STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2020.1846115

Taylor & Francis
Taylor & Francis Group

New extreme value theory for maxima of maxima

Wenzhi Cao and Zhengjun Zhang

Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT

Although advanced statistical models have been proposed to fit complex data better, the advances of science and technology have generated more complex data, e.g., Big Data, in which existing probability theory and statistical models find their limitations. This work establishes probability foundations for studying extreme values of data generated from a mixture process with the mixture pattern depending on the sample length and data generating sources. In particular, we show that the limit distribution, termed as the accelerated max-stable distribution, of the maxima of maxima of sequences of random variables with the above mixture pattern is a product of three types of extreme value distributions. As a result, our theoretical results are more general than the classical extreme value theory and can be applicable to research problems related to Big Data. Examples are provided to give intuitions of the new distribution family. We also establish mixing conditions for a sequence of random variables to have the limit distributions. The results for the associated independent sequence and the maxima over arbitrary intervals are also developed. We use simulations to demonstrate the advantages of our newly established maxima of maxima extreme value theory.

ARTICLE HISTORY
Received 8 March 2020
Revised 24 October 2020
Accepted 1 November 2020

KEYWORDS
Maximum domain of attraction; max-stable distribution; competing-maximum domain of attractions; accelerated max-stable distribution; accelerated extreme value distribution

1. Introduction

Rigorous risk analysis helps to make better decisions and prevent great failures. Extreme value theory has been a powerful tool in risk analysis and is widely applied to risk analysis in finance, insurance, health, climate, and environmental studies. In classical extreme value theory, the sequence of data is assumed to have the same marginal distribution, and the limit distribution of the maxima is in one of the extreme value types if it exists. Galambos (1978), de Haan (1993), Beirlant et al. (2004), de Haan and Ferreira (2006), Leadbetter et al. (2012) and Resnick (2013) amongst many monographs are good literatures introducing the theoretical results in the classical extreme value theory. Mikosch et al. (1997), Embrechts et al. (1999), McNeil and Frey (2000), Coles (2001), Finkenstädt and Rootzén (2004), Castillo et al. (2005), Salvadori et al. (2007) and Dey and Yan (2016) introduce many applications of extreme value method to the areas of science, engineering, nature, finance, insurance and climate. For example, in financial applications, extreme value theory is one of the tools to calculate the Value-at-Risk (VaR) and Expected Shortfall (ES) (e.g., Rocco, 2014; Tsay, 2005). Chavez-Demoulin et al. (2016) offer an extreme value theory (EVT)-based statistical approach for modelling operational risk and losses, by taking into account dependence of the parameters on covariates and time. Zhang and Smith (2010) propose the multivariate maxima of moving maxima (M4) processes and apply the method to model jumps in returns in multivariate financial time series and predict the extreme co-movements in price returns. Daouia et al. (2018) use the extreme expectiles to measure VaR and marginal expected shortfall. In the statistical inference of maximum likelihood estimation (MLE), a discussion on the properties of maximum likelihood estimators of the parameters in generalised extreme value (GEV) distribution was given by Smith (1985). In the paper, it is shown that the classical properties of the MLE hold when the shape parameter $\xi > -1/2$, but not when $\xi \le -1/2$. Bücher and Segers (2017) give a general result on the asymptotic normality of the maximum likelihood estimator for parametric models whose support may depend on the parameters.

In the age of Big Data, the advances of science and technology have been changing data generating processes in a more complex way. As a result, the data structures and dependence structures accompanied by the collected data can be very different from the existed assumptions in many commonly used models. In the literature, advanced statistical models and machine learning approaches have been proposed to fit such complex data or learn the underlying structures better. For example, the support vector machine, the deep learning method, and the random forest method have now been very well recognised and wildly used in data analysis. In extreme value analysis for more

CONTACT Zhengjun Zhang zjz@stat.wisc.edu Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA

complex data, the same marginal distribution assumption and its derived extreme value distributions can be very restrictive and lack of data fitting power. Although statistical models, e.g., Heffernan et al. (2007), Naveau et al. (2011), Tang et al. (2013), Malinowski et al. (2015), Zhang and Zhu (2016) and Idowu and Zhang (2017), have been proposed to model extreme values observed from different data sources with different populations and max-domains of attraction, their probability foundations have not been established.

The definition of the classical maximum domain of attraction cannot be applied directly to the extreme values of data drawn from different populations mixed together. Note that we are not dealing with mixtures of distributions that may belong to a maximum domain of attraction of classical extreme value distribution. In this study, we are dealing with maxima of maxima in which the maxima resulted from each population has its limit extreme value distribution and norming and centering constants and convergence rate. For example, in many real-world applications, the risks one is exposed to usually come from different resources, and the risk at a given time is decided by the dominant one, i.e., not the added risk of all risks. Let us consider a specific example: Suppose a patient suffers two severe diseases. The risk of that the patient will die over a certain time may be best described by the maximum, not the sum, of two risk variables.

This work extends the definition of the maximum domain of attraction to maxima of maxima of sequences of random variables in which the mixing patterns change along with the sample size. The accelerated max-stable distribution (accelerated extreme value distribution) is expressed as a product of the classical extreme value distributions for the maxima of maxima resulted from different distributions. Some basic properties and theoretical results are provided. It can be seen that the classical extreme value distributions are special cases of our newly established family of accelerated max-stable distributions. The results obtained can be applied to more complex data, e.g., Big Data. The new results also establish the probability foundation of previously proposed statistical models in extreme time series modeling. Those models include Heffernan et al. (2007) that introduces one scheme where the maxima are taken over random variables with different distributions, and Zhang and Zhu (2016) that models intra-daily maxima of high-frequency financial data.

The structure of this paper is as follows. In Section 2, (1) we give a brief review of the classical extreme value theory; (2) we define our maxima of maxima of sequences of random variables; (3) we use examples to demonstrate the characteristics of the maxima of maxima; (4) we establish the convergence of maxima of maxima to the accelerated max-stable distributions; (5) we illustrate density functions of the new family of accelerated max-stable distributions and evaluate moments and tail equivalence. Simulations are used to demonstrate the advantages of the accelerated max-stable distribution family in terms of the estimation accuracy of high quantiles at different levels. We also apply this new accelerated max-stable distribution to the high quantiles of the daily maxima of 330 stock returns of S&P500 companies. In Section 3, the convergence of joint probability for general thresholds and approximation errors are developed. In Section 4, theoretical results for weakly dependent sequences are derived. Section 6 concludes. Additional figures and technical proofs are included in Section Appendix.

2. Accelerated max-stable distribution for independent sequences

2.1. A brief review of classical univariate extreme value theory

In classical extreme value theory, the central result is the Fisher-Tippett theorem which specifies the form of the limit distribution for centered and normalised maxima. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) non-degenerate random variables (rvs) with common distribution function F and $M_n = \max(X_1, \ldots, X_n)$ be the sample maxima. The Fisher-Tippett theorem states that: If for some norming constants $a_n > 0$ and centering constants b_n we have

$$P\big(a_n(M_n - b_n) \le x\big) \xrightarrow{w} H(x) \tag{1}$$

for some nondegenerate H, where $\xrightarrow{w}$ stands for convergence in distribution, then H belongs to one type of the following three cumulative distribution functions (cdf's):

$$\begin{aligned}
&\text{Fréchet}: \Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ \exp\{-x^{-\alpha}\}, & x > 0, \end{cases} \quad \alpha > 0. \\
&\text{Weibull}: \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \le 0 \\ 1, & x > 0, \end{cases} \quad \alpha > 0. \\
&\text{Gumbel}: \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.
\end{aligned} \tag{2}$$

Conversely, every extreme value distribution in (2) can be a limit in (1), and in particular, when H itself is the cdf of each X_i, the limit is itself. We say that F belongs to the maximum domain of attraction of the extreme value distribution of H, and denote as $F \in \text{MDA}(H)$ when (1) holds. H is also called the max-stable distribution since for any $n = 2, 3, \ldots$, there are constants $a_n > 0$ and b_n such that $H^n(a_n x + b_n) = H(x)$. Due to this property, the equivalence of extreme value distribution or max-stable distribution in practice is mutually implied.

2.2. Maxima of maxima

Suppose that the independent mixed sequence of random variables $\{X_i\}_{i=1}^n$ is composed of k subsequences $\{X_{j,i}\}_{i=1}^{n_j}$, $j = 1, 2, \ldots, k$; $\{X_{j,i}\}_{i=1}^{n_j} \overset{\text{i.i.d.}}{\sim} F_j(x)$, $n_j \to \infty$ as $n \to \infty$ and $n = n_1 + \cdots + n_k$. Denote $M_{j,n_j} = \max(X_{j,i},\ i = 1, \ldots, n_j)$ as the maximum of the jth subsequence, $j = 1, 2, \ldots, k$. Suppose $F_j \in \mathrm{MDA}(H_j)$, where H_j is one of the three types of extreme value distributions, i.e., M_{j,n_j} has the following limit distribution with some norming constants $a_{j,n_j} > 0$ and centering constants b_{j,n_j},

$$\lim_{n\to\infty} P(a_{j,n_j}(M_{j,n_j} - b_{j,n_j}) \le x) = H_j(x). \tag{3}$$

Define $M_n = \max(M_{1,n_1}, M_{2,n_2}, \ldots, M_{k,n_k})$, i.e., M_n is the maxima of k maxima of M_{j,n_j}s. Throughout the paper, M_n is termed as the maxima of maxima. Questions can be asked: (1) whether or not (1) holds with appropriately chosen norming constants $a_n > 0, b_n$; (2) if (1) holds, whether or not $a_n > 0, b_n$ are equivalent to any of $a_{j,n_j} > 0, b_{j,n_j}$; (3) whether or not $H(x)$ is a function of $H_j(x)$; (4) if all (1)–(3) hold, which one is the best method to be used in practice. This paper intends to answer these four questions.

Practical examples related to the above defined process can be numerous. For example, (1) the maximum temperature of the US in a day can be described by the maximum of maxima of regional maximum temperatures. In each region, the maximum temperature is the maximum temperature recordings among all weather stations in the region. Considering the regions' spatial and geographical patterns, the regional maxima certainly follow different extreme value distributions from one region to another region. The US temperature maxima are the maxima of regional maxima, and should be modelled by a distribution function that is a function of the regional extreme value distribution functions. (2) Considering the daily risk of high-frequency trading in a stock market, one can partition the data into hourly data (from 9:00 am to 4:00 pm). Suppose each hourly maxima M_{j,n_j} of negative returns can be approximately modelled by an extreme value distribution of $H_j(x)$. It is clear that M_n is better modelled by a function of $H_j(x)$, $j = 1, \ldots, 7$, i.e., not a single $H_j(x)$. We use the following simple example with $k = 2$ to illustrate the idea.

Example 2.1: The sequence $\{X_i\}_{i=1}^n$ is generated by $X_i = \max(Y_i, Z_i)$, where $\{Y_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} F_1(x)$, $\{Z_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} F_2(x)$, and $F_1(x)$ and $F_2(x)$ are two distribution functions. Assuming Y_i and Z_i are independent. Then $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} F(x) = F_1(x)F_2(x)$.

Remark: The form $X_i = \max(Y_i, Z_i)$ is the simplest case in the general mixture models introduced in Zhao and Zhang (2018). It is also the simplest case in the copula structured M4 models studied by Zhang and Zhu (2016).

For illustrative purpose of Example 2.1, let's consider two scenarios. Suppose $\{Y_i^{[k]}\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\{Z_i^{[k]}\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} U[a,b]$ for $k = 1, \ldots, m$. Here $U[a,b]$ represents the uniform distribution on the interval $[a,b]$. The superscript $[k]$ stands for the kth sample sequence. In Scenario 1, Figure 1 illustrates two different simulated sequences of $\{X_i^{[k]}\}_{i=1}^n$, where $X_i^{[k]} = \max(Y_i^{[k]}, Z_i^{[k]})$, and the maxima of $M_n^{[k]} = \max(X_1^{[k]}, \ldots, X_n^{[k]})$ for $n = 100$ and a particular k, e.g., $k = 1$. Next, we repeatedly generate $m = 10{,}000$ such sequences $\{X_i^{[k]}\}_{i=1}^n$, $k = 1, \ldots, m$. By taking the maxima $M_n^{[k]} = \max(X_1^{[k]}, \ldots, X_n^{[k]})$, $k = 1, \ldots, m$, the histogram of $\{M_n^{[k]}\}_{k=1}^m$ is displayed in Figure 2(a) with $a = -2.2$ and $b = 2.2$. In Scenario 2, by replacing the marginal distribution of $Z_i^{[k]}$ with $U[-2.8, 2.8]$, the

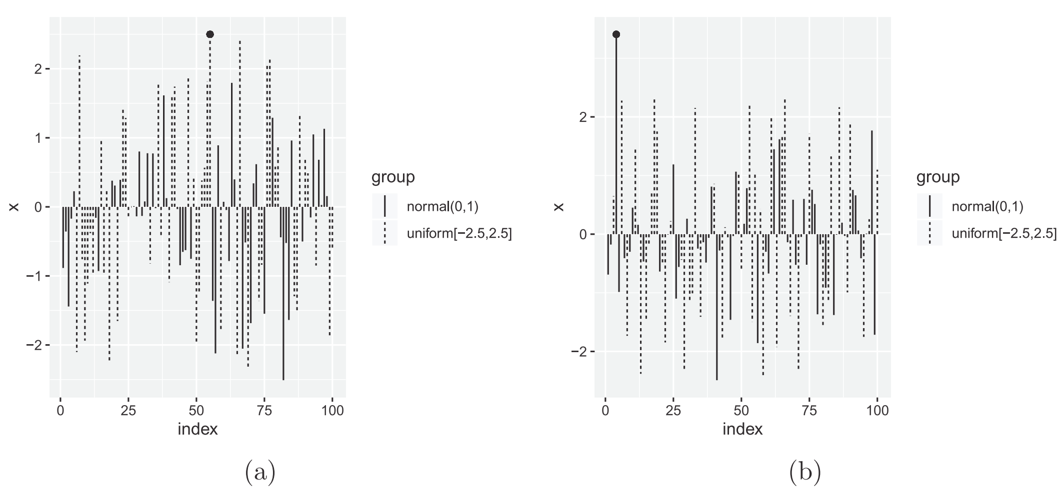

Figure 1. Simulated mixed sequences from normal and uniform distributions and their maxima (marked with black dots). In (a), the maximum is from the uniform distribution; in (b), the maximum is from $N(0, 1)$.

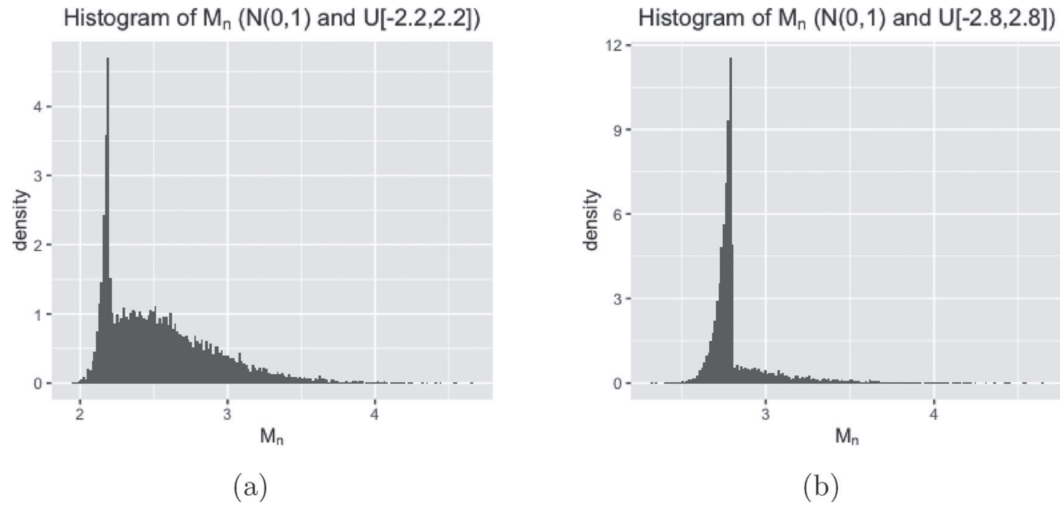

Figure 2. (a) Histogram of M_n from $N(0,1)$ and $U[-2.2,2.2]$. (b) Histogram of M_n from $N(0,1)$ and $U[-2.8,2.8]$.

histogram of $M_n^{[k]}$ is shown in Figure 2(b). It is clear that although $\{X_i^{[k]}\}_{i=1}^n$ is independent and identically distributed (i.i.d.), one can see that the distribution of M_n looks quite different from the three types of extreme value distributions.

In Example 2.1, the larger values of two paired underlying subsequences are observed while the smaller values are covered up by larger ones and are never observed. The sample sizes from the two subsequences are the same. However, in general mixed sequences the ratios of sample sizes from two subsequences n_1/n_2 can be any value between 0 and infinity and can vary as the total sample size grows. As a result, we can see many kinds of different patterns different from Figure 2.

In practice, data generating processes are naturally formed spatially and temporarily from underlying physical processes of studies. Here we provide two data generating processes in simulation.

(1) For a given sample size n, we set the numbers n_1 and n_2 satisfying $n_1 + n_2 = n$ and assume that $\lim_{n\to\infty} n_1/(n_1+n_2) \to r$, $r \in (0,1)$. Then we generate the specified n_1 and n_2 observations from two populations respectively, stack them in a sequence, and perform a random permutation of the combined sequence. In a physical process, the procedure can be designed as: we mix n_1 yellow balls and n_2 white balls in a bag. Then we draw balls sequentially. If a yellow ball is drawn, generate a number from the first population, otherwise from the second population.
(2) Alternatively, suppose a_{1,n_1}, a_{2,n_2} are norming constants defined in (3) with known population distributions in simulation, we can set $n_1 + n_2 = n$ and let $\lim_{n\to\infty} a_{1,n_1}/a_{2,n_2} = r$, and then solve n_1 and n_2 to generate the observations as the last step.

Example 2.2: Using the sampling scheme designed above. Suppose there are two sequences $\{X_{1,i}\}_{i=1}^{n_1} \overset{\text{i.i.d.}}{\sim} N(0,0.9)$ and $\{X_{2,i}\}_{i=1}^{n_2} \overset{\text{i.i.d.}}{\sim} U[-2,2]$ with $n_1 = 100$ and $n_2 = 200$. $\{X_k\}_{k=1}^{300}$ is mixed with these two sequences. Let $M_n^{[j]}$ be the maxima of the jth realisation of the sequence, $j = 1,\ldots,m$, $n = 300$. With $m = 10{,}000$, $\{M_n^{[j]}\}_{j=1}^m$ are calculated and the histogram is shown in Figure 3(a). The case of $n_1 = 200$ and $n_2 = 100$ is shown in Figure 3(b).

The histograms in Figure 3 look different from any of the three types of extreme value distributions discussed in (2). One feature is that they can be bimodal. On the other hand, the classical GEV distributions are all unimodal. Figure 3 shows two specific examples of choices of n_1 and n_2. In more general situations, the ratios of n_1 and n_2 can be any values in $(0,\infty)$. The ratio n_1/n_2 may also change as n increases. In Figures 2 and 3, the left parts of the distributions are dominated by the Weibull type induced by the uniform distribution, and the right parts resemble the Gumbel type induced by the normal distribution. The reason is that when we look at the maxima of $\{X_i\}_{i=1}^n$, there are two populations competing with each other. Taking (b) in Figure 3 as an example, the winners from $U[-2,2]$ form the steep peak on the left; and the winners from $N(0,0.9)$ form the smoother peak on the right.

Figure 4(a) shows the distribution of M_n for the sequence which is mixed with $N(0,1)$ and a Fréchet distribution. In (b), (c) and (d), they show the combinations of one Fréchet distribution and one Weibull distribution. Notice that in panel (b), the distribution looks left-skewed and is very similar to a Weibull distribution. However, with the effect of the Fréchet distribution, it actually has an infinite right endpoint.

In Figure 5, histograms of M_n are created such that the independent sequences of random variables $\{X_i\}_{i=1}^n$

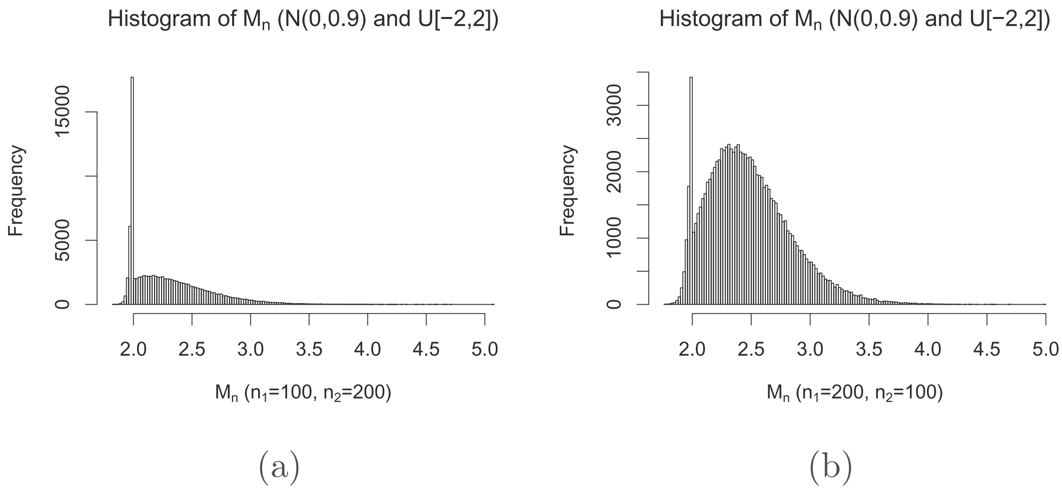

Figure 3. Histograms of combinations of M_n from $N(0, 0.9)$ and $U[-2, 2]$. (a) $n_1 = 100, n_2 = 200$. (b) $n_1 = 200, n_2 = 100$.

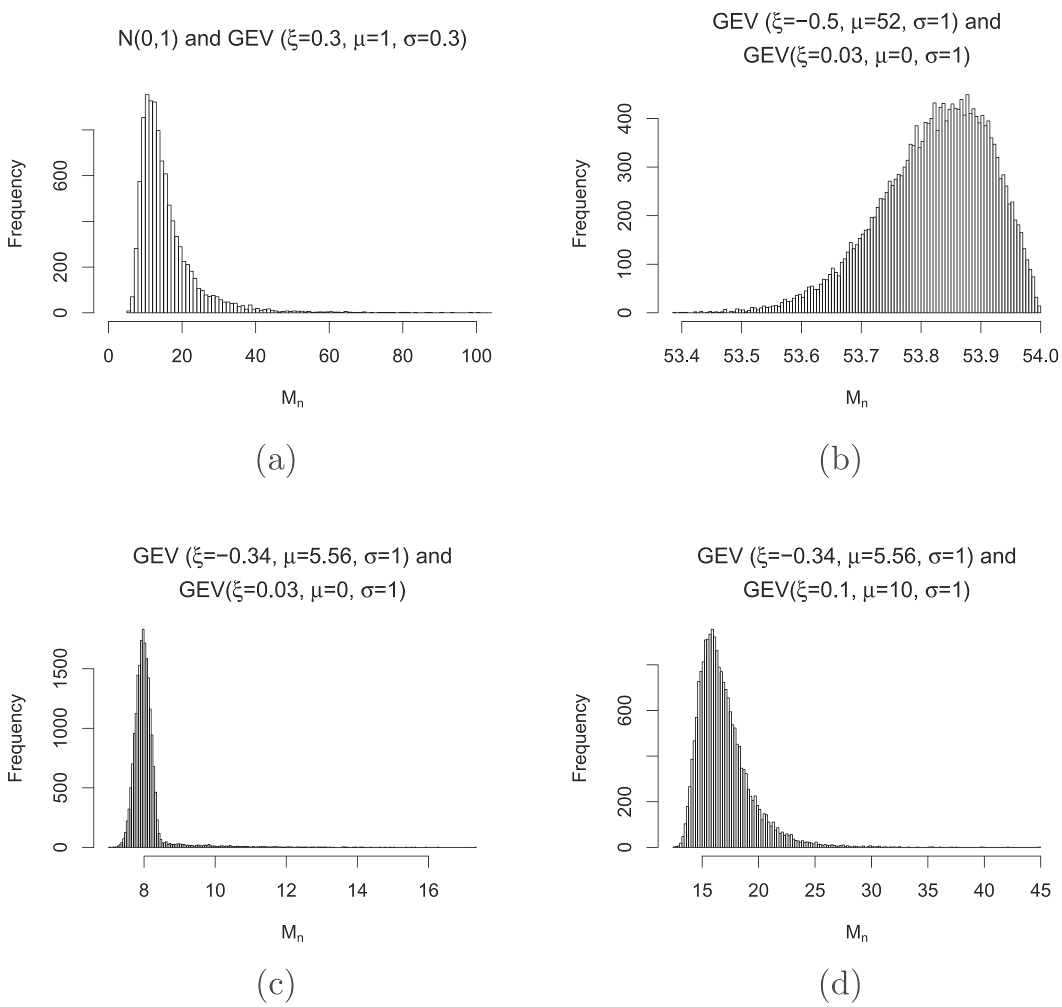

Figure 4. Histograms of M_n. (a) $N(0, 1)$ and Fréchet combination. (b)–(d) Some combinations of Fréchet and Weibull.

are generated by comparing the pairs of observations from normal and Weibull distribution. They can be unimodal or bimodel, left-skewed or right-skewed. If we use the GEV family to characterise the distributions of M_n in these examples, it may not capture the shape of the distribution properly. For example, if we look at the left part of the distribution in Figure 5(d), it resembles a Weibull distribution that has a finite right endpoint. However, because of the effect of the normal distribution on the right tail, the shape changes suddenly to be similar to a Gumbel distribution with infinite right endpoint. If we fit a GEV distribution to M_n, the left part

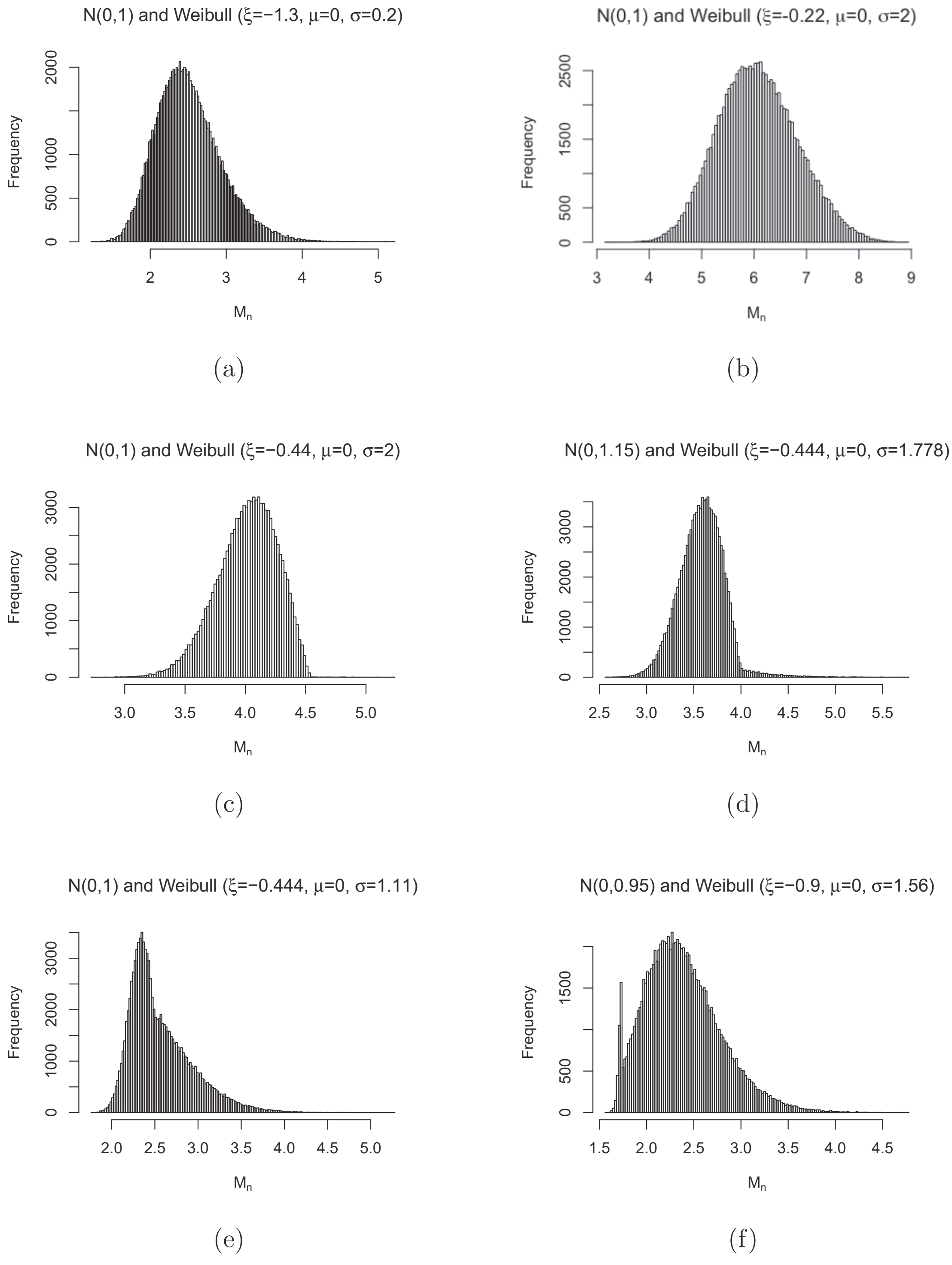

Figure 5. Histograms of M_n, with combinations of normal distribution and Weibull distribution.

with more sample data may have a large effect on the fitted distribution and we may underestimate the long tail on the right.

2.3. Convergence to the accelerated max-stable distribution

Throughout the paper, $x_F = \sup\{x; F(x) < 1\}$ is the right endpoint of a cdf F and let $\overline{F}(x) = 1 - F(x)$; $M_n = \max(M_{1,n_1}, \ldots, M_{k,n_k})$ is restricted to $k = 2$. For $k > 2$, relative results can be derived with additional notations. The following theorem shows that under certain conditions on the norming constants a_{j,n_j} and b_{j,n_j}, we can choose one set of the norming constants for the global maximum $M_n = \max(M_{1,n_1}, M_{2,n_2})$ to derive its limit distribution. Theorem 2.1 can be directly derived from Khintchine's theorem.

Theorem 2.1: *If M_{1,n_1} and M_{2,n_2} satisfy (3) for $j = 1, 2$, the limit distribution of M_n as $n \to \infty$ can be determined in the following cases:*

Case 1. *If* $\frac{a_{1,n_1}}{a_{2,n_2}} \to a > 0$, $a_{1,n_1}(b_{2,n_2} - b_{1,n_1}) \to b < +\infty$, *for some constants a and b, then*

$$P(a_{2,n_2}(M_n - b_{2,n_2}) \le x) \to H_1(ax+b)H_2(x). \tag{4}$$

Case 2. *If* $\frac{a_{1,n_1}}{a_{2,n_2}} \to 0$, $a_{1,n_1}(b_{2,n_2} - b_{1,n_1}) \to +\infty$ *then*

$$P(a_{2,n_2}(M_n - b_{2,n_2}) \le x) \to H_2(x). \tag{5}$$

Notice that the limit in Case 1 is the product of two extreme value distributions, $H_1(ax+b)H_2(x)$. Although it is in the product form, sometimes it can still be reduced to the three classical extreme value distributions. For example, $\exp\{-x^{-\alpha}\}\exp\{-(\frac{x}{2})^{-\alpha}\}$ is still a Fréchet type. However, in some situations, when the conditions in Case 1 are satisfied, the limit product form cannot be reduced to any one of the three extreme value distributions. We next present several examples to illustrate these possibilities.

Example 2.3 (Fréchet and Gumbel): Suppose $F_1(x) = \Phi_\alpha(x)$ is a Fréchet distribution function, and $F_2(x) = \Lambda(x)$ is the standard Gumbel distribution function. By choosing $a_{1,n_1} = n_1^{-1/\alpha}$, $b_{1,n_1} = 0$ $a_{2,n_2} = 1$, $b_{2,n_2} = \log n_2$ we have

$$P(M_{2,n_2} - \log n_2 \le x) \to \Lambda(x). \tag{6}$$

Then when $n_1^{1/\alpha}/\log n_2 \to \infty$, we have

$$\begin{aligned} &P(n_1^{-1/\alpha} M_n \le x) \\ &= P(n_1^{-1/\alpha} M_{1,n_1} \le x, M_{2,n_2} - \log n_2 \\ &\le n_1^{1/\alpha} x - \log n_2) \\ &\to \Phi_\alpha(x). \end{aligned}$$

Example 2.4 (Fréchet and Fréchet): Suppose $F_1(x) = \Phi_{\alpha_1}(x)$ and $F_2(x) = \Phi_{\alpha_2}(x)$ are two Fréchet distribution functions such that $\alpha_1 > \alpha_2$, which means that the tail of $F_2(x)$ is heavier than the tail of $F_1(x)$. By choosing norming constants $a_{1,n_1} = n_1^{-1/\alpha_1}$, $b_{1,n_1} = 0$ and $a_{2,n_2} = n_2^{-1/\alpha_2}$, $b_{2,n_2} = 0$ we have

$$P(n_j^{-1/\alpha_j} M_{j,n_j} \le x) = \Phi_{\alpha_j}(x), \quad j = 1, 2, \tag{7}$$

and

$$\begin{aligned} &P(n_2^{-1/\alpha_2} M_n \le x) \\ &= P\left(n_1^{-1/\alpha_1} M_{1,n_1} \le \frac{n_2^{1/\alpha_2}}{n_1^{1/\alpha_1}} x, n_2^{-1/\alpha_2} M_{2,n_2} \le x\right). \end{aligned} \tag{8}$$

If $n_2^{1/\alpha_2}/n_1^{1/\alpha_1} \to a > 0$, then

$$P(n_2^{-1/\alpha_2} M_n \le x) \to \Phi_{\alpha_1}(ax)\Phi_{\alpha_2}(x). \tag{9}$$

If $n_2^{1/\alpha_2}/n_1^{1/\alpha_1} \to +\infty$, then

$$P(n_2^{-1/\alpha_2} M_n \le x) \to \Phi_{\alpha_2}(x). \tag{10}$$

In Example 2.4, the sequence is mixed with two Fréchet distributions with different shape parameters. The limit distribution of M_n for this mixed sequence is the product of two Fréchet distributions, which is different from any of the three types of extreme value distributions.

Example 2.5 (Uniform and normal): Suppose $F_1(x)$ is the function of the uniform distribution $U[0,1]$, $F_2(x)$ is the distribution function of $N(0,1)$. By choosing

$$a_{1,n_1} = n_1, \quad b_{1,n_1} = 1, \tag{11}$$

and

$$\begin{aligned} a_{2,n_2} &= (2\log n_2)^{1/2}, \\ b_{2,n_2} &= (2\log n_2)^{1/2} \\ &\quad - \tfrac{1}{2}(2\log n_2)^{-1/2}(\log\log n_2 + \log 4\pi), \end{aligned}$$

we have

$$P(n_1(M_{1,n_1} - 1) \le x) \to e^x \tag{12}$$

for $x < 0$, and

$$P(a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \to \Lambda(x). \tag{13}$$

Then

$$\begin{aligned} &P(a_{2,n_2}(M_n - b_{2,n_2}) \le x) \\ &= P\left(n_1(M_{1,n_1} - 1) \le n_1\left(\frac{x}{a_{2,n_2}} + b_{2,n_2} - 1\right),\right. \\ &\left. a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x\right). \end{aligned}$$

Since $n_1(\frac{x}{a_{2,n_2}} + b_{2,n_2} - 1) \to +\infty$ for any x, we have

$$P(a_{2,n_2}(M_n - b_{2,n_2}) \le x) \to \Lambda(x). \tag{14}$$

Example 2.6 (Weibull and Weibull): Suppose $x_F < \infty$ and $K_1 > 0, K_2 > 0$,

$$F_1(x) = 1 - K_1(x_F - x)^{\alpha_1}, \quad x_F - K_1^{-1/\alpha_1} \le x \le x_F, \tag{15}$$

$$F_2(x) = 1 - K_2(x_F - x)^{\alpha_2}, \quad x_F - K_2^{-1/\alpha_2} \le x \le x_F, \tag{16}$$

are two polynomial functions with common finite endpoint x_F, $\alpha_1 > \alpha_2$. We can choose $a_{1,n_1} = (n_1K_1)^{1/\alpha_1}$, $b_{1,n_1} = x_F$, $a_{2,n_2} = (n_2K_2)^{1/\alpha_2}$, $b_{2,n_2} = x_F$, and

$$P((n_1K_1)^{1/\alpha_1}(M_{1,n_1} - x_F) \le x) \to \Psi_{\alpha_1}(x), \tag{17}$$

$$P((n_2K_2)^{1/\alpha_2}(M_{2,n_2} - x_F) \le x) \to \Psi_{\alpha_2}(x). \qquad (18)$$

If $\frac{(n_2K_2)^{1/\alpha_2}}{(n_1K_1)^{1/\alpha_1}} \to a > 0$, then

$$\begin{aligned} &P((n_1K_1)^{1/\alpha_1}(M_n - x_F) \le x) \\ &= P\left((n_1K_1)^{1/\alpha_1}(M_{1,n_1} - x_F), (n_2K_2)^{1/\alpha_2}\right. \\ &\quad \left.(M_{2,n_2} - x_F) \le \frac{(n_2K_2)^{1/\alpha_2}}{(n_1K_1)^{1/\alpha_1}}x\right) \\ &\quad \to \Psi_{\alpha_1}(x)\Psi_{\alpha_2}(ax). \end{aligned}$$

Example 2.7 (Normal and Pareto): Suppose $F_1(x)$ is the standard normal distribution function of $N(0,1)$, $F_2(x) = 1 - Kx^{-\alpha}$, $\alpha > 0$, $K > 0$ is a Pareto distribution function. Let

$$\begin{aligned} a_{1,n_1} &= (2\log n_1)^{1/2}, \\ b_{1,n_1} &= (2\log n_1)^{1/2} \\ &\quad - \tfrac{1}{2}(2\log n_1)^{-1/2}(\log\log n_1 + \log 4\pi), \\ a_{2,n_2} &= (Kn_2)^{-1/\alpha}, \quad b_{2,n_2} = 0. \end{aligned}$$

Then

$$\begin{aligned} &P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ &\quad \to \begin{cases} 0 & x < 0, \\ \exp(-e^{-x} - x^{-\alpha}) & x \ge 0. \end{cases} \end{aligned}$$

Furthermore, if $a_{2,n_2}b_{1,n_1} \to \infty$, then

$$P(a_{1,n_1}(M_n - b_{1,n_1}) \le x) \to \exp(-e^{-x}).$$

Example 2.8 (Cauchy and uniform distribution): $F_1(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x$ is the standard Cauchy distribution function, and $F_2(x) = x, 0 \le x \le 1$, let

$$a_{1,n_1} = \tan\frac{\pi}{n_1} \sim \frac{\pi}{n_1}, \quad b_{1,n_1} = 0, \qquad a_{2,n_2} = n_2, \quad b_{2,n_2} = 1.$$

Then

$$\begin{aligned} &P\left(\frac{\pi}{n_1}M_{1,n_1} \le x, n_2(M_{2,n_2} - 1) \le x\right) \\ &\quad \to \begin{cases} 0 & x < 0, \\ \exp(-x^{-1}) & x \ge 0, \end{cases} \end{aligned}$$

and

$$P(a_{1,n_1}(M_n - b_{1,n_1}) \le x) \to \begin{cases} 0 & x < 0, \\ \exp(-x^{-1}) & x \ge 0. \end{cases}$$

In Example 2.8, the limit distribution for the normalised M_{1,n_1} is 0 when $x < 0$, and the limit distribution for the normalised M_{2,n_2} is 1 when $x > 0$. Thus, the product is the same as the former one.

In Examples 2.3 and 2.5, we showed that when n is sufficiently large (goes to infinity), the distribution of M_n will be dominated by the subsequence whose marginal distribution has a heavier tail. In Examples 2.4 and 2.6, if the ratio $n_2^{1/\alpha_2}/n_1^{1/\alpha_1}$ converges to a constant, then one subsequence is never dominated by another, and the limit is of the product form that cannot be reduced to a classical extreme value distribution if $\alpha_1 \ne \alpha_2$.

We now introduce the accelerated max-stable distribution (AMSD) or the accelerated extreme value distribution (AEVD). We consider the convergence of the probability related to the normalised maxima M_{1,n_1} and M_{2,n_2} of two subsequences separately. By the relationship $M_n = \max(M_{1,n_1}, M_{2,n_2})$, we can use the accelerated max-stable distribution to approximate the distribution of M_n. The classical extreme value distributions will be special cases in the accelerated max-stable distribution family.

Definition 2.1: Let $H_1(x)$ and $H_2(x)$ be two max-stable distribution functions, we call $H(x) = H_1(x)H_2(x)$ the accelerated max-stable distribution (AMSD/AEVD) function, which is the product of two max-stable distribution functions. More generally, we also say that $H(x)$ belongs to the accelerated max-stable distribution family if it is the product of k max-stable distribution functions, $k \ge 2$.

Remark: If Z follows an accelerated max-stable distribution $H(x)$, then Z can be expressed as $Z = \max(Z_1, \dots, Z_k)$, where each Z_i follows a max-stable distribution. By taking maxima of $(Z_1, \dots, Z_k)$, Z_i values are accelerated by other components Z_js to get observed Z values. On the other hand, we have

$$\begin{aligned} &H_1(x)H_2(x)\cdots H_k(x) \\ &\le H_1(x)H_2(x)\cdots H_{k-1}(x) \\ &\le H_1(x)H_2(x)\cdots H_{k-2}(x) \\ &\le \cdots \le H_1(x) \end{aligned}$$

and

$$\begin{aligned} \overline{H_1(x)H_2(x)\cdots H_k(x)} &\ge \overline{H_1(x)H_2(x)\cdots H_{k-1}(x)} \\ &\ge \overline{H_1(x)H_2(x)\cdots H_{k-2}(x)} \\ &\ge \cdots \ge \overline{H_1(x)} \end{aligned}$$

where $\bar{H}(x)$ stands for the survival function, i.e.,

$$\overline{H_1(x)H_2(x)\cdots H_k(x)} = 1 - H_1(x)H_2(x)\cdots H_k(x).$$

The above inequalities may be regarded as accelerated survival rates. This observation motivates us to call the new distribution as the accelerated max-stable (extreme value) distribution. In the view of risk analysis, the systemic risk of Z is accelerated from individual risks of Z_js given a fixed confidence level.

For the independent sequence of random variables $\{X_i\}_{i=1}^{n}$ with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=2}^{n_2}$ defined as above, suppose (3) is satisfied with $j = 1, 2$ and norming constants $a_{j,n_j} > 0, b_{j,n_j}$, i.e.,

$$\lim_{n\to\infty} P(a_{j,n_j}(M_{j,n_j} - b_{j,n}) \le x) = H_j(x), \quad j = 1, 2, \tag{19}$$

then

$$P(\max(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}), a_{2,n_2}(M_{2,n_2} - b_{2,n_2})) \le x) \\ \to H(x) = H_1(x)H_2(x). \tag{20}$$

Definition 2.2: Suppose an independent sequence of random variables $\{X_i\}_{i=1}^{n}$ satisfies (19) and (20). We call the underlying distribution, F_{ni}, of X_i belongs to the competing-maximum domain of attractions of H_1 and H_2, and denote as $F_{ni} \in \mathrm{CMDA}(H_1, H_2)$.

We note that a max-stable distribution may also be decomposed into a product of two max-stable distributions. As a result, the max-stable distribution family can be thought as a family that is embedded in the accelerated max-stable distribution family. This observation can be seen in Theorem 2.1 that the limits of $P(a_{2,n_2}(M_n - b_{2,n_2}))$ under two different conditions belong to the accelerated max-stable distribution family. In other words, the accelerated max-stable distributions form an expanded family of distributions that can describe the limiting distribution of the normalised maxima for more general sequences.

For $k = 2$ and $F_{ni} \in \mathrm{CMDA}(H_1, H_2)$, AMSDs/AEVDs can have the following six possible combinations:

Case 1. $F_j \in \mathrm{MDA}(\Lambda), j = 1, 2,$

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Lambda\left(\frac{x - b_1}{a_1}\right)\Lambda\left(\frac{x - b_2}{a_2}\right).$$

Case 2. $F_1 \in \mathrm{MDA}(\Phi_{\alpha_1})$ and $F_2 \in \mathrm{MDA}(\Phi_{\alpha_2})$,

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Phi_{\alpha_1}\left(\frac{x - b_1}{a_1}\right)\Phi_{\alpha_2}\left(\frac{x - b_2}{a_2}\right).$$

Case 3. $F_1 \in \mathrm{MDA}(\Psi_{\alpha_1})$ and $F_2 \in \mathrm{MDA}(\Psi_{\alpha_2})$,

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Psi_{\alpha_1}\left(\frac{x - b_1}{a_1}\right)\Psi_{\alpha_2}\left(\frac{x - b_2}{a_2}\right).$$

Case 4. $F_1 \in \mathrm{MDA}(\Lambda)$ and $F_2 \in \mathrm{MDA}(\Phi_\alpha)$,

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Lambda\left(\frac{x - b_1}{a_1}\right)\Phi_{\alpha}\left(\frac{x - b_2}{a_2}\right).$$

Case 5. $F_1 \in \mathrm{MDA}(\Lambda)$ and $F_2 \in \mathrm{MDA}(\Psi_\alpha)$,

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Lambda\left(\frac{x - b_1}{a_1}\right)\Psi_{\alpha}\left(\frac{x - b_2}{a_2}\right).$$

Case 6. $F_1 \in \mathrm{MDA}(\Phi_{\alpha_1})$ and $F_2 \in \mathrm{MDA}(\Psi_{\alpha_2})$,

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \\ \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ \xrightarrow{w} \Phi_{\alpha_1}\left(\frac{x - b_1}{a_1}\right)\Psi_{\alpha_2}\left(\frac{x - b_2}{a_2}\right).$$

It is easy to see that the classical extreme value distributions are special cases of the AMSD family. For any $a > 0, b > 0$ satisfying $\frac{1}{a} + \frac{1}{b} = 1$, we have

$$\begin{aligned}\Lambda(x) &= \exp\{-e^{-x}\} = \exp\{-e^{-(\frac{x}{a}+\frac{x}{b})}\} \\ &= \exp\{(-e^{-x-\log a} - e^{-x-\log b})\} \\ &= \Lambda(x + \log a)\Lambda(x + \log b). \\ \Phi_\alpha(x) &= \exp\{-x^{-\alpha}\} \\ &= \exp\left\{-\left(\frac{x}{a^{-\frac{1}{\alpha}}}\right)^{-\alpha} - \left(\frac{x}{b^{-\frac{1}{\alpha}}}\right)^{-\alpha}\right\} \\ &= \Phi\left(\frac{x}{a^{-\frac{1}{\alpha}}}\right)\Phi\left(\frac{x}{b^{-\frac{1}{\alpha}}}\right). \\ \Psi_\alpha(x) &= \exp\{-(-x)^{\alpha}\} \\ &= \exp\left\{-\left(\frac{-x}{a^{\frac{1}{\alpha}}}\right)^{\alpha} - \left(\frac{-x}{b^{\frac{1}{\alpha}}}\right)^{\alpha}\right\} \\ &= \Psi\left(\frac{x}{a^{\frac{1}{\alpha}}}\right)\Psi\left(\frac{x}{b^{\frac{1}{\alpha}}}\right).\end{aligned}$$

Since $H_1(x)$ and $H_2(x)$ are max-stable distributions, for any $n_1 = 2, 3, \ldots$ and $n_2 = 2, 3, \ldots$, there are constants $a_{1,n_1} > 0, b_{1,n_1}, a_{2,n_2} > 0, b_{2,n_2}$ such that $H_1(x)H_2(x) = H_1^{n_1}(a_{1,n_1}x + b_{1,n_1})H_2^{n_2}(a_{2,n_2}x + b_{2,n_2})$.

In Equation (20), we considered the convergence of

$$P(\max(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}), a_{2,n_2}(M_{2,n_2} - b_{2,n_2})) \le x),$$

instead of the traditional $P(a_n(M_n - b_n) \le x)$. If n_1 and n_2 are sufficiently large, by (19) we have $P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x) \approx G_1(x)$ and $P(a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \approx G_2(x)$, then

$$\begin{aligned}P(M_n \le x) &= P(\max(M_{1,n_1}, M_{2,n_2}) \le x) \\ &= P(M_{1,n_1} \le x)P(M_{2,n_2} \le x)\end{aligned}$$

$$\approx G_1\big(a_{1,n_1}(x-b_{1,n_1})\big)G_2\big(a_{2,n_2}(x-b_{2,n_2})\big)$$
$$= G_1^*(x)G_2^*(x) \tag{21}$$

where G_j^* is of the same type as $G_j, j=1,2$.

To close this section, we remark that (21) is the basis of applying the newly introduced AMSD/AEVD family to real data. Based on (21), in practice, we don't need to worry about the values of n_1, n_2, a_{1,n_1}, b_{1,n_1}, a_{2,n_2}, b_{2,n_2}, as they are absorbed in $G_1^*(x)$ and $G_2^*(x)$, see also Coles (2001). In our examples, we have used some fixed numbers for n_1 and n_2. They are just for simulation convenience. When n tends to infinity, the values of n_1 and n_2 will depend on n.

The next section presents density functions and shapes from which one can see the flexibility of applying the new distribution to real data modelling.

2.4. Density functions and density plots

The density function of the accelerated max-stable distribution requires some discussion of the support region of the cumulative distribution function. We can express the two terms in the product using the form of the generalised extreme value distribution,

$$F(x) = H_{\xi_1;\mu_1,\sigma_1}(x)H_{\xi_2;\mu_2,\sigma_2}(x) = \exp\left\{-\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1} - \left[1+\xi_2\frac{x-\mu_2}{\sigma_2}\right]^{-1/\xi_2}\right\}, \tag{22}$$

where $1+\xi_1\frac{x-\mu_1}{\sigma_1} > 0$ and $1+\xi_2\frac{x-\mu_2}{\sigma_2} > 0$. We inclu de the special case $H_{0;\mu_i,\sigma_2}$ as the limit of $H_{\xi_i;\mu_i,\sigma_i}$ for $\xi_i \to 0, i=1,2$. Denote the density function as $f(x)$ and let

$$h(x) = \exp\left\{-\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1} - \left[1+\xi_2\frac{x-\mu_2}{\sigma_2}\right]^{-1/\xi_2}\right\} \times \left[\frac{1}{\sigma_1}\left(1+\xi_1\frac{x-\mu_1}{\sigma_1}\right)^{-1/\xi_1-1} + \frac{1}{\sigma_2}\left(1+\xi_2\frac{x-\mu_2}{\sigma_2}\right)^{-1/\xi_2-1}\right].$$

Since ξ_1 and ξ_2 are symmetric, we only present one of them. We have the following six cases for the density functions.

Case 1. $\xi_1 = 0, \xi_2 = 0$.

$$f(x) = \exp\left\{-e^{-\frac{x-\mu_1}{\sigma_1}} - e^{-\frac{x-\mu_2}{\sigma_2}}\right\}\left[\frac{1}{\sigma_1}e^{-\frac{x-\mu_1}{\sigma_1}} + \frac{1}{\sigma_2}e^{-\frac{x-\mu_2}{\sigma_2}}\right], \quad x \in \mathbb{R}.$$

Case 2. $\xi_1 > 0, \xi_2 > 0$, assuming $\mu_1 - \frac{\sigma_1}{\xi_1} \ge \mu_2 - \frac{\sigma_2}{\xi_2}$, then

$$f(x) = \begin{cases} h(x) & \text{if } x > \mu_1 - \frac{\sigma_1}{\xi_1}, \\ 0 & \text{if } x \le \mu_1 - \frac{\sigma_1}{\xi_1}. \end{cases}$$

Case 3. $\xi_1 < 0, \xi_2 < 0$, assuming $\mu_1 - \frac{\sigma_1}{\xi_1} \ge \mu_2 - \frac{\sigma_2}{\xi_2}$, then

$$f(x) = \begin{cases} h(x) & \text{if } x < \mu_2 - \frac{\sigma_2}{\xi_2}, \\ \exp\left\{-\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1}\right\} \times \left[\frac{1}{\sigma_1}\left(1+\xi_1\frac{x-\mu_1}{\sigma_1}\right)^{-1/\xi_1-1}\right] & \text{if } \mu_2 - \frac{\sigma_2}{\xi_2} \le x \le \mu_1 - \frac{\sigma_1}{\xi_1}, \\ 0 & \text{if } x > \mu_1 - \frac{\sigma_1}{\xi_1}. \end{cases}$$

Case 4. $\xi_1 = 0, \xi_2 > 0$.

$$f(x) = \begin{cases} \exp\left\{-e^{-\frac{x-\mu_1}{\sigma_1}} - \left[1+\xi_2\frac{x-\mu_2}{\sigma_2}\right]^{-1/\xi_2}\right\} \times \left[\frac{1}{\sigma_1}e^{-\frac{x-\mu_1}{\sigma_1}} + \frac{1}{\sigma_2}(1+\xi_2\frac{x-\mu_2}{\sigma_2})^{-1/\xi_2-1}\right] & \text{if } x > \mu_2 - \frac{\sigma_2}{\xi_2}, \\ 0 & \text{if } x \le \mu_2 - \frac{\sigma_2}{\xi_2}. \end{cases}$$

Case 5. $\xi_1 = 0, \xi_2 < 0$.

$$f(x) = \begin{cases} \exp\left\{-e^{-\frac{x-\mu_1}{\sigma_1}} - \left[1+\xi_2\frac{x-\mu_2}{\sigma_2}\right]^{-1/\xi_2}\right\} \times \left[\frac{1}{\sigma_1}e^{-\frac{x-\mu_1}{\sigma_1}} + \frac{1}{\sigma_2}(1+\xi_2\frac{x-\mu_2}{\sigma_2})^{-1/\xi_2-1}\right] & \text{if } x < \mu_2 - \frac{\sigma_2}{\xi_2}, \\ \exp\{-e^{-\frac{x-\mu_1}{\sigma_1}}\} \times \frac{1}{\sigma_1}e^{-\frac{x-\mu_1}{\sigma_1}} & \text{if } x \ge \mu_2 - \frac{\sigma_2}{\xi_2}. \end{cases}$$

Case 6. $\xi_1 > 0, \xi_2 < 0$.If $\mu_1 - \frac{\sigma_1}{\xi_1} \ge \mu_2 - \frac{\sigma_2}{\xi_2}$,

$$f(x) = \begin{cases} \exp\left\{-\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1}\right\} \times \left[\frac{1}{\sigma_1}\left(1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1-1}\right) & \text{if } x > \mu_1 - \frac{\sigma_1}{\xi_1}, \\ 0 & \text{if } x \le \mu_1 - \frac{\sigma_1}{\xi_1}. \end{cases}$$

If $\mu_1 - \frac{\sigma_1}{\xi_1} < \mu_2 - \frac{\sigma_2}{\xi_2}$,

$$f(x) = \begin{cases} \exp\left\{-\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1}\right\} \times \left[\frac{1}{\sigma_1}\left[1+\xi_1\frac{x-\mu_1}{\sigma_1}\right]^{-1/\xi_1-1}\right] & \text{if } x > \mu_2 - \frac{\sigma_2}{\xi_2}, \\ h(x) & \text{if } \mu_1 - \frac{\sigma_1}{\xi_1} \le x \le \mu_2 - \frac{\sigma_2}{\xi_2}, \\ 0 & \text{if } x < \mu_1 - \frac{\sigma_1}{\xi_1}. \end{cases}$$

In Figures 6 and 7, four density plots of Weibull-Gumbel type are shown. In Figure 8, panel (a) is the density plot of Fréchet-Fréchet type; and panel (b) is the density plot of Fréchet-Gumbel type. We can observe that they capture the shapes of the histograms shown in Figures 4 and 5.

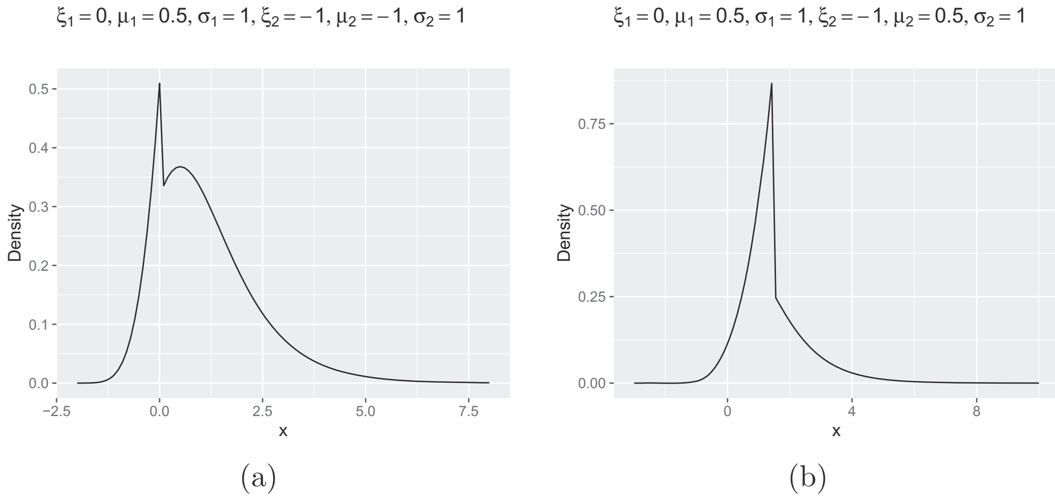

Figure 6. Density plots of the accelerated max-stable distributions with Weibull-Gumbel combinations. (a) $\xi_1 = 0, \mu_1 = 0.5, \sigma_1 = 1, \xi_2 = -1, \mu_2 = -1, \sigma_2 = 1$. (b) $\xi_1 = 0, \mu_1 = 0.5, \sigma_1 = 1, \xi_2 = -1, \mu_2 = 0.5, \sigma_2 = 1$.

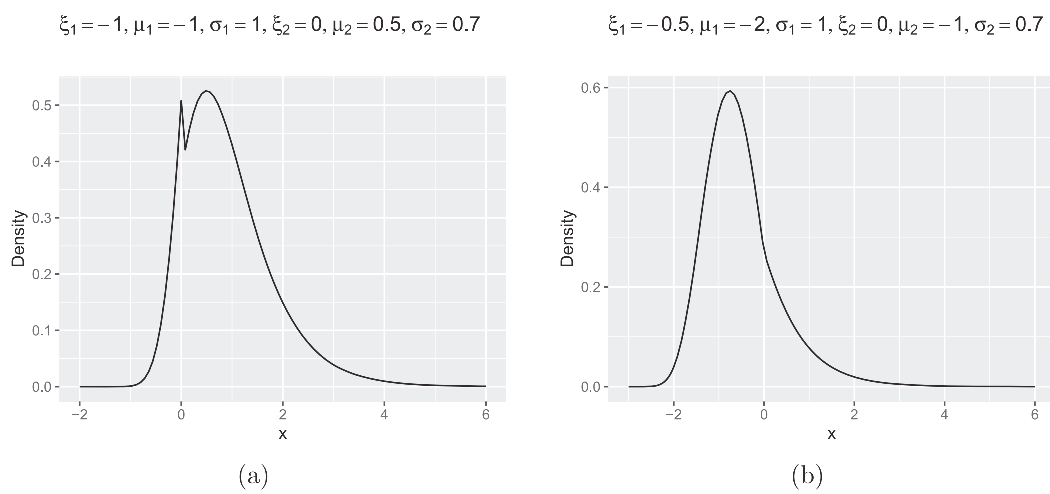

Figure 7. Density plots of the accelerated max-stable distributions with Weibull-Gumbel combinations. (a) $\xi_1 = -1, \mu_1 = -1, \sigma_1 = 1, \xi_2 = 0, \mu_2 = 0.5, \sigma_2 = 0.7$. (b) $\xi_1 = -0.5, \mu_1 = -2, \sigma_1 = 1, \xi_2 = 0, \mu_2 = -1, \sigma_2 = 0.7$.

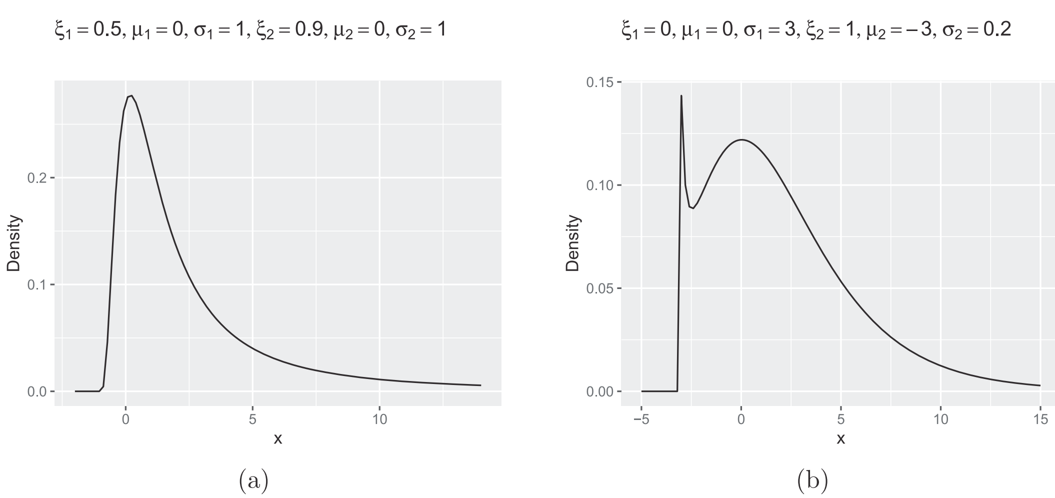

Figure 8. (a) Density plot of the accelerated max-stable distribution with Fréchet-Fréchet combinition. $\xi_1 = 0.5, \mu_1 = 0, \sigma_1 = 1, \xi_2 = 0.9, \mu_2 = 0, \sigma_2 = 1$. (b) Density plot of the accelerated max-stable distributions with Fréchet-Gumbel combinition. $\xi_1 = 0, \mu_1 = 0, \sigma_1 = 3, \xi_2 = 1, \mu_2 = -3, \sigma_2 = 0.2$.

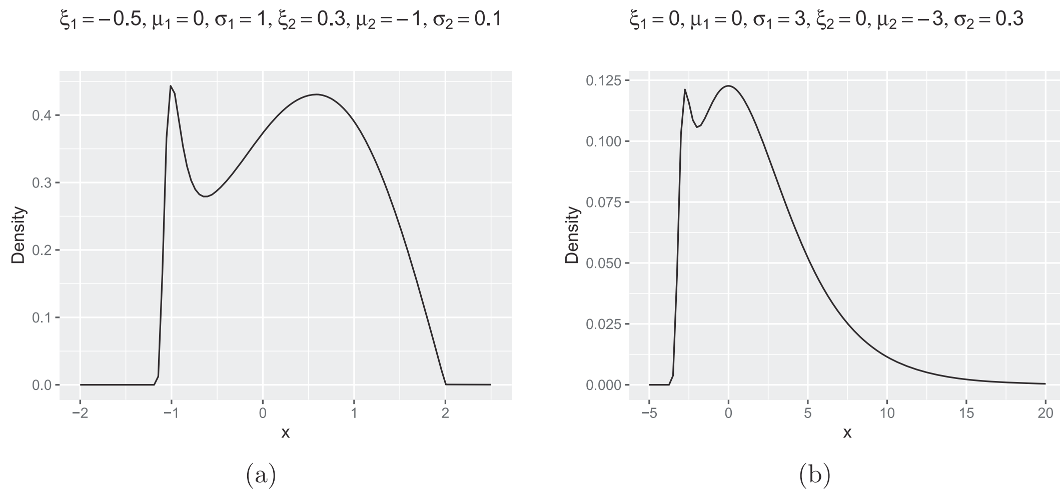

Figure 9. (a) Density plot of the accelerated max-stable distribution with Weibull-Fréchet combination. $\xi_1 = -0.5, \mu_1 = 0, \sigma_1 = 1, \xi_2 = 0.3, \mu_2 = -1, \sigma_2 = 0.1$. (b) Density plot of the accelerated max-stable distribution with Gumbel-Gumbel combinition. $\xi_1 = 0, \mu_1 = 0, \sigma_1 = 3, \xi_2 = 0, \mu_2 = -3, \sigma_2 = 0.3$.

In Figure 9(b), it is for $\xi_1 = \xi_2 = 0$, i.e., the combination of two Gumbel distributions. In this case, the density plot is bimodal, which is different from that of a Gumbel distribution. Suppose that $X_{1,i} \sim N(\mu_1, \sigma_1)$ and $X_{2,j} \sim N(\mu_2, \sigma_2)$, $1 \le i \le n_1$ and $1 \le j \le n_2$, then we have some norming constants $a_{1,n_1} > 0, b_{1,n_1}$ and $a_{2,n_2} > 0, b_{2,n_2}$ such that

$$\begin{aligned} &P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \\ &\quad \to \Lambda\left(\frac{x-\mu_1}{\sigma_1}\right)\Lambda\left(\frac{x-\mu_2}{\sigma_2}\right) \\ &\quad = \exp\{-e^{-\frac{x-\mu_1}{\sigma_1}} - e^{-\frac{x-\mu_2}{\sigma_2}}\}. \end{aligned} \tag{23}$$

Here the limit product form requires that the two scale parameters $\sigma_1 \neq \sigma_2$. Otherwise, the product $\exp\{-e^{-\frac{x-\mu_1}{\sigma}} - e^{-\frac{x-\mu_2}{\sigma}}\}$ reduces to the Gumbel type.

2.5. Tail equivalence and the existence of moments

In this section, we discuss some results of tail-equivalence, and which moments are finite for certain AMSDs/AEVDs.

Definition 2.3: Two cdf's F and H are called tail-equivalent if they have the same right endpoint, i.e., if $x_F = x_H$, and

$$\lim_{x \to x_F} \overline{F}(x)/\overline{H}(x) = c \tag{24}$$

for some constant $0 < c < \infty$.

We have the following facts.

Fact 2.1: *It is clear that the product distribution of a Weibull distribution and another type of extreme value distribution $H(x)$ is tail equivalent to $H(x)$.*

Fact 2.2: *Suppose $X \sim \Phi_{\alpha_1}\Phi_{\alpha_2}$, let $\mu_k = E(X^k)$ be the kth moment of X, then μ_k is finite only if $k < \min(\alpha_1, \alpha_2)$.*

Suppose $\alpha_1 < \alpha_2$, then Φ_{α_1} has a heavier tail than Φ_{α_2}. Let $\mu_k^{(1)}$ be the kth moment of $X \sim \Phi_{\alpha_1}$. We know that $\mu_k^{(1)} < \infty$ only if $k < \alpha_1$. This implies that $\Phi_{\alpha_1}\Phi_{\alpha_2}$ has the same right-tail heaviness as Φ_{α_1}.

Fact 2.3: *If $0 < \alpha_1 < \alpha_2$, then $\Phi_{\alpha_1}\Phi_{\alpha_2}$ and Φ_{α_1} are tail-equivalent.*

Fact 2.4: *Suppose $X \sim \Lambda(x)\Phi_\alpha(x)$. Let $\mu_k = E(X^k)$ be the kth moment of X. Then μ_k is finite only if $k < \alpha$.*

Fact 2.5: *$\Lambda(x)\Phi_\alpha(x)$ and $\Phi_\alpha(x)$ are tail-equivalent.*

Fact 2.6: *If $H_1(x)$ has a heavier tail than $H_2(x)$, then the accelerated max-stable distribution $H_1(x)H_2(x)$ is tail-equivalent to $H_1(x)$.*

3. Joint convergence and approximation errors

3.1. Convergence of joint probability for general thresholds

It may also be interesting to consider the limits of $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$ for some sequences u_{1,n_1} and u_{2,n_2} not necessarily of the form $x/a_{i,n_i} + b_{i,n_i}$ or even not dependent on x. Here n_1 and n_2 are the lengths of the two subsequences, we may write them specifically as $n_1(n)$ and $n_2(n)$ since they vary with the total length n. When choosing $u_{j,n_j} = x/a_{j,n_j} + b_{j,n_j}$ for $j = 1, 2$, it becomes the problem we discussed before. The question is:

> Which conditions on F_1 and F_2 ensure that the limit of $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$ for $n \to \infty$ exists for appropriate constants u_{1,n_1} and u_{2,n_2}?

Some conditions on tails $\bar{F}_1$ and $\bar{F}_2$ are required to ensure that $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$ converges to a non-trivial limit, i.e., a number in $(0, 1)$.

Theorem 3.1: *Suppose $\{X_i\}_{i=1}^n$ is an independent sequence of random variables which is mixed with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$ with underlying distributions $F_1(x)$ and $F_2(x)$, $n_1 \to \infty$ and $n_2 \to \infty$ as $n \to \infty$. Let $0 \le \tau < \infty$ and $\{u_{1,i}\}_{i=1}^{n_1}$ and $\{u_{2,i}\}_{i=1}^{n_2}$ are two sequences of real numbers such that*

$$n_1(1 - F_1(u_{1,n_1})) + n_2(1 - F_2(u_{2,n_2})) \to \tau \quad \text{as } n \to \infty. \tag{25}$$

Then

$$P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2}) \to e^{-\tau} \quad \text{as } n \to \infty. \tag{26}$$

Conversely, if (26) holds for some $0 \le \tau < \infty$, then so does (25).

Remark: Since $1 - F(u_{j,n_j})$ is the probability that $X_{j,i}$ exceeds level u_{j,n_j}, Equation (25) means that the expected number of exceedences of u_{1,n_1} by $\{X_{1,i}\}_{i=1}^{n_1}$ and u_{2,n_2} by $\{X_{2,i}\}_{i=1}^{n_2}$ in total converges to τ. When the sequence is generated from one distribution $F(x)$, Theorem 3.1 can be reduced to the classical result by choosing $u_{1,n_1} = u_{2,n_2} = u_n$. That is

$$n(1 - F(u_n)) \to \tau, \tag{27}$$

if and only if

$$P(M_n \le u_n) \to e^{-\tau} \tag{28}$$

as $n \to \infty$.

The following corollary gives the conditions such that we can choose one of $\{u_{1,i}\}_{i=1}^{n_1}$ and $\{u_{2,i}\}_{i=1}^{n_2}$ to be applied to M_n, and derive a similar limit of $P(M_n \le u_n)$. The condition involves both the ratio of two tail probabilities $\frac{1-F_1(u_{1,n_1})}{1-F_2(u_{2,n_2})}$ and $\frac{n_1}{n_2}$.

Corollary 3.1: *Let $0 \le \tau_1 < \infty$, $0 \le \tau_2 < \infty$. Suppose that there exist two sequences u_{1,n_1} and u_{2,n_2} such that*

$$\begin{aligned} n_1(1 - F_1(u_{1,n_1})) &\to \tau_1, \\ n_2(1 - F_2(u_{2,n_2})) &\to \tau_2. \end{aligned} \tag{29}$$

Then

$$P(M_{1,n_1} \le u_{1,n}, M_{2,n_2} \le u_{2,n}) \to e^{-\tau_1 - \tau_2}. \tag{30}$$

Moreover, if $\frac{n_2(1-F_2(u_{1,n_1}))}{n_1(1-F_1(u_{1,n_1}))} \to t$, where $0 \le t < \infty$, then

$$P(M_n \le u_{1,n_1}) \to e^{-\tau_1(1+t)}. \tag{31}$$

Specifically, if we choose $u_{1,n_1} = \frac{x}{a_{1,n_1}} + b_{1,n_1}$, $u_{2,n_2} = \frac{x}{a_{2,n_2}} + b_{2,n_2}$, and suppose that

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x) \to G_1(x), \tag{32}$$

$$P(a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \to G_2(x), \tag{33}$$

then G_1 and G_2 belong to the GEV distribution family and the limit in (31) becomes $G_1(x)G_2(x)$.

The following is an example of mixed sequence and the limit properties of the maxima of subsequences and the global maxima.

Example 3.1: Suppose $\{X_i\}_{i=1}^n$ is a sequence of random variables combining two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$. Suppose $\frac{n_1}{n} \to p$, where $0 \le p \le 1$, $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$ are i.i.d. from a Pareto distribution with $F_1(x) = 1 - Kx^{-\alpha_1}$, $\alpha_1 > 0$, $K > 0$, $x > 0$ and a Fréchet distribution with $F_2(x) = \exp(-x^{-\alpha_2})$, $\alpha_2 > 0$, $x > 0$, respectively.

Since $(1 - F_1(tx))/(1 - F_1(t)) = x^{-\alpha_1}$ for each $x > 0$, so that Type II (Fréchet) limit applies. For $u_{1,n_1} = (Kn_{1,n_1}/\tau)^{1/\alpha_1}$ we have $1 - F_1(u_{1,n_1}) = \tau/n_1$, so that

$$P\left(M_{1,n_1} \le \left(\frac{Kn_1}{\tau}\right)^{1/\alpha_1}\right) \to e^{-\tau}. \tag{34}$$

Putting $\tau = x^{-\alpha_1}$ for $x \ge 0$,

$$P((Kn_1)^{-1/\alpha_1} M_{1,n_1} \le x) \to \exp(-x^{-\alpha_1}). \tag{35}$$

On the other hand, $F_2^{n_2}(n_2^{1/\alpha_2}x) = F_2(x)$, i.e., $P(n_2^{-1/\alpha_2} M_{2,n_2} \le x) = F_2(x)$.

Then we have for $x \ge 0$,

$$P((Kn_1)^{-1/\alpha_1} M_{1,n_1} \le x, n_2^{-1/\alpha_2} M_{2,n_2} \le x) \to \exp(-x^{-\alpha_1} - x^{-\alpha_2}). \tag{36}$$

Since

$$\begin{aligned} \lim_{x\to\infty} \frac{\bar{F}_1(x)}{\bar{F}_2(x)} &= \lim_{x\to\infty} \frac{Kx^{-\alpha_1}}{1 - \exp(-x^{-\alpha_2})} \\ &= \lim_{x\to\infty} \frac{Kx^{-\alpha_1}}{x^{-\alpha_2} + O(x^{-2\alpha_2})} \\ &\to \begin{cases} 0 & \alpha_1 > \alpha_2, \\ K & \alpha_1 = \alpha_2, \\ \infty & \alpha_1 < \alpha_2. \end{cases} \end{aligned}$$

When $\alpha_1 = \alpha_2$, the condition $\frac{n_2(1-F_2(u_{1,n_1}))}{n_1(1-F_1(u_{1,n_1}))} \to \frac{1-p}{pK}$ in Corollary 3.1 is satisfied, hence

$$P(M_n \le n_1^{-1/\alpha_1} x) \to \exp\left(-x^{-\alpha_1}\left(1 + \frac{1-p}{pk}\right)\right). \tag{37}$$

Since $n_1 \sim np$, we also have

$$P(M_n \le (np)^{-1/\alpha_1} x) \to \exp\left(-x^{-\alpha_1}\left(1 + \frac{1-p}{pk}\right)\right). \tag{38}$$

3.2. Approximation error

The convergence results are usually accompanied by the question of the approximation error. Suppose $n_1(1-F_1(u_{1,n_1}))\to\tau_1$ and $n_2(1-F_2(u_{2,n_2}))\to\tau_2$, writing $\tau_{1,n_1}=n_1(1-F_1(u_{1,n_1}))$ and $\tau_{2,n_2}=n_2(1-F_2(u_{2,n_2}))$, then by Theorem 3.1 we have

$$P(M_{1,n_1}\le u_{1,n_1},M_{2,n_2}\le u_{2,n_2})\to e^{-\tau_1-\tau_2}. \qquad (39)$$

The approximation can be decomposed into several parts. We have

$$\left(1-\frac{\tau_{1,n_1}}{n_1}\right)^{n_1}\approx e^{-\tau_{1,n_1}},\qquad \left(1-\frac{\tau_{2,n_2}}{n_2}\right)^{n_2}\approx e^{-\tau_{2,n_2}},$$

and

$$e^{-\tau_{1,n_1}}\approx e^{-\tau_1},\quad e^{-\tau_{2,n_2}}\approx e^{-\tau_2}.$$

We denote

$$\begin{aligned}\Delta_{1,n_1}&=\left(1-\frac{\tau_{1,n_1}}{n_1}\right)^{n_1}-e^{-\tau_{1,n_1}},\\ \Delta'_{1,n_1}&=e^{-\tau_{1,n_1}}-e^{-\tau_1},\\ \Delta_{2,n_2}&=\left(1-\frac{\tau_{2,n_2}}{n_2}\right)^{n_2}-e^{-\tau_{2,n_2}},\\ \Delta'_{2,n_2}&=e^{-\tau_{2,n_2}}-e^{-\tau_2}.\end{aligned}$$

Then

$$\begin{aligned}P(M_{1,n_1}\le u_{1,n_1})-e^{-\tau_1}&=\Delta_{1,n_1}+\Delta'_{1,n_1},\\ P(M_{2,n_2}\le u_{2,n_2})-e^{-\tau_2}&=\Delta_{2,n_2}+\Delta'_{2,n_2}.\end{aligned}$$

The following result gives the bound for the approximation error.

Theorem 3.2: *Let $\{X_i\}_{i=1}^n$ be an independent sequence of random variables mixed with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$, which satisfies $n_1(1-F_1(u_{1,n_1}))\to\tau_1$ and $n_2(1-F_2(u_{2,n_2}))\to\tau_2$, Δ_{1,n_1}, Δ'_{1,n_1}, Δ_{2,n_2}, Δ'_{2,n_2} are defined as above, then*

$$\begin{aligned}P(M_{1,n_1}\le u_{1,n_1},M_{2,n_2}\le u_{2,n_2})-e^{-\tau_1-\tau_2}\\ \le\Delta_{1,n_1}+\Delta'_{1,n_1}+\Delta_{2,n_2}+\Delta'_{2,n_2}\end{aligned}$$

with

$$\begin{aligned}0\le-\Delta_{j,n_j}&\le\frac{\tau_{j,n_j}e^{-\tau_{j,n_j}}}{2}\cdot\frac{1}{n_j-1}\\ &\le 0.3\cdot\frac{1}{n_j-1},\quad \text{for } j=1,2,\end{aligned}$$

where the first bound is asymptotically sharp, in the sense that if $\tau_{j,n_j}\to\tau_j$ then $\Delta_{j,n_j}\sim-(\frac{\tau_je^{-\tau_j}}{2})/n_j$. Furthermore, for $\tau_j-\tau_{j,n_j}\le\log 2$,

$$\Delta'_{j,n_j}=e^{\tau_j}\{(\tau_j-\tau_{j,n_j})+\theta_j(\tau_j-\tau_{j,n_j})^2\},$$

with $0<\theta_j<1$.

If $\tau_{j,n_j}\to\tau_j$ for $u_{j,n_j}=x/a_{j,n_j}+b_{j,n_j}$, then (39) holds. By Lemma A.1, (39) holds also if a_{j,n_j} and b_{j,n_j} are replaced by different constants α_{j,n_j} and β_{j,n_j}, satisfying $\alpha_{j,n_j}/a_{j,n_j}\to 1$ and $(\beta_{j,n_j}-b_{j,n_j})/a_{j,n_j}\to 0$. However, the speed of convergence to zero of Δ'_{j,n_j} (thus the speed of $P(M_{j,n_j}\le u_{j,n_j})$ to $e^{-\tau_j}$) can be very different for different choices of norming constants.

4. Weakly dependent sequences

In this section, we extend the independent sequences to weakly dependent sequences. For a sequence of random variables $\{X_i\}_{i=1}^n$ with identical distribution, it is *stationary* if $\{X_{j_1},\dots,X_{j_n}\}$ and $\{X_{j_1+m},\dots,X_{j_n+m}\}$ have the same joint distribution for any choice of $n, j_1,\dots,j_n$, and m. For the mixed sequence, we will provide some alternatives so that the desired results still hold. We assume that the dependence between $X_{i,k}$ and $X_{i,j}$ falls off in some specific way as $|k-j|$ increases.

4.1. Review of some weakly dependent conditions

Some weakly dependent conditions in the literature can be generalised to the scenarios of mixed sequences. For m-dependent sequence $\{X_i\}_{i=1}^n$, X_i and X_j are independent if $|i-j|>m$. Another commonly used condition is the *strong mixing* condition first introduced by Rosenblatt (1956). A sequence of random variables $\{X_i\}_{i=1}^n$ is said to satisfy the strong mixing condition if for some $A\in\mathcal{F}(X_1,\dots,X_p)$ and $B\in\mathcal{F}(X_{p+k+1},X_{p+k+2},\dots)$

$$|P(A\cap B)-P(A)P(B)|<g(k)$$

for any p and k, where $g(k)\to 0$ as $k\to\infty$; $\mathcal{F}(\cdot)$ is the σ-field generated by the indicated random variables. The function $g(k)$ does not depend on the sets A and B, so the strong mixing condition is uniform.

For normal sequences, the *correlation* between X_k and X_j may be a better measure of dependence. We can also use the dependence restriction $|\mathrm{Corr}(X_k,X_j)|\le g(|k-j|)$, where $g(k)\to 0$ as $k\to\infty$.

Since the event $\{M_n\le u\}$ is the same as $\{X_1\le u, X_2\le u,\dots,X_n\le u\}$. We may restric the events on this type of event. Following Leadbetter et al. (2012), we use $F_{i_1,\dots,i_n}(u)$ to denote $P(X_{i_1}\le u,X_{i_2}\le u,\dots,X_{i_n}\le u)$. The following condition D is a weakened condition of strong mixing.

The condition D will be said to hold if for any integers $i_1<\dots<i_p$ and $j_1<\dots<j_{p'}$ for which $j_1-i_p\ge l$, and any real u,

$$|F_{i_1,\dots,i_p,j_1,\dots,j_{p'}}(u)-F_{i_1,\dots,i_p}(u)F_{j_1,\dots,j_{p'}}(u)|\le g(l) \qquad (40)$$

where $g(l)\to 0$ as $l\to\infty$.

Under the condition D, the Extremal Types Theorem also holds. Since we usually deal with the event $\{M_n\le$

$u_n\}$ for some levels $\{u_n\}$, the condition can still be weakened.The condition $D(u_n)$ is defined as follows.

The condition $D(u_n)$ will be said to hold if for any integers

$$1 \le i_1 < \cdots < i_p < j_1 < \cdots < j_{p'} \le n \tag{41}$$

for which $j_1 - i_p \ge l$, we have

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_{p'}}(u_n)| \le \alpha_{n,l} \tag{42}$$

where $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

The condition $D(u_n)$ guarantees that $\liminf P(M_n \le u_n) \ge e^{-\tau}$. We still need a further assumption to have the opposite inequality for the upper limit. Here we present the $D'(u_n)$ condition used in Watson (1954) and Loynes (1965). This condition bounds the probability of more than one exceedance among $X_1, \ldots, X_{[n/k]}$, therefore no multiple points in the point process of exceedances.

The condition $D'(u_n)$ will be said to hold for the sequence of random variables $\{X_i\}_{i=1}^n$, if

$$\limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P\{X_1 > u_n, X_j > u_n\} \to 0 \tag{43}$$

as $k \to \infty$, (where [ ] denotes the interger part).

If both conditions $D(u_n)$ and $D'(u_n)$ are satisfied, we have $P(M_n \le u_n) \to e^{-\tau}$ is equivalent to $n(1 - F(u_n)) \to \tau$ as $n \to \infty$ for $0 \le \tau < \infty$.

4.2. Weakly dependent mixed sequences

To generalise the results from non-mixed sequences to mixed sequences, we need to modify the conditions of $D(u_n)$ and $D'(u_n)$. We use $\mathbf{u}_n$ to denote the vector of levels (u_{1,n_1}, u_{2,n_2}) when the sequence $\{X_i\}_{i=1}^n$ is composed of two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=2}^{n_2}$, $n_1 + n_2 = n$. We further assume that $\frac{n_1}{n} \to p$ as $n \to \infty$, $0 \le p \le 1$, so that $\frac{n_2}{n} \to 1 - p$.

Before introducing the more general $D(\mathbf{u}_n)$ condition, we introduce some new notations. Let $u_n^{(i)} = u_{1,n_1} I(X_i \in \{X_{1,i}\}_{i=1}^{n_1}) + u_{2,n_2} I(X_i \in \{X_{2,i}\}_{i=1}^{n_2})$. Here $I(A) = 1$ indicates that the event A is true, otherwise $I(A) = 0$. The notation $u_n^{(i)}$ represents the threshold for X_i, which depends on the subsequence that X_i belongs to. For example, if $X_1 = X_{1,1}$ and $X_2 = X_{2,1}$, then $P(X_1 \le u_n^{(1)}, X_2 \le u_n^{(2)})$ represents $P(X_{1,1} \le u_{1,n_1}, X_{2,1} \le u_{2,n_2})$. After introducing this notation, we can state the condition $D(\mathbf{u}_n)$ as follows.

The condition $D(\mathbf{u}_n)$ will be said to hold for the mixed sequence of random variables $\{X_i\}_{i=1}^n$ with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$ if for any integers

$$1 \le i_1 < \cdots < i_p < j_1 < \cdots < j_{p'} \le n \tag{44}$$

for which $j_1 - i_p \ge l$, we have

$$\begin{aligned} |P(X_{i_1} \le u_n^{(i_1)}, \ldots, X_{i_p} &\le u_n^{(i_p)}, X_{j_1} \le u_n^{(j_1)}, \ldots, X_{j_{p'}} \le u_n^{(j_{p'})}) \\ &- P(X_{i_1} \le u_n^{(i_1)}, \ldots, X_{i_p} \le u_n^{(i_p)}) \\ &P(X_{j_1} \le u_n^{(j_1)}, \ldots, X_{j_{p'}} \le u_n^{(j_{p'})})| < \alpha_{n,l} \end{aligned}$$

where $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

Similarly, we can also extend the condition $D'(u_n)$ for mixed sequences, which is denoted as $D'(\mathbf{u}_n)$.

The condition $D'(\mathbf{u}_n)$ will be said to hold for the mixed sequence of random variables $\{X_i\}_{i=1}^n$ and levels $\mathbf{u}_n = (u_{1,n_1}, u_{2,n_2})$ if

$$\limsup_{n\to\infty} k \sum_{1\le i<j\le[n/k]} P(X_i > u_n^{(i)}, X_j > u_n^{(j)}) \to 0 \quad \text{as } k \to \infty \tag{45}$$

where $u_n^{(i)} = u_{1,n_1} I(X_i \in \{X_{1,i}\}_{i=1}^{n_1}) + u_{2,n_2} I(X_i \in \{X_{2,i}\}_{i=2}^{n_2})$, and [ ] denotes the integer part.

Equation (45) means that $\limsup_{n\to\infty} \sum_{1\le i<j\le[n/k]} P(X_i > u_n^{(i)}, X_j > u_n^{(j)}) = o(1/k)$. It can be observed that if $D(\mathbf{u}_n)$ holds for the mixed sequence $\{X_i\}_{i=1}^n$, then $D(u_{j,n_j})$ also holds for the subsequence $\{X_{j,i}\}_{i=1}^{n_j}$, for $j = 1, 2$. The same conclusion is also true for the condition $D'(\mathbf{u}_n)$.

After introducing the conditions $D(\mathbf{u}_n)$ and $D'(\mathbf{u}_n)$, we have the extended results for mixed sequences. We assume that the two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=2}^{n_2}$ are independent with each other. Also, for any interval I_n with l_n members, there are a_n members from $\{X_{1,i}\}_{i=1}^{n_1}$ and b_n members from $\{X_{2,i}\}_{i=2}^{n_2}$. We assume that the proportion of each subsequence $\frac{a_n}{l_n} \to p$ and $\frac{b_n}{l_n} \to 1 - p$, where $0 \le p \le 1$.

Theorem 4.1: *Let $\{X_i\}_{i=1}^n$ be a weakly dependent mixed sequence of random variables with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$, with sample size proportions $\frac{n_1}{n} \to p$ and $\frac{n_2}{n} \to 1 - p$ as $n \to \infty$, $0 \le p \le 1$. Suppose that $D(\mathbf{u}_n)$ and $D'(\mathbf{u}_n)$ hold for $\{X_i\}_{i=1}^n$, then for $0 \le \tau < \infty$,*

$$P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2}) \to e^{-\tau} \tag{46}$$

if and only if

$$n_1(1 - F_1(u_{1,n_1})) + n_2(1 - F_2(u_{2,n_2})) \to \tau. \tag{47}$$

Based on Theorem 4.1, we have the following corollary.

Corollary 4.1: *The same conclusions hold with $\tau = \infty$ (i.e., $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2}) \to 0$ if and only if $n_1(1 - F_1(u_{1,n_1})) + n_2(1 - F_2(u_{2,n_2})) \to \infty$) if the requirements that $D(\mathbf{u}_n)$, $D'(\mathbf{u}_n)$ hold are replaced by the condition that, for arbitrarily large $\tau (< \infty)$, there exists a vector of levels $\mathbf{v}_n = (v_{1,n_1}, v_{2,n_2})$ such that $v_{1,n_1} >$*

$u_{1,n_1}, v_{2,n_2} > u_{2,n_2}$, *which satisfy* $n_1(1 - F_1(v_{1,n_1})) + n_2(1 - F_2(v_{2,n_2})) \to \tau$ *with* $D(\mathbf{v}_n)$ *and* $D'(\mathbf{v}_n)$ *hold.*

Theorem 4.1 tells us the property of the joint probability $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$ given the tail properties of F_1 and F_2. $n_1(1 - F_1(u_{1,n_1})) + n_2(1 - F_2(u_{2,n_2}))$ is the mean exceedances of the two thresholds by the corresponding subsequences in total. Theorem 4.1 is the generalisation of Theorem 3.1 under the condition that the mixed sequence is weakly dependent within each subsequence.

4.3. Associated independent sequences

The 'independent sequence associated with $\{X_i\}_{i=1}^n$' can be used to study the maxima of dependent sequence. It was first introduced by Loynes (1965). For a weakly dependent sequence of random variables $\{X_i\}_{i=1}^n$, the notation $\{\widehat{X}_i\}_{i=1}^n$ is used to be the independent sequence with the same marginal distribution as $\{X_i\}_{i=1}^n$, and write $\widehat{M}_n = \max(\widehat{X}_1, \ldots, \widehat{X}_n)$. When $\{X_i\}_{i=1}^n$ is mixed with two subsequences $\{X_{1,i}\}_{i=1}^{n_1}$ and $\{X_{2,i}\}_{i=1}^{n_2}$ with different marginal distributions, we still have the associated independent subsequences $\{\widehat{X}_{1,i}\}_{i=1}^{n_1}$ and $\{\widehat{X}_{2,i}\}_{i=1}^{n_1}$, and we write $\widehat{M}_{i,n_i} = \max(\widehat{X}_{i,1}, \ldots, \widehat{X}_{i,n_i})$, for $i = 1, 2$.

The following Theorem 4.2 tells us that, under the weakly dependent conditions, $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$ and $P(\widehat{M}_{1,n_1} \le u_{1,n_1}, \widehat{M}_{2,n_2} \le u_{2,n_2})$ have the same limit if it exists. By Theorem 4.3, we can choose the same norming constant as the independent sequence to derive the same limit of $P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x)$ and $P(a_{1,n_1}(\widehat{M}_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(\widehat{M}_{2,n_2} - b_{2,n_2}) \le x)$.

Theorem 4.2: *Let* $\{X_i\}_{i=1}^n$ *be a mixed sequence of random variables with two subsequences* $\{X_{1,i}\}_{i=1}^{n_1}$ *and* $\{X_{2,i}\}_{i=1}^{n_2}$, *independent with each other. Suppose* $D(\mathbf{u}_n)$ *and* $D'(\mathbf{u}_n)$ *hold for a vector of levels* $\mathbf{u}_n = (u_{1,n_1}, u_{2,n_2})$. *Then* $P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2}) \to \theta > 0$ *if and only if* $P(\widehat{M}_{1,n_1} \le u_{1,n_1}, \widehat{M}_{2,n_2} \le u_{2,n_2}) \to \theta$. *The same holds with* $\theta = 0$ *if the condition* $D(\mathbf{u}_n)$ *and* $D'(\mathbf{u}_n)$ *are replaced by the requirement that for arbitrarily large* $\tau < \infty$ *there exists* $\mathbf{v}_n = (v_{1,n_1}, v_{2,n_2})$ *such that* $v_{1,n_1} > u_{1,n_1}, v_{2,n_2} > u_{2,n_2}$, *which satisfy* $n_1(1 - F_1(v_{1,n_1})) + n_2(1 - F_2(v_{2,n_2})) \to \tau$ *with* $D(\mathbf{v}_n)$ *and* $D'(\mathbf{v}_n)$ *hold.*

Theorem 4.3: *Suppose that* $D(\mathbf{u}_n)$ *and* $D'(\mathbf{u}_n)$ *hold for the mixed sequence of random variables* $\{X_i\}_{i=1}^n$, *with* $u_{1,n_1} = x/a_{1,n_1} + b_{1,n_1}, u_{2,n_2} = x/a_{2,n_2} + b_{2,n_2}$ *for each real x. Then*

$$P(a_{1,n_1}(M_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x) \to G(x) \tag{48}$$

if and only if

$$P(a_{1,n_1}(\widehat{M}_{1,n_1} - b_{1,n_1}) \le x, a_{2,n_2}(\widehat{M}_{2,n_2} - b_{2,n_2}) \le x) \to G(x) \tag{49}$$

for some non-degenerate continuous distribution function $G(x)$.

With the results in this section, for weakly dependent sequences with conditions $D(\mathbf{u}_n)$ and $D'(\mathbf{u}_n)$ being satisfied, we can treat them as independent sequences when studying the limit distribution of the maxima. In the next section, some numerical experiments and estimation results are presented.

5. Numerical experiments

5.1. Simulation

We study the accuracy of the accelerated max-stable distributions in estimating the high quantiles of the simulated data. They are compared to the results using the classical GEV distribution alone. To simulate the data, we first generate two sequences from two different GEV distributions with parameters ξ_1, μ_1, σ_1 and ξ_2, μ_2, σ_2, denoting them as $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, here $n = 2000$. We pair them and find their maxima, $Z_i = \max(X_i, Y_i)$, then fit the accelerated max-stable distribution and GEV distribution separately to the sequence $\{Z_i\}_{i=1}^n$ using maximum likelihood method. Using each fitted distribution, we generate a new sequence $\{Z_i^*\}_{i=1}^n$ and calculate the proportion of $\{Z_i^*\}_{i=1}^n$ that exceeds the 90th, 95th and 99th percentiles of the original sequence $\{Z_i\}_{i=1}^n$. The simulation scenarios cover all the possible combinations of three types of extreme value distributions. For each combination scenario, the process is repeated 100 times and the standard deviations of the estimated proportions are shown in the parentheses. The results are in Table 1.

From Table 1, for the 90th percentile, we can observe that accelerated max-stable distributions perform better than the GEV alone, and the exceeding proportion is closer to the theoretical value 0.1. The same is true for the 95th percentiles. For both of these two percentiles, the proportions are larger than the theoretical value 0.1 and 0.05 in general, with the GEV distribution deviating more. This observation implies that both estimations overestimate the true values. For the 99th percentiles, we observe that the differences are not large overall. With a few cases (2nd and 3rd), the accelerated max-stable distribution outperforms the GEV distribution. Also, the proportions for accelerated max-stable distributions are all larger than 0.01 and those for GEV distributions are mostly smaller than 0.01. This phenomenon implies that the accelerated max-stable distribution may overestimate the 99th percentiles. On the other hand, the GEV distribution may underestimate the 99th percentiles.

5.2. Real data

In this section, we apply both AMSD/AEVD and GEV fitting to stock data. The data contains the daily closing

Table 1. The proportions of the simulated data based on the fitted accelerated max-stable distributions and GEV distributions that exceeds the 90th, 95th and 99th percentiles of the original data Z_i.

	ξ_1	μ_1	σ_1	ξ_2	μ_2	σ_2	90th AEVD	90th GEV	95th AEVD	95th GEV	99th AEVD	99th GEV
1	0.1	0	1	−0.1	0	1	**0.1002** (0.005)	0.1071 (0.006)	**0.0502** (0.004)	0.0540 (0.004)	**0.0104** (0.002)	0.0094 (0.002)
2	0.1	0	1	−0.2	2	1	**0.0988** (0.005)	0.1191 (0.006)	**0.0496** (0.004)	0.0688 (0.006)	**0.0102** (0.002)	0.0140 (0.003)
3	0	0	1	−0.2	2	1	**0.1022** (0.009)	0.1094 (0.008)	**0.0516** (0.007)	0.0555 (0.006)	**0.0097** (0.003)	0.0088 (0.003)
4	0	0	1	−0.1	0	1	**0.1008** (0.007)	0.1039 (0.008)	**0.0507** (0.006)	0.0525 (0.006)	**0.0102** (0.003)	0.0096 (0.003)
5	-0.1	0	1	−0.2	0	1	**0.1007** (0.007)	0.1052 (0.008)	**0.0502** (0.005)	0.0534 (0.006)	0.0103 (0.003)	**0.0100** (0.003)
6	-0.3	2	1	−0.15	0	1	**0.1010** (0.008)	0.1024 (0.008)	**0.0506** (0.006)	0.0514 (0.006)	0.0104 (0.002)	**0.0098** (0.003)
7	0	10	1	0.1	0	1	**0.0997** (0.007)	0.1029 (0.008)	**0.0505** (0.006)	0.0519 (0.006)	0.0104 (0.002)	0.0096 (0.003)
8	0	30	1	0.05	0	1	**0.0996** (0.008)	0.1033 (0.008)	**0.0497** (0.005)	0.0520 (0.006)	0.0104 (0.002)	**0.0098** (0.003)
9	0.1	30	1	0.15	0	1	**0.1002** (0.007)	0.1030 (0.008)	**0.0507** (0.005)	0.0520 (0.006)	0.0106 (0.003)	**0.0099** (0.003)
10	0.05	20	1	0.08	0	1	**0.0995** (0.007)	0.1025 (0.008)	**0.0497** (0.006)	0.0517 (0.006)	0.0104 (0.003)	**0.0098** (0.003)
11	0	5	1	0	0	1	**0.0996** (0.007)	0.1031 (0.008)	**0.0502** (0.005)	0.0518 (0.006)	0.0104 (0.003)	**0.0098** (0.003)
12	0	2	1	0.2	0	1	**0.1021** (0.007)	0.1077 (0.009)	**0.0506** (0.005)	0.0548 (0.007)	0.0107 (0.003)	**0.0095** (0.003)

Notes: $\{Z_i\}_{i=1}^n$ is generated by taking the paired maxima of simulated sequences $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ from two different GEV distributions. The standard deviations of the 100 repetitions are shown in the parentheses. The numbers in bold font are the ones that are closer to the theoretical values, i.e., 0.1, 0.05, and 0.01.

prices of 330 S&P500 companies. Based on the closing prices, we calculate the daily negative log returns using the formula $r_i = -\log(\frac{p_i}{p_{i-1}})$. Here p_i represents the stock's closing price of one company on day i. For each day i, we obtain the 330 negative log returns and calculate the maximal value of them, denoting it as m_i. The time range is from 3 January 2000 to 30 December 2016, which contain 4277 trading days in the data. The histogram showing the distribution of $\{m_i\}_{i=1}^{4277}$ is in Figure 10.

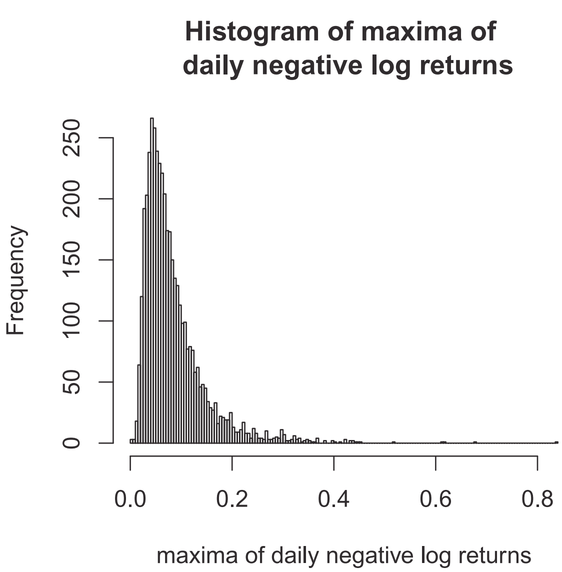

Figure 10. The histogram of the daily maxima of negative log returns of 330 stocks in the S&P500 companies list.

Table 2. The proportions of the simulated samples generated from the fitted distributions that exceed the 90th, 95th, and 99th sample percentiles of the maximal daily negative log returns.

	90th AMSD	90th GEV	95th AMSD	95th GEV	99th AMSD	99th GEV
Proportion	**0.0961**	0.0954	0.0475	**0.0521**	**0.0115**	0.0140

Note: The numbers in bold font are the ones that are closer to the theoretical values, i.e., 0.1, 0.05, and 0.01, respectively.

We find the 90th, 95th and 99th sample percentiles of $\{m_i\}_{i=1}^{4277}$, which are 0.1545, 0.2 and 0.3229, respectively. Here the daily maximal negative log returns have some time dependency. However, for the purpose of demonstration, we treat them as independent and fit the AMSD/AEVD and the GEV distribution to $\{m_i\}_{i=1}^{4277}$. Based on the fitted distributions, we generate random samples with the same size and find the proportions of the samples that exceed the three percentiles. The proportions are shown in Table 2.

Table 2 clearly reveals that the AMSD/AEVD performs better than the GEV alone. The modelling performance may be further improved if time series dependence is implemented in the model fitting, e.g., the AcF model proposed by Zhao et al. (2018) and Mao and Zhang (2018). We will leave this task as a future project.

6. Conclusions

This paper extends the classical extreme value theory to maxima of maxima of time series with mixture patterns depending on the sample size. It has been shown that the classical extreme value distributions are special cases of the accelerated max-stable (extreme value) distributions (AMSDs/AEVDs). Some basic probabilistic properties are presented in the paper. These properties can be used as the probability foundation of recently proposed statistical models for extreme observations. The AMSDs may shed the light of extreme value studies and inferences. Many of existing theories in classical extreme value literature can be renovated in a much more general setting. Many real applications, e.g., risk analysis and portfolio management, systemic risk, etc. can be reanalysed and better results can be expected. Under the newly introduced framework, many new statistical models can be introduced and explored.

Acknowledgments

The authors thank Editor Jun Shao and two referees for their valuable comments. The work by Cao was partially supported by NSF-DMS-1505367 and Wisconsin Alumni Research Foundation #MSN215758. The work by Zhang was partially supported by NSF-DMS-1505367 and NSF-DMS-2012298.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work by Cao was partially supported by NSF-DMS-1505367 and Wisconsin Alumni Research Foundation #MS N215758. The work by Zhang was partially supported by National Science Foundation NSF-DMS-1505367 and NSF-DMS-2012298.

Notes on contributors

Wenzhi Cao is a PhD student from the statistics department at the University of Wisconsin-Madison. Cao received his bachelor degree in Mathematics from Nankai University. Cao's research areas include extreme value theory and machine learning.

Zhengjun Zhang is Professor of Statistics at the University of Wisconsin. Zhang's main research areas of expertise are in financial time series and rare event modelling, virtual standard cryptocurrency, risk management, nonlinear dependence, asymmetric dependence, asymmetric and directed causal inference, gene-gene relationship in rare diseases.

References

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes: Theory and applications*. Wiley Series in Probability and Statistics. Wiley.

Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, *20*(4), 839–872. https://doi.org/10.1007/s10687-017-0292-6

Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J. M. (2005). *Extreme value and related models with applications in engineering and science*. Wiley Series in Probability and Statistics. Wiley.

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, *83*(3), 735–776. https://doi.org/10.1111/jori.v83.3

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Daouia, A., Girard, S., & Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(2), 263–292. https://doi.org/10.1111/rssb.12254

de Haan, L. (1993). Extreme value statistics. In J. Galambos, J. Lechner, & E. Simiu (Eds.), *Extreme value theory and applications* (pp. 93–122). Kluwer Academic Publisher.

de Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer Verlag.

Dey, D. K., & Yan, J. (2016). *Extreme value modeling and risk analysis: Methods and applications*. Chapman & Hall/CRC.

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, *3*(2), 30–41. https://doi.org/10.1080/10920277.1999.10595797

Finkenstädt, B., & Rootzén, H. E. (2004). *Extreme values in finance, telecommunications, and the environment*. Chapman & Hall/CRC.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. Technical Report.

Heffernan, J. E., Tawn, J. A., & Zhang, Z. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes. *Extremes*, *10*(1–2), 57–82. https://doi.org/10.1007/s10687-007-0035-1

Idowu, T., & Zhang, Z. (2017). An extended sparse max-linear moving model with application to high-frequency financial data. *Statistical Theory and Related Fields*, *1*(1), 92–111. https://doi.org/10.1080/24754269.2017.1346852

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Loynes, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *The Annals of Mathematical Statistics*, *36*(3), 993–999. https://doi.org/10.1214/aoms/1177700071

Malinowski, A., Schlather, M., & Zhang, Z. (2015). Marked point process adjusted tail dependence analysis for high-frequency financial data. *Statistics and Its Interface*, *8*(1), 109–122. https://doi.org/10.4310/SII.2015.v8.n1.a10

Mao, G., & Zhang, Z. (2018). Stochastic tail index model for high frequency financial data with Bayesian analysis. *Journal of Econometrics*, *205*(2), 470–487. https://doi.org/10.1016/j.jeconom.2018.03.019

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*(3–4), 271–300. https://doi.org/10.1016/S0927-5398(00)00012-8

Mikosch, T., Embrechts, P., & Klüppelberg, C. (1997). *Modelling extremal events for insurance and finance*. Springer Verlag.

Naveau, P., Zhang, Z., & Zhu, B. (2011). An extension of max autoregressive models. *Statistics and Its Interface*, *4*(2), 253–266. https://doi.org/10.4310/SII.2011.v4.n2.a19

Resnick, S. I. (2013). *Extreme values, regular variation and point processes*. Springer.

Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, *28*(1), 82–108. https://doi.org/10.1111/joes.2014.28.issue-1

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences USA*, *42*(1), 43–47. https://doi.org/10.1073/pnas.42.1.43

Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). *Extremes in nature: An approach using copulas*. Springer(Complexity).

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90. https://doi.org/10.1093/biomet/72.1.67

Tang, R., Shao, J., & Zhang, Z. (2013). Sparse moving maxima models for tail dependence in multivariate financial time series. *Journal of Statistical Planning and Inference*, *143*(5), 882–895. https://doi.org/10.1016/j.jspi.2012.11.008

Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.

Watson, G. S. (1954). Extreme values in samples from *m*-dependent stationary stochastic processes. *The Annals of Mathematical Statistics*, *25*(4), 798–800. https://doi.org/10.1214/aoms/1177728670

Zhang, Z., & Smith, R. L. (2010). On the estimation and application of max-stable processes. *Journal of Statistical Planning and Inference*, *140*(5), 1135–1153. https://doi.org/10.1016/j.jspi.2009.10.014

Zhang, Z., & Zhu, B. (2016). Copula structured m4 processes with application to high-frequency financial data. *Journal of Econometrics*, *194*(2), 231–241. https://doi.org/10.1016/j.jeconom.2016.05.004

Zhao, Z., & Zhang, Z. (2018). Semi-parametric dynamic max-copula model for multivariate time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(2), 409–432. https://doi.org/10.1111/rssb.12256

Zhao, Z., Zhang, Z., & Chen, R. (2018). Modeling maxima with autoregressive conditional Fréchet model. *Journal of Econometrics*, *207*(2), 325–351. https://doi.org/10.1016/j.jeconom.2018.07.004

Appendix

A.1 Proofs of Theorems and Propositions

A.1.1 Proof of Theorem 2.1

For Equation (4),

$$\begin{aligned}
&P(a_{2,n_2}(M_n - b_{2,n_2}) \le x)\\
&\quad = P(\max(a_{2,n_2}(M_{1,n_1} - b_{2,n_2}),\\
&\quad a_{2,n_2}(M_{2,n_2} - b_{2,n_2})) \le x)\\
&\quad = P(a_{2,n_2}(M_{1,n_1} - b_{2,n_2})\\
&\quad \le x, a_{2,n_2}(M_{2,n_2} - b_{2,n_2})) \le x)\\
&\quad = P\left(a_{1,n_1}(M_{1,n_1} - b_{1,n_1})\right.\\
&\quad \le a_{1,n_1}\left(\frac{x}{a_{2,n_2}} + b_{2,n_2} - b_{1,n_1}\right),\\
&\quad \left. a_{2,n_2}(M_{2,n_2} - b_{2,n_2})) \le x\right)\\
&\quad \to H_1(ax+b)H_2(x).
\end{aligned}$$

For Equation (5),

$$\begin{aligned}
&P(a_{2,n_2}(M_n - b_{2,n_2}) \le x)\\
&\quad = P\left(a_{1,n_1}(M_{1,n_1} - b_{1,n_1})\right.\\
&\quad \le a_{1,n_1}\left(\frac{x}{a_{2,n_2}} + b_{2,n_2} - b_{1,n_1}\right),\\
&\quad \left. a_{2,n_2}(M_{2,n_2} - b_{2,n_2}) \le x\right)\\
&\quad \to H_2(x).
\end{aligned}$$

A.1.2 Proof of Fact 2.2

The density of $\Phi_{\alpha_1}\Phi_{\alpha_2}$ is

$$f(x) = \begin{cases} e^{-x^{-\alpha_1}-x^{-\alpha_2}}(\alpha_1 x^{-\alpha_1-1} + \alpha_2 x^{-\alpha_2-1}) & x \ge 0,\\ 0 & x < 0. \end{cases} \tag{A1}$$

Thus

$$\begin{aligned}
\mu_k &= \int_0^\infty x^k f(x)\,\mathrm{d}(x)\\
&= \int_0^\infty x^k e^{-x^{-\alpha_1}-x^{-\alpha_2}}(\alpha_1 x^{-\alpha_1-1} + \alpha_2 x^{-\alpha_2-1})\,\mathrm{d}x. \qquad \text{(A2)}
\end{aligned}$$

Dividing the integral into two parts, we get

$$\mu_k = \int_0^1 x^k f(x)\,\mathrm{d}x + \int_1^\infty x^k f(x)\,\mathrm{d}x. \tag{A3}$$

First, let us consider $\int_0^1 x^k f(x)\,\mathrm{d}x$. Since $\lim_{x\to 0} x^k f(x) = 0$ and $x^k f(x)$ is continuous on $[0,1]$, it is bounded on $[0,1]$. This implies that

$$\int_0^1 x^k f(x)\,\mathrm{d}x < \infty. \tag{A4}$$

Next, let us consider $\int_1^\infty x^k f(x)\,\mathrm{d}x$. We have

$$\begin{aligned}
&\int_1^\infty x^k f(x)\,\mathrm{d}x\\
&\quad = \int_1^\infty e^{-x^{-\alpha_1}-x^{-\alpha_2}}(\alpha_1 x^{k-\alpha_1-1} + \alpha_2 x^{k-\alpha_2-1})\,\mathrm{d}x.
\end{aligned}$$

Notice that

$$\lim_{x\to\infty} e^{-x^{-\alpha_1}-x^{-\alpha_2}} = 1. \tag{A5}$$

Therefore $\int_1^\infty e^{-x^{-\alpha_1}-x^{-\alpha_2}}(\alpha_1 x^{k-\alpha_1-1} + \alpha_2 x^{k-\alpha_2-1})\,\mathrm{d}x < \infty$ only if $k < \alpha_1$ and $k < \alpha_2$, i.e., $k < \min(\alpha_1, \alpha_2)$.

A.1.3 Proof of Fact 2.3

We need to consider $\lim_{x\to\infty} \frac{1-e^{-x^{-\alpha_1}}e^{-x^{-\alpha_2}}}{1-e^{-x^{-\alpha_1}}}$.

Since $x^{-\alpha_1} \to 0$ and $x^{-\alpha_2} \to 0$ as $x \to \infty$, we have the Taylor expansions

$$\begin{aligned}
e^{-x^{-\alpha_1}} &= 1 - x^{-\alpha_1} + o(x^{-\alpha_1}),\\
e^{-x^{-\alpha_1}-x^{-\alpha_2}} &= 1 - (x^{-\alpha_1} + x^{-\alpha_2}) + o(x^{-\alpha_1} + x^{-\alpha_2}).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\lim_{x\to\infty} \frac{1-e^{-x^{-\alpha_1}}e^{-x^{-\alpha_2}}}{1-e^{-x^{-\alpha_1}}}\\
&\quad = \lim_{x\to\infty} \frac{(x^{-\alpha_1} + x^{-\alpha_2}) + o(x^{-\alpha_1} + x^{-\alpha_2})}{x^{-\alpha_1} + o(x^{-\alpha_1})}\\
&\quad = \lim_{x\to\infty} (1 + x^{\alpha_1-\alpha_2}) = 1. \qquad \text{(A6)}
\end{aligned}$$

This proves that $\Phi_{\alpha_1}\Phi_{\alpha_2}$ and Φ_{α_1} are tail-equivalent.

A.1.4 Proof of Fact 2.4

The density of $\Lambda(x)\Phi_\alpha(x)$ is

$$f(x) = \begin{cases} e^{-e^{-x}-x^{-\alpha}}(e^{-x} + \alpha x^{-\alpha-1}) & x \ge 0,\\ 0 & x < 0. \end{cases}$$

Thus

$$\mu_k = \int_0^\infty x^k f(x)\,\mathrm{d}(x)$$
$$= \int_0^\infty x^k e^{-e^{-x}-x^{-\alpha}}(e^{-x}+\alpha x^{-\alpha-1})\,\mathrm{d}x. \quad (A7)$$

Dividing the above equation into two parts, we get

$$\mu_k = \int_0^1 x^k f(x)\,\mathrm{d}x + \int_1^\infty x^k f(x)\,\mathrm{d}x. \quad (A8)$$

For the first part, since $\lim_{x\to 0} x^k f(x) = 0$ and $x^k f(x)$ is continuous on $[0,1]$, it is bounded on $[0,1]$. Thus $\int_0^1 x^k f(x) < \infty$.

For the second part,

$$\int_1^\infty x^k f(x) = \int_1^\infty e^{-e^{-x}-x^{-\alpha}}(e^{-x}x^k + \alpha x^{k-\alpha-1})\,\mathrm{d}x. \quad (A9)$$

Since

$$\lim_{x\to\infty} e^{-e^{-x}-x^{-\alpha}} = 1, \quad \text{and} \quad \int_1^\infty e^{-x}x^k\,\mathrm{d}x < \infty \quad \text{for} \forall k, \quad (A10)$$

we have $\int_1^\infty e^{-e^{-x}-x^{-\alpha}}(e^{-x}x^k+\alpha x^{k-\alpha-1})\,\mathrm{d}x < \infty$ if and only if $k < \alpha$.

A.1.5 Proof of Fact 2.5

We need to consider $\lim_{x\to\infty} \frac{1-e^{-e^{-x}}e^{-x^{-\alpha}}}{1-e^{-x^{-\alpha}}}$.

Since $\lim_{x\to\infty} e^{-x} \to 0$ and $\lim_{x\to\infty} x^{-\alpha} \to 0$, we have the Taylor expansions

$$e^{-e^{-x}-x^{-\alpha}} = 1 - e^{-x} - x^{-\alpha} + o(e^{-x}+x^{-\alpha}),$$
$$e^{-x^{-\alpha}} = 1 - x^{-\alpha} + o(x^{-\alpha}).$$

Thus

$$\lim_{x\to\infty} \frac{1-e^{-e^{-x}}e^{-x^{-\alpha}}}{1-e^{-x^{-\alpha}}} = \lim_{x\to\infty} \frac{e^{-x}+x^{-\alpha}+o(e^{-x}+x^{-\alpha})}{x^{-\alpha}+o(x^{-\alpha})}$$
$$= 1. \quad (A11)$$

This implies that $\Lambda(x)\Phi_\alpha(x)$ and $\Phi_\alpha(x)$ are tail-equivalent.

A.1.6 Proof of Theorem 3.1

If (25) holds, we must have

$$1 - F_1(u_{1,n_1}) \to 0,$$
$$1 - F_2(u_{2,n_2}) \to 0.$$

Then

$$n_1 \log(1-(1-F_1(u_{1,n_1}))) + n_2 \log(1-(1-F_2(u_{2,n_2})))$$
$$= -n_1(1-F_1(u_{1,n_1}))(1+o(1))$$
$$- n_2(1-F_2(u_{2,n_2}))(1+o(1))$$
$$\to -\tau \quad (A12)$$

which is equivalent to

$$P(M_{1,n_1} \le u_{1,n_1}, M_{2,n_2} \le u_{2,n_2})$$
$$= (1-(1-F_1(u_{1,n_1})))^{n_1}(1-(1-F_2(u_{2,n_2})))^{n_2}$$
$$= \exp\{n_1 \log(1-(1-F_1(u_{1,n_1})))$$
$$+ n_2 \log(1-(1-F_2(u_{2,n_2})))\}$$
$$\to e^{-\tau}.$$

Conversely, if (26) holds, which is equivalent to

$$n_1 \log(1-(1-F_1(u_{1,n_1})))$$
$$+ n_2 \log(1-(1-F_2(u_{2,n_2}))) \to -\tau, \quad (A13)$$

we must have $1-F_1(u_{1,n_1}) \to 0$ and $1-F_2(u_{2,n_2}) \to 0$. Otherwise, suppose $1-F_1(u_{1,n_1}) \not\to 0$, then there is a sequence of indexes $m_1, m_2, \ldots$ and $\epsilon > 0$ such that $1 - F_1(u_{1,m_k}) > \epsilon$ for $\forall k$. This means that

$$n_1 \log(1-(1-F_1(u_{1,m_i}))) + n_2 \log(1-(1-F_2(u_{2,n-m_i})))$$
$$< n_1 \log(1-(1-F_1(u_{1,m_i})))$$
$$< n_1 \log(1-\epsilon) \to -\infty,$$

which is contradictory to (A13). We have

$$n_1[(1-F_1(u_{1,n_1})) + o(1-F_1(u_{1,n_1}))]$$
$$+ n_2[(1-F_2(u_{2,n_2})) + o(1-F_2(u_{2,n_2}))] \to \tau \quad (A14)$$

and Equation (25) holds.

A.1.7 Proof of Corollary 3.1

Since

$$n_1(1-F_1(u_{1,n_1})) + n_2(1-F_2(u_{2,n_2})) \to \tau_1 + \tau_2, \quad (A15)$$

(30) is a direct result of Theorem 3.1.

If $\frac{n_2(1-F_2(u_{1,n_1}))}{n_1(1-F_1(u_{1,n_1}))} \to t$, then $n_2(1-F_2(u_{1,n_1})) \to t\tau_1$, where $0 \le t\tau_1 < \infty$. Therefore,

$$P(M_n \le u_{1,n_1}) = P(M_{1,n_1} \le u_{1,n_1})P(M_{2,n_2} \le u_{1,n_1})$$
$$\to e^{-\tau_1(1+t)}.$$

A.1.8 Proof of Theorem 3.2

Since $P(M_{i,n_i} \le u_{i,n_i}) = (1-\tau_{i,n_i}/n_i)^{n_i}$ and $0 \le \tau_{i,n_i} = n_1(1-F_i(u_{i,n_i})) \le n$, the result follows from Lemma A.2.

A.1.9 Proof of Theorem 4.1

For fixed k, write $n' = [n/k]$, suppose that there are n_1' members from F_1 and n_2' members from F_2 among $\{X_1, \ldots, X_{n'}\}$, $n' = n_1' + n_2'$. If (47) holds, by assumption we have $n_1' \sim pn' \sim \frac{n_1}{k}$ and $n_2' \sim (1-p)n' \sim \frac{n_2}{k}$, thus

$$n_1'(1-F_1(u_{1,n_1})) + n_2'(1-F_2(u_{2,n_2})) \to \frac{\tau}{k}. \quad (A16)$$

Since

$$P(\{M_{1,n_1'} \le u_{1,n_1}, M_{2,n_2'} \le u_{2,n_2}\})$$
$$= 1 - P(\{M_{1,n_1'} > u_{1,n_1}\} \cup \{M_{2,n_2'} > u_{2,n_2}\})$$
$$= 1 - P\left(\left(\bigcup_{i=1}^{n_1'}\{X_{1,i} > u_{1,n_1}\}\right) \cup \left(\bigcup_{j=1}^{n_2'}\{X_{2,j} > u_{2,n_2}\}\right)\right),$$

we have

$$1 - n_1'(1-F_1(u_{1,n_1})) - n_2'(1-F_2(u_{2,n_2}))$$
$$\le P(M_{1,n_1'} \le u_{1,n_1}, M_{2,n_2'} \le u_{2,n_2})$$
$$\le 1 - n_1'(1-F_1(u_{1,n_1})) - n_2'(1-F_2(u_{2,n_2})) + S_n, \quad (A17)$$

where $S_n = \sum_{1\le i<j\le n'} P(X_i > u_n^{(i)}, X_j > u_n^{(j)})$.

Condition $D'(\mathbf{u}_n)$ implies that $\limsup_{n\to\infty} S_n = o(\frac{1}{k})$ as $k \to \infty$. By (A16) and (A17), we have

$$1 - \frac{\tau}{k} \le \liminf_{n\to\infty} P(M_{1,n_1'} \le u_{1,n_1}, M_{2,n_2'} \le u_{2,n_2})$$
$$\le \limsup_{n\to\infty} P(M_{1,n_1'} \le u_{1,n_1}, M_{2,n_2'} \le u_{2,n_2})$$
$$\le 1 - \frac{\tau}{k} + o\left(\frac{1}{k}\right).$$

Since $D(\mathbf{u}_n)$ implies $D(u_{1,n_1})$ and $D(u_{2,n_2})$, Lemma A.3 holds for each subsequence. We have

$$\begin{aligned}\left(1-\frac{\tau}{k}\right)^k &\le \liminf_{n\to\infty} P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\\ &\le \limsup_{n\to\infty} P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\\ &\le \left(1-\frac{\tau}{k}+o\left(\frac{1}{k}\right)\right)^k.\end{aligned}$$

Letting $k\to\infty$, we have $\lim_{n\to\infty} P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\to e^{-\tau}$.

Conversely, if (46) holds,

$$\begin{aligned}1 &- P(M_{1,n_1'}\le u_{1,n_1}, M_{2,n_2'}\le u_{2,n_2})\\ &\le n_1'(1-F_1(u_{1,n_1}))+n_2'(1-F_2(u_{2,n_2}))\\ &\le 1-P(M_{1,n_1'}\le u_{1,n_1}, M_{2,n_2'}\le u_{2,n_2})+S_n. \qquad (A18)\end{aligned}$$

Since $P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\to e^{-\tau}$, we have $P(M_{1,n_1'}\le u_{1,n_1}, M_{2,n_2'}\le u_{2,n_2})\to e^{-\tau/k}$. By letting $n\to\infty$ in (A18),

$$\begin{aligned}1 &- e^{-\tau/k}\\ &\le \frac{1}{k}\liminf_{n\to\infty} n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\\ &\le \frac{1}{k}\limsup_{n\to\infty} n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\\ &\le 1-e^{-\tau/k}+o\left(\frac{1}{k}\right)\end{aligned}$$

from which (multiplying k on all sides and let $k\to\infty$) we have $n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\to\tau$.

A.1.10 Proof of Corollary 4.1

Suppose $n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\to\infty$, by $u_{1,n_1}<v_{1,n_1}$ and $u_{2,n_2}<v_{2,n_2}$, we have

$$\begin{aligned}P(M_{1,n_1}&\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\\ &\le P(M_{1,n_1}\le v_{1,n_1}, M_{2,n_2}\le v_{2,n_2}).\end{aligned}$$

By Theorem 4.1, $P(M_{1,n_1}\le v_{1,n_1}, M_{2,n_2}\le v_{2,n_2})\to e^{-\tau}$. Then

$$\limsup_{n\to\infty} P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})\le e^{-\tau}.$$

By letting $\tau\to\infty$, we have

$$\lim_{n\to\infty} P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})=0.$$

Conversely, we still have

$$\begin{aligned}n_1&(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\\ &\ge n_1(1-F_1(v_{1,n_1})+n_2(1-F_2(v_{2,n_2}))\to\tau.\end{aligned}$$

Since the above inequality holds for arbitrary large $\tau>0$, we must have $n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\to\infty$.

A.1.11 Proof of Theorem 4.2

For $\theta>0$, the condition $P(\widehat{M}_{1,n_1}\le u_{1,n_1}, \widehat{M}_{2,n_2}\le u_{2,n_2})\to\theta$ may be rewritten as $P(\widehat{M}_{1,n_1}\le u_{1,n_1}, \widehat{M}_{2,n_2}\le u_{2,n_2})\to e^{-\tau}$ with $\tau=-\log\theta$, this holds if and only if $n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\to\tau$. The same is true for $P(M_{1,n_1}\le u_{1,n_1}, M_{2,n_2}\le u_{2,n_2})$ by condition $D(\mathbf{u}_n)$ and $D'(\mathbf{u}_n)$. When $\theta=0$, the result follows from Corollary 4.1.

A.1.12 Proof of Theorem 4.3

If $G(x)>0$, the equivalence follows from Theorem 4.2, with $\theta=G(x)$.

If $G(x)=0$, the continuity of G shows that, if $0<\tau<\infty$,there exists x_0 such that $G(x_0)=e^{-\tau}$. $D(v_{i,n_i})$ and $D'(v_{i,n_i})$ hold for $v_{1,n_1}=x_0/a_{1,n_1}+b_{1,n_1}, v_{2,n_2}=x_0/a_{2,n_2}+b_{2,n_2}$ and $P(M_{1,n_1}\le v_{1,n_1}, M_{2,n_2}\le v_{2,n_2})\to e^{-\tau}$ or $P(\widehat{M}_{1,n_1}\le v_{1,n_1}, \widehat{M}_{2,n_2}\le v_{2,n_2})\to e^{-\tau}$ depending on the assumption made, so that $n_1(1-F_1(v_{1,n_1}))+n_2(1-F_2(v_{2,n_2}))\to\tau$. If (49) holds, then we have $n_1(1-F_1(u_{1,n_1}))+n_2(1-F_2(u_{2,n_2}))\to\infty$, thus $u_{1,n_1}<v_{1,n_1}$ and $u_{2,n_2}<v_{2,n_2}$ (since one of the inequalities must hold and also implies another). By Theorem 4.2, (48) holds. The converse direction can be proved similarly.

A.2 Lemmas

Lemma A.1 (Khintchine, Theorem 1.2.3 in Leadbetter et al. (2012)): *Let $\{F_n\}$ be a sequence of cdf's and H a nondegenerate cdf Let $a_n>0$ and b_n be constants such that*

$$F_n(a_nx+b_n)\xrightarrow{w} H(x). \qquad (A19)$$

Then for some nondegenerate cdf H_* and constants $\alpha_n>0$, β_n,

$$F_n(\alpha_nx+\beta_n)\xrightarrow{w} H_*(x) \qquad (A20)$$

if and only if

$$a_n^{-1}\alpha_n\to a \text{and} a_n^{-1}(\beta_n-b_n)\to b \qquad (A21)$$

for some $a>0$ and b, and then

$$H_*(x)=H(ax+b). \qquad (A22)$$

Lemma A.2 (Lemma 2.4.1 in Leadbetter et al. (2012)): *(1) If $0\le x\le n$ then*

$$\begin{aligned}0\le e^{-x}-\left(1-\frac{x}{n}\right)^n &\le \frac{x^2e^{-x}}{2}\cdot\frac{1}{n-1}\\ &\le 2e^{-2}\cdot\frac{1}{n-1}\\ &\le 0.3\cdot\frac{1}{n-1} \quad \text{for } n=1,2,\ldots, \qquad (A23)\end{aligned}$$

and further

$$\begin{aligned}e^{-x}&-\left(1-\frac{x}{n}\right)^n\\ &=\frac{x^2e^{-x}}{2}\frac{1}{n}\left(1+O\left(\frac{1}{n}\right)\right) \quad \text{as } n\to\infty, \qquad (A24)\end{aligned}$$

uniformly for x in bounded intervals.

(1) *If $x-y\le\log 2$ then*

$$e^{-y}-e^{-x}=e^{-x}\{(x-y)+\theta(x-y)^2\}, \qquad (A25)$$

with $0<\theta<1$.

Lemma A.3 (Lemma 3.3.2 in Leadbetter et al. (2012)): *If $D(u_n)$ holds, for a fixed integer k, we have*

$$P(M_n\le u_n)-P^k(M_{[n/k]}\le u_n)\to 0 \quad \text{as } n\to\infty. \qquad (A26)$$