The Complete Mixability and Convex Minimization Problems with Monotone Marginal Densities

Bin Wang* and Ruodu Wang†

May 2011‡

Abstract

Following the results in Rüschendorf and Uckelmann (2002), we introduce the completely mixable distributions on $\mathbb{R}$ and prove that distributions with monotone density and moderate mean are completely mixable. Using this method we solve the minimization problem $\min_{X_i \sim P} \mathbb{E} f(X_1 + \cdots + X_n)$ for convex functions f and marginal distributions P with monotone density. Our results also provide valuable implications in variance minimization, bounds for the sum of random variables and risk theory.

Key-words: complete mixability; variance minimization; multivariate dependence; monotone densities; optimal coupling.

*Department of Mathematics, Peking University, Beijing 100871, China.

†Corresponding author. School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA. Email: ruodu.wang@math.gatech.edu

‡Several minor corrections are made in 2015, 2016, 2019, thanks to comments from the readers.

1 Introduction

A classic problem in simulation and variance reduction is to minimize the variance of the sum of random variables $X_1, \cdots, X_n$ with given marginal distributions P, i.e.

$$\min_{X_i \sim P} \mathrm{Var}(X_1 + \cdots + X_n). \tag{1}$$

See Fishman [9] and Hammersley and Handscomb [11] for references about this problem. For $n = 2$ the solution is given by the antithetic variates $X_1 = F^{-}(U)$ and $X_2 = F^{-}(1-U)$ where F^{-} is the inverse cdf of P and U is uniform on [0,1]. For $n \geq 3$ the problem is generally difficult to solve. In Gaffke and Rüschendorf [10] and Rüschendorf and Uckelmann [21], their idea is to concentrate $\sum_{i=1}^n X_i$ at the expectation as much as possible. Since it is obvious $\sum_{i=1}^n X_i = c$ is an optimal solution to (1) if such constant c is possible. It raises a question: for which P, do there exist $X_i \sim P$ such that $\sum_{i=1}^n X_i$ is a constant?

In this paper, we call a marginal distribution P of random variables with a constant sum a *complete mixable* distribution. This property was studied by Gaffke and Rüschendorf [10] in the case of uniform distributions. The case of distributions with symmetric and unimodal density was studied for $n = 3$ by Knott and Smith [14], [15] and for the general case $n \geq 2$ by Rüschendorf and Uckelmann [21] using a different method. The property was also extended to multivariate distributions by Rüschendorf and Uckelmann [21]. In summary, they provided that the uniform distributions and distributions with symmetric and unimodal density are completely mixable. In this paper, we define the complete mixability with a focus on the marginal distribution, provide some nice properties of the mixability, and prove that distributions with monotone density and moderate mean are also completely mixable.

Another main contribution of this paper is that by using the complete mixability we solve the convex minimization problem

$$\min_{X_1,\cdots,X_n\sim P} \mathbb{E}f\left(X_1+\cdots+X_n\right) \tag{2}$$

in general where f is a convex function and P is a monotone distribution, i.e. a distribution with monotone density on its support. There are many special cases of (2), such as the variance minimization problem (1) and the minimum of expected product $\min_{X_i\sim \mathrm{U}[0,1]}\mathbb{E}(X_1\cdots X_n)$. Problem (2) is a lower bound problem of the Fréchet class $\mathcal{F}(P,\cdots,P)$ and it is related to various topics in statistics, risk theory, copulas and stochastic orders. We refer to Embrechts et al. [4], [5] for problems of bounds in risk theory, Nelsen [16] for copulas, Joe [12] for Fréchet classes and Shaked and Shanthikumar [22] for stochastic orders.

The rest of the paper is organized as follows. In Section 2 we introduce the completely mixability and prove our main theorem. In Section 3 we use the results in Section 2 to solve the minimization problem (2) for monotone distributions P. Applications of our main results are provided in Section 4. Some open problems are presented in Section 5. In Section 6 we draw our conclusions. Details and some of the proofs are given in Appendix.

Throughout this paper, we denote $\mathbb{U}$ the uniform distribution on [0,1]. In the notation $\mathbb{E}^P(f(X))$ P is the distribution of X, and in the notation $\mathbb{E}^Q(f(X_1,\cdots,X_n))$ Q is the joint distribution of $X_1,\cdots,X_n$.

2 The complete mixability

2.1 Definition and basic properties

Rüschendorf and Uckelmann [21] investigated *random variables with constant sums* and associated it with variance minimization problem (1). In this article, we call the marginal distribution of random variables with a constant sum a *completely mixable distribution*, as in the following definition.

Definition 2.1. Suppose n is a positive integer. A probability distribution (probability measure) P on $\mathbb{R}$ is **completely mixable with index** n if there exist n random variables $X_1, \cdots, X_n \sim P$ such that $X_1 + \cdots + X_n$ is a constant. The distribution of $(X_1, \cdots, X_n)$ is called an n**-complete mix**.

Proposition 2.1. ***(Basic properties of the complete mixability.)***

(1) (Invariance under affine transformations) Suppose the distribution of X is completely mixable with index n, then for any constants a and b, the distribution of $aX + b$ is completely mixable with index n.

(2) (Center of the complete mixability) Suppose P is completely mixable with index n, $X_i \sim P$, $i = 1, \cdots, n$ and $\mu = \frac{1}{n}(X_1 + \cdots + X_n)$ is a constant. We say μ is a center of P and P is centered at μ. If P follows the weak law of large numbers (WLLN), then μ is unique. If $\mathbb{E}^P(X)$ exists, then $\mu = \mathbb{E}^P(X)$.

(3) (Additivity 1: distribution-wise) Suppose P and Q are completely mixable with index n and centered at the same point μ. Then for any $\lambda \in [0, 1]$, $\lambda P + (1 - \lambda)Q$ is completely mixable with index n and centered at μ.

(4) (Additivity 2: index-wise) Suppose P is completely mixable with index n and Q is completely mixable with index k, then $\frac{n}{n+k}P + \frac{k}{n+k}Q$ is completely mixable with index $n + k$. As a consequence, if P is completely mixable with index n and with index k, then P is also completely mixable with index $n + k$, dn and dk for any positive integer d.

(5) (Additivity 3: random-variable-wise) Suppose P and Q are completely mixable with index n, $X \sim P$ and $Y \sim Q$ are independent. Then the distribution of $X + Y$ is

completely mixable with index n.

(6) *(Completeness) Suppose distributions P and P_i are supported in a compact set $S \subset \mathbb{R}$, P_i are completely mixable with index n, $i = 1, 2, \cdots$ and $P_k \xrightarrow{d} P$ as $k \to \infty$. Then P is completely mixable with index n.*

(7) *(A necessary condition) Suppose the distribution P is completely mixable with index n, centered at μ and $X \sim P$. Let $a = \sup\{x : \mathbb{P}(X \leq x) = 0\}$ and $b = \sup\{x : \mathbb{P}(X \leq x) < 1\}$. If one of a and b is finite, then the other one is finite, and $a + \frac{b-a}{n} \leq \mu \leq b - \frac{b-a}{n}$.*

Proof.

(1) This follows immediately from the definition.

(2) Assuming $\mathbb{E}(X_1)$ exists, taking expectation on both sides of $\mu = \frac{1}{n}(X_1 + \cdots + X_n)$ gives us $\mu = \mathbb{E}(X_1)$. Now suppose P follows WLLN. We can take independent copies of $(X_1, \cdots, X_n)$, denoted by $\{(X_{1,i}, \cdots, X_{n,i})\}_{i=1}^{\infty}$, and take their average

$$\begin{aligned} n\mu &= \frac{1}{k}\sum_{i=1}^{k}(X_{1,i} + \cdots + X_{n,i}) \\ &= \frac{1}{k}\sum_{i=1}^{k} X_{1,i} + \cdots + \frac{1}{i}\sum_{i=1}^{k} X_{n,i} \\ &= n\mathbb{E}(X_1 \mathbf{1}_{\{|X_1| \leq k\}}) + o_p(1) \end{aligned}$$

as k goes to infinity. Therefore $\mathbb{E}(X_1 \mathbf{1}_{\{|X_1| \leq k\}}) \to \mu$ and μ is unique.

(3) Suppose $X_1 + \cdots + X_n = n\mu$, $X_i \sim P$ and $Y_1 + \cdots + Y_n = n\mu$, $Y_i \sim Q$, $i = 1, \cdots, n$. Let Z be a Bernoulli(λ) random variable independent of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$. Set $Z_i = \mathbf{1}_{\{Z=1\}} X_i + \mathbf{1}_{\{Z=0\}} Y_i$, then $Z_1 + \cdots + Z_n = n\mu$ and $Z_i \sim \lambda P + (1-\lambda)Q$, $i = 1, \cdots, n$.

(4) Suppose $X_1 + \cdots + X_n = n\mu$, $X_i \sim P$, $i = 1, \cdots, n$ and $Y_1 + \cdots + Y_k = k\nu$, $Y_j \sim Q$, $j = 1, \cdots, k$. Let σ be a random permutation uniformly distributed on the set of

all $(n+k)$-permutations and independent of $X_1, \cdots, X_n, Y_1, \cdots, Y_k$. Denote

$$(Z_1, \cdots, Z_{n+k}) = \sigma(X_1, \cdots, X_n, Y_1, \cdots, Y_k),$$

then $Z_1 + \cdots + Z_{n+k} = n\mu + k\nu$ and $Z_i \sim \frac{n}{n+k}P + \frac{k}{n+k}Q$, $i = 1, \cdots, n+k$.

(5) Let $X_i \sim P$, $Y_i \sim Q$, $i = 1, \cdots, n$ such that $X_1 + \cdots + X_n$ and $Y_1 + \cdots + Y_n$ are constants. Denote $\mathbf{X} = (X_1, \cdots, X_n), \mathbf{Y} = (Y_1, \cdots, Y_n)$ and let $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ be the distributions of $\mathbf{X}$ and $\mathbf{Y}$. Let $\hat{\mathbf{X}} = (\hat{X}_1, \cdots, \hat{X}_n) \sim P_{\mathbf{X}}$ and $\hat{\mathbf{Y}} = (\hat{Y}_1, \cdots, \hat{Y}_n) \sim P_{\mathbf{Y}}$ be independent random vectors. Then we have $\hat{X}_1 + \cdots + \hat{X}_n$ and $\hat{Y}_1 + \cdots + \hat{Y}_n$ are both constants. Denoting $\hat{P}$ by the distribution of $\hat{\mathbf{X}}+\hat{\mathbf{Y}}$, the 1-marginal distribution P' of $\hat{P}$ is identical with the distribution of $X + Y$. Now $X_i + Y_i \sim P'$, $i = 1, \cdots, n$ and $\sum_{i=1}^n (X_i + Y_i)$ is a constant. Hence P' is completely mixable with index n.

(6) First note that $\mu := \mathbb{E}^P(X) = \lim_k \mathbb{E}^{P_k}(X)$ since S is compact. Denote R_k an n-complete mix with marginal distribution P_k. Since S^n is also a compact set, there is a subsequence $\{R_{k_i}\}$ such that R_{k_i} converges weakly to a distribution Q on S^n. Obviously the 1-marginal distribution of Q is the limit of 1-marginal distributions of R_k, namely P. $(X_1, \cdots, X_n) \sim Q$ will lead to $X_1 + \cdots + X_n = n\mu$. Therefore, Q is an n-complete mix and P is completely mixable with index n.

(7) Let $X_i \sim P$, $i = 1, \cdots, n$, $X_1 + \cdots + X_n = n\mu$ and suppose $a > -\infty$. Note that if $\mu < a + \frac{b-a}{n}$, then $X_1 = n\mu - (X_2 + \cdots + X_n) \leq n\mu - (n-1)a < b$, which contradicts the fact that $b = \sup\{x : \mathbb{P}(X \leq x) < 1\}$. Thus $\mu \geq a + \frac{b-a}{n}$ and $b < \infty$. The inequality $\mu \leq b - \frac{b-a}{n}$ and the case given $b < \infty$ can be obtained similarly. $\square$

One nice result for the complete mixability is given in Rüschendorf and Uckelmann [21]. We cite this result in a rewritten form in the following theorem.

Theorem 2.2. ***(Rüschendorf and Uckelmann)*** *Suppose the probability density function $p(x)$ of a distribution P is symmetric and unimodal, then P is completely mixable* *with any index greater than 1.*

Remark 2.1.

1. We conjecture that the center μ in **Proposition 2.1**(2) is always unique, i.e. for given distribution P, the center does not depend on the way we choose the index n or the random variables $X_1, \cdots, X_n$. However we did not come to a proof.

2. Rüschendorf and Uckelmann [21] also extended the complete mixability to multivariate distributions and constructed examples in some standard situations.

A few examples of completely mixable distributions are given in the following proposition.

Proposition 2.3. ***(Examples of complete mixable distributions)***

(1) P is completely mixable with index 1 if and only if P is the distribution of a constant.

(2) P is completely mixable with index 2 if and only if P is symmetric, i.e. $X \sim P$ and $a - X \sim P$ for some constant a.

(3) Suppose $r = \frac{p}{q}$ is rational, $p, q \in \mathbb{N}$, then the binomial distribution $B(n, r)$ is completely mixable with index q.

(4) The uniform distribution on any interval $[a, b]$ is completely mixable with any index greater than 1.

(5) The normal distribution and the Cauchy distribution are completely mixable with any index greater than 1.

Proof. (1) and (2) are obvious. For (3), let $S = (\underbrace{0, \cdots, 0}_{q-p}, \underbrace{1, \cdots, 1}_{p})$, σ be a random permutation uniformly distributed on the set of all q-permutations, and the random vector $\mathbf{X} = (X_1, \cdots, X_q) = \sigma(S)$. We can check that $X_i \sim B(1, r)$ for $i = 1, \cdots, q$ and

$X_1 + \cdots + X_q = p$ is a constant. Hence $B(1, r)$ is completely mixable with index q. The rest part of (3) follows from **Proposition 2.1**(5). (4) can be found in Rüschendorf and Uckelmann [21]. (5) is an application of **Theorem 2.2**. $\square$

The following theorem is the key result of this paper. It shows the complete mixability of monotone distributions on a finite interval.

Theorem 2.4. *Suppose the probability density function $p(x)$ of a distribution P is decreasing on $[0, 1]$, $p(x) = 0$ elsewhere, and $\mathbb{E}^P(X) \geq \frac{1}{n}$. Then P is completely mixable with index n.*

Before approaching the proof of this theorem, we have to introduce the mass-version of the complete mixiability and provide some necessary preliminaries.

2.2 Mass-version of the complete mixiability

In the following, a function $A : \mathbb{Z} \to \mathbb{R}_0^+$ is called a mass function.

Definition 2.2. (Simply mixable mass functions, centered at 0.) Let S be a subset of $\mathbb{Z}$. A mass function B is **simply mixable on S with index n** if its support is contained in S, and

(a) $B(i) \in \mathbb{N}_0$ for each $i \in S$. $\mathbb{N}_0$ is the set of nonnegative integers.

(b) $B(S) := \sum_{j \in S} B(j) = n$.

(c) $\sum_{j \in S} j \times B(j) = 0$.

Definition 2.3. (Completely mixable mass functions, centered at 0.) Let S be a subset of $\mathbb{Z}$. A mass function A is **completely mixable on S with index n** if $A = \sum_{i=1}^{\infty} a_i B_i$ for some $a_i \geq 0$ and B_i simply mixable on S with index n.

Remark 2.2. Suppose the support of the mass function A is contained in S. The following facts are obvious to check:

(1) Suppose $S \subset T$. If A is completely mixable on S with index n, then it is completely mixable on T with index n.

(2) If A is completely mixable on S with index n, then cA is completely mixable on S with index n for any constant $c \geq 0$.

(3) If A_i is completely mixable on S_i with index n, $i = 1, \cdots, k$ and $A_0 = A_1 + \cdots + A_k$, then A_0 is completely mixable on $\bigcup_i S_i$ with index n.

The following lemma explains why **Definition 2.3** is reasonable.

Lemma 2.5. *Suppose the mass function A is supported in a subset of S and $0 < A(S) < \infty$. A is completely mixable on S with index n, if and only if the corresponding discrete probability distribution P, i.e.*

$$P(\{i\}) = \frac{A(i)}{A(S)},$$

is completely mixable with index n.

The proof will be given in Appendix.

2.3 A combinatorial proof of Theorem 2.4

For $n = 1$ or 2, the proof is trivial since no distribution satisfies the assumption when $n = 1$, and only one distribution, namely the uniform distribution, satisfies the assumption when $n = 2$. Hence we only need to prove the case of $n \geq 3$. Since the complete mixability is invariant under affine transformations, without losing generality we assume the center to be 0.

Let d and N be positive integers, where $d = n-1 \geq 2$, and let $S_N^d := \{-N, \cdots, -1, 0, 1, \cdots, dN\}$ be a set of $(d+1)N + 1$ points.

Lemma 2.6. *Suppose the mass function A is supported in S_N^d, and the pair (A, N) satisfies* *(i) (decreasing mass)*

$$A(-N+1) \geq \cdots \geq A(0) \geq \cdots \geq A(dN) \geq 0, \tag{3}$$

(ii) (boundary condition)

$$C_N(A) = A(-N) - [d \times A(dN) + (d-1) \times A(dN-1) + \cdots + 1 \times A(dN-d+1)] \geq 0, \tag{4}$$

(iii) (zero center of mass)

$$\sum_{i=-N}^{dN} i \times A(i) = 0. \tag{5}$$

Then A is completely mixable on S_N^d with index $d+1$.

Proof. We prove this lemma by induction over N. Our idea is to write $A = \bar{A} + \sum_{i=0}^{K} b_i B_i$ such that for each i, $b_i \geq 0$, B_i is a simply mixable (on S_N^d with index $d+1$ if not specified) mass function, $\bar{A}$ is supported in S_{N-1}^d, and $(\bar{A}, N-1)$ satisfies (i) and (ii). Note that (iii) is automatically satisfied, since each simple mixable mass function B_i is centered at 0. First we need the following fact.

Lemma 2.7. *If* (3), (5) *in **Lemma 2.6** hold and $A(-N) \geq \frac{d+1}{2d} A(-N+1)$, then* (4) *holds.*

The proof will be presented in Appendix. This lemma implies that if $A(-N) \geq A(-N+1)$, (3) and (5) hold, then (4) holds. Thus, a decreasing mass function with zero center is sufficient for **Lemma 2.6**.

Now suppose **Lemma 2.6** holds for the case of $N-1$ (here $N \geq 2$).

Case 1. $C_N(A) = 0$.

If $A(-N)$=0 then (4) implies that $A(dN) = A(dN-1) = \cdots = A(dN-d+1) = 0$. Thus A is supported in S_{N-1}^d and $(A, N-1)$ satisfies (i), (ii) and (iii). Therefore A is completely mixable on S_{N-1}^d (and hence on S_N^d) with index $d+1$.

If $A(-N) > 0$, we construct B_i, $i = 0, 1, \cdots, d-1$ such that $B_i(-N) = d - i$, $B_i(-N+1) = i$, $B_i(dN-i) = 1$ and 0 otherwise. Obviously each B_i is simply mixable. Let $b_i = A(dN-i)$ and $\bar{A} = A - \sum_{i=0}^{d-1} b_i B_i$. It is straightforward to check $\bar{A}$ is still a mass function and is supported in S_{N-1}^d. Clearly $\bar{A}(i) = A(i)$ for $i = -N+2, \cdots, dN-d$, and hence (i) is satisfied by $(\bar{A}, N-1)$.

The rest work is to check (ii) $C_{N-1}(\bar{A}) \geq 0$. It is just some algebraic calculation and we leave it in Appendix. Thus $\bar{A}$ is completely mixable on S_{N-1}^d with index $d+1$. This shows $A = \bar{A} + \sum_{i=0}^{d-1} b_i B_i$ is completely mixable (on S_N^d).

Case 2. $C_N(A) > 0$.

Denote $M = M_A = \max\{i : A(i) > 0\}$. By (i) and $A(-N) > 0$, it follows that $N \leq M \leq dN$. Let q and r be integers such that

$$(d+1)N = (N+M)q + r, \quad 0 \leq r < N+M.$$

Obviously $q < d$. For $i = 0, 1, \cdots, M+N-r$, Let $B_i(-N) = d - q$, $B_i(M) = q - 1$, $B_i(r - N + i) = B_i(M - i) = 1$ and 0 elsewhere. It is easy to check each B_i is simply mixable.

Let $T = T_A = \sum_{i=0}^{M+N-r} B_i$. Then T is completely mixable, $T(-N) = (d-q)(M + N - r + 1)$, $T(M) = (q-1)(M+N-r+1) + 2$, $T(r-N) = T(r-N+1) = \cdots = T(M-1) = 2$ and 0 otherwise. We have

$$C_N(T) = \begin{cases} (d-q)(M+N-r+1), & M \leq dN - d, \\ (d-1)((d+1)N - 2r + 1) - (d-r+1)(d-r), & M > dN - d. \end{cases}$$

Thus $C_N(T) > 0$. Let $b_A = \max\{x : xT(M) \leq A(M), xC_N(T) \leq C_N(A))\}$. For each mass function A, we define an operator $\mathcal{R}A := A - b_A T_A$. Note that $C_N(\mathcal{R}A) = C_N(A) - b_A C_N(T)$. It is straightforward to check $\mathcal{R}A$ is still a mass function, $(\mathcal{R}A, N)$ satisfies (i), (ii), (iii) and either $\mathcal{R}A(M) = 0$ or $C_N(\mathcal{R}A) = 0$.

If $C_N(\mathcal{R}A) = 0$, then $\mathcal{R}A$ fits into Case 1, being completely mixable and therefore $A = \mathcal{R}A + b_A T_A$ is completely mixable.

If $C_N(\mathcal{R}A) > 0$, then $\mathcal{R}A(M) = 0$ and $M_{\mathcal{R}A} \leq M - 1$. Now we consider $\mathcal{R}^k A$, $k = 2, 3, \cdots$. Since $M_{\mathcal{R}^k A} \geq 0$ for all k as long as $\mathcal{R}^k A \neq 0$, we have $C_N(\mathcal{R}^k A) = 0$ for some k. Thus $\mathcal{R}^k A$ is completely mixable and so is $A = \mathcal{R}^k A + \sum_{i=0}^{k-1} b_{\mathcal{R}^i A} T_{\mathcal{R}^i A}$.

Now it is only left to show that the lemma holds for $N = 1$. Let T_A and M_A be defined as in Case 2. When $N = 1$, (iii) becomes $C_1(A) = 0$, therefore $C_1(T_A) = 0$ since $(T_A, 1)$ satisfies (iii). For $A(-1) = 0$, $A = 0$ on $S_1^d \setminus \{0\}$ and the lemma is trivial. For $A(-1) > 0$, let $b_A = A(M_A)/T_A(M_A)$ and $\mathcal{R}A := A - b_A T_A$. Similar to case 2, $\mathcal{R}A$ is still a mass function, $(\mathcal{R}A, 1)$ satisfies (i), (ii), (iii) and $\mathcal{R}A(M_A) = 0$. We consider $\mathcal{R}^k A$, $k = 2, 3, \cdots$ and eventually $M_{\mathcal{R}^k A} = 0$ for some k. Hence $R^k A$ is completely mixable and so is A. This completes the proof. $\square$

The following lemma is an immediate consequence of **Lemma 2.5**, **Lemma 2.6** and **Lemma 2.7**.

Lemma 2.8. *Suppose the probability mass function of a distribution P with mean 0 is decreasing on S_N^d and is 0 elsewhere, then P is completely mixable with index $d+1$.*

Let $S_N = \{-N/N, (-N+1)/N, \cdots, (dN-1)/N, dN/N\}$. For each continuous distribution P on $[-1, d]$ with mean zero and decreasing density, let $Y \sim P$. Denote $\bar{P}_N$ the distribution function of $\lfloor NY \rfloor / N$ and $\hat{P}_N$ the discrete uniform distribution on S_N. Since

$$-\frac{1}{N} \leq \int y \bar{P}_N(\mathrm{d}y) \leq 0$$

and

$$\int y \hat{P}_N(\mathrm{d}y) = \frac{d-1}{2} \geq \frac{1}{2},$$

there exists λ_N: $0 \leq \lambda_N < 2/N$ such that

$$\int y((1-\lambda_N)\bar{P}_N + \lambda_N \hat{P}_N)(\mathrm{d}y) = 0.$$

Then the distributions $\{(1-\lambda)\bar{P}_N + \lambda\hat{P}_N\}$ are decreasing on S_N, with mean zero, and converge weakly to P as $N \to \infty$. This argument shows that there exist $P_k \xrightarrow{d} P$ and each P_k is completely mixable with index $d+1$ and centered at 0. Then by **Proposition 2.1**(6), as the limit of completely mixable distributions, each continuous distribution P on $[-1, d]$ with mean 0 and decreasing density is completely mixable with index $d+1$.

Finally, by **Proposition 2.1**(1), each continuous distribution P on $[0, 1]$ with mean $\frac{1}{n}$ and decreasing density is completely mixable with index n. Just note that any decreasing density on $[0, 1]$ is also an decreasing density on $[0, a]$, hence each continuous distribution P on $[0, 1]$ with mean $\frac{a}{n}$, $a \geq 1$ and decreasing density is completely mixable with index n. This completes the proof of **Theorem 2.4**. □

As an corollary, we give the general version of **Theorem 2.4**.

Corollary 2.9. *Suppose the probability density function $p(x)$ of a distribution P is monotone on $[a, b]$ and $p(x) = 0$ elsewhere. If*

- *$p(x)$ is increasing and*
$$\mathbb{E}^P(X) \leq b - \frac{1}{n}(b-a),$$
or
- *$p(x)$ is decreasing and*
$$\mathbb{E}^P(X) \geq a + \frac{1}{n}(b-a),$$

then P is completely mixable with any index greater than or equal to n.

Remark 2.3.

1. By **Proposition 2.1**(7), the condition in **Corollary 2.9** is necessary and sufficient for a distribution P with monotone density on $[a, b]$ (where a and b are the infimum and the supremum of $\{x : p(x) > 0\}$) to be completely mixable with index n.

2. Different from Rüschendorf and Uckelmann [21], we did not construct random variables $X_1, \cdots, X_n \sim P$ such that $X_1 + \cdots + X_n$ is a constant (although we know they exist).

3 Convex minimization problems

Given a distribution P with monotone density on its support, and a convex function $f : \mathbb{R} \to \mathbb{R}$, the minimization problem (2)

$$\min_{X_1,\cdots,X_n \sim P} \mathbb{E} f(X_1 + \cdots + X_n)$$

is classic in variance minimization and simulation (see Rüschendorf and Uckelmann [21] and Hammersley and Handscomb [11]).

In the following we denote G the inverse cdf of $Y_i \sim P$, then $Y_i = G(X_i)$ for some $X_i \sim \mathbb{U}$, $i = 1, \cdots, n$ and (2) reads as

$$\min_{X_1,\cdots,X_n \sim \mathbb{U}} \mathbb{E} f(G(X_1) + \cdots + G(X_n)) = \min_{C \in \mathfrak{C}_n} \mathbb{E}^C f(G(X_1) + \cdots + G(X_n)), \qquad (6)$$

where $\mathfrak{C}_n$ is the set of all n-copulas (i.e., the joint-distribution of n U[0,1] random variables. See Nelsen [16] for a detailed introduction to copulas). Note that

1. P having an increasing (decreasing) density is equivalent to G being continuous and concave (convex). Thus both f and G have convexity in this problem and the equivalent setting for (2) is

$$\min_{X_1,\cdots,X_n \sim \mathbb{U}} \mathbb{E} f(G(X_1) + \cdots + G(X_n))$$

for $f : \mathbb{R} \to \mathbb{R}$ being convex and $G : [0,1] \to \mathbb{R}$ being concave (convex), continuous and increasing.

2. If $X \sim P$ and P has decreasing density, we can simply replace X by $-X$ (note that $f(-x)$ is also convex). Thus without loss of generality, in the following we will assume P has increasing density.

To obtain an optimal coupling for problem (6), we construct n-copulas $Q_n^P(c)$ $(n \geq 2)$ for some $0 \leq c \leq 1/n$, $(X_1, \cdots, X_n) \sim Q_n^P(c)$ satisfying

(a) For each $i = 1, \cdots, n$, the joint-density of $X_1, \cdots, X_n$ given $X_i \in [0, c]$ is uniformly supported on line segments $x_j = 1 - (n-1)x_i, \ \forall j \neq i, \ x_i \in [0, c]$; and

(b) $G(X_1) + \cdots + G(X_n)$ is a constant when $X_i \in (c, 1-(n-1)c)$ for any $i = 1, \cdots, n$.

Proposition 3.1. *Denote*

$$H(x) = G(x) + (n-1)G(1-(n-1)x). \tag{7}$$

There exists a copula $Q_n^P(c)$ satisfying (a) *and* (b) *if*

$$\int_c^{\frac{1}{n}} H(t)\mathrm{d}t \leq (\frac{1}{n} - c)H(c). \tag{8}$$

Proof. We first take random variables $Y_1, \cdots, Y_n \sim \mathrm{U}([0, c] \cup [1-(n-1)c, 1])$ such that the joint-density of $Y_1, \cdots, Y_n$ is uniformly supported on each line segment $y_j = 1-(n-1)y_i, \ \forall j \neq i, \ y_i \in [0, c]$. By **Corollary 2.9**, there exist $Z_1, \cdots, Z_n \sim \mathrm{U}[c, 1-(n-1)c]$ such that $G(Z_1) + \cdots + G(Z_n)$ is a constant since $G(Z_i)$ has an increasing density and that (8) implies

$$\mathbb{E}(G(Z_1)) \leq G(c) + \frac{n-1}{n}[G(1-(n-1)c) - G(c)].$$

Let $U \sim \mathbb{U}$ be independent of $(Y_1, \cdots, Y_n, Z_1, \cdots, Z_n)$ and $X_i = \mathbf{1}_{\{U<nc\}}Y_i + \mathbf{1}_{\{U \geq nc\}}Z_i$, then $X_i \sim \mathbb{U}$ for $i = 1, \cdots, n$. Properties (a) and (b) are satisfied by the joint distribution of $X_1, \cdots, X_n$, which shows that $Q_n^P(c)$ exists. $\square$

Remark 3.1.

1. Property (a) describes the joint distribution on the set $\bigcup_{i=1}^n \{0 \leq x_i \leq c, 1-(n-1)c \leq x_j \leq 1, j \neq i\}$, and property (b) describes it on the set $(c, 1-(n-1)c)^n$. These two sets are disjoint and their union is $[0, 1]^n$.

2. The key idea of constructing $Q_n^P(c)$ is that when X_i is small, we let other random variables $X_j,\ j \neq i$ be large. When each of X_i, $i = 1, \cdots, n$ is of medium size, we let $G(X_1)+\cdots+G(X_n)$ be a constant. This could be a good candidate of optimal coupling since the variance of $G(X_1) + \cdots + G(X_n)$ is largely reduced. Later we will show that $Q_n^P(c)$ is optimal for the smallest possible c.

3. $Q_n^P(c)$ does not always exist for arbitrary c and it may not be unique while exists. However, $\mathbb{E}^{Q_n^P(c)} f(G(X_1) + \cdots + G(X_n))$ is determined by properties (a) and (b). Therefore, in the following $Q_n^P(c)$ is just one representative in the family of copulas satisfying (a) and (b).

4. It is easy to check that when $Q_2^P(c)$ exists, it is exactly the Fréchet-Hoeffding lower bound $W_2(u, v) = (u + v - 1)_+$.

We denote c_n the smallest c such that $Q_n^P(c)$ exists and let $Q_n^P := Q_n^P(c_n)$. Note that $c_n = 0$ if and only if P is completely mixable with index n. In the following we will find c_n and show the minimality of Q_n^P. Figure 3.1 gives the support of one Q_3^P for $P = \mathbb{U}$. In this case, $c_n = 0$ and P is completely mixable. Figure 3.2 gives the support of one Q_3^P for $P = -\text{Expo}(1)$ (see also Section 4.1). Note that such Q_3^P may not be unique.

Proposition 3.2. *The smallest possible c is given by*

$$c_n = \min\{c \in [0, \frac{1}{n}] : \int_c^{\frac{1}{n}} H(t)\mathrm{d}t \leq (\frac{1}{n} - c)H(c)\}. \tag{9}$$

Proof. Suppose $Q_n^P(c)$ exists. By (b), when any of $X_i \in (c, 1-(n-1)c)$, $G(X_1)+\cdots+G(X_n)$ is a constant, namely

$$\begin{aligned} G(X_1) + \cdots + G(X_n) &= \mathbb{E}(G(X_1) + \cdots + G(X_n)|c \leq X_i \leq 1-(n-1)c) \\ &= \frac{n}{1-nc}\int_c^{1-(n-1)c} G(t)\mathrm{d}t. \end{aligned}$$

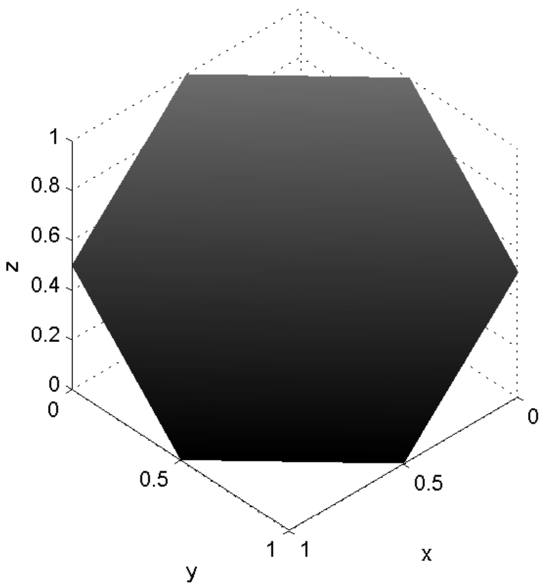

Figure 3.1: The support of one Q_3^P, $P = \mathbb{U}$.

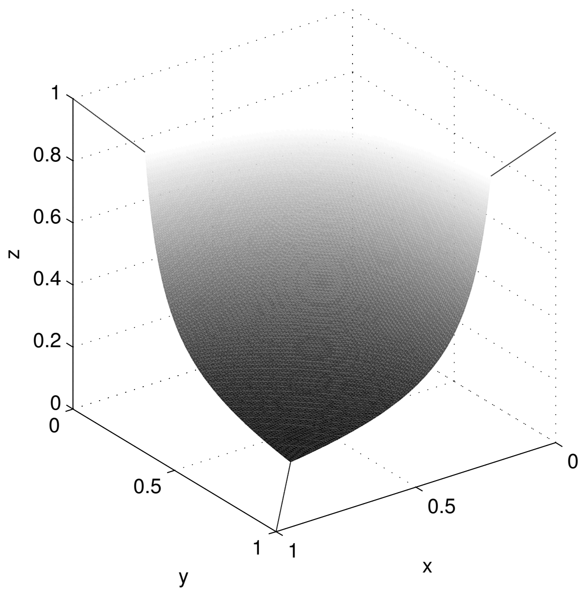

Figure 3.2: The support of one Q_3^P, $P = -\mathrm{Expo}(1)$.

Noting that the conditional distribution of $G(X_i)$ on the set $\{X_i \in (c, 1-(n-1)c)\}$ is completely mixable, by **Proposition 2.1**(7) its conditional mean is less than or equal to $G(c)/n + (n-1)G(1-(n-1)c)/n$. Thus we have a necessary condition on c,

$$\int_c^{1-(n-1)c} G(t)\mathrm{d}t \leq (\frac{1}{n} - c)[G(c) + (n-1)G(1-(n-1)c)]. \tag{10}$$

Together with (7), we obtain (8) from (10).

Note that $H(x)$ is concave on $[0, \frac{1}{n}]$ since $G(x)$ is concave. Hence the set of c satisfying (10) is a closed interval $[\hat{c}_n, \frac{1}{n}]$. (8) becomes $\hat{c}_n \leq c \leq \frac{1}{n}$ and therefore $c_n \geq \hat{c}_n$. By **Proposition 3.1** we know $Q_n^P(\hat{c}_n)$ exists and thus $c_n = \hat{c}_n$. $\square$

Now we have c_n and $Q_n^P = Q_n^P(c_n)$. We will next show the minimality of Q_n^P, where the following lemma (see **Theorem 3.A.5** in Shaked and Shanthikumar [22]) will be used.

Lemma 3.3. *Suppose X and Y with distribution functions F_1, F_2 respectively satisfy $\mathbb{E}X = \mathbb{E}Y$ and for any c in $[0,1]$, $\int_0^c F_1^-(t)dt \geq \int_0^c F_2^-(t)dt$, where $F_1^-(t) = \sup\{x : F_1(x) < t\}$ and $F_2^-(t) = \sup\{y : F_2(y) < t\}$. Then for any convex function f, $\mathbb{E}(f(X)) \leq \mathbb{E}(f(Y))$.*

Theorem 3.4. *Suppose P is a distribution with increasing density and G is the inverse cdf of P, then for any convex function f,*

$$\min_{Z_1,\cdots,Z_n \sim P} \mathbb{E}f(Z_1 + \cdots + Z_n) = \mathbb{E}^{Q_n^P} f\left(G(X_1) + \cdots + G(X_n)\right). \tag{11}$$

Proof. Let $(X_1, \cdots, X_n) \sim Q_n^P$ and $Z_i = G(Y_i)$ where $Y_i \sim \mathbb{U}$, $i = 1, \cdots, n$. Denote $X = G(X_1) + \cdots + G(X_n)$ and $Y = G(Y_1) + \cdots + G(Y_n)$. Let F_1 and F_2 be the cdf of X and Y respectively, $F_1^-(t) = \sup\{x : F_1(x) < t\}$ and $F_2^-(t) = \sup\{y : F_2(y) < t\}$. We will show that for any $c \in [0,1]$,

$$\int_0^c F_1^-(t)\mathrm{d}t \geq \int_0^c F_2^-(t)\mathrm{d}t.$$

To obtain this, denote $A_X(u) = \bigcup_i \{X_i < u\}$, $A_Y(u) = \bigcup_i \{Y_i < u\}$ and let $W(u) = \mathbb{P}(A_Y(u))$. Obviously $u \le W(u) \le nu$ and W is invertible. For $c \in [0, nc_n]$, let $u^\star = W^{-1}(c)$, it then follows that $c \ge u^\star \ge c/n$ and $\{Y_i \in [0, c/n]\} \subset \{Y_i \in [0, u^\star]\} \subset A_Y(u^\star)$.

By the definition of Q_n^P, for each i, $\{X_i \in [0, c/n] \cup [1 - (n-1)c/n, 1]\} = A_X(c/n)$. Note that $X_i \stackrel{d}{=} Y_i \sim \mathbb{U}$ and $\mathbb{P}(A_X(c/n)) = \mathbb{P}(A_Y(u^\star)) = c$, therefore

$$\mathbb{P}(A_Y(u^\star) \setminus \{Y_i \in [0, c/n]\}) = c - c/n = \mathbb{P}(Y_i \in [1 - (n-1)c/n, 1]).$$

Since G is increasing and the above two sets are equally measured, we have

$$\mathbb{E}[\mathbf{1}_{\{Y_i \in [1-(n-1)c/n, 1]\}} G(Y_i)] \ge \mathbb{E}[\mathbf{1}_{A_Y(u^\star) \setminus \{Y_i \in [0, c/n]\}} G(Y_i)].$$

It follows that

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_{A_X(c/n)} G(X_i)) &= \mathbb{E}[(\mathbf{1}_{\{X_i \in [0, c/n]\}} + \mathbf{1}_{\{X_i \in [1-(n-1)c/n, 1]\}}) G(X_i)] \\
&= \mathbb{E}[(\mathbf{1}_{\{Y_i \in [0, c/n]\}} + \mathbf{1}_{\{Y_i \in [1-(n-1)c/n, 1]\}}) G(Y_i)] \\
&\ge \mathbb{E}[(\mathbf{1}_{\{Y_i \in [0, c/n]\}} + \mathbf{1}_{A_Y(u^\star) \setminus \{Y_i \in [0, c/n]\}}) G(Y_i)] \\
&= \mathbb{E}(\mathbf{1}_{A_Y(u^\star)} G(Y_i)).
\end{aligned}$$

Thus we have

$$\mathbb{E}(\mathbf{1}_{A_X(c/n)} X) \ge \mathbb{E}(\mathbf{1}_{A_Y(u^\star)} Y). \tag{12}$$

Note that $H(x)$ is concave and differentiable. By the definition of c_n, the mean of $H(x)$ on $[c_n, \frac{1}{n}]$ is $H(c_n)$. With $H(x)$ being concave, we have $H'(c_n) \ge 0$ and thus $H(x)$ is increasing on $[0, c_n]$. Note that on the set $A_X(c_n)$,

$$X = \sum_{i=1}^n \mathbf{1}_{\{X_i < c_n\}} [G(X_i) + (n-1) G(1 - (n-1) X_i)] = \sum_{i=1}^n \mathbf{1}_{\{X_i < c_n\}} H(X_i),$$

and the events $\{X_i < c_n\}$ $i = 1, \cdots, n$ are disjoint. It follows that for $t \le H(c_n)$, $F_1(t) = \mathbb{P}(X \le t) = n\mathbb{P}(H(X_1) \le t)$. Thus for $c \le nc_n$, $F_1^-(c) = H(c/n)$ and

$$\mathbb{E}(\mathbf{1}_{A_X(c/n)} X) = n \int_0^{c/n} H(t) \mathrm{d}t = \int_0^c H(t/n) \mathrm{d}t = \int_0^c F_1^-(t) \mathrm{d}t. \tag{13}$$

Also note that

$$\mathbb{E}(\mathbf{1}_{A_Y(u^\star)}Y) \geq \int_0^c F_2^-(t)\mathrm{d}t \tag{14}$$

since $\mathbb{P}(A_Y(u^\star)) = c$. It follows from (12), (13) and (14) that for any $c \in [0, nc_n]$,

$$\int_0^c F_1^-(t)\mathrm{d}t \geq \int_0^c F_2^-(t)\mathrm{d}t.$$

For $c \in (nc_n, 1]$, note that $H_1(x) := \int_0^x F_1^-(t)\mathrm{d}t$ and $H_2(x) := \int_0^x F_2^-(t)\mathrm{d}t$ are convex functions and $\mathbb{E}(X) = \mathbb{E}(Y)$ thus $H_1(1) = H_2(1)$. Furthermore we have $F_1^-(t)$ is a constant when $t \geq c_n$ since Q_n^P satisfies (b). By the facts that $H_1(c_n) \geq H_2(c_n)$, $H_1(1) = H_2(1)$, H_1 is linear over $[nc_n, 1]$ and H_1, H_2 are convex, we conclude

$$\int_0^c F_1^-(t)\mathrm{d}t \geq \int_0^c F_2^-(t)\mathrm{d}t$$

for any $c \in [0, 1]$. By **Lemma 3.3** we obtain

$$\mathbb{E}f(G(Y_1) + \cdots + G(Y_n)) \leq \mathbb{E}^{Q_n^P} f\left(G(X_1) + \cdots + G(X_n)\right)$$

and it completes the proof. □

Remark 3.2.

1. In stochastic orderings, the above result is interpreted in the following way: suppose $Y_1, \cdots, Y_n, Z_1, \cdots, Z_n \sim P$ and $Z_1, \cdots, Z_n$ have copula Q_n^P, then

$$Z_1 + \cdots + Z_n \leq_{\mathrm{cx}} Y_1 + \cdots + Y_n \leq_{\mathrm{cx}} nY_1.$$

   Thus $Z_1 + \cdots + Z_n$ is the lower bound in the convex order on the sum $Y_1 + \cdots + Y_n$ with given marginal distributions $Y_i \sim P$. This completes the result of bounds in the convex order on the sum with given monotone marginal distributions. For an overview of the stochastic orderings, see Shaked and Shanthikumar [22].

2. The optimal copula Q_n^P solving (2) depends only on the marginal distribution P, but not on the convex function f.

3. Although we are able to show the existence and minimality, we are unable to write the function Q_n^P explicitly.

Theorem 3.5. *We have*

$$\min_{Y_1,\cdots,Y_n\sim P} \mathbb{E}f\left(Y_1+\cdots+Y_n\right) = n\int_0^{c_n} f(H(x))\mathrm{d}x + (1-nc_n)f(H(c_n)), \tag{15}$$

where $H(x)$ and c_n are defined as in (7) *and* (9)*.*

Proof. By **Theorem 3.4**,

$$\begin{aligned}
& \min_{Y_1,\cdots,Y_n\sim \mathbb{P}} \mathbb{E}f\left(Y_1+\cdots+Y_n\right) \\
= \;& \mathbb{E}^{Q_n^P} f\left(G(X_1)+\cdots+G(X_n)\right) \\
= \;& n\mathbb{E}^{Q_n^P}[f(G(X_1)+\cdots+G(X_n))\mathbf{1}_{\{X_1\in[0,c_n]\}}] \\
& + \mathbb{E}^{Q_n^P}[f(G(X_1)+\cdots+G(X_n))\mathbf{1}_{\{X_1\in[c_n,1-(n-1)c_n]\}}] \\
= \;& n\mathbb{E}^{\mathbb{U}}[f(H(X_1))\mathbf{1}_{\{X_1\in[0,c_n]\}}] + \mathbb{E}^{\mathbb{U}}[f(H(c_n))\mathbf{1}_{\{X_1\in[c_n,1-(n-1)c_n]\}}] \\
= \;& n\int_0^{c_n} f(H(x))\mathrm{d}x + (1-nc_n)f(H(c_n)).
\end{aligned}$$

□

Corollary 3.6. *If the density of P is monotone and supported in a finite interval $[a,b]$, then*

$$\min_{X_1,\cdots,X_n\sim P} \mathbb{E}f\left(X_1+\cdots+X_n\right) = f(n\mathbb{E}^P(X))$$

for n sufficiently large.

Proof. We have $a < \mathbb{E}^P(X) < b$ since P is a continuous distribution. Hence there exists N such that $b - \frac{1}{n}(b-a) > \mathbb{E}^P(X)$ for $n \geq N$. By **Corollary 2.9** we know P is completely mixable with index n and centered at $\mathbb{E}^P(X)$. Thus we have

$$\mathbb{E}[f(n\mathbb{E}^P(X))] \geq \min_{X_1,\cdots,X_n\sim P} \mathbb{E}f\left(X_1+\cdots+X_n\right) \geq f(n\mathbb{E}^P(X))$$

by Jensen's inequality. This shows that

$$\min_{X_1,\cdots,X_n\sim P} \mathbb{E}f\left(X_1+\cdots+X_n\right) = f(n\mathbb{E}^P(X))$$

for n sufficiently large. □

4 Applications

4.1 The minimum of $\mathbb{E}(X_1X_2\cdots X_n)$, $X_i \sim \mathrm{U}[0,1]$

Let us look at the problem

$$\Lambda_n := \min_{X_1,\cdots,X_n\sim\mathbb{U}} \mathbb{E}(X_1X_2\cdots X_n). \tag{16}$$

Problem (16) has a long history. For $n=3$ and $X,Y,Z\sim\mathbb{U}$, Rüschendorf [19] found $1/24$ as a lower bound for $\mathbb{E}(XYZ)$, but apparently the bound is not sharp. Baiocchi [1] constructed a discretization of X, Y and Z and applied a linear programming to approximate the minimum, which leads to a value ≈ 0.06159. Bertino [2] obtained an upper bound ≈ 0.05481 for Λ_3, by manually taking the limit of one class of discretizations of X,Y,Z. He conjectured that this upper bound was the true value of Λ_3. Recently, Nelsen and Ubeda-Flores [17] introduced the coefficients of directional dependence, whose lower bound has not been found and equals a function of the lower bound for $\mathbb{E}(XYZ)$.

This problem is a special case of problem (2). By letting P be the distribution of $\log(X)$, $X\sim\mathbb{U}$ (namely, $P=-\mathrm{Expo}(1)$) and $f(x)=\exp(x)$, we can use **Theorem 3.4** and **Theorem 3.5** to solve (16). In fact Figure 3.2 illustrates the support of Q_3^P in this problem.

Corollary 4.1. *We have*

$$\begin{aligned}\Lambda_n =&\mathbb{E}^{Q_n^P}(X_1\cdots X_n)\\ =&\frac{1}{(n-1)^2}\left(\frac{1}{n+1}-(1-(n-1)c_n)^n+\frac{n}{n+1}(1-(n-1)c_n)^{n+1}\right)\\ &+(1-nc_n)c_n(1-(n-1)c_n)^{n-1},\end{aligned} \tag{17}$$

where c_n is the unique solution to

$$\log(1-(n-1)c)-\log(c)=n-n^2c,\ 0\le c<1/n. \tag{18}$$

It is an immediate application of **Theorem 3.4** and **Theorem 3.5**, hence we omit the proof here.

n	Λ_n	c_n	e^{-n}	$\Lambda_n e^n$
1	$1/2$	N/A	3.6788×10^{-1}	1.3591
2	$1/6$	$1/2$	1.3533×10^{-1}	1.2315
3	5.4803×10^{-2}	9.4542×10^{-2}	4.9787×10^{-2}	1.1008
4	1.9098×10^{-2}	2.5406×10^{-2}	1.8316×10^{-2}	1.0427
5	6.8604×10^{-3}	7.9597×10^{-3}	6.7379×10^{-3}	1.0182
10	4.5410×10^{-5}	4.5589×10^{-5}	4.5400×10^{-5}	1.0002
20	2.0612×10^{-9}	2.0612×10^{-9}	2.0612×10^{-9}	1.0000
50	1.9287×10^{-22}	1.9287×10^{-22}	1.9287×10^{-22}	1.0000
100	3.7201×10^{-44}	3.7201×10^{-44}	3.7201×10^{-44}	1.0000

Table 4.1: Numerical values of Λ_n

The numerical values of Λ_n for different n are presented in **Table 4.1**. One may suggest that $\Lambda_n \sim e^{-n}$ as n goes to infinity.

Corollary 4.2. *We have*

$$\Lambda_n = e^{-n} + \frac{n}{2}e^{-2n} + O(n^4 e^{-3n}).$$

See Appendix for the proof.

Remark 4.1.

1. In fact this approximating procedure can be done infinitely further. For $n = 10$, $\Lambda_{10} - e^{-10} = 1.0323 \times 10^{-8}$, $5e^{-20} = 1.0306 \times 10^{-8}$. We cam see that the approximation is already very precise.

2. Nelsen and Ubeda-Flores [17] introduced the directional dependence coefficients $\rho_n^{(\alpha_1 \cdots, \alpha_n)}$, $\alpha_i \in \{-1, 1\}$, $i = 1, \cdots, n$. The lower bound on $\rho_n^{(\alpha_1 \cdots, \alpha_n)}$ can be written

as

$$\rho_n^{(\alpha_1\cdots,\alpha_n)} \geq \min_{X_1,\cdots,X_n\sim\mathbb{U}}\{2^n\mathbb{E}(X_1\cdots X_n)-1\} = 2^n\Lambda_n - 1,$$

and our **Corollary 4.1** provides this value.

4.2 Bounds on the distribution of the sum of random variables

Suppose $\psi:\mathbb{R}^n\to\mathbb{R}$ is a measurable function. For any marginal distribution F_i, let $m_\psi(s) = \inf\{\mathbb{P}(\psi(X_1,\cdots,X_n)<s) : X_i\sim F_i,\ i=1,\cdots,n\}$. Finding $m_\psi(s)$ is related to many problems in multivariate probability and risk theory. In particular, this problem is equivalent to the worst Value-at-Risk scenarios in risk management. We refer to Gaffke and Rüschendorf [10], Rüschendorf [20], Embrechts, et al. [6], Embrechts and Puccetti [7] and [8] for detailed discussions on this topic. Unfortunately, as is mentioned in Embrechts and Puccetti [7]:

This dual optimization problem ($m_\psi(s)$) is very difficult to solve. The only explicit results known in the literature are given in Rüschendorf (1982) ([20]) *for the case of the sum of marginals being all uniformly or binomially distributed.*

By using our results in Section 2, we can solve $m_+(s) = \inf\{\mathbb{P}(X_1+\cdots+X_n<s) : X_i\sim F,\ i=1,\cdots,n\}$ for F satisfying a monotone property. For simplicity, we consider $F(x)$ on $[0,1]$.

Theorem 4.3. *Assume the cdf $F(x)$ has decreasing density on its support $[0,1]$ with mean μ and $\mathbb{E}^F(X|X\geq t)\geq t+\frac{1-t}{n}$ for any $t\in[0,1)$. Denote G the inverse of F and $\psi(t)=\mathbb{E}^F(X|X\geq G(t))$ for $t\in[0,1)$, then*

$$m_+(s) = \begin{cases} 0 & s\leq n\mu; \\ \psi^{(-1)}(s/n) & n\mu<s<n; \\ 1 & s\geq n. \end{cases}$$

Proof. $m_+(s) = 1$ for $s \geq n$ is trivial. By **Corollary 2.9**, F is completely mixable with index n. It follows that $\inf\{\mathbb{P}(X_1 + \cdots + X_n < n\mu : X_1, \cdots, X_n \sim F\} = 0$ and hence $m_+(s) = 0$ for $s \leq n\mu$.

For $n\mu < s \leq n$, let $X = X_1 + \cdots + X_n$, $X_i \sim F$ and consider the inequality

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(X\mathbf{1}_{\{X<s\}}) + \mathbb{E}(X\mathbf{1}_{\{X \geq s\}}) \\
&\geq \mathbb{E}[(X_1 + \cdots + X_n)\mathbf{1}_{\{X<s\}}] + s\mathbb{P}(X \geq s) \\
&\geq n\int_0^{\mathbb{P}(X<s)} G(t)\mathrm{d}t + s\mathbb{P}(X \geq s) \\
&= n\mu - n\int_{\mathbb{P}(X<s)}^1 G(t)\mathrm{d}t + s\mathbb{P}(X \geq s).
\end{aligned}$$

Thus for $\mathbb{P}(X \geq s) > 0$,

$$\frac{1}{\mathbb{P}(X \geq s)} \int_{\mathbb{P}(X<s)}^1 G(t)\mathrm{d}t \geq s/n,$$

which implies $\psi(\mathbb{P}(X < s)) = \mathbb{E}^F[Y|Y > G(\mathbb{P}(X < s))] \geq s/n$ and $\mathbb{P}(X < s) \geq \psi^{(-1)}(s/n)$. Also note that $\mathbb{P}(X \geq s) = 0$ implies $\mathbb{P}(X < s) = 1 \geq \psi^{(-1)}(s/n)$. It follows that

$$m_+(s) \geq \psi^{(-1)}(s/n). \tag{19}$$

Now we show the equality in (19) is attainable. Denote $a = \psi^{(-1)}(s/n)$ and consider the distribution of $G(V)$ where $V \sim \mathrm{U}[a, 1]$. Apparently it has decreasing density with mean

$$\mathbb{E}(G(V)) = \int_a^1 \frac{1}{1-a} G(t)\mathrm{d}t = \psi(a) \geq G(a) + \frac{1 - G(a)}{n}.$$

Therefore, by **Corollary 2.9** the distribution of $G(V)$ is completely mixable and there exist $V_i \sim \mathrm{U}[a, 1]$ such that $G(V_1) + \cdots + G(V_n) = n\psi(a) = s$.

Now let $Y_i = G(U)\mathbf{1}_{\{U \leq a\}} + G(V_i)\mathbf{1}_{\{U > a\}}$ where $U \sim \mathbb{U}$ and U is independent of

$(V_1, \cdots, V_n)$. We can check $Y_i \sim F$ via

$$\begin{aligned}
\mathbb{P}(Y_i \le t) &= \mathbb{P}(G(U) \le t, U \le a) + \mathbb{P}(G(V_i) \le t, U > a) \\
&= \mathbb{P}(U \le F(t), U \le a) + \mathbb{P}(V_i \le F(t))\mathbb{P}(U > a) \\
&= \mathbb{P}(U \le F(t))\mathbf{1}_{\{F(t) \le a\}} + \mathbb{P}(U \le a)\mathbf{1}_{\{F(t) > a\}} + \mathbb{P}(V_i \le F(t))\mathbb{P}(U > a) \\
&= F(t)\mathbf{1}_{\{F(t) \le a\}} + a\mathbf{1}_{\{F(t) > a\}} + (1-a)\frac{F(t)-a}{1-a}\mathbf{1}_{\{F(t) > a\}} \\
&= F(t),
\end{aligned}$$

and $\mathbb{P}(Y_1 + \cdots + Y_n < s) = \mathbb{P}(U \le a) = a$. This shows $m_+(s) \le a$. Together with (19) we have $m_+(s) = a = \psi^{(-1)}(s/n)$ for $n\mu < s < n$. $\square$

The $m_+(s)$ problem has been investigated based on the well-known duality theorem by Rüschendorf [20] (see also Embrechts and Puccetti [7]),

$$\begin{aligned}
m_\psi(s) = & 1 - \inf \Bigg\{ \sum_{i=1}^n \int f_i \mathrm{d}F_i : f_i \text{ are bounded measurable functions on } \mathbb{R} \text{ s.t.} \\
& \sum_{i=1}^n f_i(x_i) \ge 1_{[s,+\infty)}(\psi(x_1, \cdots, x_n)), \text{ for all } x_i \in \mathbb{R},\ i = 1, \cdots, n \Bigg\}.
\end{aligned}$$

In the following we give a proof based on the duality, which inserts our result in a broader context.

Proof based on duality and mass transportation.[1] $m_+(s) = 1$ for $s \ge n$ is trivial and $m_+(s) = 0$ for $s \le n\mu$ follows from the complete mixability of F. Now suppose $n\mu < s < n$. $s > 1$ since $\mu \ge 1/n$. **Theorem 4.2** in Embrechts and Puccetti [7] gives a lower bound

$$m_+(s) \ge 1 - n \inf_{r \in [0, s/n)} \frac{\int_r^{s-(n-1)r} (1 - F(t))\mathrm{d}t}{s - nr}, \tag{20}$$

and since F is supported in [0,1], we have

$$m_+(s) \ge 1 - n \inf_{r \in [0, s/n)} \frac{\int_r^1 (1 - F(t))\mathrm{d}t}{s - nr}. \tag{21}$$

[1] We are grateful to Prof. G. Puccetti and Prof. L. Rüchendorf who provided this proof. It was slightly modified to fit into our paper.

For $r \in [0, \frac{s}{n})$, $\left(\frac{\int_r^1 (1-F(t))\mathrm{d}t}{s-nr}\right)' = 0$ implies

$$g(r) := -\overline{F}(r)(s - nr) + n \int_r^1 \overline{F}(t)\mathrm{d}t = 0. \tag{22}$$

Suppose $r = r^\star$ satisfies (22), then

$$\overline{F}(r^\star) = \frac{n \int_{r^\star}^1 \overline{F}(t)\mathrm{d}t}{s - nr^\star}$$

and therefore $m_+(s) \geq F(r^\star)$ by (21). Note that $F(s/n) < 1$ by the fact that $E^F(X|X \geq t)$ exists for all $0 \leq t < 1$. $r^\star$ always exists since g is continuous, $g(0) = -s + n\mu < 0$ and

$$g(s/n) = n \int_{s/n}^1 \overline{F}(t)\mathrm{d}t > 0.$$

Integration by parts leads to

$$-\overline{F}(r^\star)(s - nr^\star) + n \int_{r^\star}^1 \overline{F}(t)\mathrm{d}t = -s\overline{F}(r^\star) + n \int_{r^\star}^1 t\mathrm{d}F(t) = 0,$$

and hence

$$s(1 - F(r^\star)) = n\mathbb{E}^F(X|X > r^\star)(1 - F(r^\star)).$$

Thus $s/n = \psi(F(r^\star))$ since $F(r^\star) < 1$. Therefore $m_+(s) \geq F(r^\star) = \psi^{(-1)}(s/n)$. The rest part is to show the equality holds, which can be done by the same argument as in the above proof. □

Remark 4.2.

1. From the proof, we can see that the bound (20) given in Embrechts and Puccetti [7] is sharp for F in **Theorem 4.3**.

2. The optimal coupling corresponding to the minimum probability consists of a completely mixable part and a residual part.

3. In Rüschendorf [20], $m_+(s)$ is found for uniform or binomial marginal distributions F. Our proof is similar to his method. The result in Rüschendorf [20] for the marginal U[0,1] is a special case ($F(x) = x,\ x \in [0, 1]$) of **Theorem 4.3**.

4. The regular condition $\mathbb{E}^F(X|X \geq t) \geq t + \frac{1-t}{n}$ prevents the conditional mean of X from being too close to one side. This condition is commonly satisfied by bounded distributions with monotone density for n not too small.

4.3 Stop-loss premiums of the total risk

Let $X_1, X_2, \cdots, X_n \geq 0$ be n individual risks with the same marginal distributions P. Their stop-loss premium is defined as $\mathbb{E}[(X_1 + \cdots + X_n - t)_+]$ where $t \geq 0$ is a constant and $(\cdot)_+ = \max\{\cdot, 0\}$. See Kaas, et al. [13] for references of this topic. An important problem in variance reduction is to determine the minimum of the stop-loss premium over all possible dependence structure, i.e.

$$\min_{X_1,\cdots,X_n \sim P} \mathbb{E}[(X_1 + \cdots + X_n - t)_+] = \min_{C \in \mathfrak{C}_n} \mathbb{E}^C[(G(U_1) + \cdots + G(U_n) - t)_+] \qquad (23)$$

where G is the pseudo-inverse of the cdf of $X_i \sim P$ and $\mathfrak{C}_n$ is the set of n-copulas. Our result solves (23) for monotone distributions P. By **Theorem 3.4**, we have

$$\begin{aligned} \min_{X_1,\cdots,X_n \sim P} \mathbb{E}[(X_1 + \cdots + X_n - t)_+] &= \mathbb{E}^{Q_n^P}[(G(U_1) + \cdots + G(U_n) - t)_+] \\ &= n \int_0^{c_n} [H(u) - t]_+ \mathrm{d}u + (1 - nc_n)[H(c_n) - t]_+. \end{aligned}$$

We provide a numerical result to compare the stop-loss premium $\mathbb{E}[(X_1 + X_2 + X_3 - t)_+]$ for 4 different cases when $n = 3$. Suppose P is the exponential distribution with parameter 1 and $X_1, X_2, X_3 \sim P$.

- Case 1. X_1, X_2 and X_3 are comonotonic (see Denneberg [3]), i.e. $X_1 = X_2 = X_3$ almost surely. This case gives the maximum stop-loss premium.
- Case 2. X_1, X_2 and X_3 are independent.
- Case 3. X_1, X_2 and X_3 are negatively correlated with copula $C^{(1,2,3)}$ in Yang, et al. [23] (i.e. the corresponding uniform random variables U_1, U_2 and U_3 in (23) satisfy $U_1 = 1 - U_3$ and U_2 is independent of U_1 and U_3).

- Case 4. X_1, X_2 and X_3 have copula Q_3^P. This case gives the minimum stop-loss premium.

The result is given in Figure 4.3.

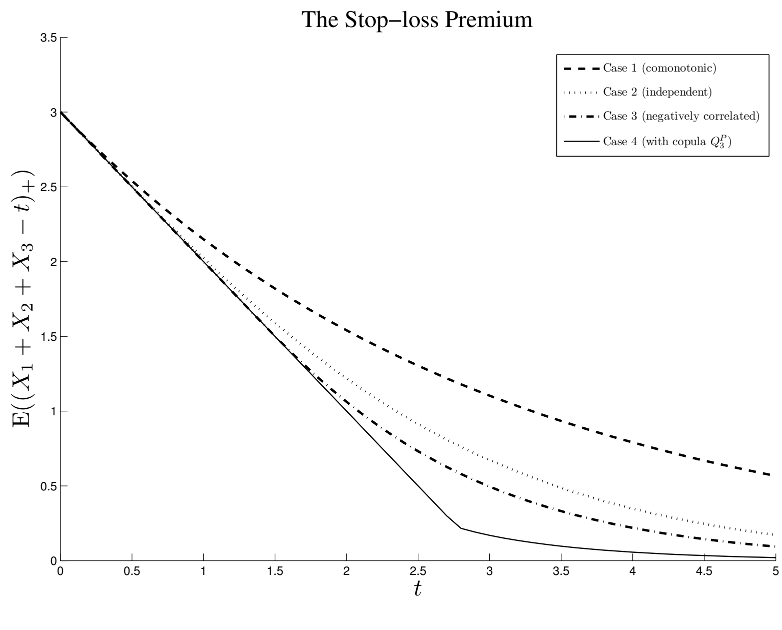

Figure 4.3: The stop-loss premium for different dependence structures

5 Open problems

There are many unsolved problems related the complete mixability and minimization problem (2). In the following we list some problems of interest.

1. Is the center of the complete mixability in **Proposition 2.1** always unique? We know it is unique when P follows WLLN.

   Embrechts and Puccetti [8] give an example of X_1, X_2, X_3 i.i.d. $\sim$ Pareto(1) (on p.123), and the distribution function of $X_1 + X_2 + X_3$ is always less than the distribution function of $3X_1$. This example shows that it is possible that when P has infinite mean, there exist $X_1, \cdots, X_n, Y_1, \cdots, Y_n \sim P$, and $X_1 + \cdots +$

$X_n > Y_1 + \cdots + Y_n$ with probability 1. However, we still do not know whether $X_1 + \cdots + X_n = \mu > \nu = Y_1 + \cdots + Y_n$ is possible for constants μ and ν.

2. **Theorem 2.2** and **Theorem 2.4** both give sufficient conditions for the complete mixability. Can we find a necessary and sufficient condition for the complete mixability?

3. For an arbitrary distribution P on $\mathbb{R}$, we can define

$$\alpha = \sup_{X_i \sim P,\ c \in \mathbb{R}} \mathbb{P}(X_1 + \cdots + X_n = c).$$

α can be considered as the measure of one kind of partial mixability. Note that $\alpha = 1$ gives the complete mixability and our Q_n^P solving (2) is actually an example of the partial mixability.

4. We only proved the existence of Q_n^P, but did not find any of them exactly. Similarly, for a completely mixable and monotone distribution P, we did not construct random variables $X_1, \cdots, X_n \sim P$ with a constant sum. It will be interesting to explicitly express $Q_n^P(x_1, \cdots, x_n) = \mathbb{P}^{Q_n^P}(X_1 \le x_1, \cdots, X_n \le x_n)$ and construct random variables $X_1, \cdots, X_n$ with a constant sum (like in Rüschendorf and Uckelmann [21]).

5. The optimal coupling Q_n^P for problem (2) does not work in the case of solving

$$\min_{X_1, \cdots, X_n \sim P} \mathbb{E}[\psi(X_1, ..., X_n)], \tag{24}$$

for a general supermodular function ψ (see e.g. Embrechts and Puccetti [8]). As a counter example, let $\psi(x_1, \cdots, x_n) = \prod_{i=1}^n x_i$ and $P = \mathbb{U}$, then (24) becomes (16) and is solved with optimal coupling $Q_n^{-\mathrm{Expo}(1)}$ (see Section 4.1), instead of $Q_n^{\mathbb{U}}$. Problem (24) is of importance in the theory of dependent measures and is still left to be solved. As a special case, for $\psi(x_1, \cdots, x_n) = \prod_{i=1}^n x_i$ and $X \sim P$

when the distribution of $\log(X)$ admits a monotone density, (24) can be solved by **Theorem 3.4**.

6 Conclusions

In this paper, we introduced the concept of the complete mixability, together with its basic properties and showed that monotone distributions with moderate mean are completely mixable. The minimum of $\mathbb{E}f(X_1 + \cdots + X_n)$ where f is a convex function and $X_i \sim P$ for monotone P was obtained. Our results also resolve some existing problems in variance reduction, bounds for the sum of random variables and individual risk models.

Acknowledgement

We are grateful to Christian Houdré, Liang Peng and Jingping Yang for their kindly guidance and valuable suggestions, Vladimir Kolchinskii for his ideas which provided great help to the proof of **Proposition 2.1**, Roger Nelsen for initially bringing up to the authors problem (16) which inspired this paper, Xiaoying Han and Cathy Jacobson for helping with the writing. We especially thank Giovanni Puccetti and Ludger Rüschendorf for the dual proof of **Theorem 4.3** and constructive comments which have significantly improved the organization, reference citation and details of this paper.

7 Appendix

Proof of Lemma 2.5

By *Remark* 2.2(2), without loss of generality we can assume $A(S) = 1$, i.e. A is a probability mass function.

$\Rightarrow$: Suppose the mass function A is completely mixable on S with index n. By definition there exist $\{B_i\}_{i=1}^K$ satisfying (a), (b) and (c) in **Definition 2.2** and $A = \sum_{i=1}^K a_i B_i$, $a_i \geq 0$. For each B_i, denote $S_{i,k} = \{j \in S : B_i(j) = k\}$, $k = 1, 2 \cdots, n$. Denote a vector $V_i = (j_1, \cdots, j_n)$, $j_1 \leq \cdots \leq j_n$, where in the sequence $j_1, \cdots, j_n$ each number in $S_{i,k}$ appear k times. Let σ be a random permutation uniformly distributed on the set of all n-permutations and let δ be a random number with $\mathbb{P}(\delta = i) = na_i$, $i = 1, \cdots, K$ and independent of σ. Note that $\sum_{i=1}^K na_i = \sum_{i=1}^K a_i B_i(S) = A(S) = 1$.

Now let the random vector $(X_1, \cdots, X_n) = \sigma(V_\delta)$, then $X_1 + \cdots + X_n = 0$ and

$$\begin{aligned}
\mathbb{P}(X_j = i) &= \mathbb{P}((\sigma(V_\delta))_j = i) \\
&= \sum_{l=1}^K \mathbb{P}((\sigma(V_l))_j = i)\mathbb{P}(\delta = l) \\
&= \sum_{l=1}^K \frac{B_l(i)}{n} \times na_l \\
&= A(i).
\end{aligned}$$

$\Leftarrow$: Suppose $X_1 + \cdots + X_n = 0$, $X_i \sim P$, $i = 1, \cdots, n$. Denote $\mathbf{X} = (X_1, \cdots, X_n)$. Then $\mathbf{X}$ takes value in S^n. For each possible value $\mathbf{a}_i = (a_{i1}, \cdots a_{in})$ of $\mathbf{X}$, $i = 1, \cdots, K$, $K \leq \infty$ we construct mass functions B_i, such that $B_i(m) = \#\{j : a_{ij} = m\}$. It is obvious that each B_i satisfies (a), (b) and (c) in **Definition 2.2**. Let $A = \sum_{i=1}^K \mathbb{P}(\mathbf{X} = \mathbf{a}_i)B_i/n$, by definition A is completely mixable on S with index n, and

$$A(j) = \sum_{i=1}^K \mathbb{P}(\mathbf{X} = \mathbf{a}_i)B_i(j)/n = \sum_{i=1}^K \mathbb{P}(\mathbf{X} = \mathbf{a}_i)\frac{1}{n}\sum_{k=1}^n \mathbb{E}(\mathbf{1}_{\{X_k = j\}}|\mathbf{X} = \mathbf{a}_i) = P(\{j\}).$$

Thus A is the mass function corresponding to distribution P. $\square$

Proof of Lemma 2.7

(5) reads as

$$N \times A(-N) + \cdots + 1 \times A(-1) = 1 \times A(1) + \cdots + dN \times A(dN). \tag{25}$$

The left-hand side of (25) is

$$
\begin{aligned}
N \times A(-N) + \cdots + 1 \times A(-1) &\leq N \times A(-N) + \frac{(N-1)N}{2} \times A(-N+1) \\
&\leq \left(N + \frac{N(N-1)}{2}\frac{2d}{d+1}\right) A(-N) \\
&= \frac{N(dN+1)}{d+1} \times A(-N).
\end{aligned}
$$

The right-hand side of (25) is

$$
\begin{aligned}
& 1 \times A(1) + \cdots + dN \times A(dN) \\
\geq\ & \frac{(dN-d+1)(dN-d+2)}{2} \times A(dN-d+1) \\
& +(dN-d+2) \times A(dN-d+2) + \cdots + dN \times A(dN) \qquad (26) \\
\geq\ & \frac{N(dN+1)}{d+1} \times (1 \times A(dN-d+1) + 2 \times A(dN-d+2) + \cdots + d \times A(dN)) \qquad (27)
\end{aligned}
$$

The last inequality is due to the fact that $A(dN-d+1) \geq \cdots \geq A(dN)$, the summation of all coefficients in (26) equals that in (27) and for each i and the summation of all coefficients from term $A(dN-d+1)$ to $A(dN-d+i)$ in (25) is greater than that in (27). Therefore we get

$$
1 \times A(dN-d+1) + 2 \times A(dN-d+2) + \cdots + d \times A(dN) \leq A(-N),
$$

and thus $C_N(A) \geq 0$. □

Proof of $C_{N-1}(\bar{A}) \geq 0$

Note that $\bar{A}(-N+1) = A(-N+1) - \sum_{i=1}^{d-1} iA(dN-i)$. Comparing the left-hand side and right-hand side of (25), we get

$$
\begin{aligned}
& N \times A(-N) + \frac{N(N-1)}{2} \times A(-N+1) \\
\geq\ & \text{LHS of (25)} \\
=\ & \text{RHS of (25)} \\
\geq\ & \frac{1}{2}(dN-d+1)(dN-d+2)A(dN-d+1) + \sum_{i=2}^{d}(dN-d+i)A(dN-d+i).
\end{aligned}
$$

Plugging $C_N(A) = 0$ in and after simplification (here we divide both sides by $N - 1$, hence $N \geq 2$ is needed), the above inequality reads as

$$N \times A(-N+1) \geq 2\sum_{j=1}^{d-2} jA(dN-j) + (d^2N - d^2 + 3d - 2)A(dN-d+1).$$

Since $A(dN-1) \leq A(dN-2) \leq \cdots \leq A(dN-d+1)$, we can calculate

$$\begin{aligned}
A(-N+1) &\geq \frac{1}{N}\left(2\sum_{j=1}^{d-1} jA(dN-j) + (d^2(N-1)+d)A(dN-d+1)\right) \\
&\geq \frac{2}{N}\left(\sum_{j=1}^{d-1} jA(dN-j) + \frac{d^2(N-1)+d}{d(d-1)}\sum_{j=1}^{d-1} jA(dN-j)\right) \\
&\geq \frac{2}{N}\left(1 + \frac{(N-1)d+1}{d-1}\right)\sum_{j=1}^{d-1} jA(dN-j) \\
&= \frac{2d}{d-1}\sum_{j=1}^{d-1} jA(dN-j).
\end{aligned}$$

This leads to

$$\bar{A}(-N+1) \geq A(-N+1) - \frac{d-1}{2d}A(-N+1) \geq \frac{d+1}{2d}A(-N+2) = \frac{d+1}{2d}\bar{A}(-N+2).$$

By **Lemma 2.7** we know $(\bar{A}, N-1)$ satisfies (ii). $\square$

Proof of Corollary 4.2

In the following we let P_n be the unique solution to

$$\log P = \frac{nP - n}{n + P - 1}, \quad P > 1. \tag{28}$$

One can show (28) has unique solution other than $P = 1$ by the following argument. Let $f(x) = \log x - n + \frac{n^2}{n+x-1}$. Then $f'(x) = \frac{1}{x} - \frac{n^2}{(n+x-1)^2}$, hence $f'(x)$ only has one root other than $x = 1$. This shows $f(x) = 0$ has at most one root other than $x = 1$. Note that $f(2) < 0$ and $f(e^n) > 0$, thus it has unique root other than $x = 1$.

Let $c_n = \frac{1}{P_n+n-1}$ ($P_n = \frac{1-(n-1)c_n}{c_n}$) and plug it in (28), we get c_n is the unique solution to (18).

For any $0 < \eta < 1$,

$$f(\eta e^n) = \log \eta + \frac{n^2}{n + \eta e^n - 1} < 0$$

for large n, hence there is a solution to $f(x) = 0$ between ηe^n and e^n. Since P_n is the solution, we know $P_n \sim e^n$, therefore $c_n = \frac{1}{P_n+n-1} \sim e^{-n}$.

Furthermore, it follows from $\log(P_n/e^n) = -n^2/(n + P_n - 1)$ and $P_n \sim e^n$ that

$$\begin{aligned}
P_n/e^n &= 1 - \frac{n^2}{P_n + n - 1} + \frac{n^4}{2(P_n + n - 1)^2} + O(\frac{n^6}{(P_n + n - 1)^3}) \\
&= 1 - \frac{n^2}{e^n} + \frac{n^2(P_n + n - 1 - e^n)}{e^n(P_n + n - 1)} + \frac{n^4}{2(P_n + n - 1)^2} + O(\frac{n^6}{e^{3n}}) \\
&= 1 - \frac{n^2}{e^n} + \frac{n^2(-n^2 + n - 1)}{e^{2n}} + O(\frac{n^6}{e^{3n}}) + \frac{n^4}{2e^{2n}} + O(\frac{n^6}{e^{3n}}) + O(\frac{n^6}{e^{3n}}) \\
&= 1 - n^2 e^{-n} + \frac{-n^4 + 2n^3 - 2n^2}{2} e^{-2n} + O(n^6 e^{-3n}).
\end{aligned}$$

Consequently

$$\begin{aligned}
c_n &= e^{-n} + (\frac{1}{P_n + n - 1} - e^{-n}) \\
&= e^{-n} + \frac{e^n - (P_n + n - 1)}{e^n(P_n + n - 1)} \\
&= e^{-n} + (n^2 - n + 1)e^{-2n} + O(n^4 e^{-3n}),
\end{aligned}$$

and

$$\begin{aligned}
\Lambda_n =& n \int_0^{c_n} x(1 - (n-1)x)^{n-1} \mathrm{d}x + (1 - nc_n)c_n(1 - (n-1)c_n)^{n-1} \\
=& n \int_0^{c_n} x(1 - (n-1)x)^{n-1} \mathrm{d}x + c_n[1 - ((n-1)^2 + n)c_n + O(n^3 c_n^2)] \\
=& n \int_0^{c_n} x(1 - (n-1)x)^{n-1} \mathrm{d}x + c_n - (n^2 - n + 1)c_n^2 + O(n^3 c_n^3). \\
=& \frac{n}{2} c_n^2 + O(n^3 c_n^3) + c_n - (n^2 - n + 1)c_n^2 + O(n^3 c_n^3). \\
=& e^{-n} + \frac{n}{2} e^{-2n} + O(n^4 e^{-3n}).
\end{aligned}$$

□

References

[1] Baiocchi, F. (1991). *Un problema di ottimizzatione nella classe di Fréchet.* Unpublished thesis, Facoltà di Scienze Statistiche, Università La Sapienza, Roma.

[2] Bertino, S. (1994). *The minimum of the expected value of the product of three random variables in the Frèchet class.* Journal of the Italian Statistical Society, 1994-6, 201-211.

[3] Denneberg, D. (1994). *Non-additive Measure and Integral.* Kluwer Academic Publishers, Boston.

[4] Embrechts, P., Lindskog, F. and McNeil, A. (2001). *Modelling dependence with copulas and applications to risk management*, in: Handbook of Heavy Tailed Distributions in Finance. S.T. Rachev (Ed.). Elsevier Science B.V. 2003.

[5] Embrechts, P., McNeil, A. and Straumann, D. (2002). *Correlation and Dependence in Risk Management: Properties and Pitfalls*, in: Risk Management: Value at Risk and Beyond, M. Dempster (Ed.), Cambridge University Press, Cambridge, 176-223.

[6] Embrechts, P., Höing, A. and Juri, A. (2003). *Using copulae to bound the value-at-risk for functions of dependent risks.* Finance and Stoch 7, 145-167.

[7] Embrechts, P and Puccetti, G. (2006). *Bounds for functions of dependent risks.* Finance and Stoch 10, 341-352.

[8] Embrechts, P. and Puccetti, G. (2010). *Risk Aggregation.* In: Copula Theory and Its Applications, P.Jaworski et al. (eds.), Lecture Notes in Statistics 198, Springer, Berlin/Heidelberg.

[9] Fishman, G.S. (1972). *Variance reduction in simulation studies.* J. Statist. Comp. and Simulation 1, 173-182.

[10] Gaffke, N. and Rüschendorf, L. (1981). *On a class of extremal problems in statistics.* Math. Operationsforschung Statist., Series Optimization, Vol. 12, 123-135.

[11] Hammersley, I.M and Handscomb, D.C. (1964). *Monte Carlo Methods.* Methuen, London.

[12] Joe, H. (1997). *Multivariate models and dependence concepts.* Chapman & Hall, London, 1997.

[13] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001). *Modern Actuarial Risk Theory.* Kluwer Academic Publishers, Boston.

[14] Knott, M. and Smith, C.S. (1998). *On multivariate variance reduction.* Unpublished results.

[15] Knott, M. and Smith, C.S. (2006). *Choosing joint distributions so that the variance of the sum is small.* Journal of Multivariate Analysis 97, 1757-1765.

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

[17] Nelsen, R. and Úbeda-Flores, M. (2010). *Directional dependence in multivariate distributions.* Unpublished results.

[18] Rachev, S.T. and Rüschendorf, L. (1998) *Mass Transportation Problems.* Volume I: Theory. Springer.

[19] Rüschendorf, L. (1980). *Inequalities for the expectation of Λ-monotone functions.* Z. Wahrsch. 54, 341-554.

[20] Rüschendorf, L. (1982). *Random variables with maximum sums.* Adv. Appl. Probab. 14, 623-632.

[21] 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.

[22] Shaked, M., Shanthikumar, J.G. (2007). *Stochastic orders*. Springer Series in Statistics.

[23] Yang, J., Qi, Y. and Wang, R. (2009). *A class of multivariate copulas with bivariate Fréchet marginal copulas*. Insurance: Mathematics and Economics 45(1), 139-147.