Sum of Arbitrarily Dependent Random Variables

Ruodu Wang*

September 15, 2014

Abstract

In many classic problems of asymptotic analysis, it appears that the scaled average of a sequence of F-distributed random variables converges to G-distributed limit in some sense of convergence. In this paper, we look at the classic convergence problems from a novel perspective: we aim to characterize all possible limits of the sum of a sequence of random variables under different choices of dependence structure. We show that under general tail conditions on two given distributions F and G, there always exists a sequence of F-distributed random variables such that the scaled average of the sequence converges to a G-distributed limit almost surely. We construct such a sequence of random variables via a structure of conditional independence. The results in this paper suggest that with the common marginal distribution fixed and dependence structure unspecified, the distribution of the sum of a sequence of random variables can be asymptotically of any shape.

1 Introduction

Consider a convergence problem of the following type:

$$\frac{\sum_{i=1}^{n} X_i - a_n}{b_n} \to X \quad \text{in some sense of convergence as } n \to \infty, \tag{1.1}$$

where X, X_i, $i \in \mathbb{N}$ are random variables, and a_n, b_n, $n \in \mathbb{N}$ are real numbers, typically with $b_n \to \infty$. Convergence problems of type (1.1) are arguably the most classic problems in

probability, dating back to the first time probability theory was established in *Ars Conjectandi*. The laws of large numbers and central limit theorems all belong to the type of (1.1).

The classic setting for (1.1) consists of three key elements:

(i) the marginal distributions of X_i, $i \in \mathbb{N}$, typically chosen as identical, say F, with some conditions;

(ii) the dependence structure in the sequence $\{X_i, \ i \in \mathbb{N}\}$, typically chosen as independence, conditional independence, time series models (AR, MA, etc...), or other specific structures;

(iii) the limiting random variable X or its distribution, and the type of convergence.

Note that the real numbers b_n and a_n, however, are chosen for normalization purpose when the conditions in (i) and (ii) are given; hence we do not emphasize their influence on the convergence in (1.1). In the following, we briefly recall some of the basic examples of (1.1). Our inspiration comes from those rather simple examples, and the reader is kindly advised to bear the triviality for the moment.

(a) The most fundamental example in probability theory would be the Strong Law of Large Numbers: in (i) we set F to have finite mean; in (ii) we set the dependence structure as independence; with normalizing constants $b_n = n$ and $a_n = 0$, in (iii) we obtain that the almost sure (a.s.) limit is the mean of F.

(b) From the viewpoint of statistics, the most prominent example would probably be the Central Limit Theorem: in (i) we set F to have finite variance σ^2; in (ii) we set the dependence structure as independence; with normalizing constants $b_n = \sigma\sqrt{n}$ and $a_n = n\mathbb{E}[X_1]$, in (iii) we obtain that the weak limit is a standard normal distribution.

(c) Another example, appearing extensively in Bayesian statistical inference, is about the conditional independence: in (i) we set F to have a finite mean; in (ii), we introduce a random variable Θ, and conditional on $\Theta = \theta \in \mathbb{R}$, let X_i, $i \in \mathbb{N}$ be iid with distribution F_θ; with normalizing constants $b_n = n$ and $a_n = 0$, in (iii) we obtain that the a.s. limit is a random variable μ_Θ, where μ_θ is the mean of F_θ, $\theta \in \mathbb{R}$.

(d) The last example is about stable distributions: in (i) we set F to have a tail order $\alpha \leqslant 2$, which implies $\mathbb{E}[|X_1|^{\alpha+\varepsilon}] = \infty$ and $\mathbb{E}[|X_1|^{\alpha-\varepsilon}] < \infty$ for all $\varepsilon > 0$; in (ii) we set the dependence structure as independence; with normalizing constants $b_n = n^{1/\alpha}$ and $a_n = n\mathbb{E}[X_1]\mathrm{I}_{\{\alpha>1\}}$, in (iii) we obtain that the weak limit is an α-stable distribution. This is a fundamental result in the study of stable distributions, with the Central Limit Theorem as a special case. Moreover, and more importantly, we obtain that if the dependence structure in (ii) is chosen as iid, then the limit in (iii) has to be α-stable no matter how we choose F in (i).

The reader might have noticed that examples (a)-(c) have a lot in common: we fix (i) and (ii), and then obtain (iii). In example (d), more is studied: when the dependence in (ii) is fixed as independence, we can determine what distribution in (iii) is a possible limit. This can be interpreted as a compatibility problem of the choices in (ii) and (iii). We remark that if the sum of the sequence in (1.1) is replaced by the maximum of the sequence, it leads to the study of extreme value distributions and their domains of attraction; see e.g. Chapter 1 of de Haan and Ferreira (2006).

The above discussion motivates us to look at the convergence problem in (1.1) from another perspective: what if (i) and (iii) are given, and (ii) is unknown (a typical setting called *uncertain dependence*)? That is, consider the compatibility problem of (i) and (iii). Let us phrase the problem precisely below.

(Q) For univariate distributions F and G, determine whether there exist a sequence of random variables $\{X_i,\ i \in \mathbb{N}\}$, $X_i \sim F$, a random variable $X \sim G$, and real sequences $a_n \in \mathbb{R}$, $b_n \in \mathbb{R}^+$, $n \in \mathbb{N}$ such that

$$\frac{S_n - a_n}{b_n} \to X \quad \text{in some sense of convergence as } n \to \infty, \tag{1.2}$$

where $S_n = \sum_{i=1}^n X_i$.

For instance, if G is chosen as degenerate, then F can be any distribution with finite α-moment, $\alpha > 0$, with dependence structure in (ii) being independence and b_n chosen as $\max\{n^{1/\alpha+\varepsilon}, n\}$, for some $\varepsilon > 0$; if G is chosen as a normal distribution, then F can be any distribution with finite second moment, with dependence structure in (ii) being independence and b_n chosen as $bn^{1/2}$ for some $b \in \mathbb{R}$; for any G, F can be chosen as G itself, with dependence structure in (ii) being comonotonicity, i.e. $X_1, X_2, \ldots$ are a.s. identical, and $b_n = n$.

Note that in question **(Q)**, only the *shape* of the distributions F and G matters to the compatibility, not the location or scale of them. Moreover, note that in all the above examples (a)-(d), typically only moment conditions or tail conditions of F affect the limit in (iii). Based on such an observation, one might naturally expect that the answer to **(Q)** is generally positive, with only conditions on the moments or tail-behavior of F and G. In this paper, we will show that this is indeed true.

Question **(Q)** concerns a fundamental question in multivariate dependence theory. Needless to say dependence structures are of crucial importance in statistical analysis; great challenges exist in both the modeling and the statistical inference of dependence structures. More specifically, Question **(Q)** belongs in the area of *distributions with fixed margins*, that is, the study on the probabilistic properties of a random vector $\mathbf{X} := (X_1, \ldots, X_n)$ with $F_1, \ldots, F_n$ fixed as its marginal distributions, while the dependence structure of $\mathbf{X}$ can vary. Tchen (1980) studied

3

inequalities for the joint distribution of $\mathbf{X}$ in the case of $n = 2$, which originated from the seminal work of Hoeffding (1940) and Fréchet (1951) in this field. In general, the range of functionals of $\mathbf{X}$ with varying dependence structure is still an open question; only limited explicit results are available in the literature. We refer to Joe (1997) for many unsolved problems on distributions with fixed margins. For instance, Chaganty and Joe (2006) studied the range for correlation matrices of $\mathbf{X}$ when $F_1, \ldots, F_n$ are chosen as Bernoulli random variables; even in that rather simple case, a full characterization of the correlation matrices is far away from being clear. A collection of research developments in the area of distributions with fixed margins can be found in DallAglio et al. (1991) and Cuadras and Fortiana (2002). Research in this area often overlaps with the study of dependence concepts and copulas; the interested reader is referred to Joe (1997) and Nelsen (2006) for an overview on dependence concepts, copulas and their relation to problems of distributions with fixed margins.

A particularly interesting and important topic in this area, also relevant to this paper, is the probabilistic behavior of $S_n := X_1 + \cdots + X_n$. Makarov (1981) and Rüschendorf (1982) considered the distribution function F_{S_n} of S_n for $n = 2$ and obtained sharp bounds on F_{S_n}. Later developments of bounds on F_{S_n} for $n \geqslant 3$ can be found in Embrechts and Puccetti (2006), Wang and Wang (2011), and Wang et al. (2013), with no universal solution achieved yet. An overview of this topic with its connection to mass-transportation theory is available in Rüschendorf (2013). Characterizing the set of all possible distributions of S_n stands even more challenging and remains open; see Bernard et al. (2014) for instance. The study of question $(\mathbf{Q})$ addresses the set of possible distributions of S_n after scaling, in an asymptotic manner as $n \to \infty$. Given the relatively large number of unsolved problems and considerable mathematical difficulties in this area, relevant work has mainly focused on cases of finite n. No existing research on the asymptotic analysis of questions of type $(\mathbf{Q})$ is available up to our knowledge. Note that in classic time series analysis, some linear or other parametric structures are typically assumed for the sequence $\{X_i, \ i \in \mathbb{N}\}$, and hence research in that direction is mathematically different from question $(\mathbf{Q})$.

In this paper, we give an answer to question $(\mathbf{Q})$. In Section 2, we show that a stronger version of $(\mathbf{Q})$,

$(\mathbf{Q'})$ there exist $X_i, i \in \mathbb{N}$, $X_i \sim F$, $X_G \sim G$, some $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$, s.t. $\dfrac{S_n}{bn} - a \overset{\text{a.s.}}{\to} X_G,$

holds for all non-degenerate distributions F with finite mean and bounded distributions G, and hence the answer to $(\mathbf{Q})$ is positive with a broad generality. The dependence structure for $(\mathbf{Q'})$ in this paper is a special construction of conditional independence. In Section 3 we show that conditions on the asymptotic behavior of F and G are required for $(\mathbf{Q'})$ to hold when G is unbounded. We obtain a sufficient and necessary condition for $(\mathbf{Q'})$ when F is regular varying,

and some sufficient conditions for **(Q')** for general distributions F and G. In Section 4 we draw our conclusions and highlight some future research directions. We hope that the results in this paper would clearly deliver the following message:

*With marginal distributions of a sequence given, and its dependence structure unspecified, the asymptotic distribution of the sum of the sequence can be of **arbitrary** shape.*

In other words, even when marginal distributions are accurate, a misspecified dependence assumption could lead to a completely problematic asymptotic behavior in a classic problem of statistical analysis.

A direct application of the main results in this paper can be found in Quantitative Risk Management, where statistical inference with *dependence uncertainty* is an important issue to consider; see McNeil et al. (2005). For instance, in modeling operational risk, individual risks (risks arising from different types of business lines) are typically non-Gaussian distributed; modeling and statistical inference are usually quite reliable for each of the individual risks. However, there is often not enough joint data to make reliable inference for the (non-Gaussian) dependence structure between individual risks. One may use conservative assumptions on the dependence structure to model risk aggregation (sum) based on marginal data, and find corresponding quantities for capital requirements or stress testing purposes. However, it is well-known that some commonly used "conservative dependence structures" such as the comonotonicity is not meaningful in order to obtain a conservative estimate of a quantile (called a Value-at-Risk in Finance) of S_n; we refer to Embrechts et al. (2013) and Embrechts et al. (2014) for detailed illustrations on this topic. The results in this paper show that such conservative modeling of risk aggregation has to consider all possible shapes of the distribution of the sum, when the number of individual risks is large. For the purpose of the current paper, we will focus on the theoretical aspects of the underlying problem; applications will be left for future discussion.

Throughout the paper, we consider a standard atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In a standard atomless probability space, there exist random vectors with any distribution. We denote by $\mathcal{F}$ the set of all univariate distributions on $\mathbb{R}$, and by $\mathcal{F}^p$ the set of univariate distributions on $\mathbb{R}$ with finite p-th moment, $p \in (0, \infty]$. For simplicity, for any distribution F, we use X_F for an F-distributed random variable. We denote the pseudo-inverse function of any non-decreasing function F by F^{-1}, that is

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geqslant t\}, \quad t \in (0,1], \quad \text{and} \quad F^{-1}(0) = \sup\{x \in \mathbb{R} : F(x) = 0\}.$$

We also denote the essential supremum of a random variable X by $\sup(X)$, i.e.

$$\sup(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) = 1\}.$$

To be consistent with the laws of large numbers, and for some technical ease, in this paper we focus on the cases when F has finite mean.

2 Sum of arbitrarily dependent random variables

First, we formally define a property considered in question **(Q')**, which is the main objective of this paper.

Definition 2.1. $F \in \mathcal{F}$ is said to be *shape-compatible* with $G \in \mathcal{F}$ if there exist a sequence of random variables $\{X_i \sim F, \ i \in \mathbb{N}\}$, a random variable $X_G \sim G$, and real numbers $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$, such that as $n \to \infty$,

$$\frac{S_n}{bn} - a \stackrel{\text{a.s.}}{\to} X_G, \tag{2.1}$$

where and throughout the rest of the paper, $S_n = \sum_{i=1}^{n} X_i$. We denote by $F \hookrightarrow G$ if F is shape-compatible with to G. We say $F \hookrightarrow G$ at rate b, where b is given in (2.1).

In the above definition, we call F an *initial* distribution, and G a *target* distribution. It is clear that if $F \hookrightarrow G$ then **(Q)** in the introduction also holds true for F and G. By definition, the property $F \hookrightarrow G$ is invariant under linear transformation. In this section, we assume the initial distribution F is has finite mean. The main result in this section addresses the generality of (2.1), summarized in the following theorem.

Theorem 2.2. *Suppose $F \in \mathcal{F}^1$ is non-degenerate. Then $F \hookrightarrow G$ for all bounded distributions G.*

We proceed to prove this theorem step by step. We first give a useful lemma. In many places below, we will assume the means of F and G are equal without loss of generality, so that the normalizing constant a in (2.1) can be set as zero.

Lemma 2.3. *For any $F \in \mathcal{F}^1$ and $G \in \mathcal{F}$, if $F \hookrightarrow G$ at rate b, then $F \hookrightarrow G$ at rate c for all $0 < c \leqslant b$.*

Proof. By definition, there exist $X_i \sim F$, $i \in \mathbb{N}$, $X_G \sim G$, $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$ such that

$$\frac{\sum_{i=1}^{n} X_i}{bn} - a \stackrel{\text{a.s.}}{\to} X_G.$$

We can assume the means of F and G are both zero (implying $a = 0$). We will construct a sequence $\{Y_i, i \in \mathbb{N}\}$ such that the averaged partial sum converges a.s. to X_G at the rate $c \leqslant b$. The idea is to show that part of the sequence $\{Y_i, i \in \mathbb{N}\}$ can be treated as zero, and the rest part can be taken as $\{X_i, \ i \in \mathbb{N}\}$. Denote $d = b/c \geqslant 1$, and let $d_i = \lfloor di \rfloor \geqslant i, i \in \mathbb{N}$. Also denote for $n \in \mathbb{N}$,

$$t(n) = \#\{d_j \leqslant n : j \in \mathbb{N}\} \leqslant n.$$

It is easy to see that

$$\lim_{n\to\infty} \frac{t(n)}{n} = \frac{1}{d} = \frac{c}{b}.$$

Let Z_i, $i \in \mathbb{N}$ be iid random variables with distribution F. Let

$$Y_i = \begin{cases} X_j & i \in \{d_j \leqslant n : j \in \mathbb{N}\}; \\ Z_{i-t(i)} & \text{otherwise.} \end{cases}$$

We can check

$$\frac{\sum_{i=1}^n Y_i}{cn} = \frac{\sum_{i=1}^{t(n)} X_i}{cn} + \frac{\sum_{i=1}^{n-t(n)} Z_i}{cn}.$$

By the law of large numebrs, we have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n-t(n)} Z_i}{cn} = 0 \quad \text{a.s.}$$

Therefore, we have

$$\frac{\sum_{i=1}^n Y_i}{cn} = \frac{\sum_{i=1}^{t(n)} X_i}{cn} + \frac{\sum_{i=1}^{n-t(n)} Z_i}{cn} = \frac{\sum_{i=1}^{t(n)} X_i}{t(n)b} + \frac{\sum_{i=1}^{n-t(n)} Z_i}{cn} \overset{\text{a.s.}}{\to} X_G.$$

Since each of X_i, Z_i, $i \in \mathbb{N}$ is distributed as F, Y_i is also distributed as F, $i \in \mathbb{N}$. The sequence $\{Y_i, i \in \mathbb{N}\}$ satisfies

$$\frac{\sum_{i=1}^n Y_i}{cn} \overset{\text{a.s.}}{\to} X_G,$$

i.e. $F \hookrightarrow G$ at rate c. □

The following lemma establishes the shapability from F to any two-point distributions. Part (a) is crucial to the proof of the main theorem in this section, and part (b) will be used later in Section 3.

Lemma 2.4. *Suppose $F \in \mathcal{F}^1$ is non-degenerate, then $F \hookrightarrow G$ for all two-point distributions. Moreover, if the means of F, G are both zero, then*

(a) $F \hookrightarrow G$ at rate $c = \mathbb{E}[|X_F|]/(2\sup|X_G|)$;

(b) if F is also a two-point distribution, on $\{r_1, r_2\}$, $r_1 < r_2$, and G is supported on $\{q_1, q_2\}$, $q_1 < q_2$, then $F \hookrightarrow G$ at rate $c = \min\{r_1/q_1, r_2/q_2\}$.

Proof. Without loss of generality we assume F and G both have mean zero. Let q_1 and q_2, with $q_1 < q_2$, be the two points at which G is supported, that is, for some $p \in (0,1)$, $\mathbb{P}(X_G = q_1) = p$, $\mathbb{P}(X_G = q_2) = 1 - p$ and $q_1 p + q_2(1-p) = 0$. It is immediate that $q_2/q_1 = p/(p-1)$.

Denote by F_p the distribution of $F^{-1}(U_p)$, where $U_p \sim \mathrm{U}[p,1]$ and by G_p the distribution of $F^{-1}(V_p)$, where $V_p \sim \mathrm{U}[0,p]$. Now let X_i, $i \in \mathbb{N}$ be iid random variables distributed as F_p and Y_i, $i \in \mathbb{N}$ be iid random variables distributed as G_p. It is easy to see that

$$\mathbb{E}[X_1] = \frac{1}{1-p}\mathbb{E}[(X_F - F^{-1}(p))_+] + F^{-1}(p),$$

7

$$\mathbb{E}[Y_1] = -\frac{1}{p}\mathbb{E}[(F^{-1}(p) - X_F)_+] + F^{-1}(p),$$

and

$$(1 - p)\mathbb{E}[X_1] + p\mathbb{E}[Y_1] = \mathbb{E}[X_F] = 0.$$

It follows that $\mathbb{E}[X_1]/\mathbb{E}[Y_1] = p/(p - 1) = q_2/q_1$. Also note that $\mathbb{E}[X_1] > 0$ and $\mathbb{E}[Y_1] < 0$, since F is not degenerate. Let A be a random event, with $\mathbb{P}(A) = 1 - p$, independent of $\{X_i, \ i \in \mathbb{N}\}$ and $\{Y_i \ i \in \mathbb{N}\}$. Denote

$$Z_i = \mathrm{I}_A X_i + \mathrm{I}_{A^c} Y_i.$$

Take

$$b = \frac{\mathbb{E}[X_1]}{q_2} = \frac{\mathbb{E}[Y_1]}{q_1},$$

and by the law of large numbers we have

$$\frac{\sum_{i=1}^n Z_i}{bn} = \mathrm{I}_A \frac{\sum_{i=1}^n X_i}{n} \frac{q_2}{\mathbb{E}[X_1]} + \mathrm{I}_{A^c} \frac{\sum_{i=1}^n Y_i}{n} \frac{q_1}{\mathbb{E}[Y_1]} \xrightarrow{\text{a.s.}} \mathrm{I}_A q_2 + \mathrm{I}_{A^c} q_1 \stackrel{\mathrm{d}}{=} G.$$

Thus $F \hookrightarrow G$ at rate b. Further,

(a) if $p \geqslant F(0)$, then $\mathbb{E}[X_1] \geqslant \mathbb{E}[(X_F)_+]$, and $b = \mathbb{E}[X_1]/q_2 \geqslant \mathbb{E}[(X_F)_+]/q_2$. If $p < F(0)$, then $\mathbb{E}[Y_1] \leqslant \mathbb{E}[(X_F)_-]$, and $b = \mathbb{E}[Y_1]/q_1 \geqslant -\mathbb{E}[(X_F)_-]/|q_1|$. Note that

$$\mathbb{E}[(X_F)_+] - \mathbb{E}[(X_F)_-] = \mathbb{E}[|X_F|]$$

and

$$\mathbb{E}[(X_F)_+] + \mathbb{E}[(X_F)_-] = \mathbb{E}[X_F] = 0.$$

Thus,

$$\mathbb{E}[(X_F)_+] = -\mathbb{E}[(X_F)_-] = \frac{\mathbb{E}[|X_F|]}{2}.$$

In summary, $b \geqslant \mathbb{E}[|X_F|]/(2\max\{q_2, |q_1|\}) = \mathbb{E}[|X_F|]/(2\sup|X_G|) = c$, and hence $F \hookrightarrow G$ at rate c.

(b) When F is also a two-point distribution, it is obvious that we have either $X_1 = r_2$ or $Y_1 = r_1$. In either cases, $b = r_2/q_2$ or $b = r_1/q_1$. Hence, $b \geqslant \min\{r_2/q_2, r_1/q_1\} = c$ and $F \hookrightarrow G$ at rate c.

$\square$

The next lemma states that if F and G can both be decomposed into a mixture of a collection of distributions with mean zero, say F_a, G_a, $a \in A$ with $F_a \hookrightarrow G_a$ for each a, then $F \hookrightarrow G$. This would be the key step of proving Theorem 2.2.

Lemma 2.5. *Suppose that A is a non-empty set, distributions $F_a \in \mathcal{F}^1$ have the same mean for $a \in A$, and distributions $G_a \in \mathcal{F}^1$ have the same mean for $a \in A$. Suppose $F_a \hookrightarrow G_a$ at rate $c_a \geqslant \varepsilon$ for each $a \in A$ and some $\varepsilon > 0$. Let H be any probability measure on A, then $F := \int_A F_a \mathrm{d}H(a)$ and $G := \int_A G_a \mathrm{d}H(a)$ are distributions and $F \hookrightarrow G$ at rate ε.*

Proof. It is obvious that F and G are distributions. Without loss of generality we assume F_a and G_a, $a \in A$ all have the same mean. By Lemma 2.3, for each $a \in A$, $F_a \hookrightarrow G_a$ at rate ε, and thus there exist $X_i^{(a)} \sim F_a$, $i \in \mathbb{N}$ and $W^{(a)} \sim G_a$ such that $T_n^{(a)} := \frac{1}{\varepsilon n} \sum_{i=1}^{n} X_i^{(a)} \overset{\mathrm{a.s.}}{\to} W^{(a)}$, as $n \to \infty$. Let $Y : \Omega \to A$ be a random variable with distribution H. Let Z_i, $i \in \mathbb{N}$ be a random variable defined as

$$Z_i(\omega) = (X_i^{(Y)})(\omega) = X_i^{(Y(\omega))}(\omega), \quad \omega \in \Omega,$$

for which we write $Z_i = X_i^{(Y)}$. Using the same notation, let $W = W^{(Y)}$ and $T_n = \frac{1}{\varepsilon n} \sum_{i=1}^{n} Z_i$. It follows that

$$\mathbb{P}(Z_i \leqslant t) = \int_A \mathbb{P}(X_i^{(a)} \leqslant t) \mathrm{d}H(a) = \int_A F_a(t) \mathrm{d}H(a) = F(t), \quad t \in \mathbb{R},$$

$$\mathbb{P}(W \leqslant t) = \int_A \mathbb{P}(W^{(a)} \leqslant t) \mathrm{d}H(a) = \int_A G_a(t) \mathrm{d}H(a) = G(t), \quad t \in \mathbb{R},$$

and

$$\mathbb{P}\left(\lim_{n \to \infty} T_n = W\right) = \int_A \mathbb{P}\left(\lim_{n \to \infty} T_n^{(a)} = W^{(a)}\right) \mathrm{d}H(a) = \int_A \mathrm{d}H(a) = 1.$$

Thus we obtain $F \hookrightarrow G$. $\qquad\square$

Remark 2.6. If we choose $F_a = F$, then we obtain that the set of distributions with mean zero and with which F are shape-compatible is a convex set.

The following lemma gives a representation of any distributions with finite mean as a mixture of two-point distributions. This lemma together with Lemmas 2.4-2.5 will eventually lead to a complete proof of Theorem 2.2.

Lemma 2.7. *Each distribution F with mean μ has the following representation:*

$$F = \int_0^1 F_a \mathrm{d}a,$$

where F_a, $a \in (0, 1)$ are two-point (not necessarily distinct) distributions with mean μ.

Proof. We show this lemma by construction. We first assume $\mathbb{P}(X_F = \mu) = 0$. Let

$$H_F(s) = \int_0^s (F^{-1}(t) - \mu) \mathrm{d}t, \quad s \in [0, 1], \tag{2.2}$$

and denote $\nu = F(\mu)$. It is easy to see that the function H_F is bounded, continuous on $[0, 1]$, strictly decreasing on $[0, \nu]$, strictly increasing on $[\nu, 1]$, $H_F(0) = H_F(1) = 0$, and the minimum

value of $H(s)$ is attained at $c := H_F(\nu) < 0$. Moreover, H_F is a convex function and hence is almost everywhere (a.e.) differentiable on $[0, 1]$. For each $s \in [c, 0]$, define

$$A_F : [c, 1] \to [0, \nu], \;\; A_F(H_F(s)) = s, \;\; s \in [0, \nu], \tag{2.3}$$

and

$$B_F : [c, 1] \to [\nu, 1], \;\; B_F(H_F(s)) = s, \;\; s \in [\nu, 1], \tag{2.4}$$

i.e. A_F and B_F are the inverse functions of H_F on the two intervals $[0, \nu]$ and $[\nu, 1]$, respectively. Note that since H_F is continuous and convex, and has an a.e. non-zero derivative, A_F and B_F are continuous and a.e. differentiable on $[c, 0]$. Let

$$K_F(s) = \begin{cases} 1 & s \geqslant 0, \\ B_F(s) - A_F(s) & c \leqslant s < 0, \\ 0 & s < c. \end{cases} \tag{2.5}$$

$K_F(s)$ is continuous, increasing, $K_F(c) = 0$, and $K_F(0) = 1$, hence it is a continuous distribution function on $[c, 0]$. For $s \in (c, 0)$, define

$$W_1(s) = F^{-1}(A_F(s)), \quad W_2(s) = F^{-1}(B_F(s)),$$

and

$$u_F(s) = \frac{\mu - W_1(s)}{W_2(s) - W_1(s)}.$$

It is easy to see that $u_F(s) \in [0, 1]$ for $s \in (c, 0)$.

Now, for $s \in (c, 0)$, let G_s be a two-point distribution on $\{W_1(s), W_2(s)\}$ with probability $\{1 - u_F(s), u_F(s)\}$ respectively. We can check that the mean of G_s is μ:

$$W_1(s)(1 - u(s)) + W_2(s)u(s) = W_1(s)\frac{W_2(s) - \mu}{W_2(s) - W_1(s)} + W_2(s)\frac{\mu - W_1(s)}{W_2(s) - W_1(s)} = \mu.$$

Let

$$G = \int_c^0 G_s \mathrm{d}K_F(s).$$

We can check for $x \leqslant \mu$,

$$\begin{aligned} G(x) &= \int_c^1 G_s(x)\mathrm{d}K_F(s) \\ &= \int_c^0 (1 - u(s))\mathrm{I}_{\{W_1(s) \leqslant x\}}\mathrm{d}K_F(s) \\ &= \int_c^0 (1 - u(s))\mathrm{I}_{\{W_1(s) \leqslant x\}}\mathrm{d}K_F(s) \\ &= \int_c^0 \frac{W_2(s) - \mu}{W_2(s) - W_1(s)}\mathrm{I}_{\{W_1(s) \leqslant x\}}\mathrm{d}(B_F(s) - A_F(s)). \end{aligned}$$

Since B_F and A_F are the inverse functions of H_F, we have a.e.

$$dB_F(s) = \frac{1}{H_F'(B_F(s))}ds = \frac{1}{W_2(s) - \mu}ds,$$

and

$$dA_F(s) = \frac{1}{H_F'(A_F(s))}ds = \frac{1}{W_1(s) - \mu}ds.$$

Thus

$$
\begin{aligned}
G(x) &= \int_c^0 \frac{W_2(s) - \mu}{W_2(s) - W_1(s)} I_{\{W_1(s) \leqslant x\}} \frac{W_1(s) - W_2(s)}{(W_2(s) - \mu)(W_1(s) - \mu)} ds \\
&= \int_c^0 \frac{1}{\mu - W_1(s)} I_{\{W_1(s) \leqslant x\}} ds \\
&= -\int_0^\nu \frac{1}{\mu - F^{-1}(A_F(H_F(t)))} I_{\{W_1(H_F(t)) \leqslant x\}} dH_F(t) \\
&= -\int_0^\nu \frac{1}{\mu - F^{-1}(t)} (F^{-1}(t) - \mu) I_{\{F^{-1}(t) \leqslant x\}} dt \\
&= F(x).
\end{aligned}
$$

Similarly, we can show that $G(x) = F(x)$ for $x > \mu$. Thus,

$$F = \int_c^0 G_s dK_F(s).$$

Let $F_a = G_{K_F^{-1}(a)}$ for $a \in (0,1)$. Through substituting s by $K_F^{-1}(a)$, we obtain

$$F = \int_0^1 F_a da.$$

Finally we comment on the case $\mathbb{P}(X_F = \mu) = p > 0$. If $p = 1$ then we simply take $F_a = P_\mu$, where P_μ is the degenerate distribution at μ. If $p < 1$, then the distribution of $X_F | X_F \neq \mu$ is well-defined; we denoted it by $\hat{F}$. Note that $\hat{F}$ has a representation

$$\hat{F} = \int_0^1 \hat{F}_a da,$$

where $\hat{F}_a$, $a \in (0,1)$ are two-point distributions with mean μ. By taking $F_a = \hat{F}_{(a-p)/(1-p)}$ for $a \in (p,1)$ and $F_a = P_\mu$ for $a \in (0,p]$, we obtain the representation of F as

$$F = (1-p)\hat{F} + pP_\mu = \int_p^1 \hat{F}_{(a-p)/(1-p)} da + \int_0^p P_\mu da = \int_0^1 F_a da.$$

The proof is complete. $\qquad\square$

Now we are finally ready to complete the proof of Theorem 2.2.

Proof of Theorem 2.2. Without loss of generality we assume the means of F and G are both zero. By Lemma 2.7, G has representation

$$G = \int_0^1 G_a da,$$

11

where G_a, $a \in (0,1)$ are two-point distributions with mean zero. Since the support of G_a is bounded by the support of G, by Lemma 2.4, we have that $F \hookrightarrow G_a$ at rate $\mathbb{E}[|X_F|]/(2\sup(X_G)) > 0$. Lemma 2.5 and the representation of G lead to the conclusion that $F \hookrightarrow G$ at rate $\mathbb{E}[|X_F|]/(2\sup(X_G))$. $\qquad\square$

The constructed dependence structure among $X_1, \ldots, X_n$ in the proof of Lemmas 2.4 and 2.5 can be explained as follows. The probability space is divided into infinitely many subsets. The random variables X_i, $i \in \mathbb{N}$ are conditionally independent on each of the subset, with $\frac{1}{bn}\sum_{i=1}^n X_i$ converging to a point in the support of G; more specifically, the limiting points appear in pair through the construction in Lemma 2.7. We remark that the dependence structure which accommodates $F \hookrightarrow G$ is not unique in general; for example the conditional independence can be replaced by any type of conditional negative dependence, as long as $\frac{1}{bn}\sum_{i=1}^n X_i$ converges a.s. to its mean conditionally on the corresponding subset of the probability space.

Remark 2.8. From the proofs, we can see that Theorem 2.2 still holds if the a.s. convergence in Definition 2.1 is replaced by an L^1 convergence, since the convergence in our proof is obtained by the Strong Law of Large Numbers. The L^1 convergence however highly relies on the fact that we restrict F to have finite mean. To allow a more general framework for future study - especially to allow F to have infinite mean - we frame our discussion in terms of a.s. convergence in this paper.

3 Unbounded target distributions

3.1 Necessary conditions

In this section we consider the case when $G \in \mathcal{F}^1$ has unbounded support. Without loss of generality we can assume the means of F and G are the same in our discussion.

As we have seen above, shapability is obtained unconditionally for bounded target distributions. One might expect a similar generality for unbounded target distributions. We will first show that some conditions on the asymptotic behavior of F and G are necessary for $F \hookrightarrow G$ to hold. We first recall the definition of tail indices of a distribution.

Definition 3.1. The *right tail-index* of a distribution F is defined as

$$\alpha_F^+ = \sup\{p \in [0, \infty) : \mathbb{E}[(X_F)_+^p] < \infty\},$$

and the *left tail-index* of a distribution F is defined as

$$\alpha_F^- = \sup\{p \in [0, \infty) : \mathbb{E}[(-X_F)_+^p] < \infty\}.$$

Note that here we do not assume that F is regular varying, hence the notion of the tail indices is slightly different from that in classic extreme value theory (see, for example, de Haan and Ferreira, 2006, Chapter 3). When F is regular varying, our definition coincides with the traditional definition used in extreme value theory.

Lemma 3.2. *Suppose $F, G \in \mathcal{F}^1$ and $F \hookrightarrow G$, then $\alpha_F^+ \leqslant \alpha_G^+$ and $\alpha_F^- \leqslant \alpha_G^-$.*

Proof. By $F \hookrightarrow G$, there exist $X_i \sim F$, $i \in \mathbb{N}$, $X_G \sim G$, $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$ such that

$$\frac{\sum_{i=1}^n X_i}{bn} - a \overset{\text{a.s.}}{\to} X_G.$$

The case when $\alpha_F^+ = 1$ is trivial since $\alpha_G^+ \geqslant 1$ for all $G \in \mathcal{F}^1$; we assume $\alpha_F > 1$ in the following. For any $1 < \alpha < \alpha_F$, we have by Fatou's lemma,

$$|(X_G)_+|_\alpha \leqslant \liminf_{n \to \infty} \left| \frac{(\sum_{i=1}^n X_i)_+}{bn} \right|_\alpha \leqslant \liminf_{n \to \infty} \sum_{i=1}^n \left| \frac{(X_i)_+}{bn} \right|_\alpha = \frac{1}{b}|(X_F)_+|_\alpha < \infty.$$

This shows $\mathbb{E}[(X_G)_+^\alpha] < \infty$ for all $1 < \alpha < \alpha_F^+$; thus we have $\alpha_G^+ \geqslant \alpha_F^+$. The case for $\alpha_G^- \geqslant \alpha_F^-$ is obtained by symmetry. $\qquad\square$

Lemma 3.2 tells us that for $F \hookrightarrow G$ it is necessary for F to have a relatively heavier tail. This can be characterized more precisely using the tail integrals of F^{-1} and G^{-1}. Denote by

$$\Phi_F(t) = \int_t^1 F^{-1}(s)\mathrm{d}s, \quad \bar{\Phi}_F(t) = \int_0^t F^{-1}(s)\mathrm{d}s, \quad t \in [0, 1],$$

for any distribution $F \in \mathcal{F}^1$.

Lemma 3.3. *Suppose $F, G \in \mathcal{F}^1$ and $F \hookrightarrow G$, then*

$$\limsup_{t \to 1} \frac{\Phi_G(t)}{\Phi_F(t)} < \infty, \tag{3.1}$$

and

$$\limsup_{t \to 0} \frac{\bar{\Phi}_G(t)}{\bar{\Phi}_F(t)} < \infty, \tag{3.2}$$

Proof. We only show (3.1). Let $X_i \sim F$, $i \in \mathbb{N}$, $X_G \sim G$ and $b \in \mathbb{R}^+$ such that

$$\frac{1}{bn} \sum_{i=1}^n X_i \overset{\text{a.s.}}{\to} X_G,$$

and let F_n be the distribution function of $\frac{1}{n} \sum_{i=1}^n X_i$. Note that

$$\frac{1}{n} \sum_{i=1}^n X_i \prec_{\mathrm{cx}} X_1,$$

where $\prec_{\mathrm{cx}}$ stands for the convex order (see Shaked and Shanthikumar, 2007, Theorem 3.A.36 for this simple fact). By the consistency of $\Phi_{(\cdot)}$ with convex order, we have that $\Phi_{F_n}(t) \leqslant \Phi_F(t)$;

see Theorem 3.A.5 of Shaked and Shanthikumar (2007). It follows from $\frac{1}{bn} \sum_{i=1}^{n} X_i \overset{\text{a.s.}}{\to} X_G$ and Fatou's lemma that

$$\limsup_{t \to 1} \frac{\Phi_G(t)}{\Phi_{F_n}(t)} \leqslant b,$$

and hence (3.1) follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the following we show that for F or G being a regular varying distribution, (3.1)-(3.2) can be written as a ratio between $F^{-1}(t)$ and $G^{-1}(t)$. We denote by $\mathcal{L}$ the set of regular varying distributions F with tail indices less than one, i.e. there exist $\alpha, \beta \in (0,1)$, such that F satisfies that as $t \to 0$,

$$F^{-1}(t) \sim -t^{-\alpha} L_1(1/t),$$

and as $t \to 1$,

$$F^{-1}(t) \sim (1-t)^{-\beta} L_2(1/(1-t)),$$

where L_1, L_2 are two slowly varying functions (see de Haan and Ferreira, 2006, Section 1.2).

Lemma 3.4. *Suppose $F \in \mathcal{L}$ and $F \hookrightarrow G$, then*

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} < \infty, \tag{3.3}$$

and

$$\limsup_{t \to 0} \frac{G^{-1}(t)}{F^{-1}(t)} < \infty, \tag{3.4}$$

Proof. We only show (3.3). Suppose

$$F^{-1}(t) \sim (1-t)^{-\beta} L(t)$$

where L_1 is a slowly varying function, and $\beta \in (0,1)$. Let $X_i \sim F$, $i \in \mathbb{N}$, $X_G \sim G$, and $b \in \mathbb{R}^+$ such that

$$\frac{1}{bn} \sum_{i=1}^{n} X_i \overset{\text{a.s.}}{\to} X_G,$$

and let F_n be the distribution function of $\frac{1}{n} \sum_{i=1}^{n} X_i$. Similar arguments as in the proof of Lemma 3.3 lead to

$$F_n^{-1}(t) \leqslant \frac{1}{1-t} \Phi_{F_n}(t) \leqslant \frac{1}{1-t} \Phi_F(t).$$

By Karamata's Theorem (see de Haan and Ferreira, 2006, Theorem B.1.5), we have

$$\lim_{t \to 1} \frac{\frac{1}{1-t} \Phi_F(t)}{F^{-1}(t)} = \frac{\beta}{\beta - 1}.$$

It follows from $\frac{1}{bn} \sum_{i=1}^{n} X_i \overset{\text{a.s.}}{\to} X_G$ that

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F_n^{-1}(t)} < \infty,$$

14

and hence

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} = \frac{\beta}{\beta - 1} \limsup_{t \to 1} \frac{G^{-1}(t)}{\frac{1}{1-t}\Phi_F(t)} \leqslant \frac{\beta}{\beta - 1} \limsup_{t \to 1} \frac{G^{-1}(t)}{F_n^{-1}(t)} < \infty.$$

$\square$

Note that in the above lemma we did not assume G is also regular varying. If we assume that $F, G \in \mathcal{L}$, then (3.3)-(3.4) are immediately equivalent to (3.1)-(3.2) by Karamata's theorem.

In the next section, we show that (3.3)-(3.4) turn out to be also sufficient for $F \hookrightarrow G$ for regular varying F.

3.2   Regular varying distributions

By Lemma 2.7, we can find the representations for F and G:

$$F = \int_0^1 F_a \mathrm{d}a, \quad \text{and} \quad G = \int_0^1 G_a \mathrm{d}a,$$

where F_a, G_a, $a \in (0, 1)$ are two-point distributions with the same mean. Assume $\mathbb{P}(X_F = \mathbb{E}[X_F]) = 0$ for the moment. It is clear that $F_a \hookrightarrow G_a$ by Lemma 2.4 since F_a and G_a are both two-point distributions for each a. Hence, to establish $F \hookrightarrow G$ one needs to control the rate of $F_a \hookrightarrow G_a$ for each a. To this consideration, it suffices to have

$$\limsup_{a \to 1} \sup_{t \in [0,1]} \frac{G_a^{-1}(t)}{F_a^{-1}(t)} < \infty. \tag{3.5}$$

Note that both F_a^{-1} and G_a^{-1} only take two values, hence t in (3.5) can be chosen as $t = 0$ and $t = 1$. Also note that for $c < 1$, G_a, $a \in [0, c]$ has bounded support, and therefore we have that

$$\frac{1}{1-c} \int_0^c F_a \mathrm{d}a \hookrightarrow \frac{1}{1-c} \int_0^c G_a \mathrm{d}a \tag{3.6}$$

always holds at a positive rate (Theorem 2.2), as long as $\frac{1}{1-c} \int_0^c F_a \mathrm{d}a$ is not degenerate. Hence, we only need to control the rate in $F_a \hookrightarrow G_a$ asymptotically as $a \to 1$. Thus, the case when $\mathbb{P}(X_F = \mathbb{E}[X_F]) > 0$ would not be problematic, and we keep the assumption $\mathbb{P}(X_F = \mathbb{E}[X_F]) = 0$ in this section for the ease of discussion.

To deal with the possibly infinite support of distributions, the right tail and the left tail need to be discussed separately. We first consider the case of one-side bounded distributions.

Lemma 3.5. *Suppose $F, G \in \mathcal{F}^1$, $F^{-1}(0) > -\infty$, $G^{-1}(0) > -\infty$, and*

$$\limsup_{a \to 1} \frac{G^{-1}(B_G(K_G^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} < \infty, \tag{3.7}$$

where B_F, B_G, K_F and K_G are defined in (2.2)-(2.5). Then $F \hookrightarrow G$.

Proof. Without loss of generality we assume $\mathbb{P}(X_F = \mathbb{E}[X_F]) = 0$, and F, G have mean zero and are non-degenerate. By the representation in Lemma 2.7, one has

$$F = \int_0^1 F_a \mathrm{d}a, \quad \text{and} \quad G = \int_0^1 G_a \mathrm{d}a.$$

By Lemma 2.4 (b), we have $F_a \hookrightarrow G_a$ at rate

$$\varepsilon_a = \min\left\{ \frac{F^{-1}(A_F(K_F^{-1}(a)))}{G^{-1}(A_G(K_G^{-1}(a)))}, \frac{F^{-1}(B_F(K_F^{-1}(a)))}{G^{-1}(B_G(K_G^{-1}(a)))} \right\} > 0.$$

By (3.7), $0 > F^{-1}(0) > -\infty$, and $0 > G^{-1}(0) > -\infty$, we have

$$\liminf_{a \to 1} \varepsilon_a > 0.$$

It then follows that $\varepsilon := \inf_{a \in (c,1)} \varepsilon_a > 0$ for some $c \in (0,1)$ and $F_a \hookrightarrow G_a$ at rate ε for $a \in (c,1)$. By (3.6) and Lemma 2.5, we have that $F_a \hookrightarrow G_a$. $\qquad\square$

By Lemma 3.5, in order to show that $F \hookrightarrow G$ for one-side bounded distributions F and G, it remains to show (3.7).

Lemma 3.6. *Suppose $F \in \mathcal{L}$ or $G \in \mathcal{L}$, $F^{-1}(0) > -\infty$, $G^{-1}(0) > -\infty$, and*

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} < \infty. \tag{3.8}$$

Then $F \hookrightarrow G$.

Proof. Without loss of generality we can assume that F and G have mean zero. Denote $\Phi_F(a) = \int_a^1 F^{-1}(t)\mathrm{d}t$ and $\Phi_G(a) = \int_a^1 G^{-1}(t)\mathrm{d}t$ for $a \in (0,1)$. By the definition of B_F and K_F, we have

$$
\begin{aligned}
\frac{\mathrm{d}B_F(K_F^{-1}(a))}{\mathrm{d}a} &= \frac{1}{F^{-1}(B_F(K_F^{-1}(a)))} \times \frac{1}{K_F'(K_F^{-1}(a))} \\
&= \frac{1}{F^{-1}(B_F(K_F^{-1}(a)))} \times \frac{1}{(F^{-1}(B_F(K_F^{-1}(a))))^{-1} + (F^{-1}(A_F(K_F^{-1}(a))))^{-1}} \\
&= \frac{1}{1 + F^{-1}(B_F(K_F^{-1}(a)))/F^{-1}(A_F(K_F^{-1}(a)))}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{\mathrm{d}\Phi_F(B_F(K_F^{-1}(a)))}{\mathrm{d}a} &= F^{-1}(B_F(K_F^{-1}(a)))\frac{\mathrm{d}B_F(K_F^{-1}(a))}{\mathrm{d}a} \\
&= \frac{1}{(F^{-1}(B_F(K_F^{-1}(a))))^{-1} + (F^{-1}(A_F(K_F^{-1}(a))))^{-1}} \\
&\to F^{-1}(0),
\end{aligned}
$$

as $a \to 1$. By L'Hospital's rule, we have

$$\lim_{a \to 1} \frac{\Phi_F(B_F(K_F^{-1}(a)))}{\Phi_G(B_G(K_G^{-1}(a)))} = \frac{F^{-1}(0)}{G^{-1}(0)} \in \mathbb{R}^+. \tag{3.9}$$

16

By (3.8), one obtains

$$\limsup_{a \to 1} \frac{\Phi_G(B_F(K_F^{-1}(a)))}{\Phi_F(B_F(K_F^{-1}(a)))} < \infty, \tag{3.10}$$

and

$$\limsup_{a \to 1} \frac{\Phi_G(B_G(K_G^{-1}(a)))}{\Phi_F(B_G(K_G^{-1}(a)))} < \infty. \tag{3.11}$$

(3.9)-(3.11) lead to

$$\limsup_{a \to 1} \frac{\Phi_G(B_F(K_F^{-1}(a)))}{\Phi_G(B_G(K_G^{-1}(a)))} < \infty, \tag{3.12}$$

and

$$\limsup_{a \to 1} \frac{\Phi_F(B_F(K_F^{-1}(a)))}{\Phi_F(B_G(K_G^{-1}(a)))} < \infty, \tag{3.13}$$

Since F^{-1} or G^{-1} is a regular varying function, Φ_F or Φ_G is also a regular varying function. In either case, (3.12) or (3.13) implies

$$\limsup_{a \to 1} \frac{1 - B_F(K_F^{-1}(a))}{1 - B_G(K_G^{-1}(a))} < \infty. \tag{3.14}$$

(i) In the case of F^{-1} being regular varying, (3.14) implies

$$\limsup_{a \to 1} \frac{F^{-1}(B_G(K_G^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} < \infty,$$

from which, along with (3.8), we obtain

$$\limsup_{a \to 1} \frac{G^{-1}(B_G(K_G^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} < \infty.$$

(ii) In the case of G^{-1} being regular varying, (3.14) implies

$$\limsup_{a \to 1} \frac{G^{-1}(B_G(K_G^{-1}(a)))}{G^{-1}(B_F(K_F^{-1}(a)))} < \infty,$$

from which, along with (3.8), we obtain

$$\limsup_{a \to 1} \frac{G^{-1}(B_G(K_G^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} < \infty,$$

In both cases, (3.7) holds. By Lemma 3.6 we have $F \hookrightarrow G$. $\qquad\square$

Theorem 3.7. *Suppose $F \in \mathcal{L}$ or $G \in \mathcal{L}$,*

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} < \infty, \tag{3.15}$$

and

$$\limsup_{t \to 0} \frac{G^{-1}(t)}{F^{-1}(t)} < \infty. \tag{3.16}$$

Then $F \hookrightarrow G$.

Proof. Without loss of generality we assume that F and G have mean zero, and $\mathbb{P}(X_F = 0) = \mathbb{P}(X_G = 0) = 0$. Let $p = \mathbb{P}(X_G > 0)$, $U_p \sim U[p, 1]$, $V_p \sim U[0, p]$ be two random variables, and $b = \mathbb{E}[G^{-1}(U_p)]/\mathbb{E}[F^{-1}(U_p)] > 0$. Denote by F_p the distribution of $bF^{-1}(U_p)$, by $\hat{F}_p$ the distribution of $bF^{-1}(V_p)$, by G_p the distribution of $G^{-1}(U_p)$, and by $\hat{G}_p$ the distribution of $G^{-1}(V_p)$. By Lemma 3.6 we have $F_p \hookrightarrow G_p$. Similarly, we have $\hat{F}_p \hookrightarrow \hat{G}_p$. By noting that the mean of F_p is equal to the mean of G_p, and the mean of $\hat{F}_p$ is equal to the mean of $\hat{G}_p$, we have, by the same argument in Lemma 2.5, that

$$p\hat{F}_p + (1 - p)F_p \hookrightarrow p\hat{G}_p + (1 - p)G_p = G.$$

Since $p\hat{F}_p + (1 - p)F_p$ is the distribution of bX_F, we conclude that $F \hookrightarrow G$. $\qquad\square$

Remark 3.8. By Lemma 3.2, we can see that (3.15)-(3.16) are necessary and sufficient conditions for $F \hookrightarrow G$ to hold for $F \in \mathcal{L}$ and arbitrary G.

3.3 General distributions

In this section we drop the assumption of regular variation in F or G. It turns out that a stronger condition is sufficient for $F \hookrightarrow G$ to hold. We give two results. The first result is for one-side bounded distributions.

Proposition 3.9. *Suppose non-degenerate distributions $F, G \in \mathcal{F}^1$ have means μ_F, μ_G, respectively, $F^{-1}(0) > -\infty$, $G^{-1}(0) > -\infty$, and*

$$\limsup_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} < \frac{\mu_G - G^{-1}(0)}{\mu_F - F^{-1}(0)}. \tag{3.17}$$

Then $F \hookrightarrow G$.

Proof. Without loss of generality we assume that F and G have mean zero. The case $G^{-1}(1) < \infty$ is covered by Theorem 2.2, and in the following we assume $G^{-1}(1) = \infty$. By (3.9) and (3.15) we have

$$\limsup_{a \to 1} \frac{\Phi_G(B_F(K_F^{-1}(a)))}{\Phi_G(B_G(K_G^{-1}(a)))} < 1. \tag{3.18}$$

It is easy to check that Φ_G is a strictly decreasing function if $G^{-1}(1) = \infty$. Thus, we have $B_G(K_G^{-1}(a))) < B_F(K_F^{-1}(a)))$ for a close to 1. Note that G^{-1} is non-decreasing, it follows that

$$\limsup_{a \to 1} \frac{G^{-1}(B_G(K_G^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} \leqslant \limsup_{a \to 1} \frac{G^{-1}(B_F(K_F^{-1}(a)))}{F^{-1}(B_F(K_F^{-1}(a)))} < \frac{G^{-1}(0)}{F^{-1}(0)}.$$

Thus, (3.7) holds, and by Lemma 3.6 we conclude that $F \hookrightarrow G$. $\qquad\square$

Finally, we give a general sufficient condition for $F \hookrightarrow G$ based on Proposition 3.9.

18

Theorem 3.10. *Suppose $F, G \in \mathcal{F}^1$ satisfy*

$$\lim_{t \to 1} \frac{G^{-1}(t)}{F^{-1}(t)} = 0, \tag{3.19}$$

and

$$\lim_{t \to 0} \frac{G^{-1}(t)}{F^{-1}(t)} = 0. \tag{3.20}$$

Then $F \hookrightarrow G$.

Proof. Using the same notation as in the proof of Theorem 3.7, we have, by (3.19)-(3.20) and Proposition 3.9, that $F_p \hookrightarrow G_p$ and $\hat{F}_p \hookrightarrow G_p$. The argument in the proof of Theorem 3.7 leads to the conclusion that $F \hookrightarrow G$. □

Theorem 3.10 tells us that $F \hookrightarrow G$ holds as long as F has a strictly heavier tail compared to G.

4 Conclusion

In this paper, we show that for any distributions F and G under general conditions, it is possible to find a sequence of random variables with common marginal distribution F such that the scaled average of this sequence converges almost surely to the distribution G. The random variables in this sequence are conditionally independent. This result adds to the study of distributions with fixed margins. The conclusion in this paper is clear: for a classic convergence problem with marginal distribution given, the shape of the limiting distribution can be anything, depending on the dependent structure.

We list some future research directions on the compatibility problem studied in this paper:

(a) cases when the initial distribution does not have finite mean or is not regular varying;

(b) problems in which the almost sure convergence is replaced by a weak convergence and the normalizing sequences are allowed to be arbitrary;

(c) characterizing the set of dependence structures which leads to a given limit;

(d) the compatibility problem of replacing the sum by the maximum or another function of the sequence of random variables;

(e) the compatibility problem with constraints on the dependence structure of the sequence of random variables;

(f) generalization of the obtained results to the setting of sequences of random vectors;

(g) applications of the results to time series analysis.

We believe that some of the above questions are challenging and important to modern statistical analysis.

References

Bernard, C., X. Jiang, and R. Wang (2014). Risk aggregation with dependence uncertainty. *Insurance Math. Econom. 54* (1), 93 – 108.

Chaganty, N. R. and H. Joe (2006). Range of correlation matrices for dependent bernoulli random variables. *Biometrika 93*, 197–206.

Cuadras, C. M. and J. E. Fortiana (eds.) (2002). *Distributions with Given Marginals and Statistical Modeling.* Kluwer, Netherlands.

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

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory. An Introduction.* Berlin: Springer.

Embrechts, P. and G. Puccetti (2006). Bounds for functions of multivariate risks. *J. Mult. Analysis 97*(2), 526–547.

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

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

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

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

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

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

McNeil, A.J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*; Princeton University Press: Princeton, NJ, USA.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer. Second edition.

Rüschendorf, L. (1982). Random variables with maximum sums. *Adv. in Appl. Probab. 14*(3), 623–632.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Heidelberg: Springer.

Shaked, M. and J. G. Shanthikumar (2007). *Stochastic Orders*. New York: Springer.

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

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

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