Chapter 12

Extreme Value Theory

This chapter is dedicated to tail (or extreme) risk modeling. Tail risk recovers two notions. The first one is related to rare events, meaning that a severe loss may occur with a very small probability. The second one concerns the magnitude of a loss that is difficult to reconciliate with the observed volatility of the portfolio. Of course, the two notions are connected, but the second is more frequent. For instance, stock market crashes are numerous since the end of the eighties. The study of these rare or abnormal events needs an appropriate framework to analyze their risk. This is the subject of this chapter. In a first section, we consider order statistics, which are very useful to understand the underlying concept of tail risk. Then, we present the extreme value theory (EVT) in the unidimensional case. Finally, the last section deals with the correlation issue between extreme risks.

12.1 Order statistics

12.1.1 Main properties

Let $X_1, \ldots, X_n$ be *iid* random variables, whose probability distribution is denoted by $\mathbf{F}$. We rank these random variables by increasing order:

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n-1:n} \leq X_{n:n}$$

$X_{i:n}$ is called the i^{th} order statistic in the sample of size n. We note $x_{i:n}$ the corresponding random variate or the value taken by $X_{i:n}$. We have:

$$\begin{aligned}
\mathbf{F}_{i:n}(x) &= \Pr\{X_{i:n} \leq x\} \\
&= \Pr\{\text{at least } i \text{ variables among } X_1, \ldots, X_n \text{ are less or equal to } x\} \\
&= \sum_{k=i}^{n} \Pr\{k \text{ variables among } X_1, \ldots, X_n \text{ are less or equal to } x\} \\
&= \sum_{k=i}^{n} \binom{n}{k} \mathbf{F}(x)^k (1 - \mathbf{F}(x))^{n-k} \qquad (12.1)
\end{aligned}$$

We note f the density function of $\mathbf{F}$. We deduce that the density function of $X_{i:n}$ has the following expression:

$$\begin{aligned}
f_{i:n}(x) \quad = \quad & \sum_{k=i}^{n} \binom{n}{k} k\mathbf{F}(x)^{k-1} (1 - \mathbf{F}(x))^{n-k} f(x) - \\
& \sum_{k=i}^{n-1} \binom{n}{k} \mathbf{F}(x)^k (n-k) (1 - \mathbf{F}(x))^{n-k-1} f(x) \qquad (12.2)
\end{aligned}$$

It follows that:

$$\begin{aligned}
f_{i:n}(x) &= \sum_{k=i}^{n} \frac{n!}{(k-1)!(n-k)!} \mathbf{F}(x)^{k-1}(1-\mathbf{F}(x))^{n-k} f(x) - \\
&\quad \sum_{k=i}^{n-1} \frac{n!}{k!(n-k-1)!} \mathbf{F}(x)^{k}(1-\mathbf{F}(x))^{n-k-1} f(x) \\
&= \sum_{k=i}^{n} \frac{n!}{(k-1)!(n-k)!} \mathbf{F}(x)^{k-1}(1-\mathbf{F}(x))^{n-k} f(x) - \\
&\quad \sum_{k=i+1}^{n} \frac{n!}{(k-1)!(n-k)!} \mathbf{F}(x)^{k-1}(1-\mathbf{F}(x))^{n-k} f(x) \\
&= \frac{n!}{(i-1)!(n-i)!} \mathbf{F}(x)^{i-1}(1-\mathbf{F}(x))^{n-i} f(x) \qquad (12.3)
\end{aligned}$$

Remark 142 *When k is equal to n, the derivative of $(1-\mathbf{F}(x))^{n-k}$ is equal to zero. This explains that the second summation in Equation (12.2) does not include the case $k=n$.*

Example 126 *If $X_1, \ldots, X_n$ follow a uniform distribution $\mathcal{U}_{[0,1]}$, we obtain:*

$$\begin{aligned}
\mathbf{F}_{i:n}(x) &= \sum_{k=i}^{n} \binom{n}{k} x^k (1-x)^{n-k} \\
&= \mathcal{IB}(x; i, n-i+1)
\end{aligned}$$

where $\mathcal{IB}(x;\alpha,\beta)$ is the regularized incomplete beta function[1]*:*

$$\mathcal{IB}(x;\alpha,\beta) = \frac{1}{\mathfrak{B}(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$$

We deduce that $X_{i:n} \sim \mathcal{B}(i, n-i+1)$. It follows that the expected value of the order statistic $X_{i:n}$ is equal to:

$$\begin{aligned}
\mathbb{E}[X_{i:n}] &= \mathbb{E}[\mathcal{B}(i, n-i+1)] \\
&= \frac{i}{n+1}
\end{aligned}$$

We verify the stochastic ordering:

$$j > i \Rightarrow \mathbf{F}_{i:n} \succ \mathbf{F}_{j:n}$$

Indeed, we have:

$$\begin{aligned}
\mathbf{F}_{i:n}(x) &= \sum_{k=i}^{n} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} \\
&= \sum_{k=i}^{j-1} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} + \sum_{k=j}^{n} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} \\
&= \mathbf{F}_{j:n}(x) + \sum_{k=i}^{j-1} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k}
\end{aligned}$$

meaning that $\mathbf{F}_{i:n}(x) \geq \mathbf{F}_{j:n}(x)$. In Figure 12.1, we illustrate this property when the random variables $X_1, \ldots, X_n$ follow the normal distribution $\mathcal{N}(0,1)$. We verify that $\mathbf{F}_{i:n}(x)$ increases with the ordering value i.

[1] It is also the Beta probability distribution $\mathcal{IB}(x;\alpha,\beta) = \Pr\{\mathcal{B}(\alpha,\beta) \leq x\}$.

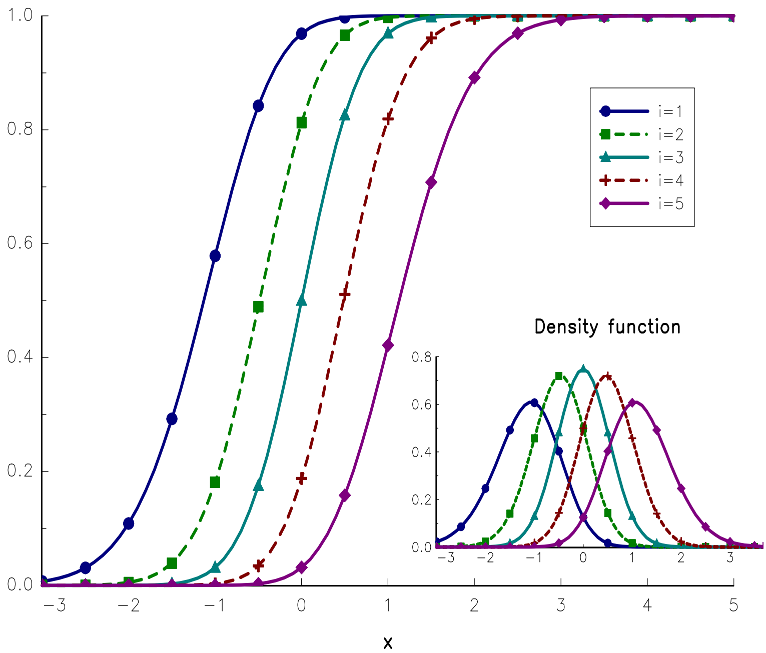

FIGURE 12.1: Distribution function $\mathbf{F}_{i:n}$ when the random variables $X_1, \ldots, X_n$ are Gaussian

12.1.2 Extreme order statistics

Two order statistics are particularly interesting for the study of rare events. They are the lowest and highest order statistics:

$$X_{1:n} = \min(X_1, \ldots, X_n)$$

and:

$$X_{n:n} = \max(X_1, \ldots, X_n)$$

We can find their probability distributions by setting $i = 1$ and $i = n$ in Formula (12.1). We can also retrieve their expression by noting that:

$$\begin{aligned}
\mathbf{F}_{1:n}(x) = \Pr\{\min(X_1, \ldots, X_n) \leq x\} &= 1 - \Pr\{\min(X_1, \ldots, X_n) \geq x\} \\
&= 1 - \Pr\{X_1 \geq x, X_2 \geq x, \ldots, X_n \geq x\} \\
&= 1 - \prod_{i=1}^{n} \Pr\{X_i \geq x\} \\
&= 1 - \prod_{i=1}^{n} (1 - \Pr\{X_i \leq x\}) \\
&= 1 - (1 - \mathbf{F}(x))^n
\end{aligned}$$

and:

$$\begin{aligned}
\mathbf{F}_{n:n}(x) = \Pr\{\max(X_1, \ldots, X_n) \leq x\} &= \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} \\
&= \prod_{i=1}^{n} \Pr\{X_i \leq x\} \\
&= \mathbf{F}(x)^n
\end{aligned}$$

We deduce that the density functions are equal to:

$$f_{1:n}(x) = n(1 - \mathbf{F}(x))^{n-1} f(x)$$

and

$$f_{n:n}(x) = n\mathbf{F}(x)^{n-1} f(x)$$

Let us consider an example with the Gaussian distribution $\mathcal{N}(0,1)$. Figure 12.2 shows the evolution of the density function $f_{n:n}$ with respect to the sample size n. We verify the stochastic ordering: $n > m \Rightarrow \mathbf{F}_{n:n} \succ \mathbf{F}_{m:m}$.

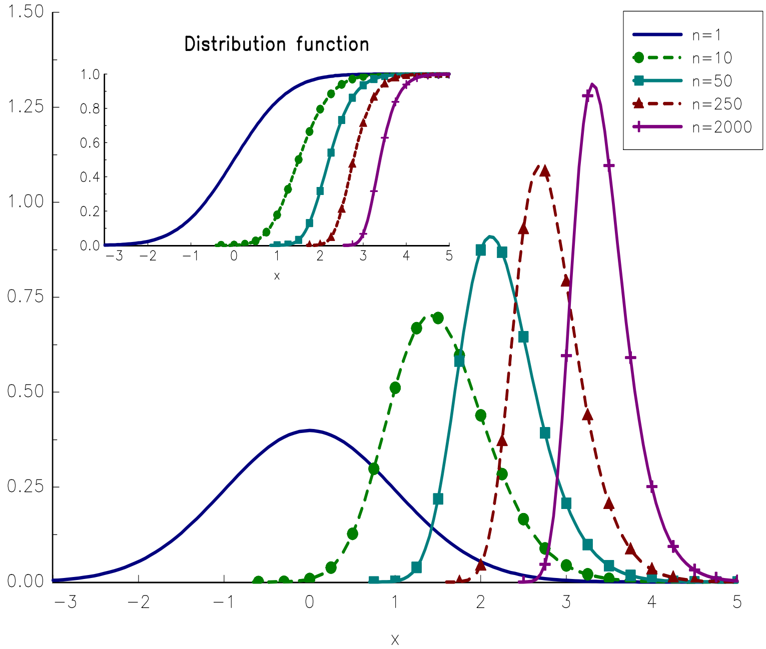

FIGURE 12.2: Density function $f_{n:n}$ of the Gaussian random variable $\mathcal{N}(0,1)$

Let us now illustrate the impact of the probability distribution tails on order statistics. We consider the daily returns of the MSCI USA index from 1995 to 2015. We consider three hypotheses:

$\mathcal{H}_1$ Daily returns are Gaussian, meaning that:

$$R_t = \hat{\mu} + \hat{\sigma} X_t$$

where $X_t \sim \mathcal{N}(0,1)$, $\hat{\mu}$ is the empirical mean of daily returns and $\hat{\sigma}$ is the daily standard deviation.

$\mathcal{H}_2$ Daily returns follow a Student's t distribution[2]:

$$R_t = \hat{\mu} + \hat{\sigma}\sqrt{\frac{\nu - 2}{\nu}} X_t$$

where $X_t \sim \mathbf{t}_\nu$. We consider two alternative assumptions: $\mathcal{H}_{2a} : \nu = 3$ and $\mathcal{H}_{2b} : \nu = 6$.

[2]We add the factor $\sqrt{\frac{\nu - 2}{\nu}}$ in order to verify that $\operatorname{var}(R_t) = \hat{\sigma}^2$.

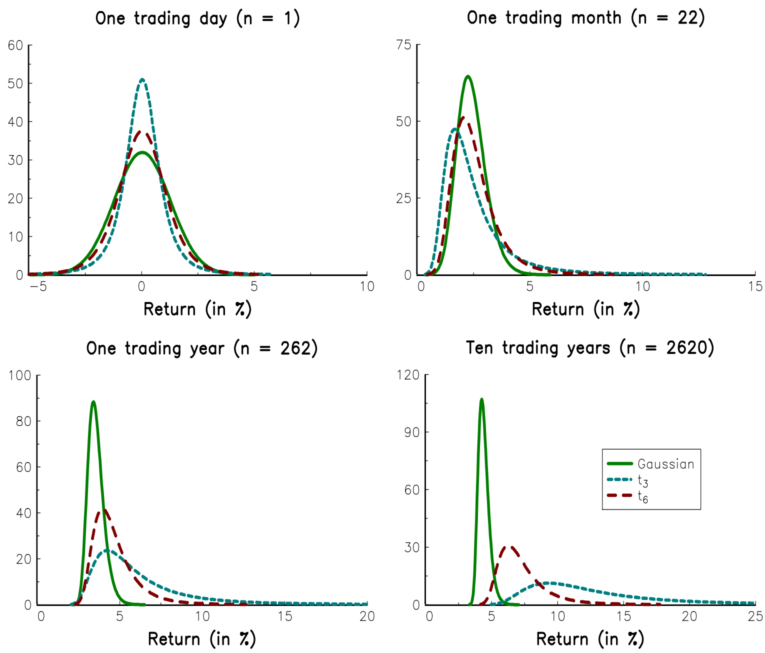

FIGURE 12.3: Density function of the maximum order statistic (daily return of the MSCI USA index, 1995-2015)

We represent the probability density function of $R_{n:n}$ for several values of n in Figure 12.3. When n is equal to one trading day, $R_{n:n}$ is exactly the daily return. We notice that it is difficult to observe the impact of the probability distribution tail. However, when n increases, the impact becomes more and more important. Order statistics allow amplifying local phenomena of probability distributions. In particular, extreme order statistics are a very useful tool to analyze left and right tails.

Remark 143 *The limit distributions of minima and maxima are given by the following results:*

$$\begin{aligned}\lim_{n\to\infty}\mathbf{F}_{1:n}(x) &= \lim_{n\to\infty} 1-(1-\mathbf{F}(x))^n \\ &= \begin{cases} 0 & \text{if } \mathbf{F}(x)=0 \\ 1 & \text{if } \mathbf{F}(x)>0 \end{cases}\end{aligned}$$

and:

$$\begin{aligned}\lim_{n\to\infty}\mathbf{F}_{n:n}(x) &= \lim_{n\to\infty}\mathbf{F}(x)^n \\ &= \begin{cases} 0 & \text{if } \mathbf{F}(x)<1 \\ 1 & \text{if } \mathbf{F}(x)=1 \end{cases}\end{aligned}$$

We deduce that the limit distributions are degenerate as they only take values of 0 *and* 1*. This property is very important, because it means that we cannot study extreme events by considering these limit distributions. This is why the extreme value theory is based on another convergence approach of extreme order statistics.*

12.1.3 Inference statistics

The common approach to estimate the parameters θ of the probability density function $f(x;\theta)$ is to maximize the log-likelihood function of a given sample $\{x_1,\ldots,x_T\}$:

$$\hat{\theta} = \arg\max \sum_{t=1}^{T} \ln f(x_t;\theta)$$

In a similar way, we can consider the sample[3] $\{x'_1,\ldots,x'_{n_S}\}$ of the order statistic $X_{i:n}$ and estimate the parameters θ by the method of maximum likelihood:

$$\hat{\theta}_{i:n} = \arg\max \boldsymbol{\ell}_{i:n}(\theta)$$

where:

$$\begin{aligned}
\boldsymbol{\ell}_{i:n}(\theta) &= \sum_{s=1}^{n_S} \ln f_{i:n}(x'_s;\theta) \\
&= \sum_{s=1}^{n_S} \ln \frac{n!}{(i-1)!\,(n-i)!} \mathbf{F}(x'_s;\theta)^{i-1} \left(1-\mathbf{F}(x'_s;\theta)\right)^{n-i} f(x'_s;\theta)
\end{aligned}$$

The computation of the log-likelihood function gives:

$$\begin{aligned}
\boldsymbol{\ell}_{i:n}(\theta) = {} & n_S \ln n! - n_S \ln (i-1)! - n_S \ln (n-i)! + \\
& (i-1)\sum_{s=1}^{n_S} \ln \mathbf{F}(x'_s;\theta) + (n-i)\sum_{s=1}^{n_S} \ln\left(1-\mathbf{F}(x'_s;\theta)\right) + \\
& \sum_{s=1}^{n_S} \ln f(x'_s;\theta)
\end{aligned}$$

By definition, the traditional ML estimator is equal to new ML estimator when $n=1$ and $i=1$:

$$\hat{\theta} = \hat{\theta}_{1:1}$$

In the other cases $(n>1)$, there is no reason that the two estimators coincide exactly:

$$\hat{\theta}_{i:n} \neq \hat{\theta}$$

However, if the random variates are drawn from the distribution function $X \sim \mathbf{F}(x;\theta)$, we can test the hypothesis $\mathcal{H} : \hat{\theta}_{i:n} = \theta$ for all n and $i \leq n$. If two estimates $\hat{\theta}_{i:n}$ and $\hat{\theta}_{i':n'}$ are very different, this indicates that the distribution function is certainly not appropriate for modeling the random variable X.

Let us consider the previous example with the returns of the MSCI USA index. We assume that the daily returns can be modeled with the Student's t distribution:

$$\frac{R_t - \mu}{\sigma} \sim \mathbf{t}_\nu$$

The vector of parameters to estimate is then $\theta = (\mu,\sigma)$. In Tables 12.1, 12.2 and 12.3, we report the values taken by the ML estimator $\hat{\sigma}_{i:n}$ obtained by considering several order statistics and three values of ν. For instance, the ML estimate $\hat{\sigma}_{1:1}$ in the case of the $\mathbf{t}_1$ distribution is equal to 50 bps. We notice that the values taken by $\hat{\sigma}_{i:n}$ are not very stable

[3] The size of the sample n_S is equal to the size of the original sample T divided by n.

TABLE 12.1: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_1$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	50									
2	48	49								
3	44	54	44							
4	41	53	53	41						
5	38	52	55	51	37					
6	35	51	56	56	48	33				
7	32	49	55	56	55	45	29			
8	31	48	53	55	54	50	43	26		
9	29	46	55	56	57	55	49	40	25	
10	28	43	53	58	57	56	53	48	37	20

TABLE 12.2: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_6$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	88									
2	89	87								
3	91	91	85							
4	95	92	89	87						
5	98	99	87	90	88					
6	101	104	95	88	92	89				
7	101	112	100	88	94	95	89			
8	102	116	103	89	85	89	98	89		
9	105	121	117	97	85	86	94	101	88	
10	105	123	120	108	91	87	92	99	104	88

TABLE 12.3: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_\infty$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	125									
2	125	124								
3	136	116	129							
4	147	116	112	140						
5	155	133	103	114	150					
6	163	142	118	107	122	157				
7	171	152	125	105	117	134	162			
8	175	165	130	106	99	111	139	170		
9	180	174	155	122	95	99	128	152	171	
10	183	182	162	136	110	100	111	127	155	181

with respect to i and n. This indicates that the three probability distribution functions ($\mathbf{t}_1$, $\mathbf{t}_6$ and $\mathbf{t}_\infty$) are not well appropriate to represent the index returns. In Figure 12.4, we have reported the corresponding annualized volatility[4] calculated from the order statistics $R_{i:10}$. In the case of the $\mathbf{t}_1$ distribution, we notice that it is lower for median order statistics than extreme order statistics. The $\mathbf{t}_1$ distribution has then the property to overestimate extreme events. In the case of the Gaussian (or $\mathbf{t}_\infty$) distribution, we obtain contrary results. The Gaussian distribution has the property to underestimate extreme events. In order to compensate this bias, the method of maximum likelihood applied to extreme order statistics will overestimate the volatility.

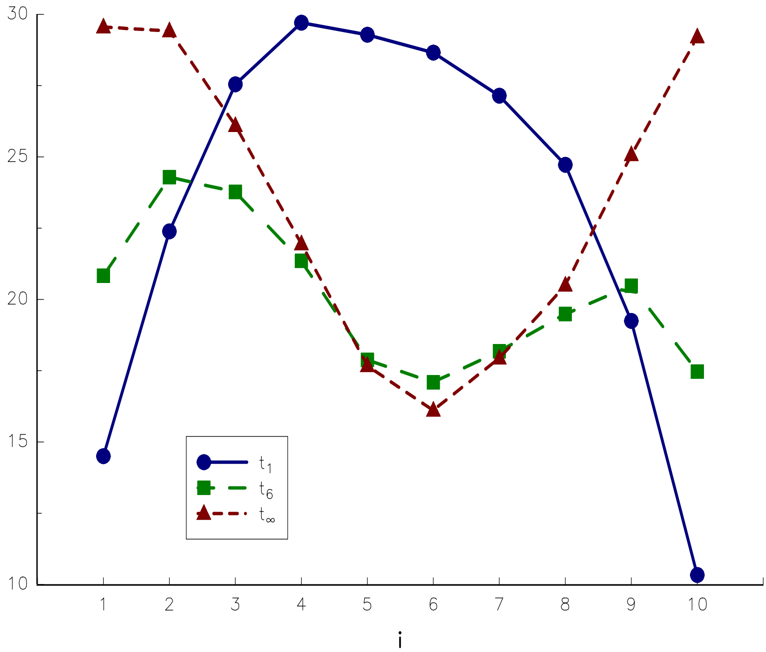

FIGURE 12.4: Annualized volatility (in %) calculated from the order statistics $R_{i:10}$

Remark 144 *The approach based on extreme order statistics to calculate the volatility is then a convenient way to reduce the under-estimation of the Gaussian value-at-risk.*

12.1.4 Extension to dependent random variables

Let us now assume that $X_1, \ldots, X_n$ are not *iid*. We note $\mathbf{C}$ the copula of the corresponding random vector. It follows that:

$$\begin{aligned}
\mathbf{F}_{n:n}(x) &= \Pr\{X_{n:n} \leq x\} \\
&= \Pr\{X_1 \leq x, \ldots, X_n \leq x\} \\
&= \mathbf{C}(\mathbf{F}_1(x), \ldots, \mathbf{F}_n(x))
\end{aligned}$$

[4]The annualized volatility takes the value $\sqrt{260} \cdot c \cdot \hat{\sigma}_{i:n}$ where the constant c is equal to $\sqrt{\nu/(\nu-2)}$. In the case of the $\mathbf{t}_1$ distribution, c is equal to 3.2.

and:

$$\begin{aligned}\mathbf{F}_{1:n}(x) &= \Pr\{X_{1:n} \le x\} \\ &= 1 - \Pr\{X_{1:n} \ge x\} \\ &= 1 - \Pr\{X_1 \ge x, \ldots, X_n \ge x\} \\ &= 1 - \breve{\mathbf{C}}(1 - \mathbf{F}_1(x), \ldots, 1 - \mathbf{F}_n(x))\end{aligned}$$

where $\breve{\mathbf{C}}$ is the survival copula associated to $\mathbf{C}$.

Remark 145 *In the case of the product copula and identical probability distributions, we retrieve the previous results:*

$$\begin{aligned}\mathbf{F}_{n:n}(x) &= \mathbf{C}^{\perp}(\mathbf{F}(x), \ldots, \mathbf{F}(x)) \\ &= \mathbf{F}(x)^n\end{aligned}$$

and:

$$\begin{aligned}\mathbf{F}_{1:n}(x) &= 1 - \mathbf{C}^{\perp}(1 - \mathbf{F}(x), \ldots, 1 - \mathbf{F}(x)) \\ &= 1 - (1 - \mathbf{F}(x))^n\end{aligned}$$

If we are interested in other order statistics, we use the following formula given in Georges *et al.* (2001):

$$\mathbf{F}_{i:n}(x) = \sum_{k=i}^{n} \left[\sum_{l=i}^{k} (-1)^{k-l} \binom{k}{l} \sum_{\mathbf{v}(\mathbf{F}_1(x),\ldots,\mathbf{F}_n(x)) \in \mathcal{Z}(n-k,n)} \mathbf{C}(u_1, \ldots, u_n) \right]$$

where:

$$\mathcal{Z}(m,n) = \left\{ \mathbf{v} \in [0,1]^n \mid v_i \in \{u_i, 1\}, \sum_{i=1}^{n} \mathbf{1}\{v_i = 1\} = m \right\}$$

In order to understand this formula, we consider the case $n = 3$. We have[5]:

$$\begin{aligned}\mathbf{F}_{1:3}(x) &= \mathbf{F}_1(x) + \mathbf{F}_2(x) + \mathbf{F}_3(x) - \\ &\quad \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), 1) - \mathbf{C}(\mathbf{F}_1(x), 1, \mathbf{F}_3(x)) - \mathbf{C}(1, \mathbf{F}_2(x), \mathbf{F}_3(x)) + \\ &\quad \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x)) \\ \mathbf{F}_{2:3}(x) &= \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), 1) + \mathbf{C}(\mathbf{F}_1(x), 1, \mathbf{F}_3(x)) + \mathbf{C}(1, \mathbf{F}_2(x), \mathbf{F}_3(x)) - \\ &\quad 2\mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x)) \\ \mathbf{F}_{3:3}(x) &= \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x))\end{aligned}$$

We verify that:

$$\mathbf{F}_{1:3}(x) + \mathbf{F}_{2:3}(x) + \mathbf{F}_{3:3}(x) = \mathbf{F}_1(x) + \mathbf{F}_2(x) + \mathbf{F}_3(x)$$

The dependence structure has a big impact on the distribution of order statistics. For instance, if we assume that $X_1, \ldots, X_n$ are *iid*, we obtain:

$$\mathbf{F}_{n:n}(x) = \mathbf{F}(x)^n$$

[5] Because $\mathbf{C}(\mathbf{F}_1(x), 1, 1) = \mathbf{F}_1(x)$.

If the copula function is the upper Fréchet copula, this result becomes:

$$\begin{aligned}\mathbf{F}_{n:n}(x) &= \mathbf{C}^{+}(\mathbf{F}(x),\ldots,\mathbf{F}(x)) \\ &= \min(\mathbf{F}(x),\ldots,\mathbf{F}(x)) \\ &= \mathbf{F}(x)\end{aligned}$$

This implies that the occurrence probability of extreme events is lower in this second case.

We consider n Weibull default times $\boldsymbol{\tau}_i \sim \mathcal{W}(\lambda_i, \gamma_i)$. The survival function is equal to $\mathbf{S}_i(t) = \exp(-\lambda_i t^{\gamma_i})$. The hazard rate $\lambda_i(t)$ is then $\lambda_i \gamma_i t^{\gamma_i - 1}$ and the expression of the density is $f_i(t) = \lambda_i(t)\,\mathbf{S}_i(t)$. If we assume that the survival copula is the Gumbel-Hougaard copula with parameter $\theta \geq 1$, the survival function of the first-to-default time is equal to:

$$\begin{aligned}\mathbf{S}_{1:n}(t) &= \exp\left(-\left((-\ln \mathbf{S}_1(t))^{\theta} + \ldots + (-\ln \mathbf{S}_n(t))^{\theta}\right)^{1/\theta}\right) \\ &= \exp\left(-\left(\sum\nolimits_{i=1}^{n} \lambda_i^{\theta} t^{\theta\gamma_i}\right)^{1/\theta}\right)\end{aligned}$$

We deduce the expression of the density function:

$$\begin{aligned}f_{1:n}(t) = &\left(\sum\nolimits_{i=1}^{n} \lambda_i^{\theta} t^{\theta\gamma_i}\right)^{1/\theta-1} \cdot \left(\sum\nolimits_{i=1}^{n} \gamma_i \lambda_i^{\theta} t^{\theta\gamma_i - 1}\right) \cdot \\ &\exp\left(-\left(\sum\nolimits_{i=1}^{n} \lambda_i^{\theta} t^{\theta\gamma_i}\right)^{1/\theta}\right)\end{aligned}$$

In the case where the default times are identically distributed, the first-to-default time is a Weibull default time: $\boldsymbol{\tau}_{1:n} \sim \mathcal{W}\left(n^{1/\theta}\lambda, \gamma\right)$. In Figure 12.5, we report the density function $f_{1:10}(t)$ for the parameters $\lambda = 3\%$ and $\gamma = 2$. We notice that the parameter θ of the copula function has a big influence on the first-to-default time. The case $\theta = 1$ corresponds to the product copula and we retrieve the previous result:

$$\mathbf{S}_{1:n}(t) = \mathbf{S}(t)^{n}$$

When the Gumbel-Hougaard is the upper Fréchet copula ($\theta \to \infty$), we verify that the density function of $\boldsymbol{\tau}_{1:n}$ is this of any default time $\boldsymbol{\tau}_i$.

12.2 Univariate extreme value theory

The extreme value theory consists in studying the limit distribution of extreme order statistics $X_{1:n}$ and $X_{n:n}$ when the sample size tends to infinity. We will see that the limit distribution converges to three probability distributions. This result will help to evaluate stress scenarios and to build a stress testing framework.

Remark 146 *In what follows, we only consider the largest order statistic $X_{n:n}$. Indeed, the minimum order statistic $X_{1:n}$ can be defined with respect to the maximum order statistic $Y_{n:n}$ by setting $Y_i = -X_i$:*

$$\begin{aligned}X_{1:n} &= \min(X_1,\ldots,X_n) \\ &= \min(-Y_1,\ldots,-Y_n) \\ &= -\max(Y_1,\ldots,Y_n) \\ &= -Y_{n:n}\end{aligned}$$

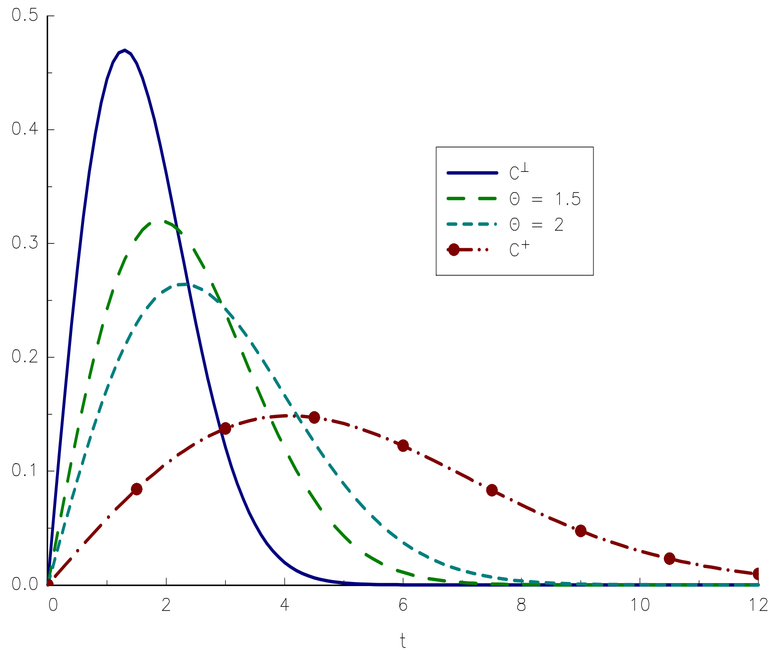

FIGURE 12.5: Density function of the first-to-default time $\boldsymbol{\tau}_{1:10}$

12.2.1 Fisher-Tippet theorem

We follow Embrechts *et al.* (1997) for the formulation of the Fisher-Tippet theorem. Let $X_1, \ldots, X_n$ be a sequence of *iid* random variables, whose distribution function is $\mathbf{F}$. If there exist two constants a_n and b_n and a non-degenerate distribution function $\mathbf{G}$ such that:

$$\lim_{n \to \infty} \Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} = \mathbf{G}(x) \tag{12.4}$$

then $\mathbf{G}$ can be classified as one of the following three types[6]:

Type I	(Gumbel)	$\boldsymbol{\Lambda}(x) = \exp\left(-e^{-x}\right)$
Type II	(Fréchet)	$\boldsymbol{\Phi}_\alpha(x) = \mathbb{1}(x \geq 0) \cdot \exp\left(-x^{-\alpha}\right)$
Type III	(Weibull)	$\boldsymbol{\Psi}_\alpha(x) = \mathbb{1}(x \leq 0) \cdot \exp\left(-(-x)^{\alpha}\right)$

The distribution functions $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_\alpha$ et $\boldsymbol{\Psi}_\alpha$ are called extreme value distributions. The Fisher-Tippet theorem is very important, because the set of extreme value distributions is very small although the set of distribution functions is very large. We can draw a parallel with the normal distribution and the sum of random variables. In some sense, the Fisher-Tippet theorem provides an extreme value analog of the central limit theorem.

[6] In terms of probability density functions, we have:

$$g(x) = \begin{cases} \exp\left(-x - e^{-x}\right) & \text{(Gumbel)} \\ \mathbb{1}(x \geq 0) \cdot \alpha x^{-(1+\alpha)} \cdot \exp\left(-x^{-\alpha}\right) & \text{(Fréchet)} \\ \mathbb{1}(x \leq 0) \cdot \alpha(-x)^{\alpha-1} \cdot \exp\left(-(-x)^{\alpha}\right) & \text{(Weibull)} \end{cases}$$

Let us consider the case of exponential random variables, whose probability distribution is $\mathbf{F}(x) = 1 - \exp(-\lambda x)$. We have[7]:

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{F}_{n:n}(x) &= \lim_{n\to\infty} \left(1 - e^{-\lambda x}\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{ne^{-\lambda x}}{n}\right)^n \\
&= \lim_{n\to\infty} \exp\left(-ne^{-\lambda x}\right) \\
&= 0
\end{aligned}$$

We verify that the limit distribution is degenerate. If we consider the affine transformation with $a_n = 1/\lambda$ et $b_n = (\ln n)/\lambda$, we obtain:

$$\begin{aligned}
\Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \Pr\left\{X_{n:n} \leq a_n x + b_n\right\} \\
&= \left(1 - e^{-\lambda(a_n x + b_n)}\right)^n \\
&= \left(1 - e^{-x - \ln n}\right)^n \\
&= \left(1 - \frac{e^{-x}}{n}\right)^n
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\mathbf{G}(x) &= \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \\
&= \exp\left(-e^{-x}\right)
\end{aligned}$$

It follows that the limit distribution of the affine transformation is not degenerate. In Figure 12.6, we illustrate the convergence of $\mathbf{F}^n(a_n x + b_n)$ to the Gumbel distribution $\boldsymbol{\Lambda}(x)$.

Example 127 *If we consider the Pareto distribution, we have:*

$$\mathbf{F}(x) = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

The normalizing constants are $a_n = x_- n^{1/\alpha}$ and $b_n = 0$. We obtain:

$$\begin{aligned}
\Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \left(1 - \left(\frac{x_- n^{1/\alpha} x}{x_-}\right)^{-\alpha}\right)^n \\
&= \left(1 - \frac{x^{-\alpha}}{n}\right)^n
\end{aligned}$$

We deduce that the law of the maximum tends to the Fréchet distribution:

$$\lim_{n\to\infty} \left(1 - \frac{x^{-\alpha}}{n}\right)^n = \exp\left(-x^{-\alpha}\right)$$

[7] Because we have:

$$\begin{aligned}
\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \\
&= \exp(x)
\end{aligned}$$

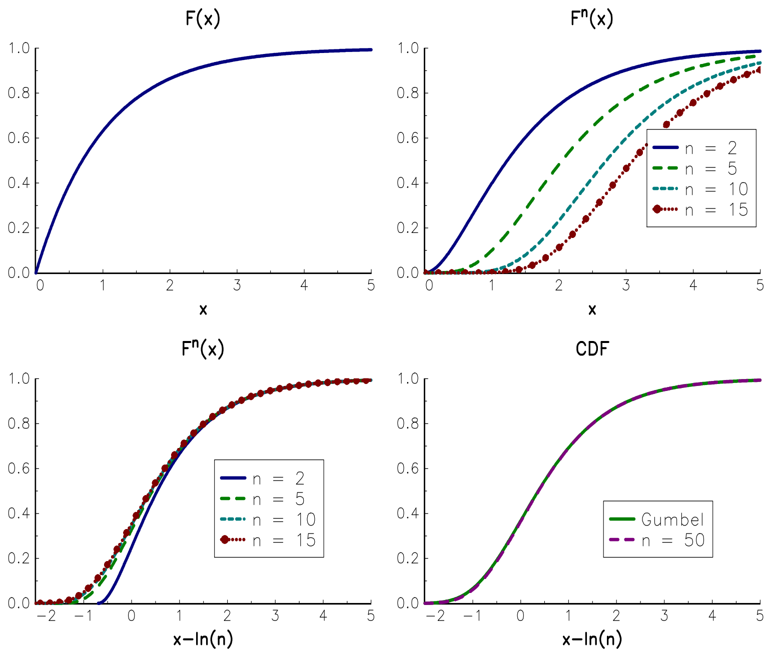

FIGURE 12.6: Max-convergence of the exponential distribution $\mathcal{E}(1)$ to the Gumbel distribution

Example 128 *For the uniform distribution, the normalizing constants become* $a_n = n^{-1}$ *and* $b_n = 1$ *and we obtain the Weibull distribution with* $\alpha = 1$*:*

$$\begin{aligned}\lim_{n\to\infty} \Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \left(1 + \frac{x}{n}\right)^n \\ &= \exp(x)\end{aligned}$$

12.2.2 Maximum domain of attraction

The application of the Fisher-Tippet theorem is limited because it can be extremely difficult to find the normalizing constants and the extreme value distribution for a given probability distribution $\mathbf{F}$. However, the graphical representation of $\mathbf{\Lambda}$, $\mathbf{\Phi}_\alpha$ and $\mathbf{\Psi}_\alpha$ given in Figure 12.7 already provides some information. For instance, the Weibull probability distribution concerns random variables that are right bounded. This is why it has less interest in finance than the Fréchet or Gumbel distribution functions[8]. We also notice some differences in the shape of the curves. In particular, the Gumbel distribution is more '*normal*' than the Fréchet distribution, whose shape and tail depend on the parameter α (see Figure 12.8).

We say that the distribution function $\mathbf{F}$ belongs to the max-domain of attraction of the distribution function $\mathbf{G}$ and we write $\mathbf{F} \in \mathrm{MDA}(\mathbf{G})$ if the distribution function of the normalized maximum converges to $\mathbf{G}$. For instance, we have already seen that $\mathcal{E}(\lambda) \in \mathrm{MDA}(\mathbf{\Lambda})$. In what follows, we indicate how to characterize the set $\mathrm{MDA}(\mathbf{G})$ and which normalizing constants are[9].

[8]However, the Weibull probability distribution is related to the Fréchet probability distribution thanks to the relationship $\mathbf{\Psi}_\alpha(x) = \mathbf{\Phi}_\alpha\left(-x^{-1}\right)$.

[9]Most of the following results come from Resnick (1987).

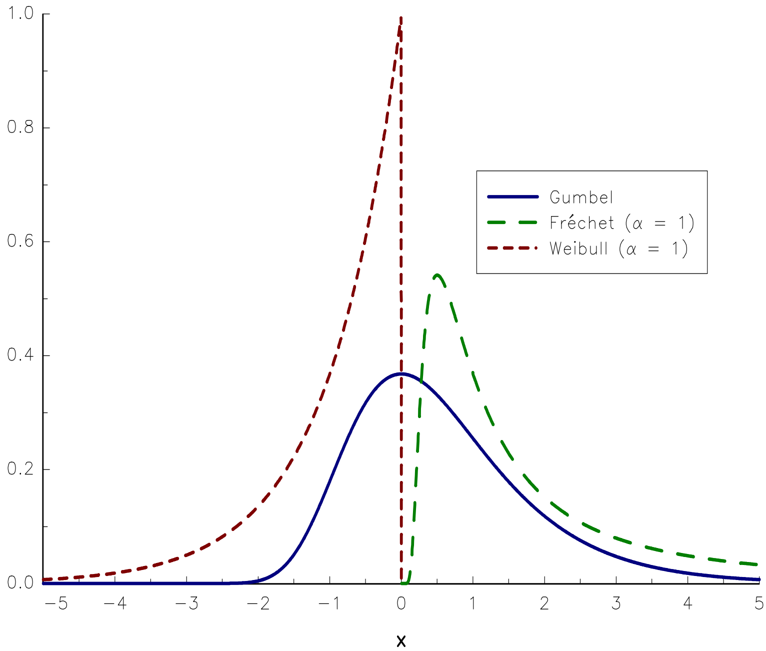

FIGURE 12.7: Density function of $\mathbf{\Lambda}$, $\mathbf{\Phi}_1$ and $\mathbf{\Psi}_1$

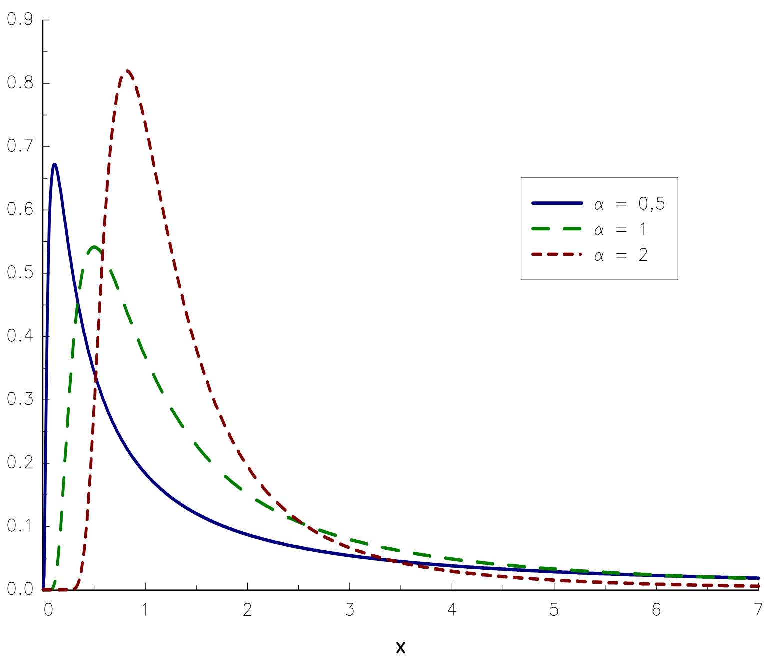

FIGURE 12.8: Density function of the Fréchet probability distribution

12.2.2.1 MDA of the Gumbel distribution

$\mathbf{F} \in \mathrm{MDA}(\mathbf{\Lambda})$ if and only if there exists a function $h(t)$ such that:

$$\lim_{t \to x_0} \frac{1 - \mathbf{F}(t + x \cdot h(t))}{1 - \mathbf{F}(t)} = \exp(-x)$$

where $x_0 \leq \infty$. The normalizing constants are then $a_n = h\left(\mathbf{F}^{-1}\left(1 - n^{-1}\right)\right)$ and $b_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right)$.

The previous characterization of $\mathrm{MDA}(\mathbf{\Lambda})$ is difficult to use because we have to define the function $h(t)$. However, we can show that if the distribution function $\mathbf{F}$ is C^2, a sufficient condition is:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = -1$$

For instance, in the case of the exponential distribution, we have $\mathbf{F}(x) = 1 - \exp(-\lambda x)$, $\partial_x \mathbf{F}(x) = \lambda \exp(-\lambda x)$ and $\partial_x^2 \mathbf{F}(x) = -\lambda^2 \exp(-\lambda x)$. We verify that:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = \lim_{x \to \infty} \frac{\exp(-\lambda x) \cdot \left(-\lambda^2 \exp(-\lambda x)\right)}{(\lambda \exp(-\lambda x))^2} = -1$$

If we consider the Gaussian distribution $\mathcal{N}(0,1)$, we have $\mathbf{F}(x) = \Phi(x)$, $\partial_x \mathbf{F}(x) = \phi(x)$ and $\partial_x^2 \mathbf{F}(x) = -x\phi(x)$. Using L'Hospital's rule, we deduce that:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = \lim_{x \to \infty} -\frac{x \cdot \Phi(-x)}{\phi(x)} = -1$$

12.2.2.2 MDA of the Fréchet distribution

We say that a function f is regularly varying with index α and we write $f \in \mathrm{RV}_\alpha$ if we have:

$$\lim_{t \to \infty} \frac{f(t \cdot x)}{f(t)} = x^\alpha$$

for every $x > 0$. We can then show the following theorem: $\mathbf{F} \in \mathrm{MDA}(\mathbf{\Phi}_\alpha)$ if and only if $1 - \mathbf{F} \in \mathrm{RV}_{-\alpha}$, and the normalizing constants are $a_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right)$ and $b_n = 0$.

Using the previous theorem, we deduce that the distribution function $\mathbf{F} \in \mathrm{MDA}(\mathbf{\Phi}_\alpha)$ if it satisfies the following condition:

$$\lim_{t \to \infty} \frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} = x^{-\alpha}$$

If we apply this result to the Pareto distribution, we obtain:

$$\begin{aligned}
\lim_{t \to \infty} \frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} &= \lim_{t \to \infty} \frac{(t \cdot x / x_-)^{-\alpha}}{(t / x_-)^{-\alpha}} \\
&= x^{-\alpha}
\end{aligned}$$

We deduce that $1 - \mathbf{F} \in \mathrm{RV}_{-\alpha}$, $\mathbf{F} \in \mathrm{MDA}(\mathbf{\Phi}_\alpha)$, $a_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right) = x_- n^{1/\alpha}$ and $b_n = 0$.

Remark 147 *The previous theorem suggests that:*

$$\frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} \approx x^{-\alpha}$$

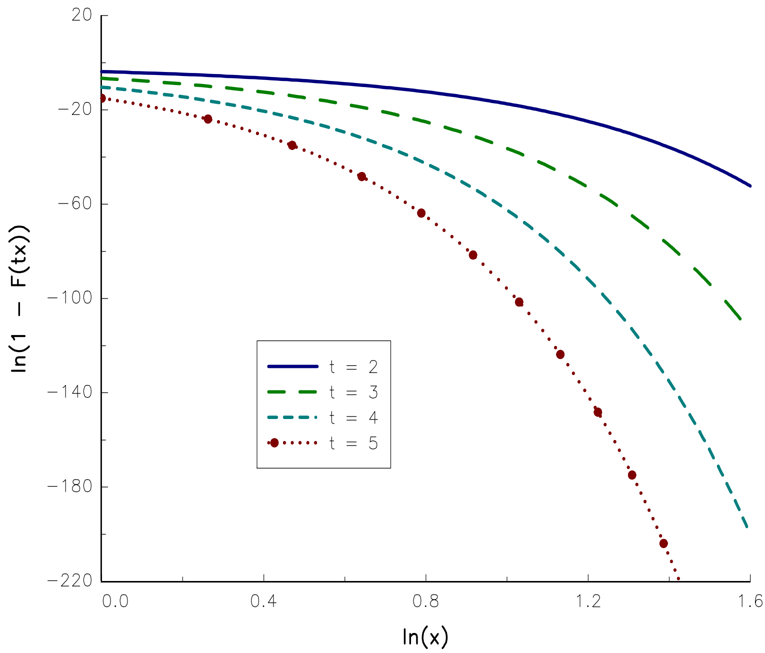

FIGURE 12.9: Graphical validation of the regular variation property for the normal distribution $\mathcal{N}(0,1)$

when t is sufficiently large. This means that we must observe a linear relationship between $\ln(x)$ and $\ln(1-\mathbf{F}(t\cdot x))$:

$$\ln(1-\mathbf{F}(t\cdot x)) \approx \ln(1-\mathbf{F}(t)) - \alpha\ln(x)$$

This property can be used to check graphically if a given distribution function belongs or not to the maximum domain of attraction of the Fréchet distribution. For instance, we observe that $\mathcal{N}(0,1) \notin \mathrm{MDA}(\mathbf{\Phi}_\alpha)$ in Figure 12.9, because the curve is not a straight line.

12.2.2.3 MDA of the Weibull distribution

For the Weibull distribution, we can show that $\mathbf{F} \in \mathrm{MDA}(\mathbf{\Psi}_\alpha)$ if and only if $1-\mathbf{F}\left(x_0 - x^{-1}\right) \in \mathrm{RV}_{-\alpha}$ and $x_0 < \infty$. The normalizing constants are $a_n = x_0 - \mathbf{F}^{-1}\left(1-n^{-1}\right)$ and $b_n = x_0$.

If we consider the uniform distribution with $x_0 = 1$, we have:

$$\mathbf{F}\left(x_0 - x^{-1}\right) = 1 - \frac{1}{x}$$

and:

$$\begin{aligned}
\lim_{t\to\infty}\frac{1-\mathbf{F}\left(1-t^{-1}x^{-1}\right)}{1-\mathbf{F}\left(1-t^{-1}\right)} &= \lim_{t\to\infty}\frac{t^{-1}x^{-1}}{t^{-1}} \\
&= x^{-1}
\end{aligned}$$

We deduce that $\mathbf{F} \in \mathrm{MDA}(\mathbf{\Psi}_1)$, $a_n = 1 - \mathbf{F}^{-1}\left(1-n^{-1}\right) = n^{-1}$ and $b_n = 1$.

TABLE 12.4: Maximum domain of attraction and normalizing constants of some distribution functions

Distribution	$\mathbf{G}(x)$	a_n	b_n
$\mathcal{E}(\lambda)$	$\mathbf{\Lambda}$	λ^{-1}	$\lambda^{-1}\ln n$
$\mathcal{G}(\alpha,\beta)$	$\mathbf{\Lambda}$	β^{-1}	$\beta^{-1}(\ln n + (\alpha-1)\ln(\ln n) - \ln\Gamma(\alpha))$
$\mathcal{N}(0,1)$	$\mathbf{\Lambda}$	$(2\ln n)^{-1/2}$	$\dfrac{4\ln n - \ln 4\pi - \ln(\ln n)}{2\sqrt{2\ln n}}$
$\mathcal{LN}(\mu,\sigma^2)$	$\mathbf{\Lambda}$	$\sigma(2\ln n)^{-1/2}b_n$	$\exp\left(\mu+\sigma\left(\dfrac{4\ln n - \ln 4\pi + \ln(\ln n)}{2\sqrt{2\ln n}}\right)\right)$
$\mathcal{P}(\alpha,x_-)$	$\mathbf{\Phi}_\alpha$	$x_- n^{1/\alpha}$	0
$\mathcal{LG}(\alpha,\beta)$	$\mathbf{\Phi}_\beta$	$\dfrac{\left(n(\ln n)^{\alpha-1}\right)^{1/\beta}}{\Gamma(\alpha)}$	0
t_ν	$\mathbf{\Phi}_\nu$	$\mathbf{T}_\nu^{-1}\left(1-n^{-1}\right)$	0
$\mathcal{U}_{[0,1]}$	$\mathbf{\Psi}_1$	n^{-1}	1
$\mathcal{B}(\alpha,\beta)$	$\mathbf{\Psi}_\alpha$	$\left(\dfrac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta+1)}\right)^{-1/\beta}$	1

Source: Embrechts *et al.* (1997).

12.2.2.4 Main results

In Table 12.4, we report the maximum domain of attraction and normalizing constants of some well-known distribution functions.

Remark 148 *Let $\mathbf{G}(x)$ be the non-degenerate distribution of $X_{n:n}$. We note a_n and b_n the normalizing constants. We consider the linear transformation $Y = cX + d$ with $c > 0$. Because we have $Y_{n:n} = cX_{n:n} + d$, we deduce that:*

$$\begin{aligned}
\mathbf{G}(x) &= \lim_{n\to\infty} \Pr\{X_{n:n} \le a_n x + b_n\} \\
&= \lim_{n\to\infty} \Pr\left\{\frac{Y_{n:n}-d}{c} \le a_n x + b_n\right\} \\
&= \lim_{n\to\infty} \Pr\{Y_{n:n} \le a_n c x + b_n c + d\} \\
&= \lim_{n\to\infty} \Pr\{Y_{n:n} \le a'_n x + b'_n\}
\end{aligned}$$

where $a'_n = a_n c$ and $b'_n = b_n c + d$. This means that $\mathbf{G}(x)$ is also the non-degenerate distribution of $Y_{n:n}$, and a'_n and b'_n are the normalizing constants. For instance, if we consider the distribution function $\mathcal{N}(\mu,\sigma^2)$, we deduce that the normalizing constants are:

$$a_n = \sigma(2\ln n)^{-1/2}$$

and:

$$b_n = \mu + \sigma \left(\frac{4 \ln n - \ln 4\pi + \ln(\ln n)}{2\sqrt{2 \ln n}} \right)$$

The normalizing constants are uniquely defined. In the case of the Gaussian distribution $\mathcal{N}(0,1)$, they are equal to $a_n = h(b_n) = b_n / \left(1 + b_n^2\right)$ and $b_n = \Phi^{-1}\left(1 - n^{-1}\right)$. In Table 12.4, we report an approximation which is not necessarily unique. For instance, Gasull *et al.* (2015) propose the following alternative value of b_n:

$$b_n \approx \sqrt{\ln\left(\frac{n^2}{2\pi}\right) - \ln\left(\ln\left(\frac{n^2}{2\pi}\right)\right) + \frac{\ln(0.5 + \ln n^2) - 2}{\ln n^2 - \ln 2\pi}}$$

and show that this solution is more accurate than the classical approximation.

12.2.3 Generalized extreme value distribution

12.2.3.1 Definition

From a statistical point of view, the previous results of the extreme value theory are difficult to use. Indeed, they are many issues concerning the choice of the distribution function, the normalizing constants or the convergence rate as explained by Coles (2001):

> "*The three types of limits that arise in Theorem 12.2.1 have distinct forms of behavior, corresponding to the different forms of tail behaviour for the distribution function* **F** *of the* X_i. *This can be made precise by considering the behavior of the limit distribution* **G** *at* x_+, *its upper end-point. For the Weibull distribution* x_+ *is finite, while for both the Fréchet and Gumbel distributions* $x_+ = \infty$. *However, the density of* **G** *decays exponentially for the Gumbel distribution and polynomially for the Fréchet distribution, corresponding to relatively different rates of decay in the tail of* **F**. *It follows that in applications the three different families give quite different representations of extreme value behavior. In early applications of extreme value theory, it was usual to adopt one of the three families, and then to estimate the relevant parameters of that distribution. But there are two weakness: first, a technique is required to choose which of the three families is most appropriate for the data at hand; second, once such a decision is made, subsequent inferences presume this choice to be correct, and do not allow for the uncertainty such a selection involves, even though this uncertainty may be substantial*".

In practice, the statistical inference on extreme values takes another route. Indeed, the three types can be combined into a single distribution function:

$$\mathbf{G}(x) = \exp\left(- \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-1/\xi} \right)$$

defined on the support $\Delta = \left\{ x : 1 + \xi\sigma^{-1}(x - \mu) > 0 \right\}$. It is known as the '*generalized extreme value*' distribution and we denote it by $\mathcal{GEV}(\mu, \sigma, \xi)$. We obtain the following cases:

- the limit case $\xi \to 0$ corresponds to the Gumbel distribution;
- $\xi = -\alpha^{-1} > 0$ defines the Fréchet distribution;
- the Weibull distribution is obtained by considering $\xi = -\alpha^{-1} < 0$.

We also notice that the parameters μ and σ are the limits of the normalizing constants b_n and a_n. The corresponding density function is equal to:

$$g(x) = \frac{1}{\sigma} \cdot \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-(1+\xi)/\xi} \cdot \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

It is represented in Figure 12.10 for various values of the parameters. We notice that μ is a parameter of localization, σ controls the standard deviation and ξ is related to the tail of the distribution. The parameters can be estimated using the method of maximum likelihood and we obtain:

$$\boldsymbol{\ell}_t = -\ln\sigma - \left(\frac{1+\xi}{\xi}\right)\ln\left(1+\xi\left(\frac{x_t-\mu}{\sigma}\right)\right) - \left(1+\xi\left(\frac{x_t-\mu}{\sigma}\right)\right)^{-1/\xi}$$

where x_t is the observed maximum for the t^{th} period.

We consider again the example of the MSCI USA index. Using daily returns, we calculate the block maximum for each period of 22 trading days. We then estimate the GEV distribution using the method of maximum likelihood. For the period 1995-2015, we obtain $\hat{\mu} = 0.0149$, $\hat{\sigma} = 0.0062$ and $\hat{\xi} = 0.3736$. In Figure 12.11, we compared the estimated GEV distribution with the distribution function $\mathbf{F}_{22:22}(x)$ when we assume that daily returns are Gaussian. We notice that the Gaussian hypothesis largely underestimates extreme events as illustrated by the quantile function in the table below:

α	90%	95%	96%	97%	98%	99%
Gaussian	3.26%	3.56%	3.65%	3.76%	3.92%	4.17%
GEV	3.66%	4.84%	5.28%	5.91%	6.92%	9.03%

For instance, the probability is 1% to observe a maximum daily return during a period of one month larger than 4.17% in the case of the Gaussian distribution and 9.03% in the case of the GEV distribution.

12.2.3.2 Estimating the value-at-risk

Let us consider a portfolio w, whose mark-to-market value is $P_t(w)$ at time t. We recall that the P&L between t and $t+1$ is equal to:

$$\begin{aligned}\Pi(w) &= P_{t+1}(w) - P_t(w) \\ &= P_t(w) \cdot R(w)\end{aligned}$$

where $R(w)$ is the daily return of the portfolio. If we note $\hat{\mathbf{F}}$ the estimated probability distribution of $R(w)$, the expression of the value-at-risk at the confidence level α is equal to:

$$\mathrm{VaR}_\alpha(w) = -P_t(w) \cdot \hat{\mathbf{F}}^{-1}(1-\alpha)$$

We now estimate the GEV distribution $\hat{\mathbf{G}}$ of the maximum of $-R(w)$ for a period of n trading days[10]. We have to define the confidence level α_{GEV} when we consider block minima of daily returns that corresponds to the same confidence level α when we consider daily returns. For that, we assume that the two exception events have the same return period, implying that:

$$\frac{1}{1-\alpha} \times 1 \text{ day} = \frac{1}{1-\alpha_{\mathrm{GEV}}} \times n \text{ days}$$

[10] We model the maximum of the opposite of daily returns, because we are interested in extreme losses, and not in extreme profits.

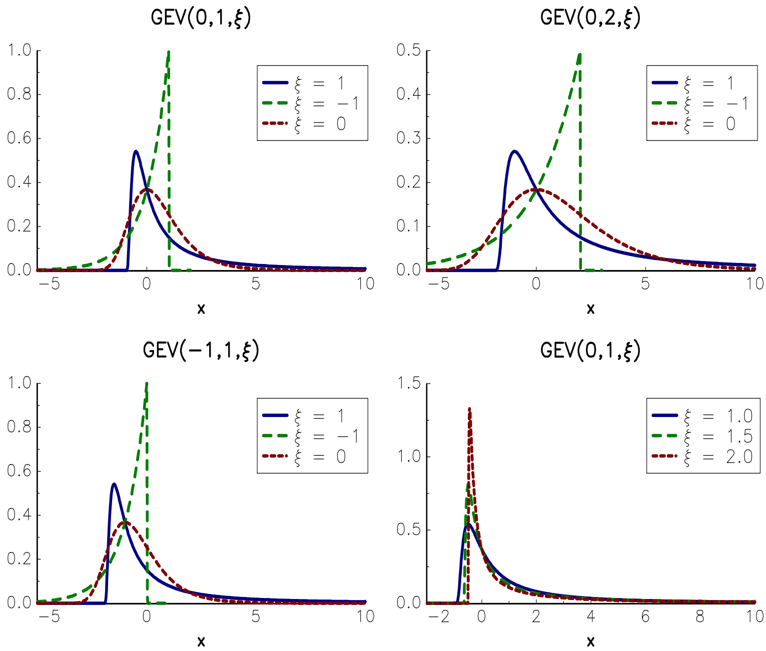

FIGURE 12.10: Probability density function of the GEV distribution

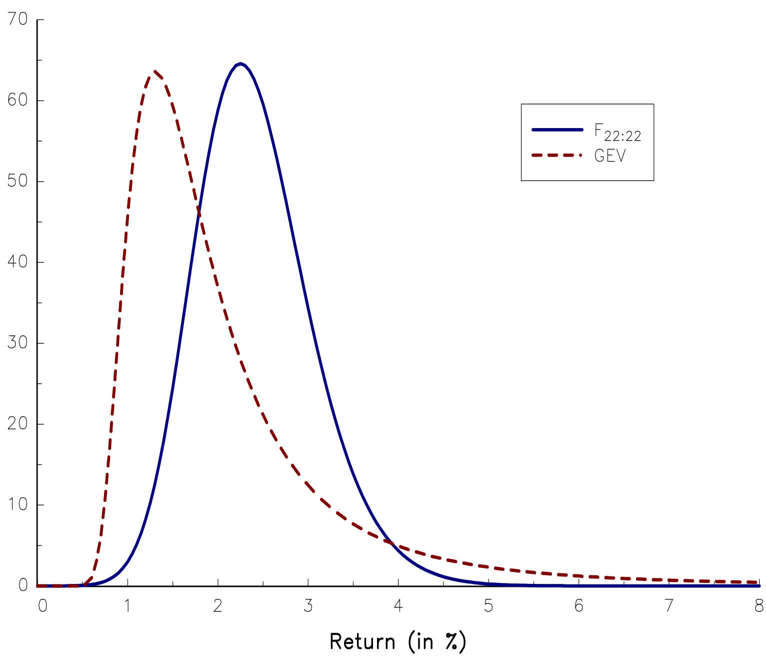

FIGURE 12.11: Probability density function of the maximum return $R_{22:22}$

We deduce that:

$$\alpha_{\text{GEV}} = 1 - (1 - \alpha) \cdot n$$

It follows that the value-at-risk calculated with the GEV distribution is equal to[11]:

$$\text{VaR}_\alpha (w) = P(t) \cdot \hat{\mathbf{G}}^{-1} (\alpha_{\text{GEV}})$$

We consider four portfolios invested in the MSCI USA index and the MSCI EM index: (1) long on the MSCI USA, (2) long on the MSCI EM index, (3) long on the MSCI USA and short on the MSCI EM index and (4) long on the MSCI EM index and short on the MSCI USA index. Using daily returns from January 1995 to December 2015, we estimate the daily value-at-risk of these portfolios for different confidence levels α. We report the results in Table 12.5 for Gaussian and historical value-at-risk measures and compare them with those calculated with the GEV approach. In this case, we estimate the parameters of the extreme value distribution using block maxima of 22 trading days. When we consider a 99% confidence level, the lowest value is obtained by the GEV method followed by Gaussian and historical methods. For a higher quantile, the GEV VaR is between the Gaussian VaR and the historical VaR. The value-at-risk calculated with the GEV approach can therefore be interpreted as a parametric value-at-risk, which is estimated using only tail events.

TABLE 12.5: Comparing Gaussian, historical and GEV value-at-risk measures

VaR	α	Long US	Long EM	Long US Short EM	Long EM Short US
Gaussian	99.0%	2.88%	2.83%	3.06%	3.03%
	99.5%	3.19%	3.14%	3.39%	3.36%
	99.9%	3.83%	3.77%	4.06%	4.03%
Historical	99.0%	3.46%	3.61%	3.37%	3.81%
	99.5%	4.66%	4.73%	3.99%	4.74%
	99.9%	7.74%	7.87%	6.45%	7.27%
GEV	99.0%	2.64%	2.61%	2.72%	2.93%
	99.5%	3.48%	3.46%	3.41%	3.82%
	99.9%	5.91%	6.05%	5.35%	6.60%

12.2.4 Peak over threshold

12.2.4.1 Definition

The estimation of the GEV distribution is a '*block component-wise*' approach. This means that from a sample of random variates, we build a sample of maxima by considering blocks with the same length. This implies a loss of information, because some blocks may contain several extreme events whereas some other blocks may not be impacted by extremes. Another approach consists in using the '*peak over threshold*' (POT) method. In this case, we are interested in estimating the distribution of exceedance over a certain threshold u:

$$\mathbf{F}_u(x) = \Pr \{X - u \leq x \mid X > u\}$$

[11] The inverse function of the probability distribution $\mathcal{GEV}(\mu, \sigma, \xi)$ is equal to:

$$\mathbf{G}^{-1}(\alpha) = \mu - \frac{\sigma}{\xi} \left(1 - (-\ln \alpha)^{-\xi}\right)$$

where $0 \leq x < x_0 - u$ and $x_0 = \sup \{x \in \mathbb{R} : \mathbf{F}(x) < 1\}$. $\mathbf{F}_u(x)$ is also called the conditional excess distribution function. It is also equal to:

$$\begin{aligned}
\mathbf{F}_u(x) &= 1 - \Pr\{X - u \leq x \mid X \leq u\} \\
&= 1 - \left(\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)}\right) \\
&= \frac{\mathbf{F}(u+x) - \mathbf{F}(u)}{1 - \mathbf{F}(u)}
\end{aligned}$$

Pickands (1975) showed that, for very large u, $\mathbf{F}_u(x)$ follows a generalized Pareto distribution (GPD): $\mathbf{F}_u(x) \approx \mathbf{H}(x)$ where[12]:

$$\mathbf{H}(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}$$

The distribution function $\mathcal{GPD}(\sigma, \xi)$ depends on two parameters: σ is the scale parameter and ξ is the shape parameter.

Example 129 *If* $\mathbf{F}$ *is an exponential distribution* $\mathcal{E}(\lambda)$*, we have:*

$$\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} = \exp(-\lambda x)$$

This is the generalized Pareto distribution when $\sigma = 1/\lambda$ *and* $\xi \to 0$.

Example 130 *If* $\mathbf{F}$ *is a uniform distribution, we have:*

$$\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} = 1 - \frac{x}{1-u}$$

It corresponds to the generalized Pareto distribution with the following parameters: $\sigma = 1 - u$ *and* $\xi = -1$.

In fact, there is a strong link between the block maxima approach and the peak over threshold method. Suppose that $X_{n:n} \sim \mathcal{GEV}(\mu, \sigma, \xi)$. It follows that:

$$\mathbf{F}^n(x) \approx \exp\left\{-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}$$

We deduce that:

$$n \ln \mathbf{F}(x) \approx -\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}$$

Using the approximation $\ln \mathbf{F}(x) \approx -(1 - \mathbf{F}(x))$ for large x, we obtain:

$$1 - \mathbf{F}(x) \approx \frac{1}{n}\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}$$

We find that $\mathbf{F}_u(x)$ is a generalized Pareto distribution $\mathcal{GPD}(\tilde{\sigma}, \xi)$:

$$\begin{aligned}
\Pr\{X > u + x \mid X > u\} &= \frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} \\
&= \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}
\end{aligned}$$

[12]If $\xi \to 0$, we have $\mathbf{H}(x) = 1 - \exp(-x/\sigma)$.

where:

$$\tilde{\sigma} = \sigma + \xi (u - \mu)$$

Therefore, we have a duality between GEV and GPD distribution functions:

> "*[...] if block maxima have approximating distribution* $\mathbf{G}$, *then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter* ξ *is equal to that of the corresponding GEV distribution. Choosing a different, but still large, block size* n *would affect the values of the GEV parameters, but not those of the corresponding generalized Pareto distribution of threshold excesses:* ξ *is invariant to block size, while the calculation of* $\tilde{\sigma}$ *is unperturbed by the changes in* μ *and* σ *which are self-compensating*" (Coles, 2001, page 75).

The estimation of the parameters (σ, ξ) is not obvious because it depends on the value taken by the threshold u. It must be sufficiently large to apply the previous theorem, but we also need enough data to obtain good estimates. We notice that the mean residual life $e(u)$ is a linear function of u:

$$\begin{aligned} e(u) &= \mathbb{E}[X - u \mid X > u] \\ &= \frac{\sigma + \xi u}{1 - \xi} \end{aligned}$$

when $\xi < 1$. If the GPD approximation is valid for a value u_0, it is therefore valid for any value $u > u_0$. To determine u_0, we can use a mean residual life plot, which consists in plotting u against the empirical mean excess $\hat{e}(u)$:

$$\hat{e}(u) = \frac{\sum_{i=1}^{n} (x_i - u)^+}{\sum_{i=1}^{n} \mathbb{1}\{x_i > u\}}$$

Once u_0 is found, we estimate the parameters (σ, ξ) by the method of maximum likelihood or the linear regression[13].

Let us consider our previous example. In Figure 12.12, we have reported the mean residual life plot for the left tail of the four portfolios[14]. The determination of u_0 consists in finding linear relationships. We have a first linear relationship between $u = -3\%$ and $u = -1\%$, but it is not valid because it is followed by a change in slope. We prefer to consider that the linear relationship is valid for $u \geq 2\%$. By assuming that $u_0 = 2\%$ for all the four portfolios, we obtain the estimates given in Table 12.6.

12.2.4.2 Estimating the expected shortfall

We recall that:

$$\mathbf{F}_u(x) = \frac{\mathbf{F}(u + x) - \mathbf{F}(u)}{1 - \mathbf{F}(u)} \approx \mathbf{H}(x)$$

where $\mathbf{H} \sim \mathcal{GPD}(\sigma, \xi)$. We deduce that:

$$\begin{aligned} \mathbf{F}(x) &= \mathbf{F}(u) + (1 - \mathbf{F}(u)) \cdot \mathbf{F}_u(x - u) \\ &\approx \mathbf{F}(u) + (1 - \mathbf{F}(u)) \cdot \mathbf{H}(x - u) \end{aligned}$$

---

[13] In this case, we estimate the linear model $\hat{e}(u) = a + b \cdot u + \varepsilon$ for $u \geq u_0$ and deduce that $\hat{\sigma} = \hat{a} / \left(1 + \hat{b}\right)$ and $\hat{\xi} = \hat{b} / \left(1 + \hat{b}\right)$.

[14] This means that $\hat{e}(u)$ is calculated using the portfolio loss, that is the opposite of the portfolio return.

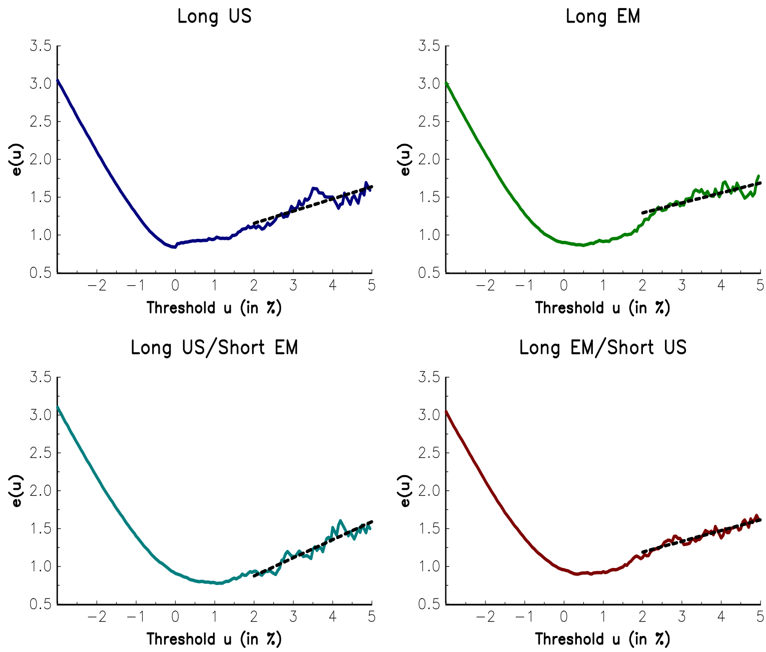

FIGURE 12.12: Mean residual life plot

TABLE 12.6: Estimation of the generalized Pareto distribution

Parameter	Long US	Long EM	Long US Short EM	Long EM Short US
$\hat{a}$	0.834	1.029	0.394	0.904
$\hat{b}$	0.160	0.132	0.239	0.142
$\hat{\sigma}$	0.719	0.909	0.318	0.792
$\hat{\xi}$	0.138	0.117	0.193	0.124

We consider a sample of size n. We note n_u the number of observations whose value x_i is larger than the threshold u. The non-parametric estimate of $\mathbf{F}(u)$ is then equal to:

$$\hat{\mathbf{F}}(u) = 1 - \frac{n_u}{n}$$

Therefore, we obtain the following semi-parametric estimate of $\mathbf{F}(x)$ for x larger than u:

$$\begin{aligned}
\hat{\mathbf{F}}(x) &= \hat{\mathbf{F}}(u) + \left(1 - \hat{\mathbf{F}}(u)\right) \cdot \hat{\mathbf{H}}(x-u) \\
&= \left(1 - \frac{n_u}{n}\right) + \frac{n_u}{n}\left(1 - \left(1 + \frac{\hat{\xi}(x-u)}{\hat{\sigma}}\right)^{-1/\hat{\xi}}\right) \\
&= 1 - \frac{n_u}{n}\left(1 + \frac{\hat{\xi}(x-u)}{\hat{\sigma}}\right)^{-1/\hat{\xi}}
\end{aligned}$$

We can interpret $\hat{\mathbf{F}}(x)$ as the historical estimate of the probability distribution tail that is improved by the extreme value theory. We deduce that:

$$\begin{aligned}
\mathrm{VaR}_\alpha &= \hat{\mathbf{F}}^{-1}(\alpha) \\
&= u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{n}{n_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right)
\end{aligned}$$

and:

$$\begin{aligned}
\mathrm{ES}_\alpha &= \mathbb{E}\left[X \mid X > \mathrm{VaR}_\alpha\right] \\
&= \mathrm{VaR}_\alpha + \mathbb{E}\left[X - \mathrm{VaR}_\alpha \mid X > \mathrm{VaR}_\alpha\right] \\
&= \mathrm{VaR}_\alpha + \frac{\hat{\sigma} + \hat{\xi}\left(\mathrm{VaR}_\alpha - u\right)}{1-\hat{\xi}} \\
&= \frac{\mathrm{VaR}_\alpha}{1-\hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi}u}{1-\hat{\xi}} \\
&= u - \frac{\hat{\sigma}}{\hat{\xi}} + \frac{\hat{\sigma}}{\left(1-\hat{\xi}\right)\hat{\xi}}\left(\frac{n}{n_u}(1-\alpha)\right)^{-\hat{\xi}}
\end{aligned}$$

We consider again the example of the four portfolios with exposures on US and EM equities. In the sample, we have 3 815 observations, whereas the value taken by n_u when u is equal to 2% is 171, 161, 174 and 195 respectively. Using the estimates given in Table 12.6, we calculate the daily value-at-risk and expected shortfall of the four portfolios. The results are reported in Table 12.7. If we compare them with those obtained in Table 12.5 on page 773, we notice that the GPD VaR is close to the GEV VaR.

TABLE 12.7: Estimating value-at-risk and expected shortfall risk measures using the generalized Pareto distribution

Risk measure	α	Long US	Long EM	Long US Short EM	Long EM Short US
	99.0%	3.20%	3.42%	2.56%	3.43%
VaR	99.5%	3.84%	4.20%	2.88%	4.13%
	99.9%	5.60%	6.26%	3.80%	6.02%
	99.0%	4.22%	4.64%	3.09%	4.54%
ES	99.5%	4.97%	5.52%	3.48%	5.34%
	99.9%	7.01%	7.86%	4.62%	7.49%

12.3 Multivariate extreme value theory

The extreme value theory is generally formulated and used in the univariate case. It can be easily extended to the multivariate case, but its implementation is more difficult. This section is essentially based on the works of Deheuvels (1978), Galambos (1987) and Joe (1997).

12.3.1 Multivariate extreme value distributions

12.3.1.1 Extreme value copulas

An extreme value (EV) copula satisfies the following relationship:

$$\mathbf{C}\left(u_1^t,\ldots,u_n^t\right)=\mathbf{C}^t\left(u_1,\ldots,u_n\right)$$

for all $t>0$. For instance, the Gumbel copula is an EV copula:

$$\begin{aligned}\mathbf{C}\left(u_1^t,u_2^t\right)&=\exp\left(-\left(\left(-\ln u_1^t\right)^\theta+\left(-\ln u_2^t\right)^\theta\right)^{1/\theta}\right)\\&=\exp\left(-\left(t^\theta\left(\left(-\ln u_1\right)^\theta+\left(-\ln u_2\right)^\theta\right)\right)^{1/\theta}\right)\\&=\left(\exp\left(-\left(\left(-\ln u_1\right)^\theta+\left(-\ln u_2\right)^\theta\right)^{1/\theta}\right)\right)^t\\&=\mathbf{C}^t\left(u_1,u_2\right)\end{aligned}$$

but it is not the case of the Farlie-Gumbel-Morgenstern copula:

$$\begin{aligned}\mathbf{C}\left(u_1^t,u_2^t\right)&=u_1^tu_2^t+\theta u_1^tu_2^t\left(1-u_1^t\right)\left(1-u_2^t\right)\\&=u_1^tu_2^t\left(1+\theta-\theta u_1^t-\theta u_2^t+\theta u_1^tu_2^t\right)\\&\neq u_1^tu_2^t\left(1+\theta-\theta u_1-\theta u_2+\theta u_1u_2\right)^t\\&\neq\mathbf{C}^t\left(u_1,u_2\right)\end{aligned}$$

The term '*extreme value copula*' suggests a relationship between the extreme value theory and these copula functions. Let $X=(X_1,\ldots,X_n)$ be a random vector of dimension n. We note $X_{m:m}$ the random vector of maxima:

$$X_{m:m}=\begin{pmatrix}X_{m:m,1}\\\vdots\\X_{m:m,n}\end{pmatrix}$$

and $\mathbf{F}_{m:m}$ the corresponding distribution function:

$$\mathbf{F}_{m:m}\left(x_1,\ldots,x_n\right)=\Pr\left\{X_{m:m,1}\leq x_1,\ldots,X_{m:m,n}\leq x_n\right\}$$

The multivariate extreme value (MEV) theory considers the asymptotic behavior of the non-degenerate distribution function $\mathbf{G}$ such that:

$$\lim_{m\to\infty}\Pr\left(\frac{X_{m:m,1}-b_{m,1}}{a_{m,1}}\leq x_1,\ldots,\frac{X_{m:m,n}-b_{m,n}}{a_{m,n}}\leq x_n\right)=\mathbf{G}\left(x_1,\ldots,x_n\right)$$

Using Sklar's theorem, there exists a copula function $\mathbf{C}\langle\mathbf{G}\rangle$ such that:

$$\mathbf{G}\left(x_1,\ldots,x_n\right)=\mathbf{C}\langle\mathbf{G}\rangle\left(\mathbf{G}_1\left(x_1\right),\ldots,\mathbf{G}_n\left(x_n\right)\right)$$

It is obvious that the marginals $\mathbf{G}_1,\ldots,\mathbf{G}_n$ satisfy the Fisher-Tippet theorem, meaning that the marginals of a multivariate extreme value distribution can only be Gumbel, Fréchet or Weibull distribution functions. For the copula $\mathbf{C}\langle\mathbf{G}\rangle$, we have the following result: $\mathbf{C}\langle\mathbf{G}\rangle$ is an extreme value copula.

With the copula representation, we can then easily define MEV distributions. For instance, if we consider the random vector (X_1, X_2), whose joint distribution function is:

$$\mathbf{F}(x_1, x_2) = \exp\left(-\left((-\ln\Phi(x_1))^\theta + (-\ln x_2)^\theta\right)^{1/\theta}\right)$$

we notice that X_1 is a Gaussian random variable and X_2 is a uniform random variable. We conclude that the corresponding limit distribution function of maxima is:

$$\mathbf{G}(x_1, x_2) = \exp\left(-\left((-\ln\mathbf{\Lambda}(x_1))^\theta + (-\ln\mathbf{\Psi}_1(x_2))^\theta\right)^{1/\theta}\right)$$

In Figure 12.13, we have reported the contour plot of four MEV distribution functions, whose marginals are $\mathcal{GEV}(0,1,1)$ and $\mathcal{GEV}(0,1,1.5)$. For the dependence function, we consider the Gumbel-Hougaard copula and calibrate the parameter θ with respect to the Kendall's tau.

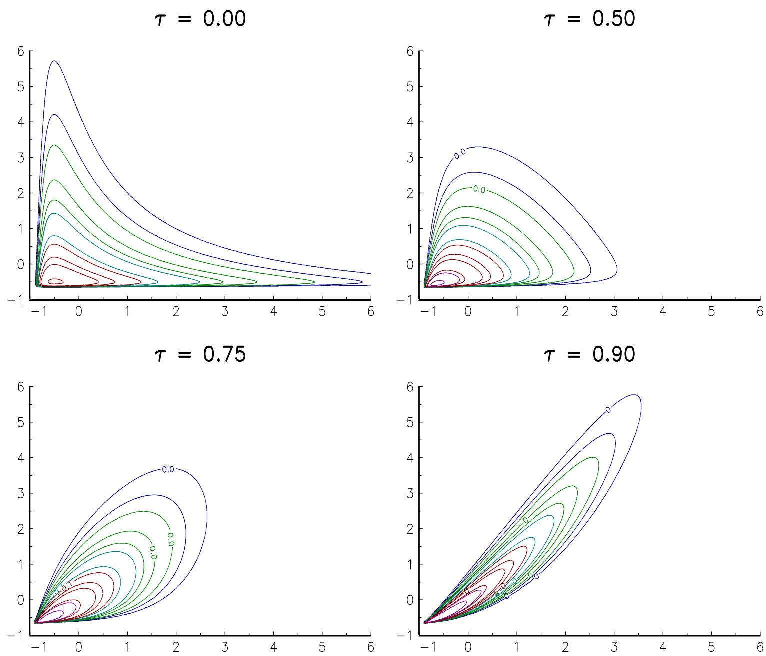

FIGURE 12.13: Multivariate extreme value distributions

12.3.1.2 Deheuvels-Pickands representation

Let $\mathbf{D}$ be a multivariate distribution function, whose survival marginals are exponential and the dependence structure is an extreme value copula. By using the relationship[15] $\mathbf{C}(u_1, \ldots, u_n) = \mathbf{C}\left(e^{-\tilde{u}_1}, \ldots, e^{-\tilde{u}_n}\right) = \mathbf{D}(\tilde{u}_1, \ldots, \tilde{u}_n)$, we have $\mathbf{D}^t(\tilde{\mathbf{u}}) = \mathbf{D}(t\tilde{\mathbf{u}})$. Therefore, $\mathbf{D}$ is a *min-stable multivariate exponential* (MSMVE) distribution.

We now introduce the Deheuvels/Pickands MSMVE representation. Let $\mathbf{D}(\tilde{\mathbf{u}})$ be a survival function with exponential marginals. $\mathbf{D}$ satisfies the relationship:

$$-\ln\mathbf{D}(t \cdot \tilde{\mathbf{u}}) = -t \cdot \ln\mathbf{D}(\tilde{\mathbf{u}}) \qquad \forall\, t > 0$$

[15]We recall that $\tilde{u} = -\ln u$.

if and only if the representation of $\mathbf{D}$ is:

$$-\ln \mathbf{D}(\tilde{\mathbf{u}}) = \int \cdots \int_{\mathcal{S}_n} \max_{1 \leq i \leq n} (q_i \tilde{u}_i) \, \mathrm{d}S(\mathbf{q}) \qquad \forall \, \tilde{\mathbf{u}} \geq \mathbf{0}$$

where $\mathcal{S}_n$ is the n-dimensional unit simplex and S is a finite measure on $\mathcal{S}_n$. This is the formulation[16] given by Joe (1997). Sometimes, the Deheuvels/Pickands representation is presented using a dependence function $B(\mathbf{w})$ defined by:

$$\begin{aligned} \mathbf{D}(\tilde{\mathbf{u}}) &= \exp\left(-\left(\sum_{i=1}^{n} \tilde{u}_i\right) B(w_1, \ldots, w_n)\right) \\ B(\mathbf{w}) &= \int \cdots \int_{\mathcal{S}_n} \max_{1 \leq i \leq n} (q_i w_i) \, \mathrm{d}S(\mathbf{q}) \end{aligned}$$

where $w_i = \left(\sum_{i=1}^{n} \tilde{u}_i\right)^{-1} \tilde{u}_i$. Tawn (1990) showed that B is a convex function and satisfies the following condition:

$$\max(w_1, \ldots, w_n) \leq B(w_1, \ldots, w_n) \leq 1 \tag{12.5}$$

We deduce that an extreme value copula satisfies the PQD property:

$$\mathbf{C}^{\perp} \prec \mathbf{C} \prec \mathbf{C}^{+}$$

In the bivariate case, the formulation can be simplified because the convexity of B and the condition (12.5) are sufficient (Tawn, 1988). We have:

$$\begin{aligned} \mathbf{C}(u_1, u_2) &= \mathbf{D}(\tilde{u}_1, \tilde{u}_2) \\ &= \exp\left(-(\tilde{u}_1 + \tilde{u}_2) B\left(\frac{\tilde{u}_1}{\tilde{u}_1 + \tilde{u}_2}, \frac{\tilde{u}_2}{\tilde{u}_1 + \tilde{u}_2}\right)\right) \\ &= \exp\left(\ln(u_1 u_2) B\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right)\right) \\ &= \exp\left(\ln(u_1 u_2) A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right) \end{aligned}$$

where $A(w) = B(w, 1-w)$. A is a convex function where $A(0) = A(1) = 1$ and satisfies $\max(w, 1-w) \leq A(w) \leq 1$.

Example 131 *For the Gumbel copula, we have:*

$$\begin{aligned} -\ln \mathbf{D}(\tilde{u}_1, \tilde{u}_2) &= \left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta} \\ B(w_1, w_2) &= \frac{\left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta}}{(\tilde{u}_1 + \tilde{u}_2)} = \left(w_1^{\theta} + w_2^{\theta}\right)^{1/\theta} \\ A(w) &= \left(w^{\theta} + (1-w)^{\theta}\right)^{1/\theta} \end{aligned}$$

[16] Note that it is similar to Proposition 5.11 of Resnick (1987), although the author does not use copulas.

We verify that a bivariate EV copula satisfies the PQD property:

$$\begin{aligned}
& \max(w, 1-w) \leq A(w) \leq 1 \\
\Leftrightarrow\quad & \max\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right) \leq A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \leq 1 \\
\Leftrightarrow\quad & \min(\ln u_1, \ln u_2) \geq \ln(u_1 u_2) \cdot A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \geq \ln(u_1 u_2) \\
\Leftrightarrow\quad & \min(u_1, u_2) \geq \exp\left(\ln(u_1 u_2) \cdot A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right) \geq u_1 u_2 \\
\Leftrightarrow\quad & \mathbf{C}^+ \succ \mathbf{C} \succ \mathbf{C}^\perp
\end{aligned}$$

When the extreme values are independent, we have $A(w) = 1$ whereas the case of perfect dependence corresponds to $A(w) = \max(w, 1-w)$:

$$\begin{aligned}
\mathbf{C}(u_1, u_2) &= \exp\left(\ln(u_1 u_2) \cdot \max\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right)\right) \\
&= \min(u_1, u_2) \\
&= \mathbf{C}^+(u_1, u_2)
\end{aligned}$$

In Table 12.8, we have reported the dependence function $A(w)$ of the most used EV copula functions.

TABLE 12.8: List of extreme value copulas

Copula	θ	$\mathbf{C}(u_1, u_2)$	$A(w)$
$\mathbf{C}^\perp$		$u_1 u_2$	1
Gumbel	$[1, \infty)$	$\exp\left(-\left(\tilde{u}_1^\theta + \tilde{u}_2^\theta\right)^{1/\theta}\right)$	$\left(w^\theta + (1-w)^\theta\right)^{1/\theta}$
Gumbel II	$[0, 1]$	$u_1 u_2 \exp\left(\theta \frac{\tilde{u}_1 \tilde{u}_2}{\tilde{u}_1 + \tilde{u}_2}\right)$	$\theta w^2 - \theta w + 1$
Galambos	$[0, \infty)$	$u_1 u_2 \exp\left(\left(\tilde{u}_1^{-\theta} + \tilde{u}_2^{-\theta}\right)^{-1/\theta}\right)$	$1 - \left(w^{-\theta} + (1-w)^{-\theta}\right)^{-1/\theta}$
Hüsler-Reiss	$[0, \infty)$	$\exp\left(-\tilde{u}_1 \vartheta(u_1, u_2; \theta) - \tilde{u}_2 \vartheta(u_2, u_1; \theta)\right)$	$w\kappa(w; \theta) + (1-w)\kappa(1-w; \theta)$
Marshall-Olkin	$[0, 1]^2$	$u_1^{1-\theta_1} u_2^{1-\theta_2} \min\left(u_1^{\theta_1}, u_2^{\theta_2}\right)$	$\max(1 - \theta_1 w, 1 - \theta_2(1-w))$
$\mathbf{C}^+$		$\min(u_1, u_2)$	$\max(w, 1-w)$

$\vartheta(u_1, u_2; \theta) = \Phi\left(\frac{1}{\theta} + \frac{\theta}{2} \ln(\ln u_1 / \ln u_2)\right)$
$\kappa(w; \theta) = \vartheta(w, 1-w; \theta)$

Source: Ghoudi *et al.* (1998).

12.3.2 Maximum domain of attraction

Let $\mathbf{F}$ be a multivariate distribution function whose marginals are $\mathbf{F}_1, \ldots, \mathbf{F}_n$ and the copula is $\mathbf{C}\langle\mathbf{F}\rangle$. We note $\mathbf{G}$ the corresponding multivariate extreme value distribution, $\mathbf{G}_1, \ldots, \mathbf{G}_n$ the marginals of $\mathbf{G}$ and $\mathbf{C}\langle\mathbf{G}\rangle$ the associated copula function. We can show that $\mathbf{F} \in \mathrm{MDA}(\mathbf{G})$ if and only if $\mathbf{F}_i \in \mathrm{MDA}(\mathbf{G}_i)$ for all $i = 1, \ldots, n$ and $\mathbf{C}\langle\mathbf{F}\rangle \in \mathrm{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$. Previously, we have seen how to characterize the max-domain of attraction in the univariate case and how to calculate the normalizing constants. These constants remains the same in the multivariate case, meaning that the only difficulty is to determine the EV copula $\mathbf{C}\langle\mathbf{G}\rangle$.

We can show that $\mathbf{C}\langle\mathbf{F}\rangle \in \mathrm{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$ if $\mathbf{C}\langle\mathbf{F}\rangle$ satisfies the following relationship:

$$\lim_{t\to\infty} \mathbf{C}^t\langle\mathbf{F}\rangle\left(u_1^{1/t},\ldots,u_n^{1/t}\right) = \mathbf{C}\langle\mathbf{G}\rangle(u_1,\ldots,u_n)$$

Moreover, if $\mathbf{C}\langle\mathbf{F}\rangle$ is an EV copula, then $\mathbf{C}\langle\mathbf{F}\rangle \in \mathrm{MDA}(\mathbf{C}\langle\mathbf{F}\rangle)$. This important result is equivalent to:

$$\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{w_1},\ldots,(1-u)^{w_n}\right)}{u} = B(w_1,\ldots,w_n)$$

In the bivariate case, we obtain:

$$\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{1-t},(1-u)^{t}\right)}{u} = A(t)$$

for all $t\in[0,1]$.

Example 132 *We consider the random vector (X_1,X_2) defined by the following distribution function:*

$$\mathbf{F}(x_1,x_2) = \left(\left(1-e^{-x_1}\right)^{-\theta} + x_2^{-\theta} - 1\right)^{-1/\theta}$$

on $[0,\infty]\times[0,1]$. The marginals of $\mathbf{F}(x_1,x_2)$ are $\mathbf{F}_1(x_1) = \mathbf{F}(x_1,1) = 1-e^{-x_1}$ and $\mathbf{F}_2(x_2) = \mathbf{F}(\infty,x_2) = x_2$. It follows that X_1 is an exponential random variable and X_2 is a uniform random variable. We know that:

$$\lim_{n\to\infty} \Pr\left(\frac{X_{n:n,1}-\ln n}{1} \le x_1\right) = \mathbf{\Lambda}(x_1)$$

and:

$$\lim_{n\to\infty} \Pr\left(\frac{X_{n:n,2}-1}{n^{-1}} \le x_2\right) = \mathbf{\Psi}_1(x_2)$$

Since the dependence function of $\mathbf{F}$ is the Clayton copula: $\mathbf{C}\langle\mathbf{F}\rangle(u_1,u_2) = \left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta}$, we have:

$$\begin{aligned}
\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{t},(1-u)^{1-t}\right)}{u} &= \lim_{u\to 0} \frac{1-(1+\theta u+o(u))^{-1/\theta}}{u} \\
&= \lim_{u\to 0} \frac{u+o(u)}{u} \\
&= 1
\end{aligned}$$

We deduce that $\mathbf{C}\langle\mathbf{G}\rangle = \mathbf{C}^{\perp}$. Finally, we obtain:

$$\begin{aligned}
\mathbf{G}(x_1,x_2) &= \lim_{n\to\infty} \Pr\left\{X_{n:n,1}-\ln n \le x_1, n\left(X_{n:n,2}-1\right)\le x_2\right\} \\
&= \mathbf{\Lambda}(x_1)\cdot\mathbf{\Psi}_1(x_2) \\
&= \exp\left(-e^{-x_1}\right)\cdot\exp(x_2)
\end{aligned}$$

If we change the copula $\mathbf{C}\langle\mathbf{F}\rangle$, only the copula $\mathbf{C}\langle\mathbf{G}\rangle$ is modified. For instance, when $\mathbf{C}\langle\mathbf{F}\rangle$ is the Normal copula with parameter $\rho<1$, then $\mathbf{G}(x_1,x_2) = \exp\left(-e^{-x_1}\right)\cdot\exp(x_2)$. When the copula parameter ρ is equal to 1, we obtain $\mathbf{G}(x_1,x_2) = \min\left(\exp\left(-e^{-x_1}\right),\exp(x_2)\right)$. When $\mathbf{C}\langle\mathbf{F}\rangle$ is the Gumbel copula, the MEV distribution becomes $\mathbf{G}(x_1,x_2) = \exp\left(-\left(e^{-\theta x_1}+(-x_2)^{\theta}\right)^{1/\theta}\right)$.

12.3.3 Tail dependence of extreme values

We can show that the (upper) tail dependence of $\mathbf{C}\langle\mathbf{G}\rangle$ is equal to the (upper) tail dependence of $\mathbf{C}\langle\mathbf{F}\rangle$:

$$\lambda^{+}\left(\mathbf{C}\langle\mathbf{G}\rangle\right)=\lambda^{+}\left(\mathbf{C}\langle\mathbf{F}\rangle\right)$$

This implies that extreme values are independent if the copula function $\mathbf{C}\langle\mathbf{F}\rangle$ has no (upper) tail dependence.

12.4 Exercises

12.4.1 Uniform order statistics

We assume that $X_1,\ldots,X_n$ are independent uniform random variables.

1. Show that the density function of the order statistic $X_{i:n}$ is:

$$f_{i:n}(x)=\frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)}x^{i-1}(1-x)^{n-i}$$

2. Calculate the mean $\mathbb{E}[X_{i:n}]$.

3. Show that the variance is equal to:

$$\operatorname{var}(X_{i:n})=\frac{i(n-i+1)}{(n+1)^2(n+2)}$$

4. We consider 10 samples of 8 independent observations from the uniform probability distribution $\mathcal{U}_{[0,1]}$:

Sample	Observation							
	1	2	3	4	5	6	7	8
1	0.24	0.45	0.72	0.14	0.04	0.34	0.94	0.55
2	0.12	0.32	0.69	0.64	0.31	0.25	0.97	0.57
3	0.69	0.50	0.26	0.17	0.50	0.85	0.11	0.17
4	0.53	0.00	0.77	0.58	0.98	0.15	0.98	0.03
5	0.89	0.25	0.15	0.62	0.74	0.85	0.65	0.46
6	0.74	0.65	0.86	0.05	0.93	0.15	0.25	0.07
7	0.16	0.12	0.63	0.33	0.55	0.61	0.34	0.95
8	0.96	0.82	0.01	0.87	0.57	0.11	0.14	0.47
9	0.68	0.83	0.73	0.78	0.27	0.85	0.55	0.57
10	0.89	0.94	0.91	0.28	0.99	0.40	0.99	0.68

   For each sample, find the order statistics. Calculate the empirical mean and standard deviation of $X_{i:8}$ for $i=1,\ldots,8$ and compare these values with the theoretical results.

5. We assume that n is odd, meaning that $n=2k+1$. We consider the median statistic $X_{k+1:n}$. Show that the density function of $X_{i:n}$ is right asymmetric if $i\leq k$, symmetric about .5 if $i=k+1$ and left asymmetric otherwise.

6. We now assume that the density function of $X_1,\ldots,X_n$ is symmetric. How are impacted the results obtained in Question 5?

12.4.2 Order statistics and return period

1. Let X and $\mathbf{F}$ be the daily return of a portfolio and the associated probability distribution. We note $X_{n:n}$ the maximum of daily returns for a period of n trading days. Using the standard assumptions, define the cumulative distribution function $\mathbf{F}_{n:n}$ of $X_{n:n}$ if we suppose that $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

2. How could we test the hypothesis $\mathcal{H}_0 : X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ using $\mathbf{F}_{n:n}$?

3. Define the notion of return period. What is the return period associated to the statistics $\mathbf{F}^{-1}(99\%)$, $\mathbf{F}_{1:1}^{-1}(99\%)$, $\mathbf{F}_{5:5}^{-1}(99\%)$ and $\mathbf{F}_{21:21}^{-1}(99\%)$?

4. We consider the random variable $X_{20:20}$. Find the confidence level α which ensures that the return period associated to the quantile $\mathbf{F}_{20:20}^{-1}(\alpha)$ is equivalent to the return period of the daily value-at-risk with a 99.9% confidence level.

12.4.3 Extreme order statistics of exponential random variables

1. We note $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$. Show that:

$$\Pr\{\boldsymbol{\tau} > t \mid \boldsymbol{\tau} > s\} = \Pr\{\boldsymbol{\tau} > t - s\}$$

   where $t > s$. Comment on this result.

2. Let $\boldsymbol{\tau}_i$ be the random variable of distribution $\mathcal{E}(\lambda_i)$. Calculate the probability distribution of $\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ and $\max(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ in the independent case. Show that:

$$\Pr\{\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) = \tau_i\} = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$$

3. Same question if the random variables $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ are comonotone.

12.4.4 Extreme value theory in the bivariate case

1. What is an extreme value (EV) copula $\mathbf{C}$?

2. Show that $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$ are EV copulas. Why $\mathbf{C}^{-}$ cannot be an EV copula?

3. We define the Gumbel-Hougaard copula as follows:

$$\mathbf{C}(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right]^{1/\theta}\right)$$

   with $\theta \geq 1$. Verify that it is an EV copula.

4. What is the definition of the upper tail dependence λ? What is its usefulness in multivariate extreme value theory?

5. Let $f(x)$ and $g(x)$ be two functions such that $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0$. If $g'(x_0) \neq 0$, L'Hospital's rule states that:

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

   Deduce that the upper tail dependence λ of the Gumbel-Hougaard copula is $2 - 2^{1/\theta}$. What is the correlation of two extremes when $\theta = 1$?

6. We define the Marshall-Olkin copula as follows:

$$\mathbf{C}(u_1, u_2) = u_1^{1-\theta_1} \cdot u_2^{1-\theta_2} \cdot \min\left(u_1^{\theta_1}, u_2^{\theta_2}\right)$$

where $(\theta_1, \theta_2) \in [0,1]^2$.

(a) Verify that it is an EV copula.

(b) Find the upper tail dependence λ of the Marshall-Olkin copula.

(c) What is the correlation of two extremes when $\min(\theta_1, \theta_2) = 0$?

(d) In which case are two extremes perfectly correlated?

12.4.5 Maximum domain of attraction in the bivariate case

1. We consider the following probability distributions:

Distribution		$\mathbf{F}(x)$
Exponential	$\mathcal{E}(\lambda)$	$1 - e^{-\lambda x}$
Uniform	$\mathcal{U}_{[0,1]}$	x
Pareto	$\mathcal{P}(\alpha, \theta)$	$1 - \left(\frac{\theta + x}{\theta}\right)^{-\alpha}$

For each distribution, we give the normalization parameters a_n and b_n of the Fisher-Tippet theorem and the corresponding limit probability distribution $\mathbf{G}(x)$:

Distribution	a_n	b_n	$\mathbf{G}(x)$
Exponential	λ^{-1}	$\lambda^{-1} \ln n$	$\mathbf{\Lambda}(x) = e^{-e^{-x}}$
Uniform	n^{-1}	$1 - n^{-1}$	$\mathbf{\Psi}_1(x-1) = e^{x-1}$
Pareto	$\theta\alpha^{-1}n^{1/\alpha}$	$\theta n^{1/\alpha} - \theta$	$\mathbf{\Phi}_\alpha\left(1 + \frac{x}{\alpha}\right) = e^{-\left(1+\frac{x}{\alpha}\right)^{-\alpha}}$

We note $\mathbf{G}(x_1, x_2)$ the asymptotic distribution of the bivariate random vector $(X_{1,n:n}, X_{2,n:n})$ where $X_{1,i}$ (resp. $X_{2,i}$) are *iid* random variables.

(a) What is the expression of $\mathbf{G}(x_1, x_2)$ when $X_{1,i}$ and $X_{2,i}$ are independent, $X_{1,i} \sim \mathcal{E}(\lambda)$ and $X_{2,i} \sim \mathcal{U}_{[0,1]}$?

(b) Same question when $X_{1,i} \sim \mathcal{E}(\lambda)$ and $X_{2,i} \sim \mathcal{P}(\theta, \alpha)$.

(c) Same question when $X_{1,i} \sim \mathcal{U}_{[0,1]}$ and $X_{2,i} \sim \mathcal{P}(\theta, \alpha)$.

2. What happen to the previous results when the dependence function between $X_{1,i}$ and $X_{2,i}$ is the Normal copula with parameter $\rho < 1$?

3. Same question when the parameter of the Normal copula is equal to one.

4. Find the expression of $\mathbf{G}(x_1, x_2)$ when the dependence function is the Gumbel-Hougaard copula.