Journal of Theoretical Probability
https://doi.org/10.1007/s10959-023-01277-2

The Distributions of the Mean of Random Vectors with Fixed Marginal Distribution

Andrzej Komisarski[1] · Jacques Labuschagne[2]

Received: 8 May 2023 / Revised: 29 June 2023 / Accepted: 10 July 2023

Abstract
Using recent results concerning non-uniqueness of the center of the mix for completely mixable probability distributions, we obtain the following result: For each $d \in \mathbb{N}$ and each non-empty bounded Borel set $B \subset \mathbb{R}^d$, there exists a d-dimensional probability distribution $\boldsymbol{\mu}$ satisfying the following: For each $n \geq 3$ and each probability distribution $\boldsymbol{\nu}$ on B, there exist d-dimensional random vectors $\mathbf{X}_{\boldsymbol{\nu},1}, \mathbf{X}_{\boldsymbol{\nu},2}, \ldots, \mathbf{X}_{\boldsymbol{\nu},n}$ such that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \boldsymbol{\nu}$ and $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ for $i = 1, 2, \ldots, n$. We also show that the assumption regarding the boundedness of the set B cannot be completely omitted, but it can be substantially weakened.

Keywords Sums of random vectors · Distributions of sums of random variables · Multivariate dependence · Complete mixability

Mathematics Subject Classification (2020) 60E05 · 60E15 · 62H05

1 Introduction

Let μ be a probability distribution on $\mathbb{R}$ and $n \in \mathbb{N}$. We say that μ is *n-completely mixable* if there exists a random vector $X = (X_1, X_2, \ldots, X_n)$ such that for each $i = 1, 2, \ldots, n$ the random variable $X_i \sim \mu$ (i.e., X_i has distribution μ) and $\sum_{i=1}^{n} X_i$

A. Komisarski and J. Labuschagne have contributed equally to this work.

✉ Andrzej Komisarski
andrzej.komisarski@wmii.uni.lodz.pl

Jacques Labuschagne
jacques.labuschagne@edu.uni.lodz.pl

1 Department of Probability Theory and Statistics, Faculty of Mathematics and Computer Science, University of Łódź, ul.Banacha 22, 90-238 Łódź, Poland

2 Doctoral School of Exact and Natural Sciences, University of Łódź, ul.Matejki 21/23, 90-237 Łódź, Poland

Published online: 25 September 2023

is a.s. constant. If such a vector X exists, then it is called a *complete mix* (or a *mix* for short). The number $C = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the *center of the mix*.

The problem of n-complete mixability is mostly theoretical in nature but it became popular because of its connections with applications, especially with risk theory. For more details, see, e.g., [1, 2, 4, 6].

Obviously, only one-point probability distributions are 1-completely mixable and each symmetric distribution is 2-completely mixable (the symmetry is the necessary and sufficient condition for 2-complete mixability). However, for $n \geq 3$ the problem of characterizing n-complete mixability is only partially solved and is still an object of thorough research (see [9] for an exhaustive list of references). In particular, the following sufficient conditions for n-complete mixability have been established:

If μ has a symmetric and unimodal density, then μ is n-completely mixable for all $n \geq 2$ ([7]).

If μ has a density, which is monotone on its support $[a, b] \subset \mathbb{R}$, then μ is n-completely mixable if and only if the expected value $E(\mu)$ is in $[a + \frac{b-a}{n}, b - \frac{b-a}{n}]$ ([8]).

If μ has a density, which is concave on its support, then μ is n-completely mixable for all $n \geq 3$ ([5]).

If μ has a density f on a finite interval $[a, b]$ and $f(x) \geq \frac{3}{n(b-a)}$, then μ is n-completely mixable ([6]).

For a long time, it was not known whether the center of the mix is always unique for an n-completely mixable probability distribution (for $n \geq 3$). The problem was recently solved in [3]. The center of the mix is not necessarily unique. In our paper, we use that fact to prove that for each $d \in \mathbb{N}$ and each non-empty bounded Borel set $B \subset \mathbb{R}^d$ there exists a d-dimensional probability distribution $\boldsymbol{\mu}$ satisfying what follows: For each $n \geq 3$ and each probability distribution $\boldsymbol{\nu}$ on B, there exist d-dimensional random vectors $\mathbf{X}_{\boldsymbol{\nu},1}, \mathbf{X}_{\boldsymbol{\nu},2}, \ldots, \mathbf{X}_{\boldsymbol{\nu},n}$ such that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \boldsymbol{\nu}$ and $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ for $i = 1, 2, \ldots, n$ (see Theorem 1). The assumption about the boundedness of the support of measure $\boldsymbol{\nu}$ can be replaced with the weaker assumption about the concentration of measure $\boldsymbol{\nu}$ (see Theorem 2 and Corollary 1).

Our results have connections with both risk theory and statistics. Namely, if we have a number of observations from some unknown probability distribution μ, and if we do not know anything about the dependence structure of these observations, then the sample mean is not a good statistic for inference about μ. For the connections with risk theory, see [1, 2, 4, 6].

In Sect. 2, we prove Theorems 1 and 2, and Corollary 1, but first we present some lemmas. The results presented in Sect. 3 show that certain generalizations of Theorem 1 are not possible.

2 The Main Results

Before proving our main results (Theorems 1 and 2), we need to present some auxiliary results.

Lemma 1 *(See [3, Example 2.6].) Let N and M be independent random variables such that $P(M=0)=P(M=1)=\frac{1}{2}$ and $P(N=2^n)=\frac{1}{2^{n+1}}$ for $n=0,1,\dots$. We define random variables $U_0=V_0=N$, $W_0=W_1=-2N$, $U_1=2(1-M)N+M$ and $V_1=2MN+1-M$. Then, $U_0\sim U_1$, $V_0\sim V_1$, and $W_0\sim W_1$ and*

$$U_0+V_0+W_0=0, \qquad U_1+V_1+W_1=1.$$

Moreover, if X is any of random variables U_0, U_1, V_0, V_1, W_0, and W_1, then

$$\frac{1}{2(x+1)} < P(|X|>x) < \frac{2}{x} \quad \text{for } x>0.$$

In particular, $E|X|=\infty$ and $E|X|^\alpha<\infty$ for each $\alpha\in(0,1)$.

Now, we use Lemma 1 and binary representations of numbers from [0, 1] to obtain the following proposition.

Proposition 1 *There exist random variables U_t, V_t, W_t with $t\in[0,1]$, such that*

$$U_t+V_t+W_t=t \quad \text{for each} t\in[0,1]$$

and the distributions of U_t, V_t, and W_t each do not depend on t.

Proof Let $(U_{0,j},V_{0,j},W_{0,j},U_{1,j},V_{1,j},W_{1,j})$ for $j=1,2,\dots$ be a sequence of independent copies of the random vector defined in Lemma 1. We define

$$U_t=\sum_{j=1}^{\infty}\frac{U_{t_j,j}}{2^j}, \qquad V_t=\sum_{j=1}^{\infty}\frac{V_{t_j,j}}{2^j}, \qquad W_t=\sum_{j=1}^{\infty}\frac{W_{t_j,j}}{2^j}, \tag{1}$$

where $t=\sum_{j=1}^{\infty}\frac{t_j}{2^j}$ is a binary representation of t, and $t_1,t_2,\dots\in\{0,1\}$.

All the above series are almost surely convergent. Indeed, by Lemma 1, we have

$$\sum_{j=1}^{\infty}P\left(\left|\frac{U_{t_j,j}}{2^j}\right|>\frac{1}{2^{j/2}}\right)=\sum_{j=1}^{\infty}P(|U_{t_j,j}|>2^{j/2})<\sum_{j=1}^{\infty}\frac{2}{2^{j/2}}<\infty,$$

and by the Borel–Cantelli lemma, we obtain that $\left|\frac{U_{t_j,j}}{2^j}\right|\le\frac{1}{2^{j/2}}$ for all but finitely many $j=1,2,\dots$. It follows that $\sum_{j=1}^{\infty}\frac{U_{t_j,j}}{2^j}$ is almost surely absolutely convergent. The same holds for $\sum_{j=1}^{\infty}\frac{V_{t_j,j}}{2^j}$ and $\sum_{j=1}^{\infty}\frac{W_{t_j,j}}{2^j}$.

The distribution of U_t does not depend on t. Indeed, if $s=\sum_{j=1}^{\infty}\frac{s_j}{2^j}$ and $t=\sum_{j=1}^{\infty}\frac{t_j}{2^j}$, then $(U_{s_j,j}: j\in\mathbb{N})\sim(U_{t_j,j}: j\in\mathbb{N})$ since both sequences are i.i.d. with the same marginal distribution. Hence, $U_s\sim U_t$. A similar argument shows that the distributions of V_t and W_t do not depend on t. For later use, we denote these distributions as follows: $U_t\sim\mu_U$, $V_t\sim\mu_V$, and $W_t\sim\mu_W$.

Finally, for each $t \in [0, 1]$ we have

$$U_t + V_t + W_t = \sum_{j=1}^{\infty} \frac{U_{t_j,j} + V_{t_j,j} + W_{t_j,j}}{2^j} = \sum_{j=1}^{\infty} \frac{t_j}{2^j} = t.$$

□

Theorem 1 *Let $d \in \mathbb{N}$ and $B \subset \mathbb{R}^d$ be a non-empty bounded Borel subset of the d-dimensional Euclidean space. There exists a d-dimensional probability distribution $\boldsymbol{\mu}$ satisfying what follows: For each $n \geq 3$ and each probability distribution $\boldsymbol{\nu}$ on B, there exist d-dimensional random vectors $\mathbf{X}_{\boldsymbol{\nu},1}, \mathbf{X}_{\boldsymbol{\nu},2}, \ldots, \mathbf{X}_{\boldsymbol{\nu},n}$ such that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \boldsymbol{\nu}$ and $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ for $i = 1, 2, \ldots, n$.*

Proof Without loss of generality, we may assume that $B \subset [0, 1]^d$. (If $B \not\subset [0, 1]^d$, then we may apply an affine transformation to each of the d coordinates of $\mathbb{R}^d$ to put B into $[0, 1]^d$.)

Let $\boldsymbol{\mu}$ be the probability distribution on $\mathbb{R}^d$ such that all d marginals of $\boldsymbol{\mu}$ are independent and equal to the convolution of the distributions μ_U, μ_V, and μ_W defined in the proof of Proposition 1 (i.e., $\boldsymbol{\mu} = \mu \otimes \cdots \otimes \mu$, where $\mu = \mu_U * \mu_V * \mu_W$).

We fix an arbitrary $n \geq 3$. Let $\boldsymbol{\nu}$ be any probability distribution on B and let $\mathbf{T} = (T^{(1)}, T^{(2)}, \ldots, T^{(d)})$ be a random vector satisfying $\mathbf{T} \sim \boldsymbol{\nu}$. Using the binary representation of $T^{(1)}, T^{(2)}, \ldots, T^{(d)}$, we define random variables $T_j^{(m)}$ with $m = 1, 2, \ldots, d$, and $j = 1, 2, \ldots$, such that $T^{(m)} = \sum_{j=1}^{\infty} \frac{T_j^{(m)}}{2^j}$ for $m = 1, 2, \ldots, d$, and random variables $T_j^{(m)}$ take values in $\{0, 1\}$. Let $(U_{0,j}^{(m),k}, V_{0,j}^{(m),k}, W_{0,j}^{(m),k}, U_{1,j}^{(m),k}, V_{1,j}^{(m),k}, W_{1,j}^{(m),k})$ (with $m = 1, 2, \ldots, d$, $k = 0, 1, \ldots, n-1$ and $j = 1, 2, \ldots$) be a system of independent copies of the random vector defined in Lemma 1. We assume that this system is independent of $\mathbf{T}$.

For $k = 0, 1, \ldots, n-1$, let $\mathbf{U}_{\mathbf{T}}^k = (U_{\mathbf{T}}^{(1),k}, U_{\mathbf{T}}^{(2),k}, \ldots, U_{\mathbf{T}}^{(d),k})$, $\mathbf{V}_{\mathbf{T}}^k = (V_{\mathbf{T}}^{(1),k}, V_{\mathbf{T}}^{(2),k}, \ldots, V_{\mathbf{T}}^{(d),k})$ and $\mathbf{W}_{\mathbf{T}}^k = (W_{\mathbf{T}}^{(1),k}, W_{\mathbf{T}}^{(2),k}, \ldots, W_{\mathbf{T}}^{(d),k})$ be given by

$$U_{\mathbf{T}}^{(m),k} = \sum_{j=1}^{\infty} \frac{U_{T_j^{(m)},j}^{(m),k}}{2^j}, \qquad V_{\mathbf{T}}^{(m),k} = \sum_{j=1}^{\infty} \frac{V_{T_j^{(m)},j}^{(m),k}}{2^j}, \qquad W_{\mathbf{T}}^{(m),k} = \sum_{j=1}^{\infty} \frac{W_{T_j^{(m)},j}^{(m),k}}{2^j}$$

for $m = 1, 2, \ldots, d$

We define the requested random vectors $\mathbf{X}_{\boldsymbol{\nu},i} = (X_{\boldsymbol{\nu},i}^{(1)}, X_{\boldsymbol{\nu},i}^{(2)}, \ldots, X_{\boldsymbol{\nu},i}^{(d)})$ as follows:

$$\mathbf{X}_{\boldsymbol{\nu},i} = \mathbf{U}_{\mathbf{T}}^{i \bmod n} + \mathbf{V}_{\mathbf{T}}^{(i+1) \bmod n} + \mathbf{W}_{\mathbf{T}}^{(i+2) \bmod n}.$$

Now, we use the property that for each $t \in \{0, 1\}$ and each m, j and k, we have $U_{t,j}^{(m),k} + V_{t,j}^{(m),k} + W_{t,j}^{(m),k} = t$. As a consequence, we obtain that for $m = 1, 2, \ldots, d$, we have

$$\frac{1}{n}(X^{(m)}_{\boldsymbol{\nu},1} + X^{(m)}_{\boldsymbol{\nu},2} + \cdots + X^{(m)}_{\boldsymbol{\nu},n}) = \frac{1}{n}\sum_{k=0}^{n-1}(U^{(m),k}_{\mathbf{T}} + V^{(m),k}_{\mathbf{T}} + W^{(m),k}_{\mathbf{T}})$$
$$= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{\infty}\frac{U^{(m),k}_{T^{(m)}_j,j} + V^{(m),k}_{T^{(m)}_j,j} + W^{(m),k}_{T^{(m)}_j,j}}{2^j} = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{\infty}\frac{T^{(m)}_j}{2^j} = T^{(m)}.$$

It follows that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) = \mathbf{T} \sim \boldsymbol{\nu}$.

It remains to show that for each $i = 1, 2, \ldots, n$, we have $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$. First, we observe that the conditional distribution of $X^{(m)}_{\boldsymbol{\nu},i} = U^{(m),i \bmod n}_{\mathbf{T}} + V^{(m),(i+1) \bmod n}_{\mathbf{T}} + W^{(m),(i+2) \bmod n}_{\mathbf{T}}$, given T, is $\mu = \mu_U * \mu_V * \mu_W$. Moreover, the coordinates $X^{(1)}_{\boldsymbol{\nu},i}$, $X^{(2)}_{\boldsymbol{\nu},i}$, ..., $X^{(d)}_{\boldsymbol{\nu},i}$ are conditionally independent (given T). It follows that $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu} = \mu \otimes \cdots \otimes \mu$, conditionally. Since this distribution does not depend on the condition, we obtain that $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ unconditionally. □

In the following results, we strengthen Theorem 1 by replacing the assumption about the boundedness of the support of measure $\boldsymbol{\nu}$ with the weaker assumption about the concentration (small tails) of measure $\boldsymbol{\nu}$.

Theorem 2 *Let $d \in \mathbb{N}$, $B_1 \subset B_2 \subset \cdots \subset \mathbb{R}^d$ be non-empty bounded Borel sets and $0 \le p_1 \le p_2 \le \ldots$ be a sequence of real numbers satisfying $\lim_{k\to\infty} p_k = 1$. There exists a d-dimensional probability distribution $\boldsymbol{\mu}$ satisfying what follows: For each $n \ge 3$ and each probability distribution $\boldsymbol{\nu}$ on $\mathbb{R}^d$ satisfying $\boldsymbol{\nu}(B_k) \ge p_k$ for $k = 1, 2, \ldots,$ there exist d-dimensional random vectors $\mathbf{X}_{\boldsymbol{\nu},1}, \mathbf{X}_{\boldsymbol{\nu},2}, \ldots, \mathbf{X}_{\boldsymbol{\nu},n}$ such that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \boldsymbol{\nu}$ and $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ for $i = 1, 2, \ldots, n$.*

Proof If $p_k = 1$ for some k, then the result follows by Theorem 1 applied to the Borel set $B = B_k$. In the sequel, we assume that $p_k < 1$ for each k. We may also assume that the sequence (p_k) is strictly increasing. Indeed, if $p_k = p_{k+1}$, then we can exclude both p_{k+1} and B_{k+1} from the respective sequences.

We put $p_0 = 0$ and $B_0 = \emptyset$. For $k = 1, 2, \ldots$ let $\boldsymbol{\mu}_k$ be the probability measure given by Theorem 1 for the Borel set $B = B_k$ and let $\boldsymbol{\mu} = \sum_{k=1}^{\infty}(p_k - p_{k-1})\boldsymbol{\mu}_k$.

Given the measure $\boldsymbol{\nu}$, we will define a sequence $(\boldsymbol{\nu}_k)$ of measures on $\mathbb{R}^d$ by formula $\boldsymbol{\nu}_k = \sum_{l=1}^{k} A_{k,l} \cdot \boldsymbol{\nu}|_{B_l \setminus B_{l-1}}$, where $A_{k,l} = \prod_{j=l}^{k-1}(\boldsymbol{\nu}(B_j) - p_j)/\prod_{j=l}^{k}(\boldsymbol{\nu}(B_j) - p_{j-1})$ for $1 \le l \le k$, and the restriction $\boldsymbol{\nu}|_B$ is defined by $\boldsymbol{\nu}|_B(A) = \boldsymbol{\nu}(A \cap B)$ for each Borel set $A \subset \mathbb{R}^d$. It is easy to verify that for each $1 \le l \le k$ we have $A_{k,l} \ge 0$, $\sum_{l=1}^{k} A_{k,l} \cdot (\boldsymbol{\nu}(B_l) - \boldsymbol{\nu}(B_{l-1})) = 1$, and $\sum_{k=l}^{\infty} A_{k,l} \cdot (p_k - p_{k-1}) = 1$. Consequently, $\boldsymbol{\nu}_k$ is a probability measure on B_k, and $\sum_{k=1}^{\infty}(p_k - p_{k-1})\boldsymbol{\nu}_k = \sum_{(l,k):1\le l\le k}(p_k - p_{k-1})A_{k,l}\boldsymbol{\nu}|_{B_l \setminus B_{l-1}} = \sum_{l=1}^{\infty} \boldsymbol{\nu}|_{B_l \setminus B_{l-1}} = \boldsymbol{\nu}$.

Now, for each $k = 1, 2, \ldots$ let $\mathbf{X}_{\boldsymbol{\nu}_k,1}, \mathbf{X}_{\boldsymbol{\nu}_k,2}, \ldots, \mathbf{X}_{\boldsymbol{\nu}_k,n}$ be d-dimensional random vectors given by Theorem 1. Moreover, let L be a random variable, independent of these vectors, such that $P(L = k) = p_k - p_{k-1}$ for $k = 1, 2, \ldots$. We define $\mathbf{X}_{\boldsymbol{\nu},i} = \mathbf{X}_{\boldsymbol{\nu}_L,i}$ for $i = 1, 2, \ldots, n$. The conditional distribution of $\mathbf{X}_{\boldsymbol{\nu},i}$ (under the condition $L = k$) is $\boldsymbol{\mu}_k$. Consequently, $\mathbf{X}_{\boldsymbol{\nu},i} \sim \sum_{k=1}^{\infty}(p_k - p_{k-1})\boldsymbol{\mu}_k = \boldsymbol{\mu}$. Similarly, $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \sum_{k=1}^{\infty}(p_k - p_{k-1})\boldsymbol{\nu}_k = \boldsymbol{\nu}$. □

We recall that for two probability distributions μ and ν on $\mathbb{R}$, we say that μ is *stochastically dominated* by ν (denoted by $\mu \leq_{st} \nu$) when their cumulative distribution functions satisfy the inequality $F_\mu(u) \geq F_\nu(u)$ for each $u \in \mathbb{R}$. Equivalently, $\mu \leq_{st} \nu$ if and only if there exist random variables $X \sim \mu$ and $Y \sim \nu$ satisfying $X \leq Y$.

Corollary 1 *Let $d \in \mathbb{N}$, $f : \mathbb{R}^d \to \mathbb{R}$ be a Borel function such that $\lim_{\mathbf{x}\to\infty} f(\mathbf{x}) = +\infty$, and ν_0 be a fixed probability distribution on $\mathbb{R}$. We consider a family $\boldsymbol{\mathcal{V}}$ consisting of probability distributions of all d-dimensional random vectors $\mathbf{Z}$ such that the distribution of $f(\mathbf{Z})$ is stochastically dominated by ν_0.*

There exists a d-dimensional probability distribution $\boldsymbol{\mu}$ satisfying what follows: For each $n \geq 3$ and each probability distribution $\boldsymbol{\nu} \in \boldsymbol{\mathcal{V}}$, there exist d-dimensional random vectors $\mathbf{X}_{\boldsymbol{\nu},1}, \mathbf{X}_{\boldsymbol{\nu},2}, \ldots, \mathbf{X}_{\boldsymbol{\nu},n}$ such that $\frac{1}{n}(\mathbf{X}_{\boldsymbol{\nu},1} + \mathbf{X}_{\boldsymbol{\nu},2} + \cdots + \mathbf{X}_{\boldsymbol{\nu},n}) \sim \boldsymbol{\nu}$ and $\mathbf{X}_{\boldsymbol{\nu},i} \sim \boldsymbol{\mu}$ for $i = 1, 2, \ldots, n$.

The case when f is the Euclidean norm in $\mathbb{R}^d$ and ν_0 is a probability distribution on $(0, \infty)$ seems to be the most useful.

Proof For $k = 1, 2, \ldots$ let $B_k = \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \leq k + a\}$, where $a \in \mathbb{R}$ is chosen in such a way that the sets B_k are non-empty and let $p_k = \nu_0((-\infty, k + a])$. By the assumption about the limit of f at infinity, the sets B_k are bounded. Applying Theorem 2 for sequences (B_k) and (p_k) completes the proof of the corollary. □

3 Some Necessary Conditions for the Center of the Mix

In this section, we present some necessary conditions that should be satisfied by each complete mix. In particular, we show that we can neither replace the [0, 1] interval in Proposition 1 nor skip the boundedness assumption about set B in Theorem 1. (However, we can weaken the assumptions about the distribution $\boldsymbol{\nu}$ as in Theorem 2 and Corollary 1.) We also generalize the following necessary condition for a probability distribution to be n-completely mixable, formulated by Wang and Wang in 2011:

Proposition 2 *([8], see also [5, 6, 9]) Suppose the probability distribution μ is n-completely mixable, centered at C. Let $a = \inf\{x : \mu((-\infty, x]) > 0\}$ and $b = \sup\{x : \mu((-\infty, x]) < 1\}$. If one of a and b is finite, then the other one is finite, and $a + \frac{b-a}{n} \leq C \leq b - \frac{b-a}{n}$.*

The following proposition and Corollary 2 significantly strengthen the above result (cf. Remark 1).

Proposition 3 *Let ν be a probability distribution on $\mathbb{R}$. If $X_1, X_2, \ldots, X_n$ are random variables satisfying $X_1 + X_2 + \cdots + X_n \sim \nu$, then for each $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $i = 1, 2, \ldots, n$ we have*

$$P(X_i > a_i) - \sum_{\substack{j=1 \\ j\neq i}}^{n} P(X_j < a_j) \leq \nu\left(\left(\sum_{j=1}^{n} a_j, \infty\right)\right) \leq \sum_{j=1}^{n} P(X_j > a_j)$$

and

$$P(X_i < a_i) - \sum_{\substack{j=1 \\ j \neq i}}^{n} P(X_j > a_j) \leq \nu\left(\left(-\infty, \sum\nolimits_{j=1}^{n} a_j\right)\right) \leq \sum_{j=1}^{n} P(X_j < a_j).$$

Proof We show the first inequality. Using the implication $(\forall_j \; X_j \leq a_j) \Rightarrow \sum_{j=1}^{n} X_j \leq \sum_{j=1}^{n} a_j$, we obtain

$$\nu\left(\left(\sum\nolimits_{j=1}^{n} a_j, \infty\right)\right) = P\left(\sum_{j=1}^{n} X_j > \sum_{j=1}^{n} a_j\right) \leq P\left(\bigcup_{j=1}^{n} (X_j > a_j)\right) \leq \sum_{j=1}^{n} P(X_j > a_j).$$

Similarly, by $(\sum_{j=1}^{n} X_j \leq \sum_{j=1}^{n} a_j, \; \forall_{j \neq i} \; -X_j \leq -a_j) \Rightarrow X_i \leq a_i$ we obtain

$$P(X_i > a_i) \leq P\left(\left(\sum_{j=1}^{n} X_j > \sum_{j=1}^{n} a_j\right) \cup \bigcup_{\substack{j=1 \\ j \neq i}}^{n} (-X_j > -a_j)\right) \leq \nu\left(\left(\sum\nolimits_{j=1}^{n} a_j, \infty\right)\right) + \sum_{\substack{j=1 \\ j \neq i}}^{n} P(X_j < a_j).$$

The proof of the second inequality is analogous. □

If we put $\nu = \delta_t$ (the one-point probability distribution concentrated at $t \in \mathbb{R}$) and $a_i = t - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j$ into Proposition 3, then we obtain the following corollary.

Corollary 2 *If $X_1, X_2, \ldots, X_n$ are random variables satisfying $X_1 + X_2 + \cdots + X_n = t \in \mathbb{R}$, then for each $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $i = 1, 2, \ldots, n$ we have*

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} P(X_j < a_j) \geq P\left(X_i > t - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j\right) \tag{2}$$

and

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} P(X_j > a_j) \geq P\left(X_i < t - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j\right). \tag{3}$$

If $\sum_{j=1}^{n} a_j < t$*, then* $\sum_{j=1}^{n} P(X_j > a_j) \geq 1$*. If* $\sum_{j=1}^{n} a_j > t$*, then* $\sum_{j=1}^{n} P(X_j < a_j) \geq 1$.

Remark 1 We will show that Corollary 2 is a generalization of Proposition 2.

Let $X_1, X_2, \ldots, X_n$ be random variables satisfying $X_1, X_2, \ldots, X_n \sim \mu$ and $\frac{1}{n}(X_1 + X_2 + \cdots + X_n) = C$. We recall that $a = \inf\{x : \mu((-\infty, x]) > 0\}$ and $b = \sup\{x : \mu((-\infty, x]) < 1\}$.

Assume that a is finite. Then, for each $j = 1, 2, \ldots, n$ we have $P(X_j < a) = \mu((-\infty, a)) = 0$. By (2) applied to $a_1 = a_2 = \cdots = a_n = a$ and $t = nC$ we obtain $P(X_i > nC - (n-1)a) = 0$. Consequently, b is finite and $b \leq nC - (n-1)a$, which is equivalent to $a + \frac{b-a}{n} \leq C$.

Now, assume that b is finite. Then, for each $i = 1, 2, \ldots, n$ we have $P(X_i > b) = \mu((b, \infty)) = 0$. By (3) applied to $a_1 = a_2 = \cdots = a_n = b$ and $t = nC$ we obtain $P(X_i < nC - (n-1)b) = 0$. Consequently, a is finite and $a \geq nC - (n-1)b$, which is equivalent to $C \leq b - \frac{b-a}{n}$.

As a consequence of Corollary 2, we immediately obtain the following corollary.

Corollary 3 *Let* $\mu_1, \mu_2, \ldots, \mu_n$ *be probability distributions. If there exist random variables* $X_1 \sim \mu_1$*,* $X_2 \sim \mu_2, \ldots, X_n \sim \mu_n$ *and* $t \in \mathbb{R}$ *such that* $X_1 + X_2 + \cdots + X_n = t$*, then*

$$-\infty < \sup\left\{\sum_{j=1}^{n} a_j : \sum_{j=1}^{n} \mu_j((-\infty, a_j)) < 1\right\} \leq t$$
$$\leq \inf\left\{\sum_{j=1}^{n} a_j : \sum_{j=1}^{n} \mu_j((a_j, \infty)) < 1\right\} < \infty.$$

Corollary 3 shows that we cannot replace the [0, 1] interval in Proposition 1 by any unbounded set. Additionally, it shows that we cannot skip the assumption that the set B is bounded in Theorem 1. (Indeed, Corollary 3 implies that the projection of B onto each coordinate has to be bounded.)

Author Contributions Andrzej Komisarski and Jacques Labuschagne wrote the main manuscript text. All authors reviewed the manuscript. The article is a part of Jacques Labuschagne's doctoral research.

Funding No funding was received to assist with the preparation of this manuscript.

Availability of Data and Materials Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dhaene, J., Denuit, M.: The safest dependence structure among risks. Insurance Math. Econ. **25**(1), 11–21 (1999)
2. Embrechts, P., Puccetti, G.: Bounds for the sum of dependent risks having overlapping marginals. J. Multivariate Anal. **101**(1), 177–190 (2009)
3. Puccetti, G., Rigo, P., Wang, B., Wang, R.: Centers of probability measures without the mean. J. Theor. Prob. **32**, 1482–1501 (2019)
4. Puccetti, G., Rüschendorf, L.: Bounds for joint portfolios of dependent risks. Stat. Risk Model. **29**(2), 107–131 (2012)
5. Puccetti, G., Wang, B., Wang, R.: Advances in complete mixability. J. Appl. Prob. **49**(2), 430–440 (2012)
6. Puccetti, G., Wang, B., Wang, R.: Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. Insurance Math. Econ. **53**(3), 821–828 (2013)
7. Rüschendorf, L., Uckelmann, L.: Variance minimization and random variables with constant sum. In: Cuadras, C.M., Fortiana, J., Rodriguez-Lallena, J.A. (eds.) Distributions with Given Marginals and Statistical Modelling, pp. 211–222. Kluwer Academic Publishers, Dordrecht (2002)
8. Wang, B., Wang, R.: The complete mixability and convex minimization problems with monotone marginal densities. J. Multivariate Anal. **102**(10), 1344–1360 (2011)
9. Wang, R.: Current open questions in complete mixability. Prob. Surv. **12**, 13–32 (2015)