mathrm{Var}(S_n) = (m_n + 1 - np)(np - m_n)^2 + (np - m_n)(m_n + 1 - np)^2 = (m_n + 1 - np)(np - m_n).$$

Suppose T is a random variable which takes integer values and has mean np. Since $(T - m_n)(T - m_n - 1) \geqslant 0$ a.s., we have that

$$\mathbb{E}[T^2] \geqslant (2m_n + 1)\mathbb{E}[T] - m_n(m_n + 1) = (2m_n + 1)np - m_n(m_n + 1),$$

and

$$\mathrm{Var}(T) \geqslant (2m_n + 1)np - m_n(m_n + 1) - (np)^2 = (m_n + 1 - np)(np - m_n).$$

As a consequence, S_n actually has the minimum variance over all possible partial sums of F-distributed random variables. Thus, $(X_k,\ k \in \mathbb{N})$ is SEND.

Remark 2.7. The bound (2.12) cannot be improved for an irrational p. Suppose p in the above example is an irrational number and let $X_1, X_2, \dots$ be a sequence of random variables from F. It is obvious that S_n is an integer, and np is an irrational number. Since $\mathbb{E}[S_n] = np$, we must have $\mathbb{P}(S_n \leqslant \lfloor np \rfloor) > 0$ and $\mathbb{P}(S_n \geqslant \lfloor np \rfloor + 1) > 0$. Thus, $\mathbb{P}(|S_n - np| \geqslant \underline{np}) > 0$. Since $\{\underline{np} : n \in \mathbb{N}\}$ is dense in $[0, 1]$, we have that for any $q \in [0, 1)$, there are infinitely many $n \in \mathbb{N}$ such that $\mathbb{P}(|S_n - np| \geqslant q) > 0$. Hence, $|S_n - np| \leqslant q$ with $q < 1$ for all $n \in \mathbb{N}$ is impossible. This also confirms that for a distribution on $[a, b]$, the bound (2.7) given in Corollary 2.3 (b) can not be improved in general.

Example 2.2. Suppose F is a uniform distribution on $[0, 1]$. Then

$$H(s) = \frac{1}{2}s(s-1), \quad s \in [0, 1],$$

$H(s)$ attains its minimum at $H(1/2) = -1/8$, and

$$A(t) = \frac{1 - \sqrt{1 - 8t}}{2}, \quad B(t) = \frac{1 + \sqrt{1 - 8t}}{2}, \quad t \in \left[-\frac{1}{8}, 0\right].$$

Y is a continuous random variable with distribution function $B(t) - A(t) = \sqrt{1 - 8t}$, $t \in \left[-\frac{1}{8}, 0\right]$. Therefore, $A(Y)$ is uniformly distributed on $[0, 1/2]$, $B(Y)$ is uniformly distributed on $[1/2, 1]$, and $A(Y) + B(Y) = 1$. It follows that $Z := F^{-1}(B(Y)) - F^{-1}(A(Y)) = B(Y) - A(Y)$ is $\mathrm{U}[0, 1]$ distributed. Therefore, $\tilde{F}$ is $\mathrm{U}[0, 1]$ and and there exists a sequence of $X_1, X_2, \dots$ with common distribution F such that for all $n \in \mathbb{N}$,

$$|S_n - n/2| \leqslant Z.$$

One can calculate that $u(t) = 1/2$ for all $t \in (-1/8, 0)$. Therefore, an END sequence can be constructed via (2.2) and (2.3) as

$$X_k = \frac{V}{2}\mathrm{I}_{\{\underline{U + k/2} \geqslant 1/2\}} + \left(1 - \frac{V}{2}\right)\mathrm{I}_{\{\underline{U + k/2} < 1/2\}}, \quad k \in \mathbb{N},$$

for some independent $\mathrm{U}[0, 1]$ random variables U, V. Equivalently, one may write $X_1 \sim \mathrm{U}[0, 1]$, and

$$X_k = \frac{1}{2} - (-1)^k \left(X_1 - \frac{1}{2}\right), \quad k = 2, 3, \dots.$$

Remark 2.8. In the above example, $\mathrm{Var}(X_1) = 1/12$ and $\mathbb{E}[Z^2] = 1/3$. By Proposition 2.6 (c), the sequence $X_1, X_2, \dots$ is SEND. Indeed, the above construction holds for all symmetric distributions F; one can always construct an SEND sequence by letting $X_1 \sim F$ and

$$X_k = \mathbb{E}[X_1] - (-1)^k (X_1 - \mathbb{E}[X_1]), \quad k = 2, 3, \dots.$$

Example 2.3. Suppose F is a Pareto distribution with index $\alpha = 2$:

$$F(x) = 1 - x^{-2}, \quad x \geqslant 1.$$

Then $F^{-1}(t) = (1-t)^{-1/2}$, $t \in (0,1)$, and $\mu = 2$.

$$H(s) = 2(1-s) - 2\sqrt{1-s}, \quad s \in [0,1],$$

$H(s)$ attains its minimum at $H(3/4) = -1/2$, and

$$A(t) = \frac{1-t-\sqrt{1+2t}}{2}, \quad B(t) = \frac{1-t+\sqrt{1+2t}}{2}, \quad t \in \left[-\frac{1}{2}, 0\right].$$

Note that $B(t) - A(t) = \sqrt{1+2t}$. Y is a continuous random variable with distribution function $B - A$, and hence the inverse distribution function of Y is $(s^2-1)/2$, $s \in (0,1)$. Thus we can write $Y = (V^2-1)/2$ where V is a U$[0,1]$ random variable. Further,

$$\begin{aligned}
A(Y) &= \frac{1-Y-\sqrt{1+2Y}}{2} = \frac{3-2V-V^2}{4}, \\
F^{-1}(A(Y)) &= \left(1 - \frac{3-2V-V^2}{4}\right)^{-1/2} = \frac{2}{1+V}, \\
B(Y) &= \frac{1-Y+\sqrt{1+2Y}}{2} = \frac{3+2V-V^2}{4}, \\
F^{-1}(B(Y)) &= \left(1 - \frac{3+2V-V^2}{4}\right)^{-1/2} = \frac{2}{1-V},
\end{aligned}$$

and

$$Z := F^{-1}(B(Y)) - F^{-1}(A(Y)) = \frac{2}{1-V} - \frac{2}{1+V} = \frac{4V}{1-V^2}.$$

It follows that

$$\tilde{F}(x) = \mathbb{P}(Z \leqslant x) = \sqrt{1 + \frac{4}{x^2}} - \frac{2}{x}, \quad x \geqslant 0,$$

and the tail of $\tilde{F}$ is Pareto-type with index 1. There exists a sequence of $X_1, X_2, \ldots$ with common distribution F such that for all $n \in \mathbb{N}$,

$$|S_n - 2n| \leqslant Z.$$

Since $Z \leqslant 2/(1-V)$, $|S_n - 2n|$ is controlled by another Pareto random variable, $2/(1-V)$, with index 1. Note that

$$u(Y) = \frac{2 - F^{-1}(A(Y))}{Z} = \frac{1-V}{2}.$$

Therefore, an END sequence can be constructed via (2.2) and (2.3) as

$$X_k = \frac{2}{1+V} \mathrm{I}_{\{\underline{U + k\frac{1-V}{2}} \geqslant \frac{1-V}{2}\}} + \frac{2}{1-V} \mathrm{I}_{\{\underline{U + k\frac{1-V}{2}} < \frac{1-V}{2}\}}, \quad k \in \mathbb{N},$$

for some independent U$[0,1]$ random variables U, V.

Remark 2.9. In the above example, F has finite $(2-\varepsilon)$-th moment for all $\varepsilon > 0$, and $\tilde{F}$ has finite $(1-\varepsilon)$-th moment for all $\varepsilon > 0$. This confirms the sharpness of the moment relation in Lemma 2.4 for one-side bounded distributions (see Remark 2.5).

3 Applications in Risk Aggregation

In quantitative risk management, when the marginal distributions of $X_1, \ldots, X_n$ are known but the joint distribution is unknown, regulators and managers are interested in the extreme values for quantities related to an aggregate position $S_n = X_1 + \cdots + X_n$ such as risk measures of S_n. In this section, we apply our main results to the extreme scenarios in risk management with dependence uncertainty.

3.1 Risk aggregation with dependence uncertainty

In the framework of risk aggregation with dependence uncertainty, it is considered that for each $i = 1, \ldots, n$ the distribution of X_i is known while the joint distribution of $\mathbf{X} := (X_1, X_2, \ldots, X_n)$ is unknown. Such setting is practical in quantitative risk management, as statistical modeling for the dependence structure (copula) is extremely difficult especially when n is relatively large. The interested reader is referred to Embrechts et al. (2014) and the references therein for research in this field. When the dependence structure is unknown, an aggregate risk S_n lives in an admissible risk class as defined below.

Definition 3.1. The *admissible risk class* is defined by the set of sums of random variables with given marginal distributions:

$$\mathfrak{S}_n(F_1, \ldots, F_n) \quad = \quad \{X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \ldots, n\}.$$

For simplicity, throughout this section, we denote by $\mathfrak{S}_n = \mathfrak{S}_n(F, \ldots, F)$. It is immediate that the study of $\mathfrak{S}_n$ is equivalent to the study of question **(A)** as mentioned in the introduction.

The following corollary is a straightforward consequence of Theorem 2.2 and Proposition 2.5.

Corollary 3.1. *Suppose F is any distribution.*

(a) If the support of F is contained in $[a, b]$, $a < b$, $a, b \in \mathbb{R}$, then

$$\max_{S \in \mathfrak{S}_n} \mathbb{P}\left(|S - \mathbb{E}[S]| \leqslant b - a\right) = 1.$$

(b) If F has finite m-th moment, $m > 1$, then uniformly in $n \in \mathbb{N}$, as $k \to \infty$,

$$\inf_{S \in \mathfrak{S}_n} \mathbb{P}\left(|S - \mathbb{E}[S]| > k\right) = o(k^{-(m-1)}).$$

In the next two sections, we will look at the extremal questions related to $\mathfrak{S}_n$.

3.2 Bounds on convex functions

Convex order (see for example, Shaked and Shanthikumar (2007), Chapter 1) describes the preference between risks from the perspective of risk-avoiding investors. As a classic result in this field, the convex ordering maximum element in $\mathfrak{S}_n$ is always obtained by the comonotonic scenario; see Dhaene et al. (2002) and Deelstra et al. (2011) for general discussions on comonotonicity and its relevance for finance and insurance. On the other hand, finding the convex ordering minimum element for admissible risks is known to be challenging and only limited results are available; see Bernard et al. (2014). For example, the infimum on $\mathbb{E}[g(S)]$ over $S \in \mathfrak{S}_n$ for a convex function g has been obtained in Wang and Wang (2011) for marginal distributions with a monotone density and Bernard et al. (2014) for distributions satisfying a condition of complete mixability.

Note that for all $S \in \mathfrak{S}_n$, $\mathbb{E}[S]$ is a constant. It is well-known that $\mathbb{E}[g(S)] \geqslant g(\mathbb{E}[S])$ by Jensen's inequality. It is then expected that the infimum on $\mathbb{E}[g(S)]$ over $S \in \mathfrak{S}_n$ is close to the value $g(\mathbb{E}[S])$. If F is n-completely mixable (Wang and Wang, 2011), the infimum is attained for the trivial case $S = \mathbb{E}[S] \in \mathfrak{S}_n$. Unfortunately, complete mixability is in general very difficult to prove, and often it is not possessed by many distributions of practical interest. Hence, we will look at a possible upper bound for $\inf_{S\in\mathfrak{S}_n} \mathbb{E}[g(S)]$ which, along with the natural bound $g(\mathbb{E}[S])$, gives quite a good estimate of $\inf_{S\in\mathfrak{S}_n} \mathbb{E}[g(S)]$.

Theorem 3.2. *Suppose F is a distribution on $[a, b]$, $a < b$, $a, b \in \mathbb{R}$, with mean μ, then for any convex function $g : \mathbb{R} \to \mathbb{R}$,*

$$g(n\mu) \leqslant \inf_{S\in\mathfrak{S}_n} \mathbb{E}[g(S)] \leqslant \frac{1}{2} g(n\mu + (b - a)) + \frac{1}{2} g(n\mu - (b - a)).$$

Proof. The first half of the inequality is due to Jensen's inequality. For the second half, by Corollary 3.1, it suffices to prove that among all distributions on $[n\mu - (b-a), n\mu + (b+a)]$ with mean $n\mu$, the Bernoulli distribution on $\{n\mu - (b - a), n\mu + (b - a)\}$ with equal probability gives the largest possible value of $\mathbb{E}[g(S)]$.

To show this, without loss of generality we assume $\mu = 0$ with $b - a = 1$. Let X be any random variable with mean 0 and support $[-1, 1]$, and let Y be a Bernoulli random variable with $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$. To show that X is smaller than Y in convex order, it suffices to show that for each $K \in [-1, 1]$, $\mathbb{E}[(X - K)_+] \leqslant \mathbb{E}[(Y - K)_+] = (1 - K)/2$.

When $\mathbb{P}(X \geqslant K) \leqslant 1/2$, we have

$$\mathbb{E}[(X - K)_+] = \mathbb{E}[(X - K)\mathrm{I}_{\{X\geqslant K\}}] \leqslant \mathbb{E}[(1 - K)\mathrm{I}_{\{X\geqslant K\}}] \leqslant \frac{1}{2}(1 - K).$$

When $\mathbb{P}(X \geqslant K) > 1/2$, we have

$$\begin{aligned}\mathbb{E}[(X-K)_+] = \mathbb{E}[(K-X)_+] + \mathbb{E}[X-K] &\leqslant \mathbb{E}[(K-X)\mathrm{I}_{\{X<K\}}] - K \\ &\leqslant \mathbb{E}[(K+1)\mathrm{I}_{\{X<K\}}] - K \\ &\leqslant \frac{1}{2}(K+1) - K \\ &= \frac{1}{2}(1-K).\end{aligned}$$

In conclusion, X is smaller than Y in convex order. Thus, $\mathbb{E}[g(X)] \leqslant \mathbb{E}[g(Y)]$ for any convex function $g : \mathbb{R} \to \mathbb{R}$. □

Remark 3.1. As a special choice of $\mathbb{E}[g(S)]$, the variance of a sequence of identically distributed random variables is of particular importance; see Section 2. The variance bound given in Proposition 2.6 (a) is stronger than the bound in Theorem 3.2 which naturally gives a bound of $(b-a)^2$ if $g(x)$ is taken as $(x-n\mu)^2$. Other quantities of the type $\mathbb{E}[g(S_n)]$, used in finance and insurance, include stop-loss premiums, European option prices, expected utilities and expected n-period returns.

3.3 Bounds on the Expected Shortfall and the Value-at-Risk

Another important class of quantities to discuss is the class of risk measures. In order to deteremine capital requirements for financial regulation, various risk measures are used in practice. Since the introduction of coherent risk measures by Artzner et al. (1999), there has been extensive research on coherent as well as non-coherent risk measures; see McNeil et al. (2005). Two commonly used capital requirement principles are the Value-at-Risk, defined as

$$\mathrm{VaR}_p(X) = \inf\{x : \mathbb{P}(X \leqslant x) \geqslant p\}, \quad p \in (0,1). \tag{3.1}$$

and the Expected Shortfall (ES), also known as the Tail Value-at-Risk (TVaR), defined as

$$\mathrm{ES}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(S)\mathrm{d}\alpha, \quad p \in [0,1). \tag{3.2}$$

In the case of risk aggregation with dependence uncertainty, finding bounds for VaR and ES becomes an important task (see for example Embrechts et al. (2013)). We will discuss the VaR case in the next section, and focus on ES for the moment. By the subadditivity of ES, the upper sharp bound $\sup_{S\in\mathfrak{S}_n}\mathrm{ES}_p(S)$ for any $p\in[0,1)$ is obtained with the comonotonic scenario, with $\sup_{S\in\mathfrak{S}_n}\mathrm{ES}_p(S) = n\mathrm{ES}_p(X)$ for $X \sim F$. On the other hand, finding the explicit minimal ES for general marginal distributions is an open question. Since the risk measure ES preserves the convex order, we have the following proposition for the smallest possible ES.

Proposition 3.3. *(a) Suppose F is a distribution on $[a,b]$, $a<b$, $a,b\in\mathbb{R}$ with mean μ, then for $p\in(0,1)$,*

$$n\mu \leqslant \inf_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S) \leqslant n\mu + (b-a).$$

(b) Suppose F is a distribution with mean μ and finite second moment, then for $p\in(0,1)$,

$$n\mu \leqslant \inf_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S) \leqslant n\mu + K,$$

for some constant K that does not depend on n but possibly depends on p.

Proof. Note that $\inf_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S) \geqslant \inf_{S\in\mathfrak{S}_n} \mathbb{E}[S] = n\mu$. The other half of part (a) comes directly from Corollary 3.1. For part (b), by Theorem 2.2, we have that

$$\inf_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S) = \inf_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S_n - n\mu) + n\mu \leqslant \mathrm{ES}_p(Z) + n\mu,$$

where $Z\sim\tilde{F}$. Since F has finite second moment, Z has finite mean, and therefore $\mathrm{ES}_p(Z)$ is finite and does not depend on n. This completes the proof. □

Remark 3.2. Proposition 3.3 gives estimates for the smallest possible $\mathrm{ES}_p(S)$ with dependence uncertainty. When n is large and $\mu\neq 0$, the estimation errors are small compared to the major term $n\mu$. Similar arguments will give asymptotic estimates for any convex risk measures.

The popular quantile-based risk measure VaR is not a convex or coherent risk measure, hence a separate discussion is necessary. Both the maximum and the minimum of VaR with dependence uncertainty are in general unavailable analytically. For a general discussion on the bounds on VaR aggregation and numerical approximations, see Embrechts et al. (2014).

Recall that $F^{-1}(p) = \inf\{x : F(x)\geqslant p\}$ for $p\in(0,1]$, hence $\mathrm{VaR}_p(X) = F^{-1}(p)$ for $p\in(0,1)$ where $X\sim F$. For $1\geqslant q > p \geqslant 0$, let

$$\mu_{p,q} = \frac{1}{q-p}\int_p^q F^{-1}(t)\mathrm{d}t.$$

If F is continuous, $\mu_{p,q}$ is the mean of the conditional distribution of F on $[F^{-1}(p), F^{-1}(q)]$. Note that $\mu_{0,q}$ and $\mu_{p,1}$ might be infinite.

Theorem 3.4. *We have for $p\in(0,1)$ and any distribution F,*

$$n\mu_{p,q} - (F^{-1}(q) - F^{-1}(p)) \leqslant \sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \leqslant n\mu_{p,1}, \tag{3.3}$$

for any $q\in(p,1]$, and

$$n\mu_{0,p} \leqslant \inf_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \leqslant n\mu_{q,p} + (F^{-1}(p) - F^{-1}(q)) \tag{3.4}$$

for any $q\in[0,p)$.

In particular, if F is a distribution on $[a,b]$, $a,b\in\mathbb{R}$, then for $p\in(0,1)$,

$$n\mu_{p,1}-(b-F^{-1}(p))\leqslant \sup_{S\in\mathfrak{S}_n}\mathrm{VaR}_p(S)\leqslant n\mu_{p,1},$$

and

$$n\mu_{0,p}\leqslant \inf_{S\in\mathfrak{S}_n}\mathrm{VaR}_p(S)\leqslant n\mu_{0,p}+(F^{-1}(p)-a).$$

Proof. We will use the following equivalence lemma. A proof can be found in the Appendix of Embrechts et al. (2015), where the alternative definition of VaR will be used:

$$\mathrm{VaR}_p^*(X)=\inf\{x\in\mathbb{R}:\mathbb{P}(X\leqslant x)>p\}.$$

Lemma 3.5 (Lemma A.4 of Embrechts et al. (2015))**.** *For $p\in(0,1)$ and any F,*

$$\sup_{S\in\mathfrak{S}_n}\mathrm{VaR}_p^*(S)=\sup\{\mathrm{essinf}S:S\in\mathfrak{S}_n(F_p,\dots,F_p)\},$$

and

$$\inf_{S\in\mathfrak{S}_n}\mathrm{VaR}_p(S)=\inf\{\mathrm{esssup}S:S\in\mathfrak{S}_n(F^p,\dots,F^p)\},$$

where F_p is the distribution of $F^{-1}(p+(1-p)U)$, and F^p is the distribution of $F^{-1}(pU)$, $i=1,\cdots,n$, for a random variable U uniformly distributed on $[0,1]$.

Note the asymmetry between the supremum and infimum. We first show that

$$\sup_{S\in\mathfrak{S}_n}\mathrm{VaR}_p^*(S)=\sup\{\mathrm{essinf}S:S\in\mathfrak{S}_n(F_p,\dots,F_p)\}\geqslant n\mu_{p,q}-(F^{-1}(q)-F^{-1}(p))\qquad(3.5)$$

for $0<p<q\leqslant 1$. The case when $F^{-1}(q)=\infty$ is trivial, hence we only consider the case when $F^{-1}(q)<\infty$.

Let $F_{p,q}$ be the distribution of $F^{-1}(V)$ where $V\sim\mathrm{U}[p,q]$ for $0<p<q\leqslant 1$. By Corollary 2.3, there exist random variables $X_1,\dots,X_n$ from $F_{p,q}$ such that $X_1+\cdots+X_n\geqslant m\mu_{p,q}-(F^{-1}(q)-F^{-1}(p))$. Let Z be any random variable with distribution F_q and let C be a random event independent of $X_1,\dots,X_n,Z$, with $\mathbb{P}(C)=(q-p)/(1-p)$. Define $Y_i=X_i\mathrm{I}_C+Z(1-\mathrm{I}_C)$ for $i=1,\dots,n$. It is straightforward to check that Y_i has distribution F_p, and

$$Y_1+\cdots+Y_n\geqslant X_1+\cdots+X_n\geqslant n\mu_{p,q}-(F^{-1}(q)-F^{-1}(p)).$$

Thus, $\mathrm{essinf}(Y_1+\cdots+Y_n)\geqslant n\mu_{p,q}-(F^{-1}(q)-F^{-1}(p))$, and we obtain (3.5). Since $\mathrm{VaR}_p(X)\geqslant\mathrm{VaR}_r^*(X)$ for any $r<p$ and random variable X, we have that

$$\begin{aligned}\sup_{S\in\mathfrak{S}_n}\mathrm{VaR}_p(S_n)\geqslant\lim_{r\to p-}\sup_{S\in\mathfrak{S}_n}\mathrm{VaR}_r^*(S_n)&\geqslant\lim_{r\to p-}(n\mu_{r,q}-(F^{-1}(q)-F^{-1}(r)))\\&=n\mu_{p,q}-(F^{-1}(q)-F^{-1}(p)).\end{aligned}$$

Note that here we use the fact that F^{-1} is left-continuous. On the other hand,

$$\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \leqslant \sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p^*(S) = \sup\{\mathrm{essinf} S : S \in \mathfrak{S}_n(F_p,\dots,F_p)\} \leqslant n\mu_{p,1}$$

always holds trivially. Thus we obtian (3.3). Similarly, we can show (3.4). □

Corollary 3.6. *Suppose F has finite k-th moment, $k \geqslant 1$. Then*

$$\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) = n\mu_{p,1} - o(n^{1/k}),$$

and

$$\inf_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) = n\mu_{0,p} + o(n^{1/k}).$$

Proof. Without loss of generality we assume $F^{-1}(p) \geqslant 0$ (otherwise this assumption can easily be satisfied with a shift of location). Choose $q_n = F(an^{1/k})$ in (3.3) for any constant $a > 0$ and large n such that $q_n > p$. We have that $F^{-1}(q_n) \leqslant an^{1/k}$, and

$$\begin{aligned}\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \geqslant n\mu_{p,q_n} - (F^{-1}(q_n) - F^{-1}(p)) \geqslant n\mu_{p,q_n} - an^{1/k}\\ = n\mu_{p,1} - n(\mu_{p,1} - \mu_{p,q_n}) - an^{1/k}.\end{aligned} \tag{3.6}$$

Note that for $X \sim F$,

$$\mu_{p,1}-\mu_{p,q_n} = \frac{1}{1-p}\mathbb{E}[X\mathrm{I}_{\{X\geqslant F^{-1}(p)\}}]-\frac{1}{q_n-p}\mathbb{E}[X\mathrm{I}_{\{F^{-1}(q_n)\geqslant X\geqslant F^{-1}(p)\}}] \leqslant \frac{1}{1-p}\mathbb{E}[X\mathrm{I}_{\{X\geqslant F^{-1}(q_n)\}}].$$

Since F has finite k-th moment, we have that

$$(an^{1/k})^{(k-1)}\mathbb{E}[X\mathrm{I}_{\{X\geqslant an^{1/k}\}}] \leqslant \mathbb{E}[|X|^k\mathrm{I}_{\{X\geqslant an^{1/k}\}}] \to 0,$$

and hence $\mathbb{E}[X\mathrm{I}_{\{X\geqslant an^{1/k}\}}] = o(n^{-1+1/k})$. Thus, $\frac{1}{1-p}\mathbb{E}[X\mathrm{I}_{\{X\geqslant F^{-1}(q_n)\}}] = o(n^{-1+1/k})$, which, together with (3.6), leads to

$$\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \geqslant n\mu_{p,1} - o(n^{1/k}) - an^{1/k}.$$

Since $a > 0$ is arbitrary, and $\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) \leqslant n\mu_{p,1}$, we have that

$$\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S) = n\mu_{p,1} - o(n^{1/k}).$$

The other half of the corollary is obtained similarly. □

Remark 3.3. Theorem 3.4 and Corollary 3.6 provide quite good estimates for the worst-case (best-case) VaR under dependence uncertainty. The estimation becomes accurate when n is large, as $n\mu_{p,1}$ (or $n\mu_{0,p}$) is large compared to the estimation error which is controlled within a rate of $n^{1/k}$, except for the trivial case when $\mu_{p,1} = 0$ (or $\mu_{0,p} = 0$).

The next two corollaries give the asymptotic limit of the superadditive ratio (see Embrechts et al. (2013)) for VaR and the asymptotic equivalence between worst-case VaR and worst-case ES (see Puccetti et al. (2013)).

Corollary 3.7. *For any distribution F, as $n \to \infty$,*

$$\frac{\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S)}{n} \to \mathrm{ES}_p(X)$$

where $X \sim F$.

Proof. We take $q_n = F(\sqrt{n})$ for large n such that $q_n > p$. It follows from (3.3) that

$$\mu_{p,q_n} - o\left(\frac{1}{\sqrt{n}}\right) \leqslant \frac{\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S)}{n} \leqslant \mu_{p,1}.$$

Obviously $q_n \to 1$, and hence $\mu_{p,q_n} \to \mu_{p,1}$. This completes the proof. In fact, if $\mu_{p,1} < \infty$, then F has finite mean and Corollary 3.7 follows directly from Corollary 3.6 by taking $k = 1$. □

Remark 3.4. The fraction $\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S)/(n\mathrm{VaR}_p(X))$ for $\mathrm{VaR}_p(X) > 0$ is called the (worst) superadditive ratio of VaR (see Embrechts et al. (2013)). It measures the amount of possible extra capital requirement needed in a diversification strategy, and hence this quantity is of independent interest in quantitative risk management. Corollary 3.7 gives the limit as $\mathrm{ES}_p(X)/\mathrm{VaR}_p(X)$ without assuming any condition on F. Note that here $\mathrm{ES}_p(X)$ can be infinite. Hence, whenever $\mathrm{ES}_p(X) = \infty$, the superadditive ratio of VaR becomes infinity. This fact clearly shows that the "diversification benefits" commonly used in practical risk management needs to be taken with care.

Corollary 3.8. *Suppose F has finite k-th moment, $k \geqslant 1$ and non-zero* ES *at level $p \in (0,1)$, then as $n \to \infty$,*

$$\frac{\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S)}{\sup_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S)} = 1 - o(n^{-1+1/k}). \tag{3.7}$$

Proof. Note that $\sup_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S) = n\mathrm{ES}_p(X) = n\mu_{p,1} \neq 0$ for $X \sim F$. Thus, the proof follows directly from Corollary 3.6. □

Remark 3.5. Corollary 3.8 implies that under the worst-case scenario of dependence, the VaR and ES risk measures are asymptotically equivalent; that is,

$$\frac{\sup_{S\in\mathfrak{S}_n} \mathrm{VaR}_p(S)}{\sup_{S\in\mathfrak{S}_n} \mathrm{ES}_p(S)} \to 1 \tag{3.8}$$

This coincides with the main results in Puccetti and Rüschendorf (2014) and Puccetti et al. (2013). Puccetti and Rüschendorf (2014) obtained (3.8) under a condition of complete mixability, which at this moment is only known to be satisfied by tail-monotone densities. Puccetti et al. (2013) gave (3.8) under a weaker condition that F has a strictly positive density and discussed some possible inhomogeneous cases. Both of the above papers assumed the continuity of F. To ensure (3.8), Corollary 3.8 only assumes that $\mathrm{ES}_p(X_1)$ is finite and non-zero, which is necessary. Hence, Corollary 3.8 establishes the weakest mathematical assumption for (3.8) to hold. In addition, Corollary 3.8 also gives the convergence rate of this asymptotic equivalence. We can see that the convergence in (3.7) is fast for the distribution F being light-tailed, and slow for F being heavy-tailed. This gives a theoretical justification of the discussion on the numerical illustrations in Section 5 of Puccetti et al. (2013), where it was observed that heavy-tailed marginal distributions in general lead to a slower convergence of (3.8), compared to the cases of light-tailed marginal distributions.

4 Final Remarks

In this paper, we introduce the notions of extreme negative dependence (END) and strong extreme negative dependence (SEND) scenarios, and showed that for each marginal distribution F with finite mean, a construction of an END sequence is always possible. With a finite third moment of F, an SEND sequence may also be obtained by the same construction. The sum of END random variables is in general concentrated around its expectation, and the difference $|S_n - n\mu|$ is controlled by a random variable that does not depend on n. We also studied asymptotic bounds for risk aggregation with dependence uncertainty and provided estimates for the worst-case and best-case risk measures VaR and ES.

There are multiple ways, other than the one in Theorem 2.2, to construct an END sequence. One possible way is through the concept of complete mixablity (Wang and Wang, 2011). However, such construction only applies to completely mixable distributions which are generally restrictive; see Puccetti et al. (2013) for existing results on complete mixability.

The concepts of END and SEND can naturally be generalized to the case of an inhomogeneous (non-identically distributed) sequence of random variables, leading to a potential direction of future research. Generalization of END and SEND in a multi-dimensional setting is also a future research direction with potential relation to QMC simulation.

Acknowledgement

The authors would like to thank Paul Embrechts (ETH Zurich), Thomas Mikosch (U Copenhagen), Giovanni Puccetti (U Firenze), Tiantian Mao (U Waterloo), Vygantas Paulauskas (Vilnius U) and an anonymous referee for helpful comments which have essentially improved the paper. R. Wang acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC Grant No. 435844) and the Forschungsinstitut für Mathematik

(FIM) at ETH Zurich during his visits in Zurich in 2013 and 2014.

References

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Ambroso, L. and Gigli, N. (2013). A users guide to optimal transport. In *Modelling and Optimisation of Flows on Networks*, pages 1–155, Springer, New York.

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

Billingsley, P. (1999). *Convergence of Probability Measures*, Second Edition, Wiley.

Deelstra, G., Dhaene J. and Vanmaele, M. (2011). An overview of comonotonicity and its applications in finance and insurance, in: *Advanced Mathematical Methods for Finance,* Di Nunno, G. and Øksendal B. (Eds.), Springer, 2011.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vynche, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics* **31**(1), 3-33.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750-2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, forthcoming.

Glasserman, P. (2006) *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools.* Princeton University Press, New Jersey.

Nelsen, R. (2006). *An Introduction to Copulas.* Second Edition. Springer, New York.

Puccetti, G. and Rüschendorf, L. (2014). Asymptotic equivalence of conservative value-at-risk- and expected shortfall-based capital charges. *Journal of Risk*, 16(3), 1–19.

Puccetti, G., Wang, B. and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, **53**(3), 821-828.

Rüschendorf, L. (2013). *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios.* Springer.

Rüschendorf, L. and Uckelmann, L. (2002). Variance minimization and random variables with constant sum, in: *Distributions with Given Marginals.* Cuadras, et al. (Eds.), Kluwer, 211–222.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders.* Springer Series in Statistics.

Tchen, A. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, **8**(4): 814–827.

Villani, C. (2009). *Optimal Transport: Old and New.* Springer, New York.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems for monotone marginal distributions. *Journal of Multivariate Analysis*, **102**, 1344–1360.