Detecting complete and joint mixability

Giovanni Puccetti[1] and Ruodu Wang[2]

[1] *School of Economics and Management, University of Firenze, 50127 Firenze, Italy*
[2] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada*

September 25, 2014

Abstract

We introduce the Mixability Detection Procedure (MDP) to check whether a set of d distribution functions is jointly mixable at a given confidence level. The procedure is based on newly established results regarding the convergence rate of the minimal variance problem within the class of joint distribution functions with given marginals. The MDP is able to detect the complete mixability of an arbitrary set of distributions, even in those cases not covered by theoretical results. Stress-tests against borderline cases show that the MDP is fast and reliable.

Key words: Joint mixability, Complete mixability, Degree of mixability, Variance reduction, Rearrangement algorithm.
AMS 2010 Subject Classification: 60E05, 65C50, 65C60, 62E17.

1 Introduction and motivation of the paper

The definition of complete mixability for a univariate distribution has first been given in Wang and Wang (2011) and then extended to an arbitrary set of distributions in Wang et al. (2013).

Definition 1.1 (Wang and Wang (2011))**.** A univariate distribution function F is called *d-completely mixable* (d-CM) if there exist d random variables $X_1, \dots, X_d$ identically distributed as F having constant sum a.s., that is satisfying

$$\mathbb{P}(X_1 + \cdots + X_d = dc) = 1,$$

for some $c \in \mathbb{R}$.

Definition 1.2 (Wang et al. (2013))**.** The d univariate distribution functions $F_1, \dots, F_d$ are said to be *jointly mixable* (JM) if there exist d random variables $X_1, \dots, X_d$ such that $X_j \overset{\mathrm{d}}{=} F_j, 1 \le j \le d$, and

$$\mathbb{P}(X_1 + \cdots + X_d = C) = 1,$$

for some $C \in \mathbb{R}$.

It is straightforward that, if F in Definition 1.1 has finite first moment μ, then $c = \mu$, and if each F_j in Definition 1.2 has finite first moment μ_j, then $C = \sum_{j=1}^{d} \mu_j$. The concept of *risks with a constant sum* goes back to Gaffke and Rüschendorf (1981), where the complete mixability of a set of uniform distributions was showed. The same notion appears in Rüschendorf and Uckelmann (2002a), Müller and Stoyan (2002, Section 8.3.1) and Knott and Smith (2006) in the context of variance minimization or as the safest aggregate risk, with a focus on random variables.

The notions of complete and joint mixability have recently gathered a lot of interest since they are related to the existence of a least element with respect to convex order within the set

$$\mathfrak{S}(F_1, \dots, F_d) := \{X_1 + \cdots + X_d : X_j \overset{\mathrm{d}}{=} F_j, 1 \le j \le d\}$$

consisting of all sums of random variables with given marginal distributions $F_1, \dots, F_d$. In general the characterization of $\mathfrak{S}(F_1, \dots, F_d)$: is known to be an open question for $d \geq 2$ (see Bernard et al. (2014)); is equivalent to the study of joint mixability for $d \geq 3$, by simply observing that $S + C \in \mathfrak{S}(F_1, \dots, F_d)$ for some $C \in \mathbb{R}$ is equivalent to $F_1, \dots, F_d, \tilde{F}_S$ are JM, where $\tilde{F}_S$ is the distribution of $-S$. Recall that a random variable X is smaller than Y in convex order, denoted by $X \leq_{cx} Y$, if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for all convex functions f such that the expectations exist. When two random variables have the same mean, as within the set $\mathfrak{S}(F_1, \dots, F_d)$, convex order is equivalent to *increasing convex* order (also known as *stop-loss* order) as defined in Müller and Stoyan (2002).

Let U be a $\mathrm{U}(0,1)$ random variable. It is well known (see for instance Tchen (1980)) that the greatest element wrt convex order in $\mathfrak{S}(F_1, \dots, F_d)$ is given by the *comonotonic* sum $F_1^{-1}(U) + \dots + F_d^{-1}(U)$, where

$$F_j^{-1}(p) = \begin{cases} \inf\{x \in \mathbb{R} : F_j(x) > p\}, \text{if } p \in [0,1), \\ \inf\{x \in \mathbb{R} : F_j(x) \geq 1\}, \text{if } p = 1, \end{cases}$$

is the generalized inverse (or quantile function) of $F_j, 1 \leq j \leq d$; see Dhaene et al. (2002) for more details on the concept of comonotonicty and several related results. In fact, we have that

$$X_1 + \dots + X_d \leq_{cx} F_1^{-1}(U) + \dots + F_d^{-1}(U),$$

for any $X_j \stackrel{\mathrm{d}}{=} F_j, 1 \leq j \leq d$. When there are only two random variables, i.e. $d = 2$, the $\leq_{cx}$-least element in $\mathfrak{S}(F_1, \dots, F_d)$ is known to be the *countermonotonic* sum $F_1^{-1}(U) + F_2^{-1}(1-U)$, i.e.

$$F_1^{-1}(U) + F_2^{-1}(1-U) \leq_{cx} X_1 + X_2,$$

for any $X_1 \stackrel{\mathrm{d}}{=} F_1$ and $X_2 \stackrel{\mathrm{d}}{=} F_2$. When $d > 2$, the problem of determining the existence of a least element in $\mathfrak{S}(F_1, \dots, F_d)$ is much more complicated as the notion of a countermonotonic sum with given marginals cannot be generalized to higher dimensions; this was studied in Dall'Aglio (1972), and we refer to Bernard et al. (2014) and Cheung and Lo (2014) for recent discussions.

It is a trivial observation that if $F_1, \dots, F_d$ have finite means $\mu_1, \dots, \mu_d$ and are JM, the least element in $\mathfrak{S}(F_1, \dots, F_d)$ is given by $\mu_1 + \dots + \mu_d$, i.e.

$$\mu_1 + \dots + \mu_d \leq_{cx} X_1 + \dots + X_d,$$

for any $X_j \stackrel{\mathrm{d}}{=} F_j, 1 \leq j \leq d$. Existence of $\leq_{cx}$-least elements on sums and the corresponding conditions of complete/joint mixability are involved in a variety of optimization problems in the theory of optimal couplings, as for example:

(i) Assume that $F_1, \dots, F_d$ have finite first moment $\mu_1, \dots, \mu_d$ with $\mu = \sum_{j=1}^d \mu_j$. For a (strictly) convex function $f : \mathbb{R} \to \mathbb{R}$, we have by Jensen's inequality that

$$\inf\{\mathbb{E}\left[f(X_1 + \dots + X_d)\right]; X_j \stackrel{\mathrm{d}}{=} F_j,\ 1 \leq j \leq d\} \geq f(\mu), \tag{1.1}$$

and equality holds if (and only if) $F_1, \dots, F_d$ are JM.

(ii) Assume that $F_1, \dots, F_d$ are continuous and have finite first moment. Let $X_j \stackrel{\mathrm{d}}{=} F_j, 1 \leq j \leq d$, and, for $a \in [0,1]$, define the function

$$\Psi(a) = \sum_{j=1}^{d} \mathbb{E}[X_j | X_j \geq F_j^{-1}(a)].$$

For any $s \geq \mu$, we have

$$M(s) = \sup\{P(X_1 + \dots + X_d \geq s); X_j \sim F_j,\ 1 \leq j \leq d\} \leq 1 - \Psi^{-}(s), \tag{1.2}$$

where $\Psi^{-}(s) = \sup\{t \in [0,1] : \Psi(t) \leq s\}$ and the sup is attained if and only if the conditional distributions of $(X_j | X_j \geq F_j^{-1}(\Psi^{-}(s)))$ are JM.

Problems (1.1) and (1.2) have relevant applications in quantitative risk management, where they are needed to assess the model risk associated to the computation of capital charges for a portfolio of losses for regulatory issues. For instance, problem (1.1) is related to the computation of bounds on the expected value of a supermodular function (Wang and Wang (2011); Puccetti and Rüschendorf (2015)) and on the expected shortfall of a sum of random variables (Puccetti (2013)). When f in (1.1) is chosen as $f(x) = (x - \mu)^2$, (1.1) becomes a variance minimization problem, which is fundamental in variance reduction and simulation; see for example Glasserman (2004). Problem (1.2), as well as its inf version, is crucial to determine upper and lower sharp bounds on the Value-at-Risk (VaR), i.e. the quantile, of a sum of random variables; see Wang et al. (2013); Puccetti and Rüschendorf (2013). For a general introduction on the regulatory motivations of these problems within quantitative risk management, we refer to Embrechts et al. (2013).

In this paper we restrict to study complete and joint mixability for sets of distributions having a bounded support. Indeed, a one-sided distribution (e.g. $F^{-1}(0) > -\infty$ and $F^{-1}(1) = \infty$) cannot be completely mixable; see Proposition 2.1, (7) in Wang and Wang (2011). However, the existence of the convex order least element and the computation of sharp bounds on the distribution function/VaR of a sum heavily rely on joint mixability assumptions even when the fixed marginal components are one-sided. In fact, to obtain a convex order least element in $\mathfrak{S}(F_1, \dots, F_d)$ when the marginals are one-sided, one however needs the distributions to be CM/JM conditionally on a bounded interval; see Wang and Wang (2011) and Bernard et al. (2014) for more details. Furthermore, optimal couplings attaining the maximal value for the quantile/VaR of a sum show a completely/jointly mixable part for distributions of interest in quantitative risk management; see for instance the discussion and the figures in Puccetti and Rüschendorf (2013) and Embrechts et al. (2013).

In view of these applications, it would be of great interest to characterize the class of completely and jointly mixable distributions. So far in the literature, only partial characterizations of the class of completely mixable distributions are known, mainly in the form of sufficient conditions to complete mixability. As straightforward examples, Binomial, Uniform, Gaussian and Cauchy distributions are d-completely mixable for some dimensions d, see Proposition 2.3 in Wang and Wang (2011). We summarize some other characterizations that we will use in the remainder of the paper.

Proposition 1.1 (Rüschendorf and Uckelmann (2002b))**.** *Any continuous distribution function having a symmetric and unimodal density is d-CM, for any $d \geq 2$.*

Proposition 1.2 (Wang and Wang (2011))**.** *Suppose F is a distribution function on the real interval $[a, b]$, $a = F^{-1}(0)$ and $b = F^{-1}(1)$, having mean μ. A necessary condition for F to be d-CM is that*

$$a + (b - a)/d \leq \mu \leq b - (b - a)/d. \tag{1.3}$$

If F is also continuous with a monotone density on $[a, b]$, condition (1.3) is also sufficient.

Proposition 1.3 (Puccetti et al. (2012))**.** *Any continuous distribution on a bounded interval $[a, b]$ having a concave density is d-CM.*

Proposition 1.4 (Puccetti et al. (2013))**.** *Assume $d \geq 3$. Any continuous distribution function F on a bounded interval $[a, b]$, $a < b$, having a density f satisfying*

$$f(x) \geq \frac{3}{d(b - a)}, \quad \textit{for all } x \in [a, b],$$

is d-CM.

As a full characterization of completely mixable distribution is still out of reach, there are even less results concerning sufficient conditions for joint mixable distributions. The only available ones are given in the recent preprint Wang and Wang (2014). It was also mentioned in Wang et al. (2013) that it seems to be extremely difficult to find general sufficient conditions for JM distributions.

In this paper, we introduce a novel detection procedure in order to check whether a set of arbitrary distribution functions is jointly mixable (which includes completely mixability as a special case). First, in Section 2, we establish some new results concerning the convergence rate of the minimal variance problem within the class of joint distribution functions with given marginals. Based on these results, we introduce the Mixability Detection Procedure (MDP) in Section 3. The MDP can be used to detect complete and joint mixability in those cases not covered by theoretical results. In Section 4, numerical examples are provided via stress-tests against borderline cases, showing that the MDP is fast and reliable. Section 5 concludes the paper by proposing directions for further developments and improvements.

2 Theoretical results

The concept of joint mixability (see Definition 1.2) is directly connected with the minimization of the variance of a sum of random variables with given univariate marginals. For a set of d univariate distribution functions $F_1, \dots, F_d$, define the *minimal variance problem* as

$$\underline{\sigma}^2(F_1, \dots, F_d) := \inf \left\{ \mathrm{Var}\left(\sum_{j=1}^{d} X_j \right) : X_j \overset{\mathrm{d}}{=} F_j, 1 \le j \le d \right\}. \tag{2.1}$$

Note that, throughout the paper, we denote by $\mathrm{Var}(X)$ or by $\mathrm{Var}(F)$ the variance of a random variable $X \overset{\mathrm{d}}{=} F$ and by $\mathrm{var}(\boldsymbol{x})$ the variance of the components of a vector $\boldsymbol{x} \in \mathbb{R}^N$. We can rewrite problem (2.1) as

$$\underline{\sigma}^2(F_1, \dots, F_d) = \inf_{C \in \mathcal{C}_d} \left\{ \mathrm{Var}\left(\sum_{j=1}^{d} F_j^{-1}(U_j) \right) : (U_1, \dots, U_d) \overset{\mathrm{d}}{=} C \right\}, \tag{2.2}$$

where $\mathcal{C}_d$ denotes the set of d-dimensional *copulas*, i.e. of all d-variate distribution functions having $\mathrm{U}(0,1)$ marginals. A simple continuity-compactness argument given in Rüschendorf (1983) shows that the inf in (2.1) is attained and therefore the following immediate result holds.

Proposition 2.1. *$\underline{\sigma}^2(F_1, \dots, F_d) = 0$ holds if and only if $F_1, \dots, F_d$ are JM.*

We now show that if two sets of d distributions are close to each other, then also the solutions of the corresponding minimal variance problems are. To measure the distance between distributions, we use the L_2-Wasserstein distance, introduced in Dobrushin (1970), which has the following simple form for univariate distributions. We denote by $|X|_p$, $p \ge 1$, the L_p norm of a random variable X, i.e. $|X|_p = (\mathbb{E}[|X|^p])^{1/p}$.

Definition 2.1. The L_2-Wasserstein distance between two univariate distribution functions F and G is defined as

$$W_2(F, G) = \left(\int_0^1 (F^{-1}(p) - G^{-1}(p))^2 dp \right)^{1/2} = \left| F^{-1}(U) - G^{-1}(U) \right|_2,$$

where U is $\mathrm{U}(0,1)$ distributed.

Given two sets of distribution functions $F_1, \dots, F_d$ and $G_1, \dots, G_d$, we define

$$\underline{\sigma}_F^2 = \underline{\sigma}^2(F_1, \dots, F_d) \quad \text{and} \quad \underline{\sigma}_G^2 = \underline{\sigma}^2(G_1, \dots, G_d).$$

Theorem 2.2. *Let $F_1, \dots, F_d$ and $G_1, \dots, G_d$ be two sets of d bounded distribution functions, and write $c := \sum_{j=1}^{d} W_2(F_j, G_j)$. We have that*

$$|\underline{\sigma}_F - \underline{\sigma}_G| \le c. \tag{2.3}$$

Proof. Assume that $(U_1^*, \ldots, U_d^*) \overset{\mathrm{d}}{=} C^*$ attains $\underline{\sigma}_G^2$. Using the elementary fact that $\mathbb{E}[X-x]^2$ is minimised at $x = \mathbb{E}(X)$ for any random variable X, we can write

$$\begin{aligned}
\underline{\sigma}_F^2 &\leq \operatorname{Var}\left(\sum_{j=1}^d F_j^{-1}(U_j^*)\right) \\
&= \mathbb{E}\left[\sum_{j=1}^d F_j^{-1}(U_j^*) - \mathbb{E}\left(\sum_{j=1}^d F_j^{-1}(U_j^*)\right)\right]^2 \\
&\leq \mathbb{E}\left[\sum_{j=1}^d F_j^{-1}(U_j^*) - \mathbb{E}\left(\sum_{j=1}^d G_j^{-1}(U_j^*)\right)\right]^2 \\
&= \mathbb{E}\left[\sum_{j=1}^d \left(F_j^{-1}(U_j^*) - G_j^{-1}(U_j^*)\right) + \sum_{j=1}^d \left(G_j^{-1}(U_j^*) - \mathbb{E}[G_j^{-1}(U_j^*)]\right)\right]^2. \qquad (2.4)
\end{aligned}$$

Writing $F_+ = \sum_{j=1}^d \left(F_j^{-1}(U_j^*) - G_j^{-1}(U_j^*)\right)$ and $G_+ = \sum_{j=1}^d \left(G_j^{-1}(U_j^*) - \mathbb{E}[G_j^{-1}(U_j^*)]\right)$, from (2.4) we obtain that

$$\underline{\sigma}_F^2 \leq \mathbb{E}(F_+^2) + 2\mathbb{E}(F_+ G_+) + \mathbb{E}(G_+^2).$$

By the Minkowski inequality,

$$\mathbb{E}(F_+^2) = |F_+|_2^2 = \left|\sum_{j=1}^d \left(F_j^{-1}(U_j^*) - G_j^{-1}(U_j^*)\right)\right|_2^2 \leq \left(\sum_{j=1}^d \left|F_j^{-1}(U_j^*) - G_j^{-1}(U_j^*)\right|_2\right)^2 = c^2.$$

Note that $\mathbb{E}[G_+^2] = |G_+|_2^2 = \underline{\sigma}_G^2$. By Hölder's inequality,

$$\mathbb{E}(F_+ G_+) \leq |F_+ G_+|_1 \leq |F_+|_2 |G_+|_2 \leq c\, \underline{\sigma}_G$$

We finally obtain that

$$\underline{\sigma}_F^2 \leq c^2 + 2c\,\underline{\sigma}_G + \underline{\sigma}_G^2 = (c + \underline{\sigma}_G)^2, \qquad (2.5)$$

Exchanging the set of marginal distributions in the above proof, we obtain analogously that

$$\underline{\sigma}_G^2 \leq c^2 + 2c\,\underline{\sigma}_F + \underline{\sigma}_F^2 = (c + \underline{\sigma}_F)^2, \qquad (2.6)$$

The inequality $|\underline{\sigma}_F - \underline{\sigma}_G| \leq c$ directly follows from (2.5) and (2.6). □

A standard way of approximating an arbitrary distribution by a discrete one is given in the following definition.

Definition 2.2. Given a distribution function F and an integer N, we define the N-*discrete* distribution associated to F as

$$F_N(x) := \frac{1}{N} \sum_{r=0}^{N-1} \mathbf{1}_{[q_r, +\infty)}(x),$$

where the jump points $q_0, \ldots, q_{N-1}$ are the quantiles of F defined by $q_r := F^{-1}(r/N)$, $0 \leq r \leq N-1$.

The N-discrete distribution $F_N(x)$ is a discrete distributions giving probability mass $1/N$ to the (r/N)-quantiles of F. It immediately follows from its definition that

$$F_N^{-1}(p) = F^{-1}\left(\frac{\lfloor Np \rfloor}{N}\right), 0 \leq p < 1. \qquad (2.7)$$

The L_2-Wasserstein distance between a bounded, continuous distribution and its N-discrete counterpart decreases as $O(1/N)$, as next lemma shows.

Lemma 2.3. *Assume that F is a bounded continuous distribution on $[a,b]$ with strictly positive density f on $[a,b]$. Define*

$$a_F := \sqrt{\frac{1}{3}\int_a^b 1/f(x)dx} \quad \text{and} \quad c_F := \sqrt{\frac{1}{3}} \sup_{x\in(a,b)} 1/f(x) < \infty.$$

If F_N is the N-discrete distribution of F, for all $N \in \mathbb{N}$ we have that

$$W_2(F, F_N) \le c_F/N.$$

Moreover, as $N \to \infty$,

$$W_2(F, F_N) \sim a_F/N \le c_F/N.$$

Proof. Recalling (2.7) and using the mean value theorem (the positive density implies differentiability of the inverse function) we have, for any $p \in [0,1]$ in which $F^{-1}(p) \neq F_N^{-1}(p)$, that

$$F^{-1}(p) - F_N^{-1}(p) = F^{-1}(p) - F^{-1}\left(\frac{\lfloor Np\rfloor}{N}\right) = (F^{-1})'(\xi_p)\left(p - \frac{\lfloor Np\rfloor}{N}\right),$$

for some $\xi_p \in (\lfloor Np\rfloor/N, p)$. For some $\eta_i \in (\frac{i-1}{N}, \frac{i}{N})$, $1 \le i \le N$, we have

$$\begin{aligned} N^2 W_2^2(F,F_N) &= N^2\int_0^1 (F^{-1}(p) - F_N^{-1}(p))^2 dp \\ &= N^2 \sum_{i=1}^N \int_{(i-1)/N}^{i/N} \left(F^{-1}(p) - F^{-1}\left(\frac{i}{N}\right)\right)^2 dp \\ &= N^2 \sum_{i=1}^N \int_{(i-1)/N}^{i/N} \left((F^{-1})'(\eta_i)\left(p - \frac{i}{N}\right)\right)^2 dp \\ &= N^2 \sum_{i=1}^N \left((F^{-1})'(\eta_i)\right)^2 \int_{(i-1)/N}^{i/N} \left(p - \frac{i}{N}\right)^2 dp \\ &= N^2 \sum_{i=1}^N \left((F^{-1})'(\eta_i)\right)^2 \int_0^{1/N} x^2 dx = \frac{1}{3N}\sum_{i=1}^N \left((F^{-1})'(\eta_i)\right)^2. \end{aligned} \tag{2.8}$$

From (2.8), we obtain

$$N^2W_2^2(F,F_N) = \frac{1}{3N}\sum_{i=1}^N \left((F^{-1})'(\eta_i)\right)^2 \le \frac{1}{3}\sup_{1\le i\le n}\left((F^{-1})'(\eta_i)\right)^2 \le \frac{1}{3}\sup_{1\le i\le n}\left(\frac{1}{f(F^{-1}(\eta_i))}\right)^2 \le c_F^2,$$

thus

$$W_2(F, F_N) \le c_F/N.$$

From (2.8) we also obtain, as $N \to \infty$, that

$$\begin{aligned} N^2W_2^2(F,F_N) = \frac{1}{3N}\sum_{i=1}^N \left((F^{-1})'(\eta_i)\right)^2 &\to \frac{1}{3}\int_0^1 \left((F^{-1})'(x)\right)^2 dx \\ &= \frac{1}{3}\int_0^1 \left(f(F^{-1}(x))\right)^{-1} dF^{-1}(x) = \frac{1}{3}\int_a^b f(t)^{-1}dt = a_F^2, \end{aligned}$$

thus

$$W_2(F, F_N) \sim a_F/N.$$

Combining the first result $W_2(F,F_N) \le c_F/N$ with $W_2(F,F_N) \sim a_F/N$ we also obtain that $a_F \le c_F$. □

Lemma 2.3 and Theorem 2.2 directly imply the following theorem, which is the main theoretical result of our paper. Given the N-discrete distributions $F_{N,1}, \dots, F_{N,d}$ associated to $F_1, \dots, F_d$, we define

$$\underline{\sigma}_N^2 := \underline{\sigma}^2(F_{N,1}, \dots, F_{N,d}).$$

Theorem 2.4. *Let $F_1, \dots, F_d$ be a set of distribution functions. Assume that each F_j is a continuous distribution on the interval $[a_j, b_j]$, with a strictly positive density on $[a_j, b_j]$. Define the positive constants*

$$k := \sum_{j=1}^{d} a_{F_j} \quad \text{and} \quad K := \sum_{j=1}^{d} c_{F_j}.$$

Then, we have that

$$|\underline{\sigma}_F - \underline{\sigma}_N| \leq K/N, \tag{2.9}$$

and

$$-(K/N)^2 - 2(K/N)\,\underline{\sigma}_F \leq \underline{\sigma}_F^2 - \underline{\sigma}_N^2 \leq (K/N)^2 + 2(K/N)\,\underline{\sigma}_N. \tag{2.10}$$

Moreover, (2.9) and (2.10) hold asymptotically if K is replaced by k, i.e.

$$|\underline{\sigma}_F - \underline{\sigma}_N| \lesssim k/N, \tag{2.11}$$

and

$$-(k/N)^2 - 2(k/N)\,\underline{\sigma}_F \lesssim \underline{\sigma}_F^2 - \underline{\sigma}_N^2 \lesssim (k/N)^2 + 2(k/N)\,\underline{\sigma}_N. \tag{2.12}$$

Proof. The inequality (2.9) ((2.11)) follows by Theorem 2.2 and Lemma 2.3 applied with $G_j = F_{N,j}$ and $c = \sum_{j=1}^{d} c_{F_j}/N$ ($c = \sum_{j=1}^{d} a_{F_j}/N$). The inequalities in (2.10) and in (2.12) follow similarly from (2.5) and (2.6). □

Remark 2.1. Since $a_F \leq c_F$ for any F, we have that $k \leq K$. In the applications to follow in Section 3 and 4, we will use (2.11) and (2.12) because the use of k (instead of K) typically leads to a more accurate estimation of $|\underline{\sigma}_F - \underline{\sigma}_N|$. The values of N used are in the order of $N \geq 10^5$ and this give reasons to use the asymptotically sharper inequalities (2.11) and (2.12) instead of the strict inequalities (2.9) and (2.10). We will further remark on this choice in Remark 3.2.

Theorem 2.4 has a series of interesting implications.

Corollary 2.5. *Under the assumptions of Theorem 2.4, we have that*

(i)

$$\lim_{N\to\infty} \underline{\sigma}_N^2 = \underline{\sigma}_F^2,$$

(ii)

$$|\underline{\sigma}_N^2 - \underline{\sigma}_F^2| = \begin{cases} O(1/N^2) & \text{if } F_1, \dots, F_d \text{ are } d\text{-JM}, \\ O(1/N) & \text{otherwise}, \end{cases}$$

(iii) If $F_1, \dots, F_d$ are not JM, then $\underline{\sigma}_N > K/N$ (and hence $\underline{\sigma}_N > k/N$) for N sufficiently large.

Proof. (i) follows from (2.9) since K is a positive constant which does not depend on N. (ii) and (iii) follow directly from (2.10), (i) and Proposition 2.1.

3 A procedure to detect complete and joint mixability

Based on Corollary 2.5, we introduce a procedure to check whether a set of arbitrary distributions $F_1, \dots, F_d$ is jointly mixable. To this aim, we need to define a measure for the degree of mixability of a distribution or of a set of distribution functions. In the following, we assume that the distributions $F_1, \dots, F_d$ have finite second moments, and at least one of them is not degenerate (i.e. $\sum_{j=1}^{d} \mathrm{Var}(F_j) > 0$).

Definition 3.1. The *degree of mixability* of the set of d marginal distributions $F_1, \dots, F_d$ is defined as

$$\gamma(F_1, \dots, F_d) := \frac{\underline{\sigma}^2(F_1, \dots, F_d)}{\sum_{j=1}^{d} \mathrm{Var}(F_j)}.$$

In the homogeneous case in which $F_j = F,\ 1 \le j \le d$, we speak about the degree of d-mixability of the distribution F.

Definition 3.2. The *degree of d-mixability* of the distribution F is defined as

$$\gamma_d(F) := \gamma(F, \dots, F) = \frac{\underline{\sigma}^2(F, \dots, F)}{d\,\mathrm{Var}(F)}.$$

Definitions 3.1 and 3.2 are justified by the following properties.

Proposition 3.1. *The degree of mixability enjoys the following properties:*

(i) For any set of d distribution functions $F_1, \dots, F_d$, we have

$$\gamma(F_1, \dots, F_d) \in [0, 1].$$

(ii) $\gamma(F_1, \dots, F_d) = 0$ if and only if $F_1, \dots, F_d$ are JM.

(iii) $\gamma(F_1, \dots, F_d) = 1$ if and only if all the F_j's but one are degenerate.

(iv) Let $X_j \overset{\mathrm{d}}{=} F_j$ and G_j be the distribution of $a_j + bX_j, 1 \le j \le d$, for some $a_1, \dots, a_d, b \in \mathbb{R}$, $b \neq 0$. Then

$$\gamma(F_1, \dots, F_d) = \gamma(G_1, \dots, G_d).$$

Proof. (i) If $U_1^\Pi, \dots, U_d^\Pi$ are d independent $\mathrm{U}(0,1)$ random variables, we have

$$0 \le \underline{\sigma}^2(F_1, \dots, F_d) \le \mathrm{Var}\left(\sum_{j=1}^d F_j^{-1}(U_j^\Pi)\right) = \sum_{j=1}^d \mathrm{Var}\left(F_j^{-1}(U_j^\Pi)\right) = \sum_{j=1}^d \mathrm{Var}(F_j),$$

i.e. $0 \le \gamma(F_1, \dots, F_d) \le 1$. (ii) follows directly from Proposition 2.1. (iii) Suppose that all the F_j's but one are degenerate; say F_1 is the non-degenerate distribution. Let $X_j \sim F_j,\ j = 1, \dots, d$. It follows immediately that $\mathrm{Var}(X_1 + \cdots + X_d) = \mathrm{Var}(X_1) = \sum_{j=1}^d \mathrm{Var}(X_j)$, and $\gamma(F_1, \dots, F_d) = 1$. Vice versa, it is well known that for two distributions F_1 and F_2 with at least two points in their respective essential supports, the comonotonic sum $F_1^{-1}(U) + F_2^{-1}(U)$ has a strictly higher variance than the countermonotonic sum $F_1^{-1}(U) + F_2^{-1}(1-U)$; see for instance the proof of Theorem 2.5 in Puccetti and Scarsini (2010). Thus, if the variance of the sum is invariant over all possible bivariate distributions having fixed marginals, then at least one of the two has to be degenerate. The case $d > 2$ is similar. (iv) Assume $(U_1^*, \dots, U_d^*)$ attains $\underline{\sigma}^2(G_1, \dots, G_d)$. Using the fact that $G_j^{-1} = bF_j^{-1} + a_j,\ 1 \le j \le d$, we obtain that

$$\begin{aligned}\gamma(G_1, \dots, G_d) = \frac{\mathrm{Var}\left(\sum_{j=1}^d G_j^{-1}(U_j^*)\right)}{\sum_{j=1}^d \mathrm{Var}(a_j + bX_j)} &= \frac{\mathrm{Var}\left(\sum_{j=1}^d (bF_j^{-1}(U_j^*) + a_j)\right)}{\sum_{j=1}^d b^2 \mathrm{Var}(X_j)} \\ &= \frac{b^2 \mathrm{Var}\left(\sum_{j=1}^d F_j^{-1}(U_j^*)\right)}{\sum_{j=1}^d b^2 \mathrm{Var}(F_j)} \ge \gamma(F_1, \dots, F_d).\end{aligned}$$

$\gamma(F_1, \dots, F_d) \ge \gamma(G_1, \dots, G_d)$ follows similarly from the same argument, noticing the fact that F_i can also be written as a transformation of G_i of the same type. $\square$

Remark 3.1. Any norm $||\cdot||$ of $X - \mathbb{E}[X]$ is a possible candidate to replace $\mathrm{Var}(\cdot)$ in the definition of the degree of mixability. We chose the variance norm $\mathbb{E}[X - \mathbb{E}(X)]^2$ for: (i) obvious interpretation and computational ease; (ii) relevance in variance reduction; and (iii) that $f(x) = (x - \mu)^2$ is an analytical function which yields convenience in analysis.

The computation of the degree of mixability of a set of marginal distributions requires the computation of the quantity $\underline{\sigma}_F^2$. So far in the literature, the only analytical methods to compute $\underline{\sigma}_F^2$ work only in the homogeneous case when $F_j = F,\ 1 \leq j \leq d$, see Theorem 3.5 in Wang and Wang (2011). All these analytical methods require an a-priori knowledge of the complete mixability of F, hence are not useful to check if F is completely mixable nor they can be applied to the arbitrary marginals case.

However, if one is able to compute or at least to approximate numerically the quantity $\underline{\sigma}_N^2$, we introduce the *mixability detection procedure* to check if a set of distributions is jointly mixable. The procedure is based on the results stated in Theorem 2.4 and Corollary 2.5.

Mixability Detection Procedure (MDP)

Given the d distribution functions $F_1, \dots, F_d$, fix a (large) integer N and let $F_{N,1}, \dots, F_{N,d}$ be the corresponding N-discrete distributions. Compute the quantity $\underline{\sigma}_N$ and $k = \sum_{j=1}^d a_{F_j}$;

(A) if $\underline{\sigma}_N > k/N$, then $F_1, \dots, F_d$ are *not* jointly mixable;

(B) if $\underline{\sigma}_N < k/N$, then we have that

$$\gamma(F_1, \dots, F_d) \leq \hat{\gamma},$$

where

$$\hat{\gamma} = \frac{(\underline{\sigma}_N + k/N)^2}{\sum_{j=1}^d \mathrm{Var}(F_j)}.$$

In this latter case we say that $F_1, \dots, F_d$ are JM at the confidence level $\gamma = \hat{\gamma}$.

Definition 3.3. The d distributions $F_1, \dots, F_d$ are said to be jointly mixable (JM) at the confidence level $\hat{\gamma} \in (0,1)$ if

$$\gamma(F_1, \dots, F_d) \leq \hat{\gamma}.$$

The distribution F is said to be d-completely mixable (CM) at the confidence level $\hat{\gamma} \in (0,1)$ if

$$\gamma_d(F) \leq \hat{\gamma}.$$

Note that the outcome (A) in the above procedure is conclusive wrt joint mixability. Indeed, on the basis of Corollary 2.5 (iii), if $\underline{\sigma}_N > k/N$ the set of underlying distributions is not JM and a range for the degree of mixability is given by

$$\frac{(\underline{\sigma}_N - k/N)^2}{\sum_{j=1}^d \mathrm{Var}(F_j)} \leq \gamma(F_1, \dots, F_d) \leq \frac{(\underline{\sigma}_N + k/N)^2}{\sum_{j=1}^d \mathrm{Var}(F_j)}. \tag{3.1}$$

In the case (B), however, one cannot state that $F_1, \dots, F_d$ are JM as the condition $\underline{\sigma}_N \leq k/N$ could be violated at a larger N. In practice, one should fix a confidence level $\gamma \in (0,1)$ and iterate the procedure above for increasing values of N until either the condition $\underline{\sigma}_N \leq k/N$ is violated or the underlying distributions are found to be JM at the given level of confidence. Of course, this latter case does not imply perfect joint mixability (i.e. $\gamma(F_1, \dots, F_d) = 0$). From Definition 3.3 it follows that a set of distributions is JM if and only if it is JM at *any* positive confidence level. However, the level γ should represent a threshold under which the set of distributions is regarded as indistinguishable from a truly JM one. If computational resources allow ($N \to \infty$), our procedure will eventually be able to distinguish any set of non-JM distributions to a JM one.

Remark 3.2. The MDP can analogously be defined with the constant k replaced by $K = \sum_{j=1}^d c_{F_j}$. We find in practice that the values of k are much smaller and give better estimation of $|\underline{\sigma}_F - \underline{\sigma}_N|$.

4 Numerical verifications and stress-testing

To the best of our knowledge, in order to compute $\underline{\sigma}^2(F_{N,1},\dots,F_{N,d})$ for a set of arbitrary distribution functions $F_1,\dots,F_d$, the only method available in the literature so far, which handles large d at a reasonable speed, is the Rearrangement Algorithm (RA) first presented in Puccetti and Rüschendorf (2012). The RA is a numerical procedure based on the iterative rearrangements of the column of a matrix containing the support of the $F_{N,j}$. It has been recently shown to find very good approximations of sharp lower and upper bounds on the expected value of a supermodular function of d random variables having fixed marginal distributions; see Puccetti and Rüschendorf (2015). As the minimal variance problem belongs to the domain of application of the RA, we briefly describes below how it can be used for the computation of $\underline{\sigma}_N^2$.

Rearrangement Algorithm (RA) to compute $\underline{\sigma}^2(F_{N,1},\dots,F_{N,d})$.

1. Fix an integer N and a set of d distribution functions $F_1,\dots,F_d$.
2. Define the $N\times d$ matrix $\boldsymbol{X}=(x_{i,j})$ as
$$x_{i,j}=F_j^{-1}\left(\frac{i-1}{N}\right),\ 1\le i\le N,\ 1\le j\le d. \tag{4.1}$$
3. Permute randomly the elements in each column of $\boldsymbol{X}$.
4. Iteratively rearrange the j-th column of $\boldsymbol{X}$ so that it becomes oppositely ordered to the sum of the other columns, for $1\le j\le d$.
5. Repeat Step 4. until no further changes are possible. A matrix $\boldsymbol{X}^*$ is found. Let $\boldsymbol{y}^*\in\mathbb{R}^N$ be the vector having as components the componentwise sums of each row of $\boldsymbol{X}^*$, i.e.
$$\boldsymbol{y}_i^*:=\sum_{j=1}^d x_{i,j}^*,\quad 1\le i\le N.$$
6. Define $s_N=\mathrm{var}(\boldsymbol{y}^*)$. In practice we find that
$$\underline{\sigma}^2(F_{N,1},\dots,F_{N,d})\le s_N \tag{4.2}$$
and
$$s_N\overset{N\to\infty}{\simeq}\underline{\sigma}^2(F_{N,1},\dots,F_{N,d}). \tag{4.3}$$

Remark 4.1. We discuss some points regarding the RA applied to our framework. For any further details, we refer the interested reader to the papers Puccetti and Rüschendorf (2012) and Puccetti and Rüschendorf (2015).

(i) There does not exist an analytical proof that the limit result (4.3) holds for all initial configurations of the algorithm. However, non-convergence seems to be confined to pathological cases, i.e. to special choices of the starting matrix of the algorithm; see Remark 6.2 in Embrechts et al. (2013). Using the randomization Step 3., we found the algorithm to provide excellent approximations with moderately large values of N for all marginal distributions we use in the applicative section below. Moreover, there is a growing literature in the field of quantitative risk management which uses the RA for a variety of optimization purposes and the algorithm seems always to provide excellent results. We refer for instance to the papers Embrechts et al. (2013); Bernard et al. (2014); Puccetti (2013). However, a full proof of the convergence of the RA, along with corresponding regularity conditions, remains an open problem.

(ii) The bound in (4.2) holds true for any N. Thus, by using the RA with the MDP in Section 3 one can always find a confidence level at which a set of arbitrary distribution functions is JM. More care has to be taken when the condition in (B) is satisfied. In this case the study of the convergence rate of $\underline{\sigma}_N^2$ can be helpful, as an extra tool to conclude whether a set of distributions is JM.

(iii) As a numerical algorithm, the RA can be used with any type of marginal distributions. The computation time needed for one iteration of the algorithm increases linearly both in N and in d and is reported in our applications below. For a given d, at some (very large) N memory constraint prohibits the application of the algorithm. In practice one should choose N as a good compromise between computation time needed to get an estimate of $\underline{\sigma}_N^2$ and the (low) level of confidence requested for joint mixability.

In the remainder of this section, we implement the MDP introduced in Section 3 in connection with the RA in order to check whether sets of distribution functions of interest are jointly mixable. In order to test the reliability of our procedure, we first treat the homogeneous case $F_j = F,\ 1 \le j \le d$, in those cases where we know from theory if the underlying df is CM or not.

For illustrative reasons, in what follows we focus on the study of distributions of interest in quantitative risk management. Although typical distributions used in risk management are one-sided, the detection of the joint mixability in bounded intervals is crucial to finding the solutions of a variety of problems involving unbounded distributions; see our introductory section and the references therein. Even if the notions of complete/joint mixability arose within this field, the methodology described in this paper is no doubt applicable to a broader context. In the remainder of this section, all estimates for $\underline{\sigma}_N^2$ are obtained via the RA. We remark that, our procedure is open to implement with any algorithm which calculates $\underline{\sigma}^2(F_{N,1}, \ldots, F_{N,d})$, and not limited to the RA.

Homogeneous case I: the Pareto distribution

In applications to quantitative risk management such as the ones described for instance in Bernard et al. (2014), one typically needs to check the complete/joint mixability of conditional distributions over some bounded interval. For $X \overset{\mathrm{d}}{=} F$ and $a, b \in \mathbb{R}$ with $a \le b$, we denote by $F^{[a,b]}$ the conditional distribution of $(X|X \in [a, b])$.

The Pareto distribution $\mathrm{Pa}_\theta(x) = 1-(1+x)^{-\theta}, x \ge 0, \theta > 0$ is continuous with a decreasing density over its entire support. From Proposition 1.2, $\mathrm{Pa}_\theta^{[a,b]}$ is d-CM if and only if

$$\mu_\theta^{[a,b]} \ge a + (b - a)/d,$$

where $\mu_\theta^{[a,b]}$ denotes the first moment of $\mathrm{Pa}_\theta^{[a,b]}$.

As a first example we set $d = 3$ and $F_j = \mathrm{Pa}_2^{[0,b]},\ 1 \le j \le 3$. In this case, $\mathrm{Pa}_2^{[0,b]}$ is known to be 3-CM if and only if

$$0 \le b \le 1,$$

while it is not 3-CM for $b > 1$. Table 1 collects the results from the application of the MDP for values of b around the borderline value $b = 1$. At the discretization value $N = 10^5$, the MDP detects complete mixability at all values $b < 1$, at very low confidence levels in the order of magnitude of 10^{-9}. Moreover, the procedure also perfectly detects no complete mixability for values $b \ge 1.003$.

A first warning from Table 1 is to be raised by observing the fact that the $\mathrm{Pa}_2^{[0,1.002]}$ is regarded to be 3-CM at the level $6.83e-09$. We strongly remark that this assertion is, by Definition 3.3, not in contrast with the fact that the distribution is not perfectly CM in this case. It is however clear from Table 1 that the borderline value is situated below the threshold $b = 1.003$. At this point one may perform a more sensitive (and more time consuming) analysis using the procedure at a higher value of N. The results for $N = 10^6$ are reported in Table 2. The cases $b = 1.002$ and $b = 1.001$ are resolved against perfect complete mixability while at the value $b = 1.0005$ the distribution is observed to be 3-CM at level 10^{-11} (which, again, is correct). Of course, borderline cases like this cannot be avoided for distributions so close to perfect mixability to be indistinguishable by our (any) numerical procedure. If needed, one may increase the value of N to obtain higher sensitivity. However, in these borderline cases, a qualitative analysis of the convergence rate of $\underline{\sigma}_N^2$ may be conclusive. By Corollary 2.5 (ii), the sequence $\underline{\sigma}_N^2$ converges to zero as $O(1/N^2)$ if and only if the underlying distribution are JM.

In Figure 1 we plot the values of $\underline{\sigma}_N^2\, N^2$ in the Pareto homogeneous case described above for $b = 1$ (left) and $b = 1.0005$ (right). It is immediately clear that this latter case is not CM. The graphical tool can be very useful to detect joint mixability in borderline cases, but does not provide ranges for the degree of mixability like the MDP does. Finally note that a plot of the convergence rate is based on the iteration of the procedure at different values of N, thus it might be very time consuming.

$N = 10^5$	theoretical	observed	range for $\gamma_3(F)$
$b = 0.995$ (3.1 sec.)	CM	CM at level $3.45e-09$	$[0, 3.45e-09]$
$b = 0.996$ (3.2 sec.)	CM	CM at level $3.46e-09$	$[0, 3.46e-09]$
$b = 0.997$ (3.3 sec.)	CM	CM at level $3.46e-09$	$[0, 3.46e-09]$
$b = 0.998$ (3.3 sec.)	CM	CM at level $3.46e-09$	$[0, 3.46e-09]$
$b = 0.999$ (3.3 sec.)	CM	CM at level $3.47e-09$	$[0, 3.46e-09]$
$b = 1.000$ (3.3 sec.)	CM	CM at level $3.47e-09$	$[0, 3.47e-09]$
$b = 1.001$ (3.6 sec.)	NOT	CM at level $4.16e-09$	$[0, 4.16e-09]$
$b = 1.002$ (3.7 sec.)	NOT	CM at level $6.83e-09$	$[0, 6.83e-09]$
$b = 1.003$ (3.7 sec.)	NOT	NOT	$[2.07e-11, 1.16e-08]$
$b = 1.004$ (3.7 sec.)	NOT	NOT	$[1.18e-09, 1.89e-08]$
$b = 1.005$ (3.8 sec.)	NOT	NOT	$[4.67e-09, 2.94e-08]$

Table 1: Detection results for 3-complete mixability for the distribution $\mathrm{Pa}_2^{[0,b]}$, for some values of b of interest. The average computation times are estimated over 100 identical runs of the algorithm.

$N = 10^6$	theoretical	observed	range for $\gamma_3(F)$
$b = 0.9995$ (91 sec.)	CM	CM at level $3.47e-11$	$[0, 3.47e-11]$
$b = 1.0000$ (95 sec.)	CM	CM at level $3.48e-11$	$[0, 3.48e-11]$
$b = 1.0005$ (100 sec.)	NOT	CM at level $8.15e-11$	$[0, 8.15e-11]$
$b = 1.0010$ (97 sec.)	NOT	NOT	$[3.18e-11, 2.54e-10]$
$b = 1.0015$ (96 sec.)	NOT	NOT	$[2.14e-10, 6.22e-10]$
$b = 1.0020$ (101 sec.)	NOT	NOT	$[6.40e-10, 1.26e-09]$

Table 2: The same as Table 1 with a number of discretization points set at $N = 10^6$.

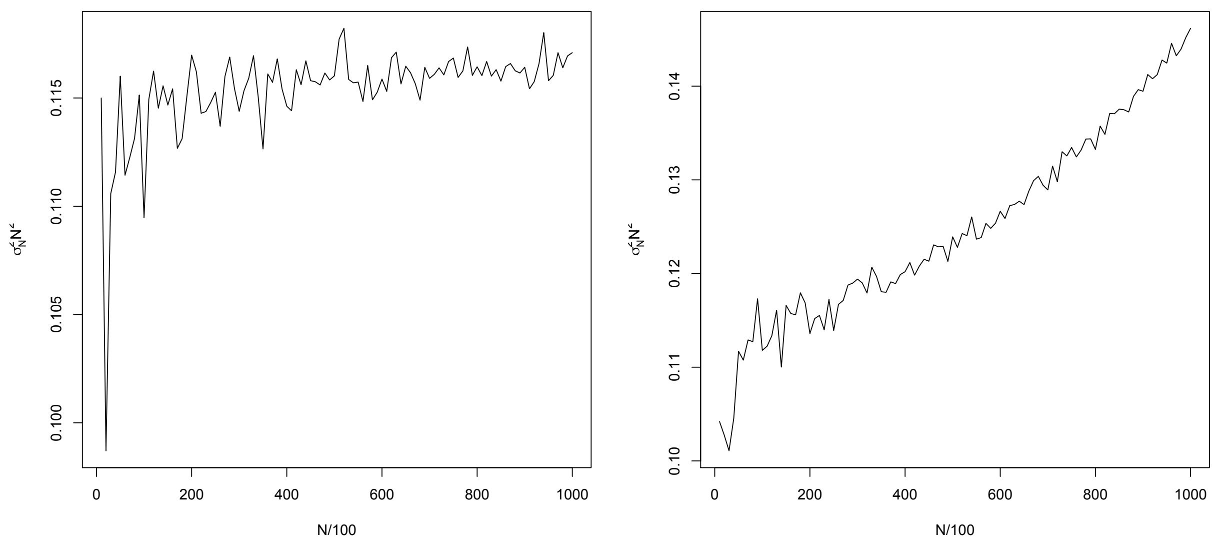

Figure 1: Plot of the function $\underline{\sigma}_N^2(F,F,F)N^2$ versus the number of discretization points N, when $F = \mathrm{Pa}_2^{[0,b]}$ with $b = 1$ (left) and $b = 1.0005$ (right).

Homogeneous case II: the Lognormal distribution

Denote by $\mathrm{LogN}_{\mu,\sigma}$ the distribution of the random variable $X = e^Y$ with $Y \stackrel{\mathrm{d}}{=} N(\mu, \sigma^2)$. As a second example we set $d = 3$, $F_j = \mathrm{LogN}_{2,1}^{[0.2,b]}, 1 \leq j \leq 3$. For $b > 2.7183$, it is easy to see that the density of the $\mathrm{LogN}_{2,1}^{[0.2,b]}$ distribution is not monotone nor concave, and does not satisfy the condition given in Proposition 1.4. As a consequence, the 3-mixability of $\mathrm{LogN}_{2,1}^{[0.2,b]}$ is not covered by any known theoretical results. By Proposition 1.2 we can only state that for $b > \hat{b} \simeq 22.65$ the moderate mean condition (1.3) is not satisfied and $\mathrm{LogN}_{2,1}^{[0.2,b]}$ is not 3-CM. Table 3 collects the results from the application of the detection procedure for values of b around $\hat{b}$. At a discretization value $N = 10^5$, the procedure detects no perfect mixability at all values $b > \hat{b}$, but also at some lower thresholds. This indicates that the moderate mean condition is not sufficient for a distribution with non-monotone density to be completely mixable. The borderline value seems to be smaller than the threshold $b = 21.8$. A more sensitive analysis at the level $N = 10^6$ indicates that the distribution is no more 3-CM at some threshold between $b = 21.55$ and $b = 21.6$. The graphical analysis of the convergence rate in Figure 2 confirms that for $b = 21.55$ we have complete mixability, while for $b = 21.6$ the distribution is not completely mixable.

$N = 10^5$	theoretical	observed	range for $\gamma_3(F)$
$b = 21.2$ (4.5 sec.)	UNKNOWN	CM at level $4.21e-09$	$[0, 4.21e-09]$
$b = 21.5$ (7.9 sec.)	UNKNOWN	CM at level $4.58e-09$	$[0, 1.31e-06]$
$b = 21.8$ (6.8 sec.)	UNKNOWN	NOT	$[5.00e-06, 5.51e-06]$
$b = 22.1$ (6.2 sec.)	UNKNOWN	NOT	$[3.35e-05, 3.48e-05]$
$b = 22.4$ (6.0 sec.)	UNKNOWN	NOT	$[9.04e-05, 9.26e-05]$
$b = 22.7$ (5.8 sec.)	NOT	NOT	$[1.78e-04, 1.81e-04]$

Table 3: Detection results for 3-complete mixability for the distribution $\mathrm{LogN}_{2,1}^{[0.2,b]}$. The average computation times are estimated over 100 identical runs of the algorithm.

$N = 10^6$	theoretical	observed	range for $\gamma_3(F)$
$b = 21.45$ (175 sec.)	UNKNOWN	CM at level $4.47e-11$	$[0, 4.47e-11]$
$b = 21.50$ (212 sec.)	UNKNOWN	CM at level $4.59e-11$	$[0, 4.59e-11]$
$b = 21.55$ (298 sec.)	UNKNOWN	CM at level $4.82e-11$	$[0, 4.82e-11]$
$b = 21.60$ (419 sec.)	UNKNOWN	NOT	$[4.46e-09, 6.08e-09]$
$b = 21.65$ (223 sec.)	UNKNOWN	NOT	$[3.72e-07, 3.86e-07]$
$b = 21.70$ (209 sec.)	UNKNOWN	NOT	$[1.35e-06, 1.38e-06]$

Table 4: The same as Table 3 with a number of discretization points set at $N = 10^6$.

It is also interesting to perform a comparative study on the same distribution with different values of d. As an example, we study the distribution $\mathrm{LogN}_{2,1}^{[0.2,30]}$. By the discussion above, this distribution is not 3-CM. However, using Proposition 1.4 we can state that it is d-CM for any $d \geq 35$. The application of the detection procedure and the analysis of the convergence rate however indicate that the distribution is d-CM for all values $d \geq 4$. We report in Table 5 the analysis for dimensions d of interest.

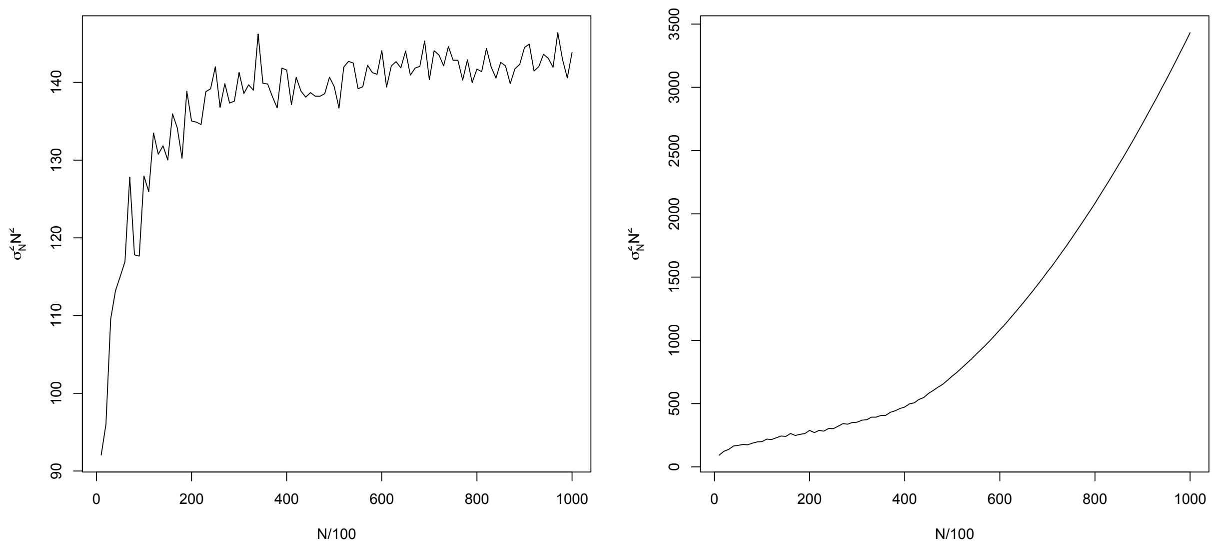

Figure 2: Plot of the function $\underline{\sigma}_N^2(F,F,F)N^2$ versus the number of discretization points N, when $F = \mathrm{LogN}_{2,1}^{[0.2,b]}$ with $b = 21.55$ (left) and $b = 21.6$ (right).

$N = 10^5$	theoretical	observed	range for $\gamma_d(F)$
$d = 3$ (4.3 sec.)	NOT	NOT	$[1.23e-02, 1.23e-02]$
$d = 4$ (2.8 sec.)	UNKNOWN	CM at level $[6.77e-09]$	$[0, 6.77e-09]$
$d = 5$ (3.3 sec.)	UNKNOWN	CM at level $[8.21e-09]$	$[0, 8.21e-09]$
$d = 6$ (3.6 sec.)	UNKNOWN	CM at level $[9.68e-09]$	$[0, 9.68e-09]$
$d = 10$ (6.3 sec.)	UNKNOWN	CM at level $[1.57e-08]$	$[0, 1.57e-08]$
$d = 20$ (14 sec.)	UNKNOWN	CM at level $[3.08e-08]$	$[0, 3.08e-08]$

Table 5: Detection results for d-complete mixability for the distribution $\mathrm{LogN}_{2,1}^{[0.2,30]}$. The average computation times are estimated over 100 identical runs of the algorithm.

Inhomogeneous case: mixed marginals with monotone densities

As a third example we set an inhomogeneous framework with monotone densities. It was shown in Wang and Wang (2014) that d continuous distributions $F_1, \dots, F_d$ with decreasing densities on their respective supports are JM if and only if the following extended moderate mean condition is satisfied:

$$\sum_{j=1}^{d} a_j + \max_{j=1,\dots,d}(b_j - a_j) \leq \sum_{j=1}^{d} \mu_j \leq \sum_{j=1}^{d} b_j - \max_{j=1,\dots,d}(b_j - a_j), \tag{4.4}$$

where the interval $[a_j, b_j]$ and μ_j represent the essential support and, respectively, the first moment of F_j, $1 \leq j \leq d$.

In the case $d = 3$, $F_1 = \mathrm{Pa}_2^{[0,2]}$, $F_2 = \Phi^{[0,2]}$ (Φ is the distribution of a standard Gaussian random variable), $F_3 = \mathrm{Exp}_1^{[0,2]}$ and $F_4 = \mathrm{LogN}_{2,1}^{[3,b]}$, it is easy to compute that (4.4) is satisfied for $b \in [\underline{b}, \overline{b}]$, with

$$\underline{b} \simeq 3.180719 \quad \text{and} \quad \overline{b} \simeq 6.582307.$$

Tables 6 and 7 collect the results from the application of the detection procedure for values of b around the borderline values above. The analyses of the convergence rates of $\underline{\sigma}_N^2$ in Figures 3 and 4 are again conclusive. For this example, observations are coherent with the fact that condition (4.4) is sufficient to guarantee joint mixability of distributions with decreasing densities. This numerical example was obtained earlier than the proof stated in Wang and Wang (2014) and actually shows that the detection procedure can be of invaluable help to the future developments of sufficient conditions for complete and joint mixability. Moreover, we remark again that the MDP is the only tool available to check the joint mixability of a set of *arbitrary* distributions.

5 Conclusions and future developments

In this paper, we introduce a novel procedure, called the Mixability Detection Procedure (MDP), to check whether a set of d distribution functions is jointly mixable at a given confidence level. The MDP is based on newly established results regarding the convergence rate of the minimal variance problem within the class of joint distribution functions with given marginals.

The application of the MDP to a given set of distributions has two possible outcomes: either the set of distributions is observed to be JM at some given confidence level, or non-perfect mixability is detected and a deterministic range for the so-called degree of mixability of the set of distributions is given. The same procedure can be analogously used to detect complete mixability in the homogeneous case where all the distributions are identical.

The application of the procedure needs a numerical method to evaluate the minimal variance of a sum of random variables with fixed marginal distributions. To this aim, in this paper we use the so-called rearrangement algorithm, with excellent results. Stress-tests against borderline cases show that the MDP is fast and reliable: we never experienced cases in which a mixable distributions has been observed in numerics to be not mixable. In borderline cases, we also provide an extra qualitative analysis which is typically conclusive.

So far in the literature, the procedure introduced in this paper represents the first method to check the complete or joint mixability of an *arbitrary* set of bounded distribution functions, which is an essential requirement in various problems of optimal coupling, mass transportations and with respect the existence of the convex order least element within the class of all sums of random variables with fixed marginal distributions.

In view of the sufficient condition in Proposition 1.4, the detection of mixability is usually performed for small dimensions d. However, the RA used in connection with our procedure is capable of dealing also with high dimensional sets of distributions, even for values of $d \simeq 1000$. Future research on sufficient conditions to complete and joint mixability will no doubt benefit and take inspiration from the numerical findings provided by the detection procedure. The analysis of computation times seems to suggest that the algorithm requires a larger number of iterations in borderline cases. This should be taken into account in the search for a general proof of convergence of the RA.

$N = 10^5$	theoretical	observed	range for $\gamma(F_1, \ldots, F_4)$
$b = 3.14$ (6.3 sec.)	NOT (conjectured)	NOT	$[2.18e-06, 2.61e-06]$
$b = 3.15$ (6.0 sec.)	NOT (conjectured)	NOT	$[8.88e-07, 1.17e-06]$
$b = 3.16$ (5.8 sec.)	NOT (conjectured)	NOT	$[2.40e-07, 3.96e-07]$
$b = 3.17$ (5.4 sec.)	NOT (conjectured)	NOT	$[1.88e-08, 7.69e-08]$
$b = 3.18$ (5.0 sec.)	NOT (conjectured)	JM at level $5.67e-09$	$[0, 5.67e-09]$
$b = 3.19$ (4.1 sec.)	JM (conjectured)	JM at level $5.49e-09$	$[0, 5.49e-09]$
$b = 3.20$ (4.2 sec.)	JM (conjectured)	JM at level $5.59e-09$	$[0, 5.59e-09]$
$b = 6.57$ (4.4 sec.)	JM (conjectured)	JM at level $5.49e-09$	$[0, 5.49e-09]$
$b = 6.58$ (4.8 sec.)	JM (conjectured)	JM at level $5.49e-09$	$[0, 5.50e-09]$
$b = 6.59$ (5.4 sec.)	NOT (conjectured)	NOT	$[4.41e-10, 2.53e-08]$
$b = 6.60$ (5.4 sec.)	NOT (conjectured)	NOT	$[5.90e-08, 1.45e-07]$
$b = 6.61$ (5.5 sec.)	NOT (conjectured)	NOT	$[2.91e-07, 4.59e-07]$
$b = 6.62$ (5.5 sec.)	NOT (conjectured)	NOT	$[7.98e-07, 1.06e-06]$

Table 6: Detection results for joint mixability of the $d = 4$ distributions $F_1 = \mathrm{Pa}_2^{[0,2]}$, $F_2 = \Phi^{[0,2]}$, $F_3 = \mathrm{Exp}_1^{[0,2]}$ and $F_4 = \mathrm{LogN}_{2,1}^{[3,b]}$. The average computation times are estimated over 100 identical runs of the algorithm.

$N = 10^6$	theoretical	observed	range for $\gamma(F_1, \ldots, F_4)$
$b = 3.179$ (147 sec.)	NOT (conjectured)	NOT	$[3.95e-11, 4.14e-10]$
$b = 3.180$ (139 sec.)	NOT (conjectured)	JM at level $1.13e-10$	$[0, 1.13e-10]$
$b = 3.181$ (135 sec.)	JM (conjectured)	JM at level $5.52e-11$	$[0, 5.53e-11]$
$b = 6.582$ (137 sec.)	JM (conjectured)	JM at level $5.50e-11$	$[0, 5.50e-11]$
$b = 6.583$ (144 sec.)	NOT (conjectured)	JM at level $8.87e-11$	$[0, 8.87e-11]$
$b = 6.584$ (140 sec.)	NOT (conjectured)	NOT	$[5.75e-12, 2.63e-10]$

Table 7: The same as Table 6 with a number of discretization points set at $N = 10^6$.

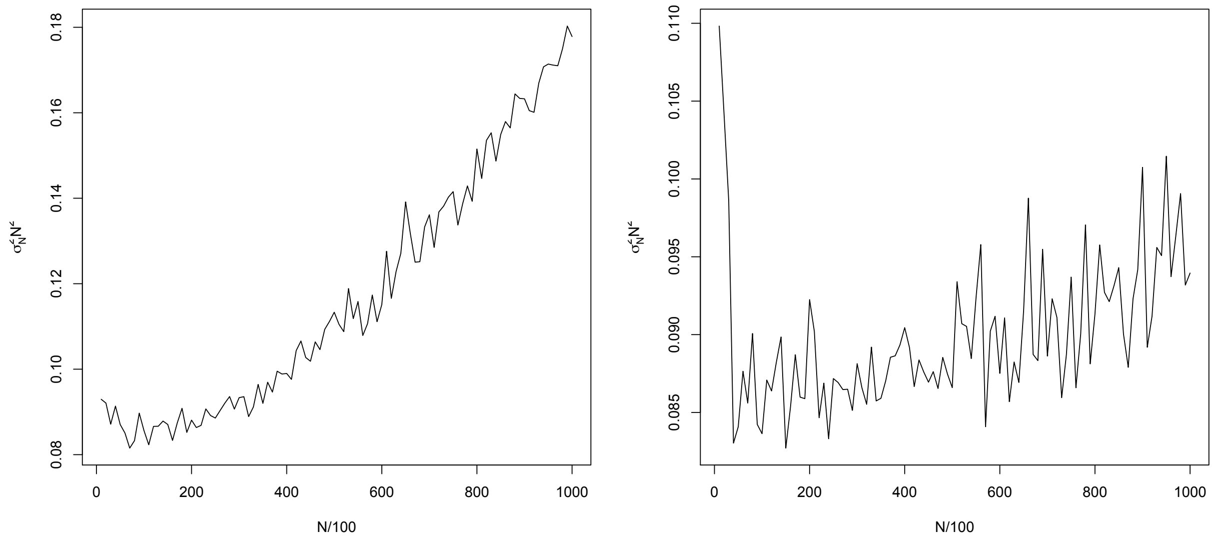

Figure 3: Plot of the function $\underline{\sigma}_N^2(F,F,F)N^2$ in the case $F_1 = \mathrm{Pa}_2^{[0,2]}$, $F_2 = \Phi^{[0,2]}$, $F_3 = \mathrm{Exp}_1^{[0,2]}$ and $F_4 = \mathrm{LogN}_{2,1}^{[3,b]}$ with $b = 3.180$ (left) and $b = 3.181$ (right).

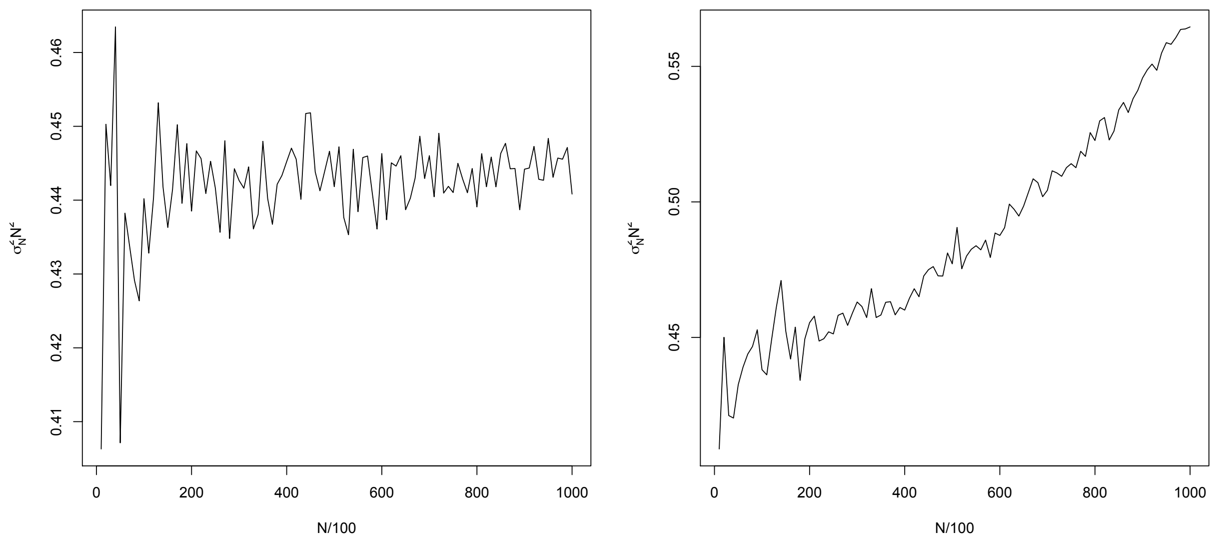

Figure 4: The same as Figure 3 with $b = 6.582$ (left) and $b = 6.583$ (right).

Finally, we remark that our procedure is an open structure in the sense that it can be used also with more precise methods than the rearrangement algorithm and more accurate ranges for the degree of mixability, if and when they will be available in the future.

Acknowledgement

The authors would like to thank an anonymous referee for valuable comments on an earlier version of this paper. Giovanni Puccetti acknowledges grant under the call PRIN 2010-2011 from MIUR within the project *Robust decision making in markets and organisations*. Ruodu Wang acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

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

Bernard, C., L. Rüschendorf, and S. Vanduffel (2014). Value-at-risk bounds with variance constraints. Preprint, available at http://ssrn.com/abstract=2342068.

Cheung, K. C. and A. Lo (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance Math. Econ. 55*, 180 – 190.

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica 9*, 131–150.

Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom. 31*(1), 3–33.

Dobrushin, R. (1970). Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl. 15*(3), 458–486.

Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and VaR aggregation. *J. Bank. Financ. 37*(8), 2750 – 2764.

Gaffke, N. and L. Rüschendorf (1981). On a class of extremal problems in statistics. *Math. Operationsforsch. Statist. Ser. Optim. 12*(1), 123–135.

Glasserman, P. (2004). *Monte Carlo methods in financial engineering.* New York: Springer-Verlag.

Knott, M. and C. S. Smith (2006). Choosing joint distributions so that the variance of the sum is small. *J. Multivariate Anal. 97*, 1757–1765.

Müller, A. and D. Stoyan (2002). *Comparison Methods for Stochastic Models and Risks.* Chichester: Wiley.

Puccetti, G. (2013). Sharp bounds on the expected shortfall for a sum of dependent random variables. *Stat. Probabil. Lett. 83*(4), 1227–1232.

Puccetti, G. and L. Rüschendorf (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comput. Appl. Math. 236*(7), 1833–1840.

Puccetti, G. and L. Rüschendorf (2013). Sharp bounds for sums of dependent risks. *J. Appl. Probab. 50*(1), 42–53.

Puccetti, G. and L. Rüschendorf (2015). Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals. Fortchcoming in *Commun. Stat. Simulat.*

Puccetti, G. and M. Scarsini (2010). Multivariate comonotonicity. *J. Multivariate Anal. 101*(1), 291 – 304.

Puccetti, G., B. Wang, and R. Wang (2012). Advances in complete mixability. *J. Appl. Probab. 49*(2), 430–440.

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

Rüschendorf, L. (1983). Solution of a statistical optimization problem by rearrangement methods. *Metrika 30*, 55–61.

Rüschendorf, L. and L. Uckelmann (2002a). On the n-coupling problem. *J. Multivariate Anal. 81*(2), 242–258.

Rüschendorf, L. and L. Uckelmann (2002b). Variance minimization and random variables with constant sum. In *Distributions with given marginals and statistical modelling*, pp. 211–222. Dordrecht: Kluwer Acad. Publ.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Ann. Probab. 8*(4), 814–827.

Wang, B. and R. Wang (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal. 102*(10), 1344–1360.

Wang, B. and R. Wang (2014). Joint mixability. Preprint, University of Waterloo.

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