Joint Mixability

Bin Wang* and Ruodu Wang†

23 July 2015

Abstract

Many optimization problems in probabilistic combinatorics and mass transportation impose fixed marginal constraints. A natural and open question in this field is to determine all possible distributions of the sum of random variables with given marginal distributions; the notion of joint mixability is introduced to address this question. A tuple of univariate distributions is said to be jointly mixable if there exist random variables, with respective distributions, such that their sum is a constant. We obtain necessary and sufficient conditions for the joint mixability of some classes of distributions, including uniform distributions, distributions with monotone densities, distributions with unimodal-symmetric densities, and elliptical distributions with the same characteristic generator. Joint mixability is directly connected to many open questions on the optimization of convex functions and probabilistic inequalities with marginal constraints. The results obtained in this paper can be applied to find extreme scenarios on risk aggregation under model uncertainty at the level of dependence.

MSC2000 subject classification: Primary: 60C05; secondary: 60E15.

OR/MS subject classification: Primary: mathematics: combinatorics; secondary: probability: distributions; finance: portfolio.

Keywords: joint mixability; multivariate dependence; convex optimization; optimal coupling; risk aggregation.

1 Introduction

In a standard atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, assuming $X_i \sim F_i$, $i = 1, \ldots, n$ for given univariate distributions $F_1, \ldots, F_n$, what can we say about the distribution of $S = X_1 + \cdots + X_n$, when the dependence structure among $X_1, \ldots, X_n$ is allowed to be arbitrary? That is

*Department of Mathematics, Beijing Technology and Business University, No. 11/33, Fucheng Road, Haidian District, Beijing 100048, China.

†Corresponding author. Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L3G1, Canada. Email: wang@uwaterloo.ca.

(Q) to determine whether a given distribution F is a possible distribution of S.

The simply-stated question (Q) turns out to be highly non-trivial to answer. A more general question arises if the sum S is replaced by any measurable function on $\mathbb{R}^n$ of $X_1, \ldots, X_n$. At this moment, even the case for the sum seems to be challenging enough; we focus on this special case in this paper. To frame question (Q) mathematically, we denote the set of all possible multivariate distributions with given marginal distributions, called a *Fréchet class*, as

$$\mathfrak{F}_n = \mathfrak{F}_n(F_1, \ldots, F_n) = \{\text{distribution of } (X_1, \ldots, X_n) : X_i \sim F_i,\ i = 1, \ldots, n\}, \tag{1.1}$$

and the set of all possible distributions of the sum of random variables with given marginal distributions as

$$\mathfrak{S}_n = \mathfrak{S}_n(F_1, \ldots, F_n) = \{\text{distribution of } X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \ldots, n\}. \tag{1.2}$$

The study on $\mathfrak{F}_n$ is referred to as *distributions with fixed margins*. Early results on inequalities in Fréchet classes date back to Hoeffding (1940) and Fréchet (1951). Some later milestone results on inequalities and compatibility for distributions with fixed margins include Strassen (1965), Dall'Aglio (1972) and Tchen (1980). A collection of more recent research developments in this area can be found in Dall'Aglio et al. (1991) and Cuadras and Fortiana (2002). We refer to Joe (1997) for an overview on dependence concepts and Fréchet classes.

The set of distributions $\mathfrak{F}_n$ is fully characterized, whereas the characterization of $\mathfrak{S}_n$ is not clear yet. Therefore, optimization problems over $\mathfrak{S}_n$ could be much more challenging than classic optimization problems over $\mathfrak{F}_n$. We write $S \in_{\mathrm{d}} \mathfrak{S}_n$ if the distribution of S is in $\mathfrak{S}_n$. Typical examples of optimization problems over $\mathfrak{S}_n$ include

$$\sup_{S \in_{\mathrm{d}} \mathfrak{S}_n} \rho(S) \quad \text{and} \quad \inf_{S \in_{\mathrm{d}} \mathfrak{S}_n} \rho(S) \tag{1.3}$$

for a law-determined convex functional $\rho : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is a set of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and

$$\sup_{S \in_{\mathrm{d}} \mathfrak{S}_n} \mathbb{P}(S \leqslant s) \quad \text{and} \quad \inf_{S \in_{\mathrm{d}} \mathfrak{S}_n} \mathbb{P}(S \leqslant s) \tag{1.4}$$

for a real number s. The history of the search for such optimizations problems goes back to Makarov (1981), who, in response to a question earlier raised by A.N. Kolmogorov, gave the maximal and minimal values in (1.4) for $n = 2$. The optimization of convex functions in (1.3) are answered in Tchen (1980): the maximum for any n and the minimum for $n = 2$ are solved. However, general solutions to (1.3) and (1.4) are not available for $n \geqslant 3$ except for the maximum problem in (1.3). Since such extreme values are important in quantitative risk management, significant mathematical developments are made also in the financial mathematics literature; the interested reader is referred to Embrechts et al. (2014) for history, financial implications and recent progresses on such questions. Contributions have been made using copulas and mass-transportation techniques (see Rüschendorf, 2013), but a feasible way to fully answer question (Q) is never clear. As far as we know, even for the case $n = 2$ question (Q) is still open. The optimization problem (1.3) includes discrete minimization of the maximal time spent by a set of workers (the bottleneck of a schedule); see for instance Haus (2015).

In this paper, we will partially answer question (Q) by formally introducing the theory of *joint mixability*. A vector $(X_1, \ldots, X_n)$ is called a *joint mix* if $X_1 + \cdots + X_n$ is almost surely a constant. An n-tuple of distributions $(F_1, \ldots, F_n)$ is said to be *jointly mixable* if there exists a joint mix with univariate marginal distributions $F_1, \ldots, F_n$. The concept of joint mixability was first introduced in Wang et al. (2013), as a generalization of *complete mixability* in Wang and Wang (2011) (a distribution F is *n-completely mixable* if there exists a joint mix taking values in $\mathbb{R}^n$ with marginal distributions identical to F). Earlier study of *risks with constant sum*, containing special cases of complete mixability, can be found in Gaffke and Rüschendorf (1981), Rüschendorf and Uckelmann (2002) and Müller and Stoyan (2002, Section 8.3.1). Wang et al. (2013) briefly defined joint mixability with limited mathematical development of the theory.

Question (Q) can be reformulated as: do there exist random variables $S, X_1, \ldots, X_n$ such that $S \sim F$, $X_i \sim F_i$, $i = 1, \ldots, n$ and $X_1 + \cdots + X_n - S = 0$? Equivalently, letting $\bar{F}(x) = \mathbb{P}(-S \leqslant x)$, is $(\bar{F}, F_1, \ldots, F_n)$ jointly mixable? Here without loss of generality we can choose the constant K to be 0 by shifting F. Therefore, question (Q) is equivalent to the determination of joint mixability.

Complete mixability and joint mixability describe whether it is possible to generate random variables from given distributions with a target distribution of the sum. The properties are particularly of interest in quantitative risk management, where dependence between risks is usually unknown or partially unknown. The notion of complete mixability helps to determine best- and worst-cases of capital requirements (special cases of (1.3) and (1.4)) under *dependent uncertainty*; for such discussions, the reader is referred to Wang et al. (2013), Bernard et al. (2014) and Embrechts et al. (2013, 2015). A connection of the above probabilistic problems to mass-transportation theory is found in Rüschendorf (2013). Note that the existing research mainly focused on complete mixability; results on joint mixability are very limited in the literature as it is a relatively new notion and is regarded as more challenging.

Complete mixability and joint mixability can be treated as the strongest form of negative dependence; see for instance Müller and Stoyan (2002, Section 8.3.1) and Puccetti and Wang (2015b). Existing results on complete mixability are summarized in Wang and Wang (2011) and Puccetti et al. (2012, 2013). Although it has been mentioned with applications, non-trivial examples of jointly mixable distributions are only found in Wang et al. (2013) where necessary and sufficient conditions for the joint mixability of normal distributions are given. A numerical procedure to check joint mixability is presented in Puccetti and Wang (2015a). In this paper, we develop the theory of joint mixability, and provide necessary and sufficient conditions for the joint mixability of uniform distributions, distributions with monotone densities, distributions with unimodal-symmetric densities, and elliptical distributions with the same characteristic generator. The most valuable and technical contribution of this paper is Theorem 3.2 which characterizes the joint mixability of distributions with monotone densities. It generalizes Theorem 2.4 of Wang and Wang (2011), which is regarded as the most relevant result on complete mixability for risk management (see Embrechts et al., 2014, Section 3).

The rest of the paper is organized as follows. In Section 2 we introduce the definition, some examples, necessary conditions and some technical properties of joint mixability. In Section 3 we present our main results on sufficient conditions of joint mixability for some classes of distributions. In particular, the main technical result in this paper provides sufficient and necessary conditions for the joint mixability of a tuple of distributions with monotone densities. Section 4 is dedicated to some probabilistic inequalities and applications related to our main results. The proofs of the main theorems are put in Section 5. Section 6 draws a conclusion.

2 Joint mixability

Throughout the paper, we consider a standard atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $L^p := L^p(\Omega, \mathcal{A}, \mathbb{P})$, $p \in [0, \infty]$ be the set of all real-valued random variables on that probability space with finite p-moment. For simplicity, we identify probability distributions with its cumulative distributions functions (cdf) in this paper. For a cdf F, we write $X \sim F$ to denote $F(x) = \mathbb{P}(X \leqslant x)$, $x \in \mathbb{R}$. We also denote the generalized inverse function of F by $F^{-1}(p)$, that is $F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geqslant p\}$ for $p \in (0, 1]$, and $F^{-1}(0) = \inf\{t \in \mathbb{R} : F(t) > 0\}$. We write $X \overset{\mathrm{d}}{=} Y$ if the random variables (or vectors) X and Y have the same distribution.

2.1 Definition

Definition 2.1 (Joint mixability)**.** Suppose n is a positive integer. An n-tuple of probability distributions on $\mathbb{R}$ $(F_1, \dots, F_n)$ is *jointly mixable* (JM) if there exist n random variables $X_1 \sim F_1, \dots, X_n \sim F_n$ such that $X_1 + \cdots + X_n =: K$ is almost surely a constant. Such $K \in \mathbb{R}$ is called a *joint center* of $(F_1, \dots, F_n)$, and the random vector $(X_1, \dots, X_n)$ is called a *joint mix*. We denote by $\mathcal{J}_n(K)$ the set of all n-tuples of JM distributions with joint center K, that is,

$$\mathcal{J}_n(K) = \{(F_1, \dots, F_n) : (F_1, \dots, F_n) \text{ is JM with joint center } K\}.$$

A probability distribution F on $\mathbb{R}$ is *n-completely mixable* (n-CM) if $(F, \dots, F) \in J_n(K)$ for some K.

Remark 2.1. Equivalently, $(F_1, \dots, F_n)$ is jointly mixable if the set $\mathfrak{S}_n$ in (1.2) contains a degenerate distribution. When the means of $F_1, \dots, F_n$ are finite, the joint center K is unique and equal to the sum of the means of $F_1, \dots, F_n$.

The term *mixability* in Definition 2.1 reflects that the property concerns whether one is *able* to construct a joint *mix* with the given margins. We first give a few natural examples of joint mixes and jointly mixable distributions.

Example 2.1 (Multinomial distributions)**.** Suppose that $(X_1, \dots, X_n)$ follows from a multinomial distribution with parameters $(N; p_1, \dots, p_n)$, $N \in \mathbb{N}$, $p_1, \dots, p_n \geqslant 0$ and $\sum_{i=1}^n p_i = 1$. That is, for $(x_1, \dots, x_n) \in \mathbb{N}_0^n$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \frac{N!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n} \mathrm{I}_{\{x_1 + \dots + x_n = N\}}.$$

Obviously, $(X_1, \dots, X_n)$ is a joint mix by definition. The marginal distributions of a multinomial distribution with parameters $(N; p_1, \dots, p_n)$ are binomial distributions with respective parameters $(N; p_1), \dots, (N; p_n)$. As a consequence, an n-tuple of binomial distributions with respective parameters $(N; p_1), \dots, (N; p_n)$ where $\sum_{i=1}^n p_i = 1$ is jointly mixable.

Example 2.2 (Multivariate normal distributions)**.** Suppose that $\sigma_1, \sigma_2, \sigma_3 > 0$,

$$2 \max_{i=1,2,3} \sigma_i \leqslant \sigma_1 + \sigma_2 + \sigma_3, \tag{2.1}$$

and (X_1, X_2, X_3) follows from a multivariate normal distribution with parameters $(\mathbf{0}, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) & \sigma_3^2 \end{pmatrix}.$$

It is straightforward to check that Σ is positive semi-definite under condition (2.1) and hence the multivariate normal distribution with parameters $(\mathbf{0}, \Sigma)$ is well defined. We can verify that $\mathbb{E}[(X_1 + X_2 + X_3)^2] = 0$ which is the summation of all entries in Σ. Hence, $(X_1, \dots, X_n)$ is a joint mix, and the triplet of normal distributions $(\mathrm{N}(0, \sigma_1^2), \mathrm{N}(0, \sigma_2^2), \mathrm{N}(0, \sigma_3^2))$ is jointly mixable. Later we will see that (2.1) is necessary and sufficient for the joint mixability of distributions in a broad class.

Recent results on complete mixability can be found in Wang and Wang (2011), Puccetti et al. (2012, 2013). On the other hand, the properties and characterization of JM distributions are very limited in the literature (only Wang et al. (2013) gave a result on normal distributions). In this paper, we aim to explore the theory for joint mixability in more depth. Some of the results on joint mixability are parallel to the results on complete mixability; however we remark that most of the proofs on complete mixability cannot be naturally generalized to joint mixability, and much more work is needed. In the next section, law-determined norms will be used.

Definition 2.2 (Law-determined norm)**.** A law-determined norm $||\cdot||$ is a map from L^0 to $[0,\infty]$, such that

(i) $||aX|| = |a|\cdot||X||$ for $a \in \mathbb{R}$ and $X \in L^0$;

(ii) $||X+Y|| \leqslant ||X|| + ||Y||$ for $X, Y \in L^0$;

(iii) $||X|| = 0$ implies $X = 0$ a.s.;

(iv) $||X|| = ||Y||$ if $X \overset{\mathrm{d}}{=} Y$, $X, Y \in L^0$.

The L^p-norms $p \in [1,\infty)$, $||\cdot||_p : L^0 \to [0,\infty]$, $X \mapsto (\mathbb{E}[|X|^p])^{1/p}$ and the L^∞-norm $||\cdot||_\infty : L^0 \to [0,\infty]$, $X \mapsto \text{ess-sup}(|X|)$ are law-determined norms. Here, we allow $||\cdot||$ to take a value of ∞, which means that the non-negative functional $||\cdot||$ is not necessarily a norm in the common sense; we slightly abuse the terminology here since all natural examples are norms in respective proper spaces.

2.2 Properties

Throughout this paper, for $i = 1, \ldots, n$, we denote by μ_i the mean of F_i and let $a_i = \sup\{x : F_i(x) = 0\}$, $b_i = \inf\{x : F_i(x) = 1\}$ and $l_i = b_i - a_i$. All of the above quantities are possibly infinite.

We first give three necessary conditions for joint mixability. The three conditions seem trivial, but they are essential for several classes of distributions to be JM, as we will discuss in Section 3. To avoid possible ill-definition, we naturally allow $-\infty + \infty \leqslant K \leqslant -\infty + \infty$ for any $K \in \mathbb{R} \cup \{-\infty, \infty\}$.

Theorem 2.1 (Necessary conditions)**.** *If the n-tuple of distributions $(F_1, \ldots, F_n)$ is JM, and $\mu_1, \ldots, \mu_n$ are finite, then the following inequalities hold:*

(a) (Mean inequality)

$$\sum_{i=1}^n a_i + \max_{i=1,\ldots,n} l_i \leqslant \sum_{i=1}^n \mu_i \leqslant \sum_{i=1}^n b_i - \max_{i=1,\ldots,n} l_i. \tag{2.2}$$

(b) (Norm inequality)

$$\sum_{i=1}^n ||X_i - \mu_i|| \geqslant 2 \max_{i=1,\ldots,n} ||X_i - \mu_i||, \tag{2.3}$$

where $X_i \sim F_i$, $i = 1, \ldots, n$ and $||\cdot||$ is a law-determined norm on L^0.

Proof. For part (a), we only need to prove for the cases where at least one of $\sum_{i=1}^n a_i + \max_{i=1,\dots,n} l_i$, and $\sum_{i=1}^n b_i - \max_{i=1,\dots,n} l_i$ is finite. For part (b), we only need to show for the case where each $||X_i||$ is finite.

(a) Assume $\sum_{i=1}^n a_i + \max_{i=1,\dots,n} l_i$ is finite. Take $\varepsilon > 0$. Suppose $X_i \sim F_i$, $i = 1, \dots, n$ and $X_1 + \cdots + X_n = K \in \mathbb{R}$. Since the support of F_i is bounded away from $-\infty$, $i = 1, \dots, n$, it is easy to see that $K = \sum_{i=1}^n \mu_i$. Then for any $j = 1, \dots, n$

$$\mathrm{I}_{\{K > \sum_{i=1}^n a_i + l_j - \varepsilon\}} = \mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n a_i + l_j - \varepsilon\right) \geqslant \mathbb{P}\left(X_j > b_j - \varepsilon\right) > 0.$$

This implies

$$\sum_{i=1}^n \mu_i \geqslant \sum_{i=1}^n a_i + \max_{i=1,\dots,n} l_i.$$

The other half of the assertion is similar.

(b) Since $(F_1, \dots, F_n)$ is JM, there exists a joint mix $(Y_1, \dots, Y_n)$ where $Y_i \sim F_i$, $i = 1, \dots, n$. Then $\sum_{i=1}^n (Y_i - \mu_i) = 0$. It follows that for $j = 1, \dots, n$,

$$||Y_j - \mu_j|| = \left|\left|\sum_{i=1, i\neq j}^n (Y_i - \mu_i)\right|\right| \leqslant \sum_{i=1, i\neq j}^n ||Y_i - \mu_i||,$$

and (2.3) follows since $||\cdot||$ is law-determined. $\square$

Remark 2.2. Suppose that $(F_1, \dots, F_n)$ is JM. If $\mu_1, \dots, \mu_n$ are finite, (2.2) implies that

$$\min_{i=1,\dots,n} a_i > -\infty \;\Leftrightarrow\; \max_{i=1,\dots,n} b_i < \infty.$$

Indeed, the above assertion holds without assuming the finiteness of $\mu_1, \dots, \mu_n$. To see this, suppose that $\min_{i=1,\dots,n} a_i > -\infty$ and $b_j = \infty$ for some $j = 1, \dots, n$. Then for any $x \in \mathbb{R}$, and $X_i \sim F_i$, $i = 1, \dots, n$,

$$\mathbb{P}(X_1 + \cdots + X_n > x) \geqslant \mathbb{P}\left(X_j > x - \sum_{i=1,\ i\neq j}^n a_i\right) > 0.$$

This shows that $(X_1, \dots, X_n)$ cannot be a joint mix. The case when $\max_{i=1,\dots,n} b_i < \infty$ and $a_j = -\infty$ for some $j = 1, \dots, n$ is similar.

Theorem 2.1 (a) is a generalization of the mean condition for complete mixability, which is of crucial importance in the study of complete mixability. Note that joint mixability is location-invariant; the mean condition is not affected by affine transformations. As an additional condition for joint mixability, (b) forms a polygon inequality: the largest component can not be greater than the sum of the rest. This is geometrically intuitive, since if we treat $X_1, \dots, X_n$ as vectors, then they form an n-polygon if and only if $X_1 + \cdots + X_n = 0$.

When $(F_1, \dots, F_n)$ is JM, inequality (2.3) holds for all possible law-determined norms, which could be difficult to check theoretically. In the following corollary we give some special cases of (2.3) which can be easily checked, and sometimes they turn out to be sufficient as well. We denote by σ_i^2 the variance of F_i, $i = 1, \dots, n$; they could be infinite.

Corollary 2.2 (Necessary conditions - special cases)**.** *If the n-tuple of distributions $(F_1, \dots, F_n)$ is JM, then the following inequalities hold:*

(c) (Length inequality)

$$\sum_{i=1}^n l_i \geqslant 2 \max_{i=1,\dots,n} l_i. \tag{2.4}$$

(d) (Variance inequality)

$$\sum_{i=1}^n \sigma_i \geqslant 2 \max_{i=1,\dots,n} \sigma_i. \tag{2.5}$$

Proof. If $\max_{i=1,\dots,n} l_i < \infty$, (2.4) follows directly from (2.2) by noting that

$$\sum_{i=1}^n a_i + \max_{i=1,\dots,n} l_i \leqslant \sum_{i=1}^n b_i - \max_{i=1,\dots,n} l_i \quad \Rightarrow \quad 2 \max_{i=1,\dots,n} l_i \leqslant \sum_{i=1}^n (b_i - a_i);$$

if $\max_{i=1,\dots,n} l_i = \infty$, then (2.4) holds trivially. (2.5) follows directly from (2.3). $\square$

In some cases, (c) and (d) are equivalent (for example, when $F_1, \dots F_n$ are in a scale family with bounded support; in that case, (b) is also equivalent to (c) and (d)). We remind the readers that conditions (a)-(d) are generally not sufficient for joint mixability. Later in this paper, we will show that (a)-(d) are sufficient for some classes of distributions.

To characterize joint mixability, we begin by looking at the set of jointly mixable distributions. In the following we give some basic properties of joint mixability, which will be used to prove the main results in Section 3. The properties (i)-(v) are parallel to Proposition 2.1 in Wang and Wang (2011) and Theorems 3.1 and 3.2 in Puccetti et al. (2012) for complete mixability, and they can be checked in a relatively straightforward manner; we only give a proof for (iv) here.

Proposition 2.3. *For simplicity, we denote by $\mathbf{F} = (F_1, \dots, F_n)$ and $\mathbf{G} = (G_1, \dots, G_m)$. Here the positive integers n and m can be the same depending on the context. Suppose $K, L \in \mathbb{R}$.*

(i) If $\mathbf{F}, \mathbf{G} \in \mathcal{J}_n(K)$, then $\lambda \mathbf{F} + (1-\lambda)\mathbf{G} \in \mathcal{J}_n(K)$ for all $\lambda \in [0, 1]$.

(ii) If $\mathbf{F} \in \mathcal{J}_n(K)$ and $\mathbf{G} \in \mathcal{J}_m(L)$, then $(F_1, \dots, F_n, G_1, \dots, G_m) \in \mathcal{J}_{n+m}(K + L)$.

(iii) The element-wise weak limit of a sequence in $\mathcal{J}_n(K)$ is still in $\mathcal{J}_n(K)$.

(iv) As a consequence of (i) and (iii), a (possibly infinite) convex combination of vectors in $\mathcal{J}_n(K)$ is still in $\mathcal{J}_n(K)$.

(v) Given $r, t_1, \ldots, t_n \in \mathbb{R}$, let $F_i[r, t_i](x) = \mathbb{P}(rX_i + t_i \leqslant x)$ for $x \in \mathbb{R}$ and $i = 1, \ldots, n$. If $(F_1, \ldots, F_n) \in \mathcal{J}_n(K)$, then $(F_1[r, t_1], \ldots, F_n[r, t_n]) \in \mathcal{J}_n(rK + \sum_{i=1}^n t_i)$.

(vi) $(F_1, \ldots, F_n, G_1, \ldots, G_m)$ is JM if and only if there exists a distribution F, such that both $(F_1, \ldots, F_n, F)$ and $(G_1, \ldots, G_m, \bar{F})$ are JM, where $\bar{F}(x) = \mathbb{P}(-X \leqslant x)$, $x \in \mathbb{R}$, $X \sim F$.

Proof. Similar to properties of complete mixability, (i)-(v) can be checked directly. We only show (iv).

$\Rightarrow$: Suppose that $X_i \sim F_i$, $i = 1, \ldots, n$ and $Y_j \sim G_j$, $j = 1, \ldots, m$ and $X_1 + \cdots + X_n + Y_1 + \cdots + Y_m = K$ is a constant. Let $Y = Y_1 + \cdots + Y_m$ and denote its distribution by F. Then $X_1 + \cdots + X_n + Y = K$, that is, $(F_1, \ldots, F_n, F)$ is JM. On the other hand, $Y_1 + \cdots + Y_m - Y = 0$, that is, $(G_1, \ldots, G_m, \bar{F})$ is JM.

$\Leftarrow$: Suppose that there exists a distribution F such that $Y, Y' \sim F$, $(X_1, \ldots, X_n, Y)$ and $(Y_1, \ldots, Y_m, -Y')$ are two joint mixes. We claim that there exists a random vector $\mathbf{Z} = (Z_1, \ldots, Z_n, W, W_1, \ldots, W_m)$ where $(Z_1, \ldots, Z_n, W) \stackrel{\mathrm{d}}{=} (X_1, \ldots, X_n, Y)$ and $(W, W_1, \ldots, W_m) \stackrel{\mathrm{d}}{=} (Y', Y_1, \ldots, Y_m)$. The existence of $\mathbf{Z}$ is guaranteed by the compatibility of the two marginal constraints that are overlapping only on a single variable; see Joe (1997, Section 3.2). Moreover, $(Z_1, \ldots, Z_n, W_1, \ldots, W_m)$ is a joint mix since $(Z_1, \ldots, Z_n, W)$ and $(-W, W_1, \ldots, W_m)$ are both joint mixes. Therefore, $(F_1, \ldots, F_n, G_1, \ldots, G_m)$ is JM. $\square$

Note that in Proposition 2.3 (v), the scale parameter r has to be identical for all marginal distributions.

As mentioned earlier, existing research on joint mixability is limited. As far as we know, the only known cases are the normal distributions (Wang et al., 2013). In the next section, we will identify new classes of jointly mixable distributions and their corresponding necessary and sufficient conditions.

3 Main results

In this section, we list sufficient conditions for the joint mixability of several classes of distributions. The proofs are put in Section 5. In Section 3.1 we study distributions with monotone densities. In Section 3.2 we study symmetric distributions.

3.1 Distributions with monotone densities

The first result in this section is the joint mixability of non-identical uniform distributions, based on which the proofs of the joint mixability of monotone densities and unimodal-symmetric densities will be obtained later.

For uniform distributions we can easily verify that the four necessary conditions (2.2)-(2.5) are all equivalent. Moreover, it turns out that each of them is sufficient for a tuple of several uniform distributions to be JM.

Theorem 3.1. *Suppose that $F_1, \dots, F_n$ are n uniform distributions. Then $(F_1, \dots, F_n)$ is JM if any of (2.2)-(2.5) is satisfied.*

In the following we present the most important theorem in this paper, a necessary and sufficient condition for the joint mixability of a tuple of distributions with decreasing densities. Its proof is mathematically more involved than the other theorems in this section. According to Remark 2.2, distributions with unbounded supports and decreasing densities are not jointly mixable; hence we only study the case where they have bounded supports.

Theorem 3.2. *Suppose that $F_1, \dots, F_n$ are n distributions with decreasing densities on their respective bounded supports. Then $(F_1, \dots, F_n)$ is JM if the mean inequality (2.2) is satisfied.*

If $(X_1, \dots, X_n)$ is a joint mix, then $(-X_1, \dots, -X_n)$ is also a joint mix; thus (2.2) is also sufficient for the joint mixability of a tuple of distributions with increasing densities. We remark that Theorem 3.2, parallel to Theorem 2.4 of Wang and Wang (2011), has been a conjecture for about four years. It has been expected to hold based on some numerical evidence; see Puccetti and Wang (2015a) for numerical verifications of Theorem 3.2 through a discretized approximation. The result itself is of practical importance in risk management; see Wang et al. (2013) and Embrechts et al. (2014). Similar to the case of identical margins in Wang and Wang (2011), an explicit joint distribution function (or its copula) of a joint mix with different monotone densities seems to be very difficult to write down.

The proposition below arises in the proof of Theorem 3.2. It consists of a different type of results and is of independent interest: the existence of F with some constraints such that (F_1, F_2, F) is JM.

Proposition 3.3. *Suppose that F_1 and F_2 have decreasing density functions on their support $[0, L_1]$ and $[0, L_2]$, with mean μ_1 and μ_2, respectively. There exists a distribution F_3 with an increasing density on $[-M, 0]$, where $M = \max\{L_1, L_2, 2\mu_1 + 2\mu_2\}$, such that $(F_1, F_2, F_3) \in \mathcal{J}_3(0)$.*

Proposition 3.3 implies that for distributions F_1 and F_2 with decreasing densities on $[0, L]$, means μ_1, μ_2 and $\mu_1 + \mu_2 \leqslant L/2$, it is possible to find $X_1 \sim F_1$ and $X_2 \sim F_2$ such that $X_1 + X_2$ also has a decreasing density on $[0, L]$.

3.2 Symmetric distributions

The following theorem concerns the joint mixability of unimodal-symmetric densities, not necessarily from the same location-scale family.

Theorem 3.4. *Suppose that $F_1, \ldots, F_n$ are distributions with unimodal-symmetric densities, and mode 0. Let $p_i(x)$ be the density function of F_i and let $G_i(x) = F_i(x) - xp_i(x) - \frac{1}{2}$ for $i = 1, \ldots, n$ and $x \geqslant 0$. Then $(F_1, \ldots, F_n)$ is JM if*

$$\sum_{i=1}^{n} G_i^{-1}(a) \geqslant 2 \max_{i=1,\ldots,n} G_i^{-1}(a) \quad \text{for all } a \in (0, \tfrac{1}{2}). \tag{3.1}$$

An immediate corollary of Theorem 3.4 is that a triplet of two identical uniform distributions and a unimodal-symmetric distribution with a compatible support is JM.

Corollary 3.5. *Suppose that F is a unimodal-symmetric distribution. For $a > 0$, $(\mathrm{U}[0, a], \mathrm{U}[0, a], F)$ is JM if and only if F is supported in an interval of length at most $2a$.*

When the distributions $F_1, \ldots, F_n$ are unimodal-symmetric, and from a location-scale family with scale parameters $\theta_1, \ldots, \theta_n$, respectively, inequality (3.1) reads as

$$\sum_{i=1}^{n} \theta_i \geqslant 2 \max_{i=1,\ldots,n} \theta_i. \tag{3.2}$$

Note that (3.2) is also equivalent to the norm inequality (2.3) and it is sufficient for joint mixability, as implied by Theorem 3.4.

Corollary 3.6. *Suppose that $F_1, \ldots, F_n$ are unimodal-symmetric distributions from the same location-scale family. Then $(F_1, \ldots, F_n)$ is JM if the scale inequality (3.2) is satisfied.*

Remark 3.1. Theorem 2.1 of Rüschendorf and Uckelmann (2002) gave the complete mixability of distributions with a unimodal-symmetric density by an analytical construction. Theorem 3.4 and Corollary 3.6 generalize Theorem 2.1 of Rüschendorf and Uckelmann (2002). The mathematical approach in the latter paper does not seems to apply to the scope of joint mixability. The proof of Theorem 3.4 in this paper is based on a different technical approach.

In the next theorem, we give necessary and sufficient conditions for the joint mixability of marginal elliptical distributions. An n-elliptical distribution $E_n(\mu, \Sigma, \phi)$ is an n-variate distribution with characteristic function

$$\exp\{\boldsymbol{i}\mathbf{t}^\top \mu\}\phi(\mathbf{t}^\top \Sigma \mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^n$$

where $\boldsymbol{i}$ is the imaginary unit, $\mu \in \mathbb{R}^n$, Σ is an $n \times n$ positive semi-definite matrix, $A^\top$ represents the transpose of a matrix A, and $\phi : [0, \infty) \to \mathbb{R}$ is called a *characteristic generator*. A necessary and sufficient condition for ϕ to be a characteristic generator is given in Theorem 2 of Cambanis et al. (1981); the interested reader is referred to Fang et al. (1990) and Lindskog et al. (2003) for more details. Many commonly used multivariate distributions such as the multivariate normal distribution and the multivariate t-distribution are special cases of elliptical distributions.

Theorem 3.7. *Suppose that $\mu_i \in \mathbb{R}$, $\sigma_i \geqslant 0$, $i = 1, \dots, n$, ϕ is a characteristic generator for an n-elliptical distribution, and $F_i \sim E_1(\mu_i, \sigma_i^2, \phi)$. Then $(F_1, \dots, F_n)$ is JM if the inequality (2.5) is satisfied.*

In the above theorem, the real parameters $\sigma_1, \dots, \sigma_n$ are not necessarily the respective standard deviations of $F_1, \dots, F_n$. For $i = 1, \dots, n$, if F_i has a finite variance, then $\phi'(0+)$ exists and σ_i is $-2\phi'(0+)$ times the standard deviation of F_i; see Cambanis et al. (1981) for details.

Remark 3.2. We remark that the class of symmetric distributions is identical to the class of one-dimensional elliptical distributions; see Cambanis et al. (1981). Wang et al. (2013) gave the joint mixability of normal distributions as a special case of Theorem 3.7. For normal random vectors, having a constant sum is also related to minimal concordance order in the Fréchet class; see Joe (1990).

4 Optimization problems

In this section, we briefly study two types of optimization problems related to joint mixability as mentioned in Section 1. These results can directly be applied to the determination of conservative values of risk measures on risk aggregation, and have close connections to the bottleneck assignment problems as discussed in Haus (2015).

4.1 Optimization with respect to convex order

Convex order is a partial order based on variability between two random variables.

Definition 4.1 (Convex order)**.** Let X and Y be two random variables with finite means. X is smaller than Y *in convex order*, denoted by $X \prec_{\mathrm{cx}} Y$, if for all convex functions f,

$$\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)], \tag{4.1}$$

whenever both sides of (4.1) are well-defined.

It is immediate that $X \prec_{\mathrm{cx}} Y$ implies $\mathbb{E}[X] = \mathbb{E}[Y]$. Convex order has been extremely useful in combinatorics, optimization, probability and financial mathematics. In what follows, by saying an element in $\mathfrak{S}_n$ we actually mean a random variable whose distribution belongs to $\mathfrak{S}_n$. Finding the largest and the smallest elements in $\mathfrak{S}_n$ w.r.t. convex order would typically solve the optimization problems in (1.3), as law-determined convex functionals often respect convex order; see for instance Föllmer and Schied (2011, Section 4).

It is well-known that the convex ordering largest element in $\mathfrak{S}_n(F_1,\dots,F_n)$ is always obtained by $F_1^{-1}(U)+\cdots+F_n^{-1}(U)$ for a random variable $U \sim \mathrm{U}[0,1]$. However, it remains open in general to find the smallest element in $\mathfrak{S}_n(F_1,\dots,F_n)$ w.r.t. convex order for $n \geqslant 3$. Bernard et al. (2014) gave an example where $\mathfrak{S}_n(F_1,\dots,F_n)$ does not contain a smallest element w.r.t. convex order.

For $F_1,\dots,F_n$ with finite means, a sufficient condition for the existence of the smallest element w.r.t. convex order in $\mathfrak{S}_n(F_1,\dots,F_n)$ is joint mixability. Some sufficient conditions for joint mixability are given in Section 3. In such cases, the corresponding smallest element w.r.t. convex order is μ, the sum of the means of $F_1,\dots,F_n$.

In summary, for all $S \in_{\mathrm{d}} \mathfrak{S}_n(F_1,\dots,F_n)$,

$$\mu \prec_{\mathrm{cx}} S \prec_{\mathrm{cx}} F_1^{-1}(U)+\cdots+F_n^{-1}(U), \tag{4.2}$$

where U is a $\mathrm{U}[0,1]$ random variable. Both inequalities in (4.2) are sharp if the conditions in any of Theorems 3.1, 3.2, 3.4 and 3.7, and Corollaries 3.5 and 3.6 are satisfied.

Remark 4.1. When the Fréchet class $\mathfrak{F}_n$ admits *mutually exclusive* random vectors (with very restrictive conditions on $F_1,\dots,F_n$), mutual exclusivity leads to a smallest element in $\mathfrak{S}_n$ w.r.t. convex order; see Dhaene and Denuit (1999) and Cheung and Lo (2014).

Remark 4.2. Similar to the homogeneous results in Wang and Wang (2011, Section 3), Theorem 3.2 could be used to characterize the smallest element w.r.t. convex order in $\mathfrak{S}_n(F_1,\dots,F_n)$ when $F_1,\dots,F_n$ have monotone densities but do not satisfy the mean condition (2.2). This requires non-trivial further study; see recent results in Jakobsons et al. (2015).

4.2 Optimization on the distribution function of the sum

The joint mixability is also related to the minimal or maximal probability function of the sum over $\mathfrak{S}_n$ in (1.4). Denote

$$m_+(s)=\inf\{\mathbb{P}(S\leqslant s): S\in_{\mathrm{d}} \mathfrak{S}_n(F_1,\dots,F_n)\},\quad s\in\mathbb{R},$$

and

$$M_+(s)=\sup\{\mathbb{P}(S\leqslant s): S\in_{\mathrm{d}} \mathfrak{S}_n(F_1,\dots,F_n)\},\quad s\in\mathbb{R}.$$

The quantities $m_+(s)$ and $M_+(s)$ are in general unclear analytically for $n\geqslant 3$.

In quantitative risk management, the Value-at-Risk (VaR) of an aggregate risk is of particular interest in portfolio management. The Value-at-Risk of a random variable X at level $p \in (0,1)$ is defined as the (left-continuous) inverse distribution function

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

Quantities of interest are

$$\overline{\mathrm{VaR}}_p = \sup\{\mathrm{VaR}_p(S) : S \in_{\mathrm{d}} \mathfrak{S}_n(F_1, \dots, F_n)\}, \quad p \in (0,1),$$

and

$$\underline{\mathrm{VaR}}_p = \inf\{\mathrm{VaR}_p(S) : S \in_{\mathrm{d}} \mathfrak{S}_n(F_1, \dots, F_n)\}, \quad p \in (0,1).$$

$\overline{\mathrm{VaR}}_p$ and $\underline{\mathrm{VaR}}_p$ represent the best and worst cases of Value-at-Risk for a portfolio in presence of model uncertainty at the level of dependence. For more discussions and applications of this topic, see Embrechts and Puccetti (2006) and Embrechts et al. (2013, 2014).

Here we look at $m_+(s)$ and $\overline{\mathrm{VaR}}_p$ since the story of $M_+(s)$ and $\underline{\mathrm{VaR}}_p$ is similar. The following result is given in Wang et al. (2013). Define, for $t \in [0,1)$, $\Phi(t) = \sum_{i=1}^n \mathbb{E}[X_i | X_i \geqslant F_i^{-1}(t)]$, where $X_i \sim F_i$, $i = 1, \dots, n$, and let Φ^{-1} be the inverse of Φ. In addition, let $\Phi(1) = \sum_{i=1}^n F_i^{-1}(1)$, $\Phi^{-1}(x) = 0$ if $x < \Phi(0)$ and $\Phi^{-1}(x) = 1$ if $x \geqslant \Phi(1)$. Assuming that $F_1, \dots, F_n$ have positive densities on their supports, Φ is strictly increasing on $(0,1)$.

Proposition 4.1. *Suppose the distributions $F_1, \dots, F_n$ have positive density on their supports with finite means.*

(1) We have for $s \in \mathbb{R}$,

$$m_+(s) \geqslant \Phi^{-1}(s), \tag{4.3}$$

and for $p \in (0,1)$,

$$\overline{\mathrm{VaR}}_p \leqslant \Phi(p). \tag{4.4}$$

(2) For each fixed $s \geqslant \Phi(0)$, the equality

$$m_+(s) = \Phi^{-1}(s) \tag{4.5}$$

holds if and only if the n-tuple of the distributions of $F_1^{-1}(W), \dots, F_n^{-1}(W)$ is jointly mixable, where $W \sim \mathrm{U}[\Phi^{-1}(s), 1]$, and for $p \in (0,1)$, the equality

$$\overline{\mathrm{VaR}}_p = \Phi(p) \tag{4.6}$$

holds if and only if the n-tuple of the distributions of $F_1^{-1}(W), \dots, F_n^{-1}(W)$ is jointly mixable, where $W \sim \mathrm{U}[p, 1]$.

We remark that, although the above result was given in Wang et al. (2013), no example was presented in the latter paper for the case when $F_1, \dots, F_n$ are non-identical, since there were no known non-trivial classes of truncated distributions that are jointly mixable.

Example 4.1. Suppose the distribution F_i is a uniform distribution on $[a_i, a_i+l_i]$ for $i = 1, \dots, n$. Suppose that $l_1, \dots, l_n$ satisfy (2.4). Let $a = \sum_{i=1}^n a_i + \frac{1}{2}\sum_{i=1}^n l_i$ and $l = \sum_{i=1}^n l_i$. Then from Theorem 3.1 and Proposition 4.1,

$$\overline{\mathrm{VaR}}_t = \Phi(t) = a + \frac{tl}{2}, \quad t \in [0,1]$$

and

$$m_+(s) = 1 \wedge \left(\frac{2s-2a}{l}\right)_+, \quad s \in \mathbb{R}.$$

For the case when all $F_1, \dots F_n$ are identically a uniform distribution on $[0,1]$, Rüschendorf (1982) gave the value of $m_+(s) = 1 \wedge \left(\frac{2s-n}{n}\right)_+$ which is a special case of the above example.

Remark 4.3. For the case where $l_1, \dots, l_n$ do not satisfy (2.4), $m_+(s)$ and $\overline{\mathrm{VaR}}_p$ can be obtained via a convex order argument without involving joint mixability; for connection between convex order and smallest probability function or worst Value-at-Risk, see Bernard et al. (2014, Section 4). The questions of $\overline{\mathrm{VaR}}_p$ where $F_1, \dots, F_n$ have monotone densities in unbounded supports are recently answered based on results obtained in this paper (in particular, Theorem 3.2); see Jakobsons et al. (2015) where joint mixability is assumed as a sufficient condition for main results therein to hold.

5 Proofs of main results

In this section, for any distribution F, we denote by $\bar{F}$ the distribution of the random variable $-X$, where $X \sim F$; this notation was used in Proposition 2.3 (iv). Throughout, $\mathrm{U}[a,b]$ represents a uniform distribution on $[a,b]$, $a, b \in \mathbb{R}$.

5.1 Proof of Theorem 3.1

Proof. Since the four necessary conditions are equivalent, we will only show that the length condition (2.4) is sufficient. Without loss of generality, we assume the mean of F_i is zero for $i = 1, \dots, n$.

We first prove the theorem for $n = 3$. The following lemma comes in handy.

Lemma 5.1. *Suppose F_1, F_2, F_3 are uniform distributions with lengths l_1, l_2, l_3 and $l_1 + l_2 = l_3$. Then (F_1, F_2, F_3) is JM.*

Proof of Lemma 5.1. For a random variable $U \sim \mathrm{U}[-\frac{1}{2}, \frac{1}{2}]$, let $X_1 = l_1 U \sim F_1$, $X_2 = l_2 U \sim F_2$ and $X_3 = -l_3 U \sim F_3$ then $X_1 + X_2 + X_3 = 0$. This shows that (F_1, F_2, F_3) is JM. $\square$

Now we proceed to prove the general case for $n = 3$. Write $F_1 = \mathrm{U}[-a, a]$, $F_2 = \mathrm{U}[-b, b]$ and $F_3 = \mathrm{U}[-c, c]$, where $a, b, c > 0$. Without loss of generality, assume $a \geqslant b$. By (2.4), we have $a - b \leqslant c \leqslant a + b$. The case $c = a - b$ or $c = a + b$ is implied by Lemma 5.1. In the following we assume $a - b < c < a + b$. Denote $k = \frac{c}{a+b}$, $m = \frac{a-b}{c}$ then $0 \leqslant k, m < 1$, $k(a + b) = c$, and $mc = a - b$.

We will show that we can decompose each of F_1, F_2, F_3 into six different uniform distributions $F_{i,j}$ with length $l_{i,j}$, $i = 1, 2, 3$, $j = 1, \ldots, 6$. The decomposition is as follows.

Group	F_1	F_2	F_3	lengths
(i)	$F_{1,1} = \mathrm{U}[-a, ka]$	$F_{2,1} = \mathrm{U}[kb, b]$	$F_{3,1} = \mathrm{U}[-c, mc]$	$l_{2,1} + l_{3,1} = l_{1,1}$
(ii)	$F_{1,2} = \mathrm{U}[-ka, a]$	$F_{2,2} = \mathrm{U}[-b, -kb]$	$F_{3,2} = \mathrm{U}[-mc, c]$	$l_{2,2} + l_{3,2} = l_{1,2}$
(iii)	$F_{1,3} = \mathrm{U}[ka, a]$	$F_{2,3} = \mathrm{U}[-b, kb]$	$F_{3,3} = \mathrm{U}[-c, -mc]$	$l_{1,3} + l_{3,3} = l_{2,3}$
(iv)	$F_{1,4} = \mathrm{U}[-a, -ka]$	$F_{2,4} = \mathrm{U}[-kb, b]$	$F_{3,4} = \mathrm{U}[mc, c]$	$l_{1,4} + l_{3,4} = l_{2,4}$
(v)	$F_{1,5} = \mathrm{U}[-a, a]$	$F_{2,5} = \mathrm{U}[-b, b]$	$F_{3,5} = \mathrm{U}[-mc, mc]$	$l_{2,5} + l_{3,5} = l_{1,5}$
(vi)	$F_{1,6} = \mathrm{U}[-ka, ka]$	$F_{2,6} = \mathrm{U}[-kb, kb]$	$F_{3,6} = \mathrm{U}[-c, c]$	$l_{1,6} + l_{2,6} = l_{3,6}$

Note that the sum of the means of each triplet $(F_{1,j}, F_{2,j}, F_{3,j})$ is 0, and by Lemma 5.1, $(F_{1,j}, F_{2,j}, F_{3,j})$ is JM for $j = 1, \ldots, 6$. Now we seek a way to write each F_i as a combination of $F_{i,j}$, $i = 1, 2, 3$, $j = 1, \ldots, 6$. Let

$$q = \frac{1}{4}\left(\frac{1}{1-k^2} + \frac{1}{1-m^2} - 1\right)^{-1},$$

$q_1 = q\frac{4m^2}{1-m^2}$ and $q_2 = q\frac{4k^2}{1-k^2}$. Then

$$4q + q_1 + q_2 = \frac{1}{4}\left(\frac{1}{1-k^2} + \frac{1}{1-m^2} - 1\right)^{-1}\left(4 + \frac{4k^2}{1-k^2} + \frac{4m^2}{1-m^2}\right) = 1.$$

Denote the distributions

$$F_i^* := \sum_{j=1}^{4} qF_{i,j} + q_1 F_{i,5} + q_2 F_{i,6}.$$

It remains to check for each $i = 1, 2, 3$, $F_i^* = F_i$.

By construction, the density of F_1^* is constant on the three intervals $[-a, -ka]$, $[-ka, ka]$ and $[ka, a]$ respectively. By symmetry, it suffices to check that the density of F_1^* on $[-a, -ka]$ is $\frac{1}{2a}$. The density of F_1^* on $[-a, -ka]$ is

$$q\frac{1}{(1+k)a} + q\frac{1}{(1-k)a} + q\frac{4m^2}{1-m^2}\frac{1}{2a} = \frac{2q}{a}\left(\frac{1}{1-k^2} + \frac{1}{1-m^2} - 1\right) = \frac{1}{2a}.$$

Thus, $F_1^* = F_1$. $F_2^* = F_2$ follows from the same calculation.

By construction, F_3^* is also uniform on the intervals $[-c,-mc]$, $[-mc,mc]$ and $[mc,c]$ respectively. By symmetry, it suffices to check that the density of F_3^* on $[-c,-mc]$ is $\frac{1}{2c}$. The density of F_3^* on $[-c,-mc]$ is

$$q\frac{1}{(1+m)c}+q\frac{1}{(1-m)c}+q\frac{4k^2}{1-k^2}\frac{1}{2c}=\frac{2q}{c}\left(\frac{1}{1-m^2}+\frac{1}{1-k^2}-1\right)=\frac{1}{2c}.$$

Thus, $F_3^*=F_3$.

Finally, by Proposition 2.3 (iv), it follows from

$$F_i:=\sum_{j=1}^{4}qF_{i,j}+q_1F_{i,5}+q_2F_{i,6},\ \text{ for } i=1,2,3,$$

and

$$(F_{1,j},F_{2,j},F_{3,j})\in\mathcal{J}_3(0),\ \text{ for } j=1,\dots,6,$$

that $(F_1,F_2,F_3)\in\mathcal{J}_3(0)$.

Now we complete the theorem for the cases for $n\neq 3$. The case for $n=2$ is trivial. We show the cases for $n>3$ by induction. Suppose $n\geqslant 4$ and (2.4) is satisfied, and without loss of generality assume $l_1\leqslant l_2\leqslant\ldots\leqslant l_n$. Let $l=l_1+l_2$ and $F=\mathrm{U}[-\frac{l}{2},\frac{l}{2}]=\bar{F}$. It is easy to see that $(F,F_3,F_4,\dots,F_n)$ still satisfies the mean condition, and hence $(F,F_3,F_4,\dots,F_n)\in\mathcal{J}_{n-1}(0)$ by inductive hypothesis. By Lemma 5.1, we have that $(F_1,F_2,\bar{F})\in\mathcal{J}_3(0)$. As a consequence, we obtain that $(F_1,\dots,F_n)\in\mathcal{J}_n(0)$ using Proposition 2.3 (vi). $\square$

5.2 Proof of Theorem 3.2

Below we briefly explain the main idea behind the proof of Theorem 3.2. First, we consider distributions with decreasing step density functions. We implement mathematical induction such that in each step, the n-tuple of distributions is written as several pieces of simpler distributions which are "approximately mixable", plus a remaining n-tuple which has a reduced maximum essential support. We refer to this procedure as the *reduction*. The mathematical tools to handle approximations of mixability and combinations of distributions are presented in Lemmas 5.2 and 5.3. Lemmas 5.4 and 5.5 give distances between uniform densities and decreasing densities, and its impact on the approximation of mixability. Lemma 5.6 provides an inequality on the mean of a decreasing density which is useful to show Lemma 5.7. Lemma 5.7 is a technical result which guarantees that one can always repeat the reduction. Lemma 5.8 contains details of the reduction and confirms that the assertion in Theorem 3.2 holds for step density functions and for the case where $n=3$. The results on step density functions in Lemma 5.8 are extended to general decreasing density functions in Lemma 5.9. Proposition 3.3 shows that two decreasing densities can be combined into one, and hence it builds a bridge between the case where $n=3$ and the case where $n>3$. Lemma 5.10 concludes the main result of the theorem.

Since the proof itself involves multiple Lemmas and partial results, in this section we will use the symbols $F_1, \dots, F_n$ repeatedly in different places, and they are not the distributions mentioned in the statement of Theorem 3.2. We first introduce some handy notation.

For $K, C \in \mathbb{R}$ we denote by $\mathcal{M}_n(K, C)$ the set of n-tuples $(F_1, \dots, F_n)$, where $F_1, \dots, F_n$ are univariate distributions, for which there exist n random variables $X_1 \sim F_1, \dots, X_n \sim F_n$ such that $|X_1 + \cdots + X_n - K| \leqslant C$ almost surely. Elements in $\mathcal{M}_n(K, C)$ can be interpreted as "approximately mixable with error C". In particular, $\mathcal{M}_n(K, 0) = \mathcal{J}_n(K)$, and obviously $\mathcal{M}_n(K_1, C_1) \subset \mathcal{M}_n(K_2, C_2)$ if $C_2 - C_1 \geqslant |K_2 - K_1|$.

We define the Wasserstein L^∞-distance between two distributions F and G,

$$d(F, G) = \sup_{t \in [0,1]} \{|F^{-1}(t) - G^{-1}(t)|\}.$$

Here for notational ease we identify a probability distribution F and its cdf. In all the following lemmas, $F, F_1, \dots, F_n$ and $G, G_1, \dots, G_n$ are distributions on $\mathbb{R}$.

Lemma 5.2. *If $(F_1, \dots, F_n) \in \mathcal{M}_n(K, C)$ for some $K, C \in \mathbb{R}$, and $d(F_i, G_i) \leqslant d_i \in \mathbb{R}$ for $i = 1, \dots, n$ then $(G_1, \dots, G_n) \in \mathcal{M}_n(K, C + d_1 + \cdots + d_n)$.*

Lemma 5.3. *If $(F_1, \dots, F_n) \in \mathcal{M}_n(K, C)$ and $(G_1, \dots, G_n) \in \mathcal{M}_n(K, C)$ for some $K, C \in \mathbb{R}$, then $(\lambda F_1 + (1-\lambda)G_1, \dots, \lambda F_n + (1-\lambda)G_n) \in \mathcal{M}_n(K, C)$ for all $\lambda \in [0, 1]$.*

Lemmas 5.2 and 5.3 are straightforward to verify and we omit the proofs here.

Lemma 5.4. *If the essential support of a distribution F is $[0, L]$, $L > 0$, and F admits a decreasing density on $[0, L]$, with mean μ, then $d(F, G) = L - 2\mu$ where $G = \mathrm{U}[0, 2\mu]$.*

Proof. By definition $G^{-1}(t) = 2\mu t$, $t \in [0, 1]$ and F^{-1} is a concave function. We have that for all $a, b, \lambda \in [0, 1]$,

$$F^{-1}(\lambda a + (1-\lambda)b) \leqslant \lambda F^{-1}(a) + (1-\lambda)F^{-1}(b). \tag{5.1}$$

Moreover, $F^{-1}(0) = 0$, $F^{-1}(1) = L$. By (5.1), for a fixed $t_0 \in [0, 1]$, we have that

$$F^{-1}(t_0) \leqslant t_0 L \leqslant t_0(2\mu) + L - 2\mu = G^{-1}(t_0) + L - 2\mu.$$

Let $A = F^{-1}(t_0)$. By (5.1), $F^{-1}(t) \leqslant \frac{t}{t_0}A$ for $0 \leqslant t \leqslant t_0$ and $F^{-1}(t) \leqslant \frac{t-t_0}{1-t_0}L + \frac{1-t}{1-t_0}A$ for $t_0 \leqslant t \leqslant 1$. It follows that

$$\begin{aligned}\mu = \int_0^1 F^{-1}(t)\mathrm{d}t \leqslant \int_0^{t_0} \frac{t}{t_0} A\mathrm{d}t + \int_{t_0}^1 \left(\frac{t-t_0}{1-t_0}L + \frac{1-t}{1-t_0}A\right)\mathrm{d}t &= \frac{t_0 A}{2} + (1-t_0)\frac{A+L}{2} \\ &= \frac{A}{2} + \frac{(1-t_0)L}{2}.\end{aligned}$$

Therefore, $F^{-1}(t_0) \geqslant 2\mu - (1-t_0)L = t_0 L - (L - 2\mu) \geqslant G^{-1}(t_0) - (L - 2\mu)$. Since t_0 is arbitrary, $d(F, G) \leqslant L - 2\mu$. Finally, $d(F, G) = L - 2\mu$ follows from the fact that $F^{-1}(1) - G^{-1}(1) = L - 2\mu$. $\square$

Lemma 5.4 shows that the distance between a decreasing density and a uniform density can be controlled by $L-2\mu$. Later we will use this relationship repeatedly to approximate the mixability of decreasing densities based on the mixability of uniform densities. Lemma 5.5 introduces a class of 2-step density functions $v(\cdot;\alpha,c,\beta,L)$ and give its distance from uniform densities. This class of 2-step density functions will be used later as building blocks for general step density functions.

Lemma 5.5. *Define for real numbers $\alpha \leqslant c \leqslant \beta < L$, a distribution $V(\alpha,c,\beta,L)$ with density function*

$$v(x;\alpha,c,\beta,L) = \frac{1}{\beta-\alpha}\frac{\beta-c}{L-\alpha}\mathrm{I}_{\{x\in[\alpha,\beta)\}} + \frac{1}{L-\beta}\frac{L-\alpha-\beta+c}{L-\alpha}\mathrm{I}_{\{x\in[\beta,L]\}}.$$

Then $V(\alpha,c,\beta,L)$ has mean $(L+c)/2$ and $d(V(\alpha,c,\beta,L),\mathrm{U}[c,L]) = c-\alpha$.

Proof. It suffices to verify that $V(\alpha,c,\beta,L)$ has an increasing density on $[\alpha,L]$, the mean of $V(\alpha,c,\beta,L)$ is $(L+c)/2$, and then apply Lemma 5.4, which holds also for increasing densities by symmetry. $\square$

Lemma 5.6 gives a lower bound for the mean of a decreasing density in a given support with known density values at both endpoints.

Lemma 5.6. *Suppose f is a decreasing density function on $\mathbb{R}$, supported in $[0,L]$, $L\in\mathbb{R}$. For any real number $M \geqslant L$, it is straightforward that $f(0) \geqslant 1/M \geqslant f(M)$. Denote $A = Mf(0)$ and $B = f(M)M$, then the mean μ of f satisfies*

$$\mu \geqslant M\frac{AB+1-2B}{2(A-B)}, \quad \text{if } A > B, \tag{5.2}$$

$$\mu \geqslant \frac{M}{2A}, \tag{5.3}$$

and

$$\mu \geqslant \frac{MB}{2}. \tag{5.4}$$

Proof. Denote the distribution function $F(x) = \int_0^x f(t)\mathrm{d}t$, then $\mu = \int_0^M tf(t)\mathrm{d}t = M - \int_0^M F(x)\mathrm{d}x$.

Since $\frac{B}{M} = f(M) \leqslant f(x) \leqslant f(0) = \frac{A}{M}$ for all $x\in[0,M]$, we have that $F(x) \leqslant \frac{A}{M}x$ and $F(x) \leqslant 1-\frac{B}{M}(M-x) = 1-B+\frac{B}{M}x$. Therefore, for any $T\in[0,M]$, we have

$$\begin{aligned} M-\mu = \int_0^M F(x)\mathrm{d}x &= \int_0^T F(x)\mathrm{d}x + \int_T^M F(x)\mathrm{d}x \\ &\leqslant \frac{1}{2}\frac{A}{M}T^2 + (1-B)(M-T) + \frac{1}{2}\frac{B}{M}(M^2-T^2). \end{aligned} \tag{5.5}$$

Setting $T = \frac{1-B}{A-B}M$, we have that $M-\mu \leqslant \frac{M}{2}\frac{2A-AB-1}{A-B}$ and (5.2) follows. Setting $T = M/A$, we have that $\mu \geqslant \frac{M}{2A} + \frac{1}{2}BM(1-\frac{1}{A})^2 \geqslant \frac{M}{2A}$, and (5.3) follows. Setting $T=0$, (5.4) follows. $\square$

Lemma 5.7 contains the key technical result which ensures that the reduction can be performed repeatedly.

Lemma 5.7. *Suppose for each $i = 1, \dots, n$, f_i is a strictly decreasing density function on its support $[0, L_i]$, $L_i \in \mathbb{R}$, with mean μ_i. Let $M = \mu_1 + \dots + \mu_n$, and suppose $M \geqslant L_i$, $i = 1, \dots, n$. Denote $A_i = f_i(0)M$ and $B_i = f_i(M)M$. There exist $\lambda_1, \dots, \lambda_n \in (0, 1)$ and $D > 0$ such that*

$$\sum_{i=1}^{n} \lambda_i = 1,$$

$$\lambda_i A_i + (1 - \lambda_i) B_i = D, \tag{5.6}$$

for each $i = 1, \dots, n$, and

$$\sum_{i=1}^{n} (1 - \lambda_i) B_i \leqslant D. \tag{5.7}$$

Proof. By (5.6), for any D, it is $\lambda_i = \lambda_i(D) = \frac{D - B_i}{A_i - B_i}$ whenever $A_i \neq B_i$. It is easy to see that if $A_i = B_i$ for some i, then $A_i = B_i = 1$ and $\mu_i = M/2$ which contradicts with the fact that f_i has strictly a decreasing density. Hence $A_i > 1$ for all i. Let

$$W_0 = \sum_{i=1}^{n} \frac{1}{A_i - B_i}, \quad W_1 = \sum_{i=1}^{n} \frac{B_i}{A_i - B_i}, \quad W_2 = \sum_{i=1}^{n} \frac{A_i B_i}{A_i - B_i}.$$

Note that $\lambda_1(D), \dots, \lambda_n(D)$ are well-defined and are increasing functions of D. Solving $\lambda_1 + \dots + \lambda_n = 1$ we obtain

$$1 = W_0 D - W_1 \Leftrightarrow D = \frac{1 + W_1}{W_0}.$$

Now by Lemma 5.6,

$$\lambda_i(B_1) = \frac{B_1 - B_i}{A_i - B_i} \leqslant \frac{B_1}{A_i} \leqslant \frac{4\mu_1 \mu_i}{M^2}$$

for $i = 2, \dots, n$. Then

$$\lambda_1(B_1) + \dots + \lambda_n(B_1) \leqslant 0 + \frac{4\mu_1(\mu_2 + \dots + \mu_n)}{M^2} = \frac{4\mu_1}{M}\left(1 - \frac{\mu_1}{M}\right) < 1.$$

Thus, we have that $D > B_1 \geqslant 0$. Similarly, $D > B_i$ and $\lambda_i > 0$ for all $i = 1, \dots, n$. Note that by Lemma 5.6,

$$1 = \frac{1}{M}(\mu_1 + \dots + \mu_n) \geqslant \sum_{i=1}^{n} \frac{A_i B_i + 1 - 2B_i}{2(A_i - B_i)} = \frac{W_0 + W_2}{2} - W_1,$$

thus $1 + W_1 \geqslant \frac{W_0 + W_2}{2}$, and hence $1 + W_1 \geqslant \sqrt{W_0 W_2}$. It follows that $D = \frac{1 + W_1}{W_0} \geqslant \frac{W_2}{1 + W_1}$. We can now verify that

$$\sum_{i=1}^{n} (1 - \lambda_i) B_i = \sum_{i=1}^{n} \frac{A_i - D}{A_i - B_i} B_i = W_2 - W_1 D \leqslant D.$$

$\square$

Remark 5.1. If we only assume a (non-strict) decreasing density of f_i, $i = 1, \dots, n$, then Lemma 5.7 holds with $\lambda_i \in [0, 1]$, $i = 1, \dots, n$,.

To show Theorem 3.2 for decreasing densities, we start with the case $n = 3$ and consider *decreasing step density functions* as defined below. Suppose for each $i = 1, 2, 3$, f_i is a density function on $[a_0^i, a_{N_i}^i)$, with

$$f_i(x) = \sum_{k=1}^{N_i} h_k^i \mathrm{I}_{[a_{k-1}^i, a_k^i)}(x),$$

where $a_0^i < a_1^i < \cdots < a_{N_i}^i$, and $h_1^i > \cdots > h_{N_i}^i > 0$. Denote $L_i = a_{N_i}^i - a_0^i$. We will assume $a_k^i - a_{k-1}^i \leqslant 1$ for each i, k. The following lemma contains the essential steps to the proof of Theorem 3.2.

Lemma 5.8. *Suppose f_1, f_2, f_3 are decreasing step density functions with mean μ_1, μ_2, μ_3, respectively, and let $M = \mu_1 + \mu_2 + \mu_3 - a_0^1 - a_0^2 - a_0^3$. If $L_i \leqslant M$, $i = 1, 2, 3$, then $(F_1, F_2, F_3) \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6)$ where F_1, F_2, F_3 are the corresponding distribution functions of f_1, f_2, f_3, respectively.*

Proof. First, we define a quantity N, the number of steps plus one half of the number of "zero-steps at the right-end",

$$N(F_1, F_2, F_3) = N_1 + N_2 + N_3 + \frac{1}{2}(\mathrm{I}_{\{L_1 < M\}} + \mathrm{I}_{\{L_2 < M\}} + \mathrm{I}_{\{L_3 < M\}}).$$

Such N is uniquely determined by F_1, F_2, F_3. We will show the lemma by induction on N. Note that the smallest possible value of N is 4, in which case f_1, f_2 and f_3 are three uniform densities with $L_1 = L_2 + L_3$, and the lemma holds trivially as $(F_1, F_2, F_3) \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 0)$ in this case. In the following we will show that the case of $N = K/2$, $K \in \mathbb{N}$ can be reduced to the case of $N < K/2$.

Without loss of generality we let $a_0^1 = a_0^2 = a_0^3 = 0$. Consequently $a_{N_i} = L_i$, $i = 1, 2, 3$ and $M = \mu_1 + \mu_2 + \mu_3$. First, observe that if $L_1 + L_2 + L_3 - 2M \leqslant 6$, then by Lemma 5.4, we have

$$\sum_{i=1}^{3} d(F_i, \mathrm{U}[0, 2\mu_i]) = \sum_{i=1}^{3} (L_i - 2\mu_i) = L_1 + L_2 + L_3 - 2M \leqslant 6.$$

Note that

$$(\mathrm{U}[0, 2\mu_1], \mathrm{U}[0, 2\mu_2], \mathrm{U}[0, 2\mu_3]) \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 0),$$

and by Lemma 5.2, we have

$$(F_1, F_2, F_3) \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6).$$

Hence, in the following we consider

$$L_1 + L_2 + L_3 - 2M > 6, \tag{5.8}$$

which implies that $M \geqslant L_i > 6$ for all $i = 1, 2, 3$.

By Lemma 5.7, there exist $\lambda_1, \lambda_2, \lambda_3, D \in (0, \infty)$ such that $\lambda_1 + \lambda_2 + \lambda_3 = 1$,

$$\lambda_1 f_1(0) + (1 - \lambda_1) f_1(M) = \lambda_2 f_2(0) + (1 - \lambda_2) f_2(M) = \lambda_3 f_3(0) + (1 - \lambda_3) f_3(M) = D$$

and

$$(1 - \lambda_1) f_1(M) + (1 - \lambda_2) f_2(M) + (1 - \lambda_3) f_3(M) \leqslant D.$$

Denote $D_i = (1 - \lambda_i) f_i(M)$, $i = 1, 2, 3$, and $D_0 = D - D_1 - D_2 - D_3 \geqslant 0$.

Let $t > 0$ be a small number, such that for each i,

$$\lambda_i t \leqslant a_1^i \tag{5.9}$$

and

$$(1 - \lambda_i) t \leqslant \begin{cases} M - L_i & L_i < M, \\ M - a^i_{N_i - 1} & L_i = M. \end{cases} \tag{5.10}$$

Note that all the quantities below depends on t, and later we will determine the value of t. Consider the following three triplets of distributions:

$$U_1 := (\mathrm{U}[M - (1 - \lambda_1)t, M], \mathrm{U}[0, \lambda_2 t], \mathrm{U}[0, \lambda_3 t]) \in \mathcal{M}_3(M, 0),$$

$$U_2 := (\mathrm{U}[0, \lambda_1 t], \mathrm{U}[M - (1 - \lambda_2)t, M], \mathrm{U}[0, \lambda_3 t]) \in \mathcal{M}_3(M, 0),$$

and

$$U_3 := (\mathrm{U}[0, \lambda_1 t], \mathrm{U}[0, \lambda_2 t], \mathrm{U}[M - (1 - \lambda_3)t, M]) \in \mathcal{M}_3(M, 0).$$

Consider the triplet of measures $G = (G_1, G_2, G_3) := (F_1, F_2, F_3) - D_1 t U_1 - D_2 t U_2 - D_3 t U_3$. It is easy to check that G_i is a non-negative measure, with a step density on $[0, M]$, $i = 1, 2, 3$. The first step of G_1 has a height of $f_1(0) - \frac{D_2}{\lambda_1} - \frac{D_3}{\lambda_1} = \frac{D_0}{\lambda_1}$ on $[0, \lambda_1 t]$. Let $I_t^i = [\lambda_i t, M - t + \lambda_i t]$, $i = 1, 2, 3$. We can see that $G_1 = F_1$ on I_t^1 and hence G_1 has a decreasing step density function on I_t^i. Moreover, G_1 assigns zero measure on $[M - t + \lambda_1 t, M]$. It holds similarly for G_2 and G_3.

We consider two possibilities: if $D_0 = 0$, then G_1, G_2, G_3 all have step densities on I_t^1, I_t^2, I_t^3, respectively. Note that the lengths of I_t^1, I_t^2, I_t^3 are all $M - t$, and $D_0 = 0$ implies that at least one of L_1, L_2, L_3 is equal to M. We can take t as the largest number such that (5.9)-(5.10) hold. In that case, one of the six inequalities in (5.9)-(5.10) is an equality. Now we compare (F_1, F_2, F_3) with (G_1, G_2, G_3). At least one of G_1, G_2, G_3 has fewer steps compared to F_1, F_2, F_3. Let $\hat{G} = G/(1 - D_1 t - D_2 t - D_3 t)$ be a triplet of distributions. The sum of the means of the three components of $\hat{G}$ is still M, since F, U_1, U_2, U_3 all satisfy this property. We have that $N(\hat{G}) < N(F_1, F_2, F_3)$ and therefore $\hat{G} \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6)$ by inductive hypothesis. Hence, by Lemma 5.3,

$$(F_1, F_2, F_3) = (1 - D_1 t - D_2 t - D_3 t)\hat{G} + D_1 t U_1 + D_2 t U_2 + D_3 t U_3 \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6).$$

In the following we consider the more complicated case that $D_0 > 0$. Intuitively, one wants to further find H, U such that $G = H + U$ where the components of H has decreasing densities and fewer steps compared to those of F_1, F_2, F_3 (hence an inductive argument can be applied), and $U \in \mathcal{M}_3(M, 0)$ based on some prior results.

In the following we define a few quantities L'_i, c_i^j, α_i^j and β_i^j where i, j are two distinct elements of $\{1,2,3\}$. For i, j, k being distinct elements of $\{1,2,3\}$, let

$$L'_i = \begin{cases} L_i & L_i < M, \\ M - (1-\lambda_i)t & L_i = M, \end{cases}$$

that is, L'_i is the right-end-point of the essential support of the measure G_i. Let $c_i^j = M - L'_k - \frac{1}{2}\lambda_i t$, $\beta^j = a^j_{N_j - 2}$ (note that $N_j \geqslant 6$ since $L_j \geqslant 6$), and

$$\alpha_i^j = \max\{\lambda_j t, a_l^j : a_l^j \leqslant M - L'_k - 1,\ l = 1, \dots, N_j\}.$$

We can check that $c_i^j \geqslant \lambda_j t \geqslant 0$: if $L_k < M$, then $L'_k = L_k$, $(1-\lambda_k)t \leqslant M - L'_k$ by (5.10), and hence $c_i^j = M - L'_k - \frac{1}{2}\lambda_i t \geqslant (1-\lambda_k)t - \frac{1}{2}\lambda_i t \geqslant \lambda_j t$; if $L_k = M$, then $L' = M - (1-\lambda_i)t$, and hence $c_i^j = M - L'_k - \frac{1}{2}\lambda_i t = (1-\lambda_k)t - \frac{1}{2}\lambda_i t \geqslant \lambda_j t$.

Consider the triplet of distributions

$$U_4 = (\mathrm{U}[0, \lambda_1 t], \mathrm{U}[c_1^2, L'_2], \mathrm{U}[c_1^3, L'_3]).$$

We can check that $\mathrm{U}[c_1^2, L'_2]$, $\mathrm{U}[c_1^3, L'_3]$ have the same length $L'_2 + L'_3 + \frac{1}{2}\lambda_1 t - M$, and $U_4 \in \mathcal{M}_3(M, 0)$. We verify that by (5.8) and (5.9),

$$\begin{aligned} L'_2 + L'_3 + \frac{1}{2}\lambda_1 t - M &\geqslant L_2 + L_3 + \frac{1}{2}\lambda_1 t - (1-\lambda_2)t - (1-\lambda_3)t - M \\ &\geqslant 6 - \left(\frac{1}{2}\lambda_1 + \lambda_2 + \lambda_3\right)t \\ &> 2. \end{aligned}$$

Let (recall the definition of $V[\alpha, c, \beta, L]$ in Lemma 5.5)

$$U_5 = (\mathrm{U}[0, \lambda_1 t], V[\alpha_1^2, c_1^2, \beta^2, L'_2], V[\alpha_1^3, c_1^3, \beta^3, L'_3]).$$

It is easy to check that $\alpha_1^j \leqslant c_1^j$, $c_1^j - \alpha_1^j \leqslant 2$, and $c_1^j < \beta^j < L'_j$ (note that $L'_j - c_1^j > 2$ and $L'_j - \beta^j \leqslant 2$), $j = 2, 3$. By Lemmas 5.2 and 5.5, and the fact that $U_4 \in \mathcal{M}_3(M, 0)$, we have that $U_5 \in \mathcal{M}_3(M, 4)$. Similarly, by denoting

$$U_6 = (V[\alpha_2^1, c_2^1, \beta^1, L'_1], \mathrm{U}[0, \lambda_2 t], V[\alpha_2^3, c_2^3, \beta^3, L'_3]),$$

and

$$U_7 = (V[\alpha_3^1, c_3^1, \beta^1, L'_1], V[\alpha_3^2, c_3^2, \beta^2, L'_2], \mathrm{U}[0, \lambda_3 t]),$$

we have that $U_6, U_7 \in \mathcal{M}_3(M, 4)$.

Now we consider $H = G - D_0 t(U_5 + U_6 + U_7)$. To guarantee that H is a triplet of non-negative measures, we let t satisfy for $\{i, j, k\} = \{1, 2, 3\}$,

$$D_0 t \left(\frac{1}{L_i' - \beta^i} \frac{L_i' - \alpha_j^i - \beta^i + c_j^i}{L_i' - \alpha} + \frac{1}{L_i' - \beta^i} \frac{L_i' - \alpha_k^i - \beta^i + c_k^i}{L_i' - \alpha} \right) \leqslant h_{N_i}^i, \tag{5.11}$$

in addition to (5.9)-(5.10).

With t satisfying (5.9)-(5.11), it is easy to verify that $\hat{H} = (\hat{H}_1, \hat{H}_2, \hat{H}_3) := H/(1 - 3D_0 t - D_1 t - D_2 t - D_3 t)$ is a triplet of distributions. The sum of the means of the three components of $\hat{H}$ is still M, since G, U_5, U_6, U_7 all satisfy this property. It is straightforward that for $i = 1, 2, 3$, the essential support of $\hat{H}_i$ is $[\lambda_i t, L_i'] \subset I_t^i$ if (5.11) is a strict inequality, and it is $[\lambda_i t, a_{N_i - 1}^i] \subset I_t^i$ if (5.11) is an equality.

Now we take t as the largest number such that (5.9)-(5.11) are satisfied. In that case, at least one of the inequalities in (5.9)-(5.11) becomes an equality. We call the transformation from (F_1, F_2, F_3) to $\hat{H}$ an *operation*.

(i) If one inequality in (5.11) is an equality, then the last non-zero-step in F_i for some $i = 1, 2, 3$ vanishes after the operation, while creating at most one "zero-step at the right-end" in $\hat{H}$.

(ii) If each inequality in (5.11) is not an equality, then no extra "zero-steps at the right-end" will be created after the operation, and

  (a) if one inequality in (5.9) is an equality, then the first step in F_i for some $i = 1, 2, 3$ vanishes after the operation;

  (b) if one inequality in (5.10) is an equality, then either the last non-zero-step in F_i for some $i = 1, 2, 3$ vanishes after the operation, or the last "zero-step at the right-end" in F_i for some $i = 1, 2, 3$ vanishes in the operation.

In each case, it follows that $N(\hat{H}) < N(F_1, F_2, F_3)$ and therefore $\hat{H} \in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6)$ by inductive hypothesis. By Lemma 5.3,

$$\begin{aligned}(F_1, F_2, F_3) &= (1 - 3D_0 t - D_1 t - D_2 t - D_3 t)\hat{H} + D_0 t(U_5 + U_6 + U_7) + D_1 t U_1 + D_2 t U_2 + D_3 t U_3 \\ &\in \mathcal{M}_3(\mu_1 + \mu_2 + \mu_3, 6).\end{aligned}$$

Now we finish the proof of the lemma by induction. $\square$

One can extend the above results on decreasing step density functions to general decreasing density functions. The result is given in Lemma 5.9.

Lemma 5.9. *Suppose that distributions F_1, F_2, F_3, with supports contained in $[0,1]$ and means μ_1, μ_2, μ_3, respectively, have decreasing density functions on $[0,1]$, and $\mu_1+\mu_2+\mu_3 = 1$. Then $(F_1, F_2, F_3) \in \mathcal{J}_3(1)$.*

Proof. First assume that F_1, F_2, F_3 have strictly decreasing density functions on $[0,1]$. Let $X_1 \sim F_1$, $X_2 \sim F_2$, $X_3 \sim F_3$ and $U \sim \mathrm{U}[0,1]$ independent of X_1, X_2, X_3. For an $m \in \mathbb{N}$, denote $Y_i = \lfloor mX_i \rfloor + U$, $i = 1,2,3$. Let G_1, G_2, G_3 be the distributions of Y_1, Y_2, Y_3, respectively. Note that $\nu_m := \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \mathbb{E}[Y_3] \geqslant \mathbb{E}[mX_1] + \mathbb{E}[mX_2] + \mathbb{E}[mX_3] = m$ since the densities of F_1, F_2, F_3 are decreasing. On the other hand, $\nu_m \leqslant \mathbb{E}[mX_1+1] + \mathbb{E}[mX_2+1] + \mathbb{E}[mX_3+1] = m+3$. It is easy to check that G_i is supported on $[0,m]$ and satisfy the conditions of Lemma 5.8, and hence $(G_1, G_2, G_3) \in \mathcal{M}_3(\nu_m, 6) \subset \mathcal{M}_3(m, 9)$. Let $\hat{F}_i$ be the distribution of mX_i, $i = 1,2,3$. By $|mX_i - Y_i| \leqslant 1$, we have that $d(\hat{F}_i, G_i) \leqslant 1$ and hence by Lemma 5.2, we have that $(\hat{F}_1, \hat{F}_2, \hat{F}_3) \in \mathcal{M}_3(m, 12)$. Therefore, $(F_1, F_2, F_3) \in \mathcal{M}_3(1, \frac{12}{m})$. Since m is arbitrary, by a compactness argument, we conclude that $(F_1, F_2, F_3) \in \mathcal{M}_3(1, 0)$. If the density functions of F_1, F_2, F_3 are not strictly decreasing on $[0,1]$, we can always find a sequence of strictly decreasing densities that converge to F_1, F_2, F_3 with the same mean. By Proposition 2.3 (iii), we obtain that $(F_1, F_2, F_3) \in \mathcal{M}_3(1,0) = \mathcal{J}_3(1)$. □

Next, we show Proposition 3.3, which will serve as another key step in the proof of Theorem 3.2; it allows one to go from $n = 3$ to $n > 3$. Proposition 3.3 states that "Suppose that F_1 and F_2 have decreasing density functions on their supports $[0, L_1]$ and $[0, L_2]$, with means μ_1 and μ_2, respectively. There exists a distribution F_3 with an increasing density on $[-M, 0]$, where $M = \max\{L_1, L_2, 2\mu_1 + 2\mu_2\}$, such that $(F_1, F_2, F_3) \in \mathcal{J}_3(0)$."

Proof of Proposition 3.3. We prove this proposition for the case of step density functions. Suppose that F_1 and F_2 have step density functions f_1, f_2, respectively, such that

$$f_i(x) = \sum_{k=1}^{N_i} h_k^i \mathrm{I}_{[a_{k-1}^i, a_k^i)}(x),$$

$i = 1,2$, where $0 = a_0^i < a_1^i < \cdots < a_{N_i}^i = L_i$, and $h_1^i \geqslant \ldots \geqslant h_{N_i}^i > 0$. Here, we do not assume $a_k^i - a_{k-1}^i \leqslant 1$.

We show the lemma for step density functions by induction on $N = N_1 + N_2$. For the case $N_1 = N_2 = 1$, F_1 and F_2 are uniform distributions. Take $F_3 = \mathrm{U}[-M, 0]$, $M = L_1 + L_2 = 2\mu_1 + 2\mu_2$. By Theorem 3.1 we have that $(F_1, F_2, F_3) \in \mathcal{J}_3(0)$, that is, the lemma holds true in this case.

Now assume the lemma holds for step density functions for $N < K$. We will show the case for $N = K$ can be reduced to $N = K - 1$.

If $M = 2\mu_1 + 2\mu_2 \geqslant L_i$, $i = 1, 2$, then we again take $F_3 = \mathrm{U}[-M, 0]$, then the sum of the means of F_1, F_2 and F_3 is 0. By Lemma 5.9 we obtain that $(F_1, F_2, F_3) \in \mathcal{J}_3(0)$. Thus, the lemma holds true in this case.

In the following, without loss of generality we assume $M = L_1 \geqslant L_2$, $M > 2\mu_1 + 2\mu_2$. Similar to the proof of Lemma 5.8 we assume $h_1^i > \cdots > h_{N_i}^i > 0$, $i = 1, 2$. Consider the following set of decreasing step probability density functions (pdf):

$$Q_i = \left\{ f \text{ is pdf} : f(x) = \sum_{k=1}^{N_i} g_k^i \mathrm{I}_{[a_{k-1}^i, a_k^i)}(x) : g_1^i \geqslant \ldots \geqslant g_{N_i}^i > 0 \right\}, \quad i = 1, 2.$$

Let $\mu(\cdot)$ be the mean of a distribution function or a density function. For any $f \in Q_i$, it is obvious that $\mu(f)$ takes value in $(\frac{a_1^i}{2}, \frac{a_{N_i}^i}{2}]$. Note that $\frac{a_1^1}{2} + \frac{a_1^2}{2} \leqslant \mu_1 + \mu_2 < \frac{L_1}{2}$ and $\frac{a_{N_1}^1}{2} + \frac{a_{N_2}^2}{2} = \frac{L_1}{2} + \frac{L_2}{2} > \frac{L_1}{2}$. Therefore, there exist $g_1 \in Q_1$, $g_2 \in Q_2$ such that $\mu(g_1) + \mu(g_2) = \frac{L_1}{2} = \frac{M}{2}$. Let G_1 and G_2 be the distribution functions of g_1 and g_2, respectively. By Lemma 5.9 we have that $(G_1, G_2, G_3) \in \mathcal{J}_3(0)$ where $G_3 = \mathrm{U}[-M, 0]$. For $i = 1, 2$, write

$$g_i(x) = \sum_{k=1}^{N_i} g_k^i \mathrm{I}_{[a_{k-1}^i, a_k^i)}(x),$$

and in addition $h_{N_i+1}^i = g_{N_i+1}^i = 0$. Let

$$\lambda = \min\left\{ \frac{h_r^i - h_{r+1}^i}{g_r^i - g_{r+1}^i} : r = 1, \ldots, N_i, \ i = 1, 2 \right\}.$$

Note that $h_{r+1}^i - h_r^i > 0$ for all r and $g_{r+1}^i - g_r^i > 0$ for at least one r, so $\lambda < \infty$ is well-defined. It is easy to see that $0 < \lambda \leqslant 1$, and $\lambda = 1$ implies that $f_i = g_i$, $i = 1, 2$. This conflicts with $\mu_1 + \mu_2 < M = \mu(g_1) + \mu(g_2)$.

In the following we consider $\lambda < 1$. Let $W_i = \frac{1}{1-\lambda}(F_i - \lambda G_i)$, $i = 1, 2$. We can check that the density function of W_i is again a decreasing step function, by noting that

$$\frac{h_r^i - h_{r+1}^i}{g_r^i - g_{r+1}^i} \geqslant \lambda \quad \Leftrightarrow \quad h_r^i - \lambda g_r^i \geqslant h_{r+1}^i - \lambda g_{r+1}^i \geqslant 0, \quad r = 1, \ldots, N_i - 1, \ i = 1, 2. \tag{5.12}$$

Moreover, one of the inequalities in (5.12) is an equality by the definition of λ; this implies that for $i = 1$ or 2, W_i has at least one less step compared to F_i. Let $[0, L_1']$ and $[0, L_2']$ be the essential supports of W_1 and W_2, respectively, and $M' = \max\{L_1', L_2', 2\mu(W_1) + 2\mu(W_2)\}$. Note that $L_i' \leqslant L_i$, $i = 1, 2$, and $\mu(G_1) + \mu(G_2) > \mu_1 + \mu_2$, implying that $\mu(W_1) + \mu(W_2) < \mu_1 + \mu_2$; hence $M' \leqslant M$. By inductive hypothesis, there exists W_3 with an increasing density on $[-M', 0]$, therefore also on $[-M, 0]$, such that $(W_1, W_2, W_3) \in \mathcal{J}_3(0)$.

Finally, take $F_3 = \lambda G_3 + (1 - \lambda) W_3$, and also note that $F_i = \lambda G_i + (1 - \lambda) W_i$, $i = 1, 2$. By Proposition 2.3 (i), we have that $(F_1, F_2, F_3) \in \mathcal{J}_3(0)$ and the lemma holds true for all step density functions. By a standard compactness argument, we conclude that the lemma holds true for all decreasing densities. $\square$

Lemma 5.10. *Suppose that distributions $F_1, \dots, F_n$, with supports contained in $[0,1]$ and means $\mu_1, \dots, \mu_n$, respectively, have decreasing density functions on $[0,1]$, and $\mu_1 + \dots + \mu_n = 1$. Then $(F_1, \dots, F_n) \in \mathcal{J}_n(1)$.*

Proof. The case for $n = 2$ is trivial as the assumption that $\mu_1 + \mu_2 = 1$ implies $F_1 = F_2 = \mathrm{U}[0,1]$. The case for $n = 3$ is shown in Lemma 5.9.

Now we consider $n \geqslant 4$, and we will show the lemma by induction. Without loss of generality we assume $\mu_1 \leqslant \mu_2 \leqslant \dots \leqslant \mu_n$. Since $\mu_1 + \dots + \mu_n = 1$, we have that $\mu_1 + \mu_2 \leqslant 1/2$. By Proposition 3.3, there exists a distribution G with an increasing density on $[-1,0]$ such that $(F_1, F_2, G) \in \mathcal{J}_3(0)$. Note that $\bar{G}$ has a decreasing density on $[0,1]$, and $\mu(\bar{G}) = -\mu(G) = \mu_1 + \mu_2$. Hence, by inductive hypothesis, $(\bar{G}, F_3, \dots, F_n) \in \mathcal{J}_{n-1}(1)$. By Proposition 2.3 (vi), $(F_1, F_2, G) \in \mathcal{J}_3(0)$ and $(\bar{G}, F_3, \dots, F_n) \in \mathcal{J}_{n-1}(1)$ imply that $(F_1, \dots, F_n) \in \mathcal{J}_n(1)$. □

Finally, using Lemma 5.10, we are able to complete the proof of Theorem 3.2.

Proof of Theorem 3.2. Since joint mixability is invariant under linear transformations (Proposition 2.3 (v)), we can assume $a_1 = \dots = a_n = 0$ without loss of generality. The mean condition (2.2) reads as

$$\mu_1 + \dots + \mu_n \geqslant \max_{i=1,\dots,n} l_i.$$

Now F_i has decreasing density functions on $[0, l_i] \subset [0, \mu_1 + \dots + \mu_n]$. By Lemma 5.10 and the invariance of joint mixability under linear transformations, we conclude that $(F_1, \dots, F_n) \in \mathcal{J}_n(\mu_1 + \dots + \mu_n)$. □

5.3 Proof of Theorem 3.4

Proof. For each $i = 1, \dots, n$, it is easy to check that $G_i : \mathbb{R}^+ \to [0, \frac{1}{2}]$ is a non-decreasing function, hence $G_i^{-1}(a)$ for $a \in (0, \frac{1}{2})$ is properly defined. Let $m \in \mathbb{N}$ and $U_{i,j}^m$ be the cdf of the uniform distribution on $[-G_i^{-1}(\frac{j}{m+1}), G_i^{-1}(\frac{j}{m+1})]$ for $i = 1, \dots, n$, $j = 1, \dots, m$. By Theorem 3.1 and (3.1) we know that $(U_{1,j}^m, \dots, U_{n,j}^m) \in \mathcal{J}_n(0)$ for $j, m \in \mathbb{N}$ and $j \leqslant m$.

Define distributions

$$F_i^m = \sum_{j=1}^{m} \frac{1}{m} U_{i,j}^m$$

for $i = 1, \dots, n$, $m \in \mathbb{N}$. By Proposition 2.3 (iv) we have for $m \in \mathbb{N}$

$$(F_1^m, \dots, F_n^m) \in \mathcal{J}_n(0). \tag{5.13}$$

We will show that $(F_1^m,\dots,F_n^m)\to(F_1,\dots,F_n)$ weakly as $m\to\infty$. For $x>0$, we calculate

$$\begin{aligned}F_i^m(x)=&\sum_{j=1}^m\frac{1}{m}U_{i,j}^m(x)\\=&\frac{1}{2}+\frac{1}{2}\sum_{j=1}^m\frac{1}{m}\left(\mathrm{I}_{\{G_i^{-1}(\frac{j}{m+1})\leqslant x\}}+\mathrm{I}_{\{G_i^{-1}(\frac{j}{m+1})>x\}}\frac{x}{G_i^{-1}(\frac{j}{m+1})}\right).\end{aligned}$$

Note that

$$\lim_{m\to\infty}\sum_{j=1}^m\frac{1}{m}\mathrm{I}_{\{G_i^{-1}(\frac{j}{m+1})\leqslant x\}}=G(x),$$

and

$$\begin{aligned}\lim_{m\to\infty}\sum_{j=1}^m\frac{1}{m}\mathrm{I}_{\{G_i^{-1}(\frac{j}{m+1})>x\}}\frac{x}{G_i^{-1}(\frac{j}{m+1})}&=\int_{G_i(x)}^{\frac{1}{2}}\frac{x}{G_i^{-1}(t)}\mathrm{d}t\\&=\int_x^\infty\frac{x}{y}\mathrm{d}G(y)\\&=\int_x^\infty\frac{x}{y}(-yp'(y))\mathrm{d}y\\&=\int_x^\infty -x\mathrm{d}p(y)\\&=xp(x).\end{aligned}$$

Thus, we have for $i=1,\dots,n$,

$$\lim_{m\to\infty}F_i^m=\frac{1}{2}+\frac{1}{2}(G(x)+xp(x))=\frac{1}{2}+\int_0^x p_i(t)\mathrm{d}t=F_i(x). \tag{5.14}$$

By Proposition 2.3 (iii), (5.13) and (5.14), we conclude that $(F_1,\dots,F_n)\in\mathcal{J}_n(0)$. $\square$

5.4 Proof of Theorem 3.7

Proof. We only need to prove the case for $n\geqslant 2$. Without loss of generality, we assume $\mu_j=0$ and $\sigma_1\geqslant\sigma_2\geqslant\dots\geqslant\sigma_n$.

Generally, if $\mathbf{Y}\sim E_n(\mu,\Sigma,\phi)$, then for any $\mathbf{b}\in\mathbb{R}^n$,

$$\mathbf{b}^\top\mathbf{Y}\sim E_1(\mathbf{b}^\top\mu,\mathbf{b}^\top\Sigma\mathbf{b},\phi). \tag{5.15}$$

Denote

$$\mathcal{T}_{n-1}=\{\text{positive semi-definite } (n-1)\times(n-1) \text{ matrix with diagonal entries } \sigma_2^2,\sigma_3^2,\dots,\sigma_n^2\}.$$

Denote the $(n-1)$-vector $\mathbf{1}=(1,\dots,1)^\top$. Define a function

$$f(T)=\sigma_1-\sqrt{\mathbf{1}^\top T\mathbf{1}},\qquad T\in\mathcal{T}_{n-1}.$$

Obviously $f(T)$ is a continuous function of T over $\mathcal{T}_{n-1}$ with respect to a standard matrix metric (such as Frobenius metric). Take $T=(\sigma_2,\dots,\sigma_n)^\top(\sigma_2,\dots,\sigma_n)\in\mathcal{T}_{n-1}$. We have that

$f(T) = \sigma_1 - \sum_{j=2}^{n} \sigma_j \leqslant 0$ by (2.5). On the other hand, let $\hat{\sigma}_j = (-1)^j \sigma_j$ for $j = 2, \ldots, n$ and take $T = (\hat{\sigma}_2, \ldots, \hat{\sigma}_n)^\top (\hat{\sigma}_2, \ldots, \hat{\sigma}_n) \in \mathcal{T}_{n-1}$. We have that $f(T) = \sigma_1 - \sum_{j=2}^{n} \hat{\sigma}_j = \sum_{j=1}^{n} (-1)^{j+1} \sigma_j \geqslant 0$. Hence there exists $T_0 \in \mathcal{T}_{n-1}$ such that $f(T_0) = 0$. Now we let $(X_2, \ldots, X_n) \sim E_{n-1}(0, T_0, \phi)$, and by (5.15) we have that $X_2 + \cdots + X_n \sim E_1(0, \mathbf{1}^\top T_0 \mathbf{1}, \phi) = E_1(0, \sigma_1^2, \phi)$. Let $X_1 = -(X_2 + \cdots + X_n)$. It follows that $X_1 \sim E_1(0, \sigma_1^2, \phi)$ and $X_1 + \cdots + X_n \sim E_1(0, 0, \phi)$. Thus, $(F_1, \ldots, F_n) \in \mathcal{J}_n(0)$. $\square$

6 Conclusion

In this paper, we introduce the theory of joint mixability. We provide necessary conditions for joint mixability: (a) mean inequality, and (b) norm inequality. As special cases of (b), two extra conditions that can be easily checked, (c) length inequality and (d) variance inequality are also introduced. It is shown that any of (a)-(d) is sufficient for the joint mixability of a tuple of uniform distributions; (a) is sufficient for the joint mixability of a tuple of distributions with monotone densities; (d) is sufficient for the joint mixability of a tuple of elliptical distributions with the same characteristic generator; and a stronger condition of type (b) is sufficient for the joint mixability of a tuple of distributions with unimodal-symmetric densities. Our results partially solve long-time existing open questions in the literature of multivariate distributions with fixed margins.

This paper concerns questions of possible distributions of $X_1 + \cdots + X_n$ when $X_1, \ldots, X_n$ have fixed marginal distributions. Instead of $X_1 + \cdots + X_n$, one may more generally consider the possible distributions of $f(X_1, \ldots, X_n)$, where $f : \mathbb{R}^n \to \mathbb{R}^k$ is a general function ($k \in \mathbb{N}$). Substantial challenges are expected for this type of questions at this moment; see recent discussions in Bignozzi and Puccetti (2015).

Acknowledgement

The authors would like to thank Carole Bernard, Paul Embrechts, Harry Joe, Giovanni Puccetti, Gennady Samorodnitsky, the Area Editor, an Associate Editor, and two anonymous referees for helpful comments which have led to significant improvements to the paper. B. Wang thanks the Department of Statistics and Actuarial Science at the University of Waterloo for kind hospitality and financial support during his visit to Waterloo in 2015. R. Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC Grant No. 435844).

References

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108. 1, 4.1, 4.3

Bignozzi, V. and Puccetti, G. (2015). Studying mixability with supermodular aggregating functions. *Statistics and Probability Letters*, forthcoming. 6

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**(3), 368–385. 3.2, 3.2, 3.2

Cheung, K. C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance: Mathematics and Economics*, **55**, 180–190. 4.1

Cuadras, C. M. and Fortiana J. E. (eds.) (2002). *Distributions with given marginals and statistical modeling.* Kluwer, Netherlands. 1

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, **9**, 131–150. 1

Dall'Aglio, G., Kotz, S. and Salinetti, G. (eds.) (1991). *Advances in probability distributions with given marginals: beyond the copulas*, Springer, Netherlands. 1

Dhaene, J. and Denuit M. (1999). The safest dependence structure among risks. *Insurance: Mathematics Economics*, **25**(1), 11–21. 4.1

Embrechts, P and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, **10**, 341–352. 4.2

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764. 1, 4.2

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48. 1, 3.1, 4.2

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, forthcoming. 1

Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric multivariate and related distributions.* London: Chapman & Hall. 3.2

Föllmer, H. and Schied, A. (2011). *Stochastic finance: An introduction in discrete time.* Walter de Gruyter, Third Edition. 4.1

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. University Lyon: Sect. A, (3)* **14**, 53–77. 1

Gaffke, N. and Rüschendorf, L. (1981). On a class of extremal problems in statistics. *Math. Operationsforsch. Statist. Ser. Optimization*, **12**, 123–136. 1

Haus, U. (2015). Bounding stochastic dependence, complete mixability of matrices, and multi-dimensional bottleneck assignment problems. *Operations Research Letters*, **43**(1), 74-?9. 1, 4

Hoeffding, W. (1940). Masstabinvariante korrelationstheorie. *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik des Universitat Berlin*, **5**, 181–233. 1

Jakobsons, E., Han, X. and Wang, R. (2015). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, forthcoming. DOI: `http://dx.doi.org/10.1080/03461238.2015.1012223`. 4.2, 4.3

Joe, H. (1990). Multivariate concordance. *Journal of Multivariate Analysis*, **35**(1), 12–30. 3.2

Joe, H. (1997). *Multivariate models and dependence concepts.* London: Chapman & Hall. 1, 2.2

Lindskog, F., McNeil, A. and Schmock, U. (2003). Kendalls tau for elliptical distributions. In: Bohl, G., Nakhaeizadeh, G., Rachev, S.T., Ridder, T., Vollmer, K.H. (Eds.), *Credit risk: Measurement, evaluation and management*, 649-?76. Physica-Verlag, Heidelberg. 3.2

Makarov, G. D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and its Applications*, **26**, 803–806. 1

Müller, A. and Stoyan, D. (2002). *Comparison methods for statistical models and risks.* Wiley, England. 1

Puccetti, G., Wang, B. and Wang, R. (2012). Advances in complete mixability. *Journal of Applied Probability,* **49**(2), 430–440. 1, 2.1, 2.2

Puccetti, G., Wang, B. and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, **53**(3), 821–828. 1, 2.1

Puccetti, G. and Wang, R. (2015a). Detecting complete and joint mixability. *Journal of Computational and Applied Mathmatics*, **280**, 174–187. 1, 3.1

Puccetti, G. and Wang, R. (2015b). Extremal dependence concepts. *Statistical Science*, forthcoming. Available at SSRN: `http://ssrn.com/abstract=2524282`. 1

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability* **14**, 623–632. 4.2

Rüschendorf, L. (2013). *Mathematical risk analysis. Dependence, risk bounds, optimal allocations and portfolios*. Springer, Heidelberg. 1

Rüschendorf, L. and Uckelmann, L. (2002). Variance minimization and random variables with constant sum. In *Distributions with given marginals and statistical modelling*, pp. 211–222. Kluwer, Netherlands. 1, 3.1

Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, **36**, 423–439. 1

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability,* **8**(4), 814–827. 1, 1

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, **102**, 1344–1360. 1, 2.1, 2.2, 3.1, 4.2

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417. 1, 2.1, 2.2, 3.1, 3.2, 4.2, 4.2