Journal of Multivariate Analysis 136 (2015) 12–25

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Extreme negative dependence and risk aggregation

Bin Wang[a], Ruodu Wang[b,*]

[a] Department of Mathematics, Beijing Technology and Business University, Beijing 100048, China

[b] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L3G1, Canada

ARTICLE INFO

Article history:
Received 5 October 2014
Available online 19 January 2015

AMS 2010 subject classifications:
60F05
60E15

Keywords:
Negative dependence
Variance reduction
Sums of random variables
Central limit theorem
Risk aggregation

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.

1. Introduction

For a given univariate distribution (function) F with finite mean μ, let $X_1, X_2, \ldots$ 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 $X_1, X_2, \ldots$ 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);
(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 $(X_1, \ldots, X_n)$?

* Corresponding author.
E-mail address: wang@uwaterloo.ca (R. Wang).

http://dx.doi.org/10.1016/j.jmva.2015.01.006
0047-259X/

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 [1]): 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 [12] (for copula theory), [19] (for mass transportation), [10] (for (Q)MC simulation) and [11] (for quantitative risk management). Moreover, in [15] (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) in the case of a supermodular cost function c is well studied already in [18]. When $n \geqslant 3$, even for the variance problem (1.2), analytical solutions are unknown for general marginal distributions. See [16,21,3] for recent research on explicit solutions to (1.2) for $n \geqslant 3$ under particular assumptions, and [7] 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 [8] 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 $X_1, \ldots, X_n$ are individual risk random variables. Assume we know the marginal distributions of $X_1, \ldots, X_n$ but the joint distribution of $(X_1, \ldots, X_n)$ 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 [21,20,7]. 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 [13,14] 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 asymptotic 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 $X_1, X_2, \ldots$ 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).