1 Introduction

Extreme value theory is a branch of mathematics dealing with strong deviations from the mean of a probability distribution: it is used in many scientific areas, including engineering, hydrology, finance and insurance. While in classical statistics we mainly focus on the moments of the distribution of a single random variable, or maybe of the sum of random variables, in extreme value statistics we usually pay attention to the behaviour of the maximum/minimum of a set of random variables: for example, we can be interested in finding the distribution of the maximum of normal random variables.
So, in traditional statistics, central limit theorem states that, given a set $\{X_1, \ldots, X_n\}$ of independent and identically distributed random variables with finite mean and variance, the distribution of the random variable

$$\frac{X_1 + X_2 + \cdots + X_n - nE(X)}{\sqrt{nVAR(X)}}$$

is approximately standard normal when n approaches ∞.
In extreme-value statistics, Fisher-Tippet-Gnedenko theorem, known as Extremal Types theorem, affirms that the distribution of the maximum or of the minimum of n independent and identically distributed random variables, normalised by suitable constants a_n and b_n, will converge to an extreme-value distribution, which can be Gumbel, Frechet-Pareto or Weibull, if a limit distribution exists. In other words,

$$\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \xrightarrow{n\to\infty} G(x),$$

where $G(x)$ can be in one of the following two forms:

$$G(x) = \begin{cases} \exp((1+\gamma x)^{-\frac{1}{\gamma}}) \text{ if } \gamma \neq 0,\ 1+\gamma x > 0 \\ \exp(-\exp(-x)) \text{ if } \gamma = 0 \end{cases} .$$

In the case $\gamma > 0$, a Frechet distribution arises, if $\gamma = 0$, we end with a Gumbel random variable, otherwise the normalised maximum follows an extreme Weibull distribution. So we can immediately note that normal distribution does not appear anymore in that limit and this has an important practical implication: using normal distribution will lead to under-evaluation of the probability that extreme events, such as a powerful earthquake or a large insurance claim, will occur.
In this dissertation, we'll proceed in following way, using the same approach as Berlaint and alias [12]. In section 2, we'll briefly present extreme value theory in univariate case, proving the extremal types theorem introduced above. We'll also offer some criteria in order to determine which extreme-value distribution the maximum of a set of independent and identically distributed random variables is attracted to. In section 3, we'll focus on bivariate extreme value theory from a mathematical point of view: we'll introduce two fundamental measures, the exponent measure and the spectral measure, and also some coefficients in order to detect dependence both at finite and at asymptotic level. In section 4, by contrast, we'll focus on the statistics of bivariate extreme value theory. We'll present two common methods, the block maxima and the exceedances over threshold procedures, in order to determine the bivariate cumulative distribution function: both non-parametric and parametric approach, based on extreme-value copula, will be employed. Moreover, we'll review the most common test of extreme-value dependence among two random variables. A subsection on the estimation of asymptotic dependence coefficients is also given.

Finally, in section 5, we'll apply the methods of estimation of bivariate extreme-value distribution and of extremal coefficients presented in section 4 to a data-set of fire insurance claims collected by the Copenhagen Reinsurance Company: in particular, we'll use the extreme-value copulas estimated with excesses over a threshold method in order to determine the premium that the Copenhagen Reinsurance company would pay in the case it wants to reduce the riskiness of its fire insurance policies portfolio by signing a stop-loss or an excess-of-loss reinsurance treaty. An Appendix with the proofs of some theorems, lemmas, corollaries and propositions stated in the thesis is also added.
So we start by presenting univariate extreme value theory.

2 Univariate Extreme Value Theory

2.1 Extreme Value Distributions

As stated in the Introduction, we are mainly interested in the distribution of the random variable $X_{n,n} = \max(X_1, X_2 \dots X_n)$, so we start this chapter by considering the following fundamental theorem.

Theorem 2.1. *Let $\{X_i\}_{i=1}^n$ be a sequence of iid random variables with distribution function F: then there exist sequences a_n, b_n such that the distribution of the random variable*

$$\frac{X_{n,n} - b_n}{a_n}$$

converges to $G(x)$ as n approaches ∞, where

$$G(x) = \begin{cases} \exp((1+\gamma x)^{-\frac{1}{\gamma}}) \text{ if } \gamma \neq 0 \\ \exp(-\exp(-x)) \text{ if } \gamma = 0 \end{cases} . \tag{1}$$

Proof. Before going on, we briefly state the following lemma by Helly and Bray, whose proof is contained in [18] by Billingsley [18].

Lemma 2.2. *Given the random variables $Y_n \sim F_n$ and $Y \sim F$, then $Y_n \to Y$ if and only if, for all real, continuous and bounded functions z, $E[z(Y_n)] \xrightarrow{n\to\infty} E[z(Y)]$.*

So, our main target becomes simply to find the function $G(x)$ appearing in the following limit:

$$E\left[z\left(\frac{X_{n,n} - b_n}{a_n}\right)\right] \to \int_{-\infty}^{\infty} z(v) dG(v). \tag{2}$$

We write equation the following expectation in this way:

$$\begin{aligned} E\left[z\left(\frac{X_{n,n} - b_n}{a_n}\right)\right] &= \int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) d(F^n(x)) = \\ &= n\int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) F^{n-1}(x) d(F(x)) \end{aligned} \tag{3}$$

where we use the fact that, by elementary probability theory, $P[X_{n,n} \leq x] = F^n(x)$.
Let us now define the quantile function $Q(y) = F^{-1}(y) = \inf\{x : F(x) \geq y\}$ and the tail quantile function $U(y) = Q(1 - \frac{1}{y}) = x$. Since F is continuous, we can write $F(x) = 1 - \frac{v}{n}$, that implies that $x = F^{-1}(1 - \frac{v}{n}) = U(\frac{n}{v})$, by direct application of previous definitions. So we can rewrite the expectation in equation (3) in the following way:

$$\int_0^n z\left(\frac{U\left(\frac{n}{v}\right) - b_n}{a_n}\right) \left(1 - \frac{v}{n}\right)^{n-1} dv. \tag{4}$$

So, by looking at equation (4), we can see that, in order to get a limit for $E\left[z\left(\frac{X_{n,n}-b_n}{a_n}\right)\right]$, we need the quantity $\frac{U\left(\frac{n}{v}\right)-b_n}{a_n}$ to be convergent for any positive v: setting $x = n$ and $u = 1/v$, we need that there exist two auxiliary functions $a(x)$ and $b(x)$ such that $\frac{U(ux)-b(x)}{a(x)}$ is convergent for any positive value of u. Choosing $b(x) = U(x)$ for the sake of simplicity, we get the following limit condition:

$$\lim_{x\to+\infty} \frac{U(ux) - U(x)}{a(x)} = h_\gamma(u), \tag{5}$$

where $h_\gamma(u)$ is the limit function in the variable u we are looking for. We want to prove that :

$$h_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \gamma \neq 0 \\ \log x & \gamma = 0 \end{cases}. \tag{6}$$

Let us rewrite condition (5) in the following way:

$$\lim_{x\to\infty} \frac{U(xuv) - U(x)}{a(x)} = \lim_{x\to\infty} \frac{U(xuv) - U(xu)}{a(xu)} \frac{a(xu)}{a(x)} + \frac{U(xu) - U(x)}{a(x)} = h_\gamma(uv) \tag{7}$$

where u, v are positive real numbers : in order to make the term on the left to converge, we need the quantity in the middle to converge too, and in particular $\frac{a(xu)}{a(x)}$ must converge. Defining $\lim_{x\to+\infty} \frac{a(xu)}{a(x)} = g(u)$, we get the following equation:

$$\frac{a(xuv)}{a(x)} = \frac{a(xuv)}{a(xu)} \frac{a(xu)}{a(x)},$$

which is equivalent, as $x \to \infty$, to:

$$g(uv) = g(u)g(v).$$

So we end up with the canonical Cauchy Equation, and we now apply the following lemma, whose proof is described in the Appendix:

Lemma 2.3. *Any positive solution of the equation $g(uv) = g(u)g(v)$ is of the form $g(x) = x^\gamma$, with $\gamma \in \mathbb{R}$.*

So, the function $a(x) = x^\gamma l(x)$ satisfies the condition:

$$\lim_{x\to\infty} \frac{a(xu)}{a(x)} = \lim_{x\to\infty} \frac{(ux)^\gamma l(xu)}{x^\gamma l(u)} = u^\gamma$$

if and only if the following limit condition is verified:

$$\lim_{x\to\infty} \frac{l(xu)}{l(x)} = 1,$$

id est if and only if the function $l(x)$ is slowly varying at infinity or $a(x)$ is regularly varying with parameter γ.
Let us consider condition (7) again, writing it in the following way:

$$h_\gamma(uv) = h_\gamma(v)u^\gamma + h_\gamma(u). \tag{8}$$

We can easily check that the function $h_\gamma(x) = d(x^\gamma - 1)$, $d \in R$, satisfies equation (8). For sake of simplicity, we set $d = \frac{c}{\gamma}$, with $\gamma \neq 0$, ending up with

$$h_\gamma(x) = c\frac{x^\gamma - 1}{\gamma}.$$

Now we want to determine $h_\gamma(x)$ in the case of $\gamma = 0$: in order to make $h(x)$ continuous, we set h_0 equal to the limit of $h_\gamma(x)$ as $\gamma \to 0$:

$$\lim_{\gamma\to 0} c\frac{x^\gamma - 1}{\gamma} = c \log x.$$

Incorporating the constant c in the auxiliary function $a(x)$, we have just proved the definition of $h_\gamma(x)$ given in (6). From now on, we indicate $h_\gamma(x)$ simply by $h(x)$.
So, using the limit condition above, we can write the expectation in (4) as:

$$\begin{aligned} &E\left[z\left(\frac{X_{n,n}-U(x)}{a(x)}\right)\right] \xrightarrow{n\to\infty} \\ &\int_0^{+\infty} z\left(h\left(\frac{1}{v}\right)\right)\exp\left(-v\right)\,dv. \end{aligned} \tag{9}$$

Now we set $h\left(\frac{1}{v}\right)=u$, implying that, by definition of $h(x)$, if $\gamma\neq 0$,

$$v=(u\gamma+1)^{-\frac{1}{\gamma}}.$$

So we end up with the following form for the expectation as $n\to\infty$ according to the sign of the parameter γ:

$$\begin{cases} \int_{-\frac{1}{\gamma}}^{+\infty} z(u)d(\exp\left(-(1+\gamma u)^{-\frac{1}{\gamma}}\right) & \gamma>0 \\ \\ \int_{-\infty}^{+\infty} z(u)d(\exp\left(\exp\left(-u\right)\right) & \gamma=0 \\ \\ \int_{-\infty}^{-\frac{1}{\gamma}} z(u)d(\exp\left(-(1+\gamma u)^{-\frac{1}{\gamma}}\right) & \gamma<0. \end{cases} \tag{10}$$

The distribution function of each extreme value distribution can be seen in equation (10) in the argument of the differential of each integral: in the case $\gamma>0$, we get a Frechet Distribution, a Weibull Distribution arises when γ is negative, while with a null value of the parameter we end up with a Gumbel Distribution, proving exactly equation (1). $\square$

At this point, the following question arises: which are the distributions whose maximum is attracted by a Pareto? Or by a Gumbel? The answer is provided in the next subsections.

2.1.1 Pareto-Frechet Distribution

The Pareto-Frechet Distribution arises when the value of γ in definition (1) is positive. We start with strict Pareto Distribution, recalling that its survival function is given by $\bar{F}(x)=x^{-\alpha}=x^{-1/\gamma}$ and that its tail quantile function is defined as $U(x)=x^\gamma=x^{1/\alpha}$. So, by applying condition (5) and definition (6) for positive values of γ, we get:

$$\lim_{x\to\infty}\frac{U(xu)-U(x)}{a(x)}=\lim_{x\to\infty}\frac{x^\gamma}{a(x)}(u^\gamma-1)=h(u)=\frac{u^\gamma-1}{\gamma},$$

which is satisfied by $a(x)=\gamma x^\gamma$.
Now, we take a more general class of distributions with tail quantile function given by $U(x)=l_U(x)x^\gamma$, where $l_U(x)$ is clearly a slowly varying function, so that:

$$\frac{U(xu)-U(x)}{a(x)}=\frac{x^\gamma l_U(x)}{a(x)}\left(\frac{u^\gamma l_U(xu)}{l_U(x)}-1\right).$$

By definition of slowly varying function, the second term on the RHS goes to $u^\gamma-1$ as $x\to\infty$, so it is sufficient to set $a(x)=\gamma x^\gamma l_U(x)$ to satisfy limit condition (5), with $h(u)$ as in (6). Distributions with $U(x)=l_U(x)x^\gamma$ are known as Pareto-Type Distributions: they can be identified, also, by the survival function, which assumes the following form:

$$\bar{F}(x)=x^{-\frac{1}{\gamma}}l_F(x), \tag{11}$$

where $l_F(x)$ is again slowly varying, see Embrechts and alias [19] for the proof.
But there is a more general definition of Pareto-Type Distributions provided by the following theorem by Von Mises: in order to understand that definition, we need to state the Representation Theorem due to Karamata, that we prove in the appendix.

Theorem 2.4. *Let $l(x)$ be a slowly-varying function: then there exist two auxiliary functions $\epsilon(x)$ and $c(x)$ such that*

$$l(x) = c(x) \exp\left(\int_1^x \frac{\epsilon(u)}{u}\, du\right),$$

where $c(x) \to c$ and $\epsilon(x) \to 0$ as $x \to \infty$.

Before going on, we need to state the following corollary of theorem 2.3, whose proof is contained again in the appendix: this corollary will be used later in section 3.

Corollary 2.4.1. *Let $l(x)$ be a s.v. function, then:*

$$\lim_{x\to\infty} \frac{\log(l(x))}{\log(x)} = 0.$$

We can now state the new definition of Pareto-Type Distributions through the following theorem.

Theorem 2.5. *Let $X_1 \dots X_n$ be a sequence of independent and identically distributed random variables with $X_i \sim F$ $\forall i = 1 \dots n$ and $\inf\{x : F(x) = 1\} = +\infty$. Let $r(x) = \frac{f(x)}{1-F(x)}$ be the hazard function of X. Then F is in the domain of attraction of the Pareto Type Distribution, id est $F \in D(\phi_{\frac{1}{\gamma}})$ with parameter $\frac{1}{\gamma}$ if and only if*

$$\lim_{x\to\infty} xr(x) = \frac{1}{\gamma}.$$

Proof. Let us define $\alpha = \frac{1}{\gamma}$ and $\epsilon(x) = xr(x) - \alpha$, which implies that $r(x) = -\frac{\partial \log(1-F(x))}{\partial x} = \frac{\epsilon(x)+\alpha}{x}$. Using the fact that $P[X > x] = P[X > 1]P[X > x|X > 1],\ x > 1$ and integrating both sides, we get:

$$1 - F(x) = (1 - F(1)) \exp\left(-\int_1^x \frac{\epsilon(u) + \alpha}{u}\, du\right) = (1 - F(1)) \exp\left(-\int_1^x \frac{\epsilon(u)}{u}\, du\right) x^{-\alpha}.$$

By Theorem 2.4, the function $(1 - F(1)) \exp\left(-\int_1^x \frac{\epsilon(u)}{u}\, du\right)$ is slowly varying, implying that the survival function is the product of a power function and a s.v. function, that is exactly definition (11) of Pareto Type Distributions stated above: this proves the theorem. □

T-Distributions and Fisher Distributions belong, for example, belong to the Pareto-Type-Distributions. Let X be a T-Student Random Variable with k degrees of freedom. Then the density and the survival functions are given respectively by:

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

$$1 - F(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{k\sqrt{\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k}{2}}.$$

Then it is not difficult to check that $\lim_{x\to\infty} xr(x) = k$, meaning that X is Pareto-Type Distributed with parameter $\alpha = \frac{1}{\gamma} = k$.

By the same way of reasoning, it can be proven that $X \sim F(m,n)$, where $F(m,n)$ indicates a Fisher distribution with parameters m and n with density given by

$$f(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} x^{\frac{m}{2}-1} \left(1 + x\frac{m}{n}\right)^{-\frac{m+n}{2}} \quad x > 0,$$

is Pareto-Type Distributed with parameter $\alpha = \frac{n}{2}$.
Finally, log-gamma distribution belongs to this domain of attraction too, see Embrechts and alias [19] for the proof.

2.1.2 Weibull Distribution

In this section, we'll study the case in which γ is negative, when a Weibull Distribution arises in the asymptotic behaviour of the normalised maximum of a sequence of random variables: in particular, we'll show that there is a strict relationship between the domain of attraction of the Pareto-Frechet and of the Weibull Distributions.
Let us start this section by stating and proving the following theorem on Weibull-Type Random Variables.

Theorem 2.6. *The random variable $X \sim F$ is Weibull-Type, or $F \in D(\Psi_{-\frac{1}{\gamma}})$, with $\gamma < 0$, if and only if, as x approaches the upper bound of the support of X, the survival function and the tail quantile function can be written in the following form:*

$$1 - F\left(M - \frac{1}{x}\right) = x^{\frac{1}{\gamma}} l_F(x) \tag{12}$$

$$U(x) = M - x^{\gamma} l_U(x)$$

where $M = \inf\{x : F(x)\} = 1$ and $l_F(x), l_U(x)$ are slowly-varying.

Proof. For sake of simplicity, we prove this theorem assuming that $l_F(x) = l_U(x) = 1\ \forall x$. It is not difficult to check that $1 - F(x) = (M-x)^{-\frac{1}{\gamma}}$; moreover, since $U(x) = M - x^{\gamma}$, the function $a(x) = -\gamma x^{\gamma}$ satisfies condition (5) with $h(x)$ as in (6). So:

$$P\left[\frac{X_{n,n} - U(n)}{a(n)} \le x\right] = P\left[\frac{X_{n,n} - M + n^{\gamma}}{-\gamma n^{\gamma}} \le x\right] = 1 - P\left[X_{n,n} \le M - \gamma x n^{\gamma} - n^{\gamma}\right].$$

By applying equation (12), we can rewrite the previous probability, when x is sufficiently large, as:

$$1 - \left[1 - \frac{1}{n}(x\gamma + 1)^{-\frac{1}{\gamma}}\right]^n,$$

which, as n approaches ∞, is equivalent to:

$$1 - \exp\left(-(\gamma x + 1)^{-\frac{1}{\gamma}}\right),$$

which is exactly the cumulative distribution function of an Extreme Weibull Distribution. □

An alternative condition to define Weibull type-Distributions is through hazard function, as stated in the following Theorem by Von Mises:

Theorem 2.7. *Let us define $X \sim F$ and $M = \inf\{x : F(x)) = 1\} < \infty$. $F \in D(\Psi_{\alpha})$ if and only if $\lim_{x \to M}(M-x)r(x) = \alpha$, where $\alpha = -\frac{1}{\gamma}$ and $r(x)$ the hazard function of X.*

Proof. Let us define $\epsilon(x) = (M-x)r(x) - \alpha$, implying that $r(x) = -\frac{\partial \log(1-F(x))}{\partial x} = \frac{\epsilon(x)+\alpha}{M-x}$; integrating both sides, we easily get:

$$\begin{aligned} 1-F(x) &= (1-F(1)) \exp\left(-\int_1^x \frac{\epsilon(u)+\alpha}{M-u}\, du\right) = \\ &= (1-F(1)) \exp\left(-\int_1^x \frac{\epsilon(u)}{M-u}\, du\right)(M-1)^{-\alpha}(M-x)^{\alpha}. \end{aligned} \tag{13}$$

Using the same procedure adopted to prove theorem 2.4, it is not difficult to check that $(1-F(1)) \exp\left(-\int_1^x \frac{\epsilon(u)}{M-u}\, du\right)(M-1)^{-\alpha}$ is s.v.: then the survival function satisfies condition (12), so X is Weibull-Type Distributed. $\square$

Uniform distribution, for example, is in the Weibull domain of attraction: in fact, its survival function $1-F\left(1-\frac{1}{x}\right) = x^{-1}$ clearly satisfies condition (12) with $M=1$,$\gamma=-1$ and $l_F(x)=1$. But also Beta Distribution belongs to this domain of attraction. Let Y be Beta Distributed with parameters (p,q) and with density given by: $f(y) = \frac{y^{p-1}(1-y)^{q-1}}{B(m,n)}$. We want to compute the limit $\lim_{x\to M}(M-x)r(x)$, which, in the case of Beta Density, corresponds to:

$$\lim_{y\to 1} \frac{\frac{(1-y)^q y^{p-1}}{B(p,q)}}{1-\int_0^y \frac{(1-t)^{q-1}t^{p-1}}{B(p,q)}\, dt}$$

Applying De L'Hopital rule, the result of the limit is equal to q, implying that $Y \in D(\Psi_q)$, by direct application of Theorem 2.7 above.
Now we conclude this section by considering the relationship between Frechet-Pareto Case and Weibull case.

Theorem 2.8. *Suppose that $X \sim F$, where $F \in D(\Psi_\alpha)$. Moreover, let us define $Y = \frac{1}{M-X} \sim G$: then the distribution G is in Frechet domain of attraction, $G \in D(\phi_\alpha)$.*

Proof. By assumption, $F_X \in D(\Psi_\alpha)$. Then $1-F_X\left(M-\frac{1}{x}\right) = P\left[\frac{1}{M-x} \geq x\right] = P[Y \geq x] = 1-F_Y(x)$, which is equivalent to $x^{\frac{1}{\gamma}} l_F(x)$ by condition (12), so $Y \in D(\phi_\alpha)$ by equation (11). This proves the relationship between the two domain of attractions. $\square$

2.1.3 Gumbel Distribution

In this subsection, we take a look at the domain of attraction of Gumbel Distribution. Let us start characterising Gumbel-Type Variables by stating and proving the following theorem.

Theorem 2.9. *Let Y be a random variable with distribution F: then $F \in D(\Lambda)$ if and only if, as y approaches the upper bound of the support of Y,called M, the following limits holds:*

$$\frac{1-F(y+vb(y))}{1-F(y)} \to \exp(-v) \tag{14}$$

$$\frac{b(y+vb(y))}{b(y)} \to 1 \tag{15}$$

Proof. We set $U(xu) - U(x) = k(u,x)a(x)$, where, by condition (5), $k(u,x) \to h(u)$. Then it is not difficult to check that

$$\frac{1}{u} = \frac{1-F(U(x)+k(u,x)a(x))}{1-F(U(x))} = \frac{1-F\{y+\tilde{k}(y,u)a(U^{-1}(y))\}}{1-F(y)},$$

with $U(x) = y$. Setting $a(U^{-1}(y)) = b(y)$, we get:

$$\frac{1 - F(y + \tilde{k}(y,u)b(y))}{1 - F(y)}.$$

Finally, setting $v = \log(u) = h(u)$, since $\tilde{k}(u,y) \to \log u = v$, we obtain:

$$\lim_{y \to M} \frac{1 - F(y + \tilde{k}(y,u)b(y))}{1 - F(y)} = \lim_{y \to M} \frac{1 - F(y + vb(y))}{1 - F(y)} = \exp(-v),$$

which proves limit (14), with $b(y) = a(U^{-1}(y))$. To prove limit (15), we make previous substitutions in reverse order, getting:

$$\lim_{y \to M} \frac{b(y + b(y)v)}{b(y)} = \lim_{x \to \infty} \frac{a\{U^{-1}(U(x) + va(x))\}}{a(x)} = \lim_{x \to \infty} \frac{a(xu)}{a(x)} = 1$$

where we use the fact that $a(x)$ is regularly varying with parameter γ, as stated in section 2.1, with $\gamma = 0$ in this context. So we have proved both limits. □

However, the direct application of the previous proposition in order to check whether a distribution belongs to this domain of attraction is quite difficult: so, we follow the approach by Embrechts and alias [19], introducing the following useful theorem again by Von Mises.

Theorem 2.10. *Let $X \sim F$ and $M = \inf\{x : F(x) = 1\}$. Suppose that there exists a real number $z < M$ such that:*

$$\bar{F}(x) = c \exp\left(-\int_z^x \frac{1}{a(t)}\, dt\right), \tag{16}$$

with $a(x)$ continuous, differentiable and such that : $\lim_{x \to M} a^{'}(x) = 0$. Then $F \in D(\Lambda)$.

Proof. If (16) holds, then:

$$\frac{\bar{F}(x + ta(x))}{\bar{F}(x)} = \exp\left(-\int_x^{x+ta(x)} \frac{1}{a(u)}\, du\right).$$

Setting $v = \frac{u-x}{a(x)}$, we get:

$$\exp\left(-\int_0^t \frac{a(x)}{a(x + v(a(x))}\, dv\right)$$

But, since $a^{'}(x) \to 0$ as $x \to M$ by assumption, then, for a given $\epsilon > 0$, there exists a value $x_0(\epsilon)$ such that

$$|a((x) + va(x)) - a(x)| = \int_x^{x+va(x)} a^{'}(s)\, ds \leq \epsilon |v| a(x) \leq \epsilon |t| a(x),\ x > x_0(\epsilon)$$

implying that, for any x larger than $x_0(\epsilon)$:

$$a(x) \left| \frac{a(x + va(x))}{a(x)} - 1 \right| \leq \epsilon |t|.$$

Clearly we can make the RHS as small as we wish, then it is easy to notice that a sufficient condition in order to satisfy the inequality above for any ϵ is

$$\lim_{x \to M} \frac{a(x)}{a(x + va(x))} = 1,$$

implying:

$$\lim_{x \to M} \frac{\bar{F}(x + t(a(x))}{\bar{F}(x)} = \exp\left(-\int_0^t 1\, dv\right) = \exp(-t),$$

meaning that $F \in D(\Lambda)$ by direct application of condition (14). □

Now we state the following corollary, whose proof is described in the appendix:

Corollary 2.10.1. *Let $X \sim F$, where F is a distribution function twice differentiable on an interval (z, M), where $M = \inf\{x : F(x) = 1\}$. Then $F \in D(\Lambda)$ if and only if the following limit holds:*

$$\lim_{x \to M} \frac{\bar{F}(x)F''(x)}{f^2(x)} = -1 \tag{17}$$

Many distributions belong to Gumbel domain of attraction: we start by considering Exponential distribution with survival function given by $\bar{F}(x) = \exp(-\lambda x)$. It is not difficult to check that $a(x) = \frac{1}{\lambda}$ satisfies equation (16) and fulfils all conditions stated in the theorem 2.10, so Exponential Distribution belongs to the Gumbel Domain, as well as Weibull Distribution. In fact, the survival function of a Weibull is given by:

$$\bar{F}(x) = \exp(-cx^{\tau}) \ \tau, c > 0$$

so we need to find $a(x)$ such that :

$$\exp\left(-\int_0^x \frac{1}{a(u)}\, du\right) = \exp(-cx^{\tau}).$$

By taking logarithms and by differentiating both sides, we can see that a possible solution is given by $a(x) = (c\tau)^{-1}x^{-\tau+1}$, that satisfies all the conditions in Theorem 2.10 when τ and c are positive, as previously assumed.
Finally, also normal distribution belongs to Gumbel Domain. In fact, it is not difficult to check that, by De L'Hopital rule, $\lim_{x\to\infty} \frac{\bar{F}(x)}{f(x)} = \frac{1}{x}$, implying that, asymptotically, $\bar{F}(x) \sim \frac{f(x)}{x}$. Moreover, $F''(x) = -xf(x) < 0$. Then:

$$\lim_{x\to\infty} \frac{\bar{F}(x)\bar{F}''(x)}{f^2(x)} = -\lim_{x\to\infty} \frac{\frac{f(x)}{x} x f(x)}{f^2(x)} = -1,$$

so Normal Distribution is in Gumbel Domain by Corollary 2.10.1.
Gamma, Logistic and Log-normal Distributions belong to this domain of attraction as well, see Embrechts and alias [19].

2.2 Tail Estimation Methods

In this section, we'll present some methods in order to estimate the shape parameter γ from the data: we'll start by considering methods for Pareto-Type Models, with strict positive values of γ, and then we'll extend the analysis to methods valid for any value of γ.

2.2.1 Hill Estimator

Hill estimator can be used to estimate γ only when we're in the Frechet-Pareto Domain of attraction. By equation (11), it is not difficult to check that:

$$\lim_{t\to\infty} \frac{\bar{F}(xt)}{\bar{F}(t)} = x^{-\frac{1}{\gamma}}.$$

This is equivalent to say that, asymptotically,

$$P[X > xt | X > t] = P\left[\frac{X}{t} > x | X > t\right] = x^{-\frac{1}{\gamma}}.$$

So we basically associate to the random variable of relative excesses, $\frac{X}{t}$, with $X > t$, a strict Pareto Distribution with parameter $\frac{1}{\gamma}$. Let us denote $Y = \frac{X}{t}$,with t sufficiently large, then:

$$f_Y(y) = \frac{1}{\gamma} y^{-1-\frac{1}{\gamma}}, \ y > 1.$$

Then it is not difficult to check that the maximum likelihood estimator of γ is given by:

$$\gamma^{MLE} = \frac{\sum_{i=1}^{N} \log{(Y_i)} 1_{Y_i > 1}}{\sum_{i=1}^{N} 1_{Y_i > 1}}.$$

Now we choose as threshold t the $N-k$ largest observation of the order statistics $X_1, X_2 \dots X_N$, with $N > k$, getting:

$$\gamma^{HILL} = \frac{\sum_{i=1}^{k} (\log{(X_{N-i+1})} - \log{(X_{N-k})})}{k}$$

This estimator is consistent for γ, if $k, N \to \infty$ and $\frac{k}{N} \to 0$, as showed by Gunst and alias [22]. However, it has many drawbacks.
First of all, it is not constant over k, so it is necessary to find a way to choose the parameter carefully: this is done in Berlaint and alias [12]. Moreover, sometimes, assuming a strict Pareto Distribution for the relative excesses may be too optimistic: maybe, the survival function can be in the form of $\bar{F}(x) = x^{-\frac{1}{\gamma}} l_F(x)$, with $l_F(x) \neq 1$. This can make the Hill Estimator biased: a discussion of the asymptotic bias and variance and of the methods to reduce this bias is performed again in [12].
We move now to estimation methods valid for any real value of γ.

2.2.2 Block Maxima Method

The first method we discuss in order to estimate γ in the general case where it is a real number is the Block Maxima method: given m samples $(X_{1,1}, X_{1,2} \dots X_{1,n}) \dots (X_{m,1}, X_{m,2} \dots X_{m,n})$, we take the maximum of each sample, getting the set of maxima $(Y_1, \dots Y_m)$, where $Y_i = \max(X_{i,1} \dots X_{i,n})$. By Theorem 2.1, we know that the normalised maximum follows a generalised extreme value distribution with density function given by:

$$g(y) = \exp\left(-\left(1+\gamma\frac{y-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}} \left(1+\gamma\frac{y-\mu}{\sigma}\right)^{-\frac{1}{\gamma}-1} \frac{1}{\sigma} \text{ if } \gamma \neq 0, \ 1+\gamma\frac{y-\mu}{\sigma} > 0$$

and

$$g(y) = \exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right) \exp\left(-\frac{y-\mu}{\sigma}\right) \frac{1}{\sigma} \text{ if } \gamma = 0.$$

Then it is not difficult to understand that the log-likelihood function is given in the former case by:

$$\log(L(\mu,\sigma,\gamma)) = -\sum_{i=1}^{m}\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right)^{-\frac{1}{\gamma}} - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{m}\log\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right) - m\log(\sigma),$$

and in the latter one by:

$$\log{(L(\sigma,0,\mu))} = -m\log{(\sigma)} - \sum_{i=1}^{m}\exp\left(-\frac{Y_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\frac{Y_i-\mu}{\sigma}.$$

The maximum likelihood estimators of (μ, σ, γ) are obtained by maximising the log-likelihood functions: analytical solutions can not be computed, so it is necessary to implement numerical procedures, see Hosking for further details[14]. One of the problems we notice is that the support of Y depends on the value of the parameters in the case $\gamma \neq 0$, so consistency of the maximum likelihood estimators is not guaranteed: however, Smith [23] has proved that consistency holds in the case $\gamma > -0.5$.
Now we present the method of probability-weighted moments, that are the quantities $M_{p,r,s} = E[Y^p F^r(Y)(1-F(y))^s]$, for real parameters p, r, s. Assuming that $\gamma \neq 0$, we set $s = 0$ and $p = 1$, getting:

$$M_{1,r,0} = \int_{\mu-\frac{\sigma}{\gamma}}^{\infty} y \exp\left(-(r+1)\left(1+\gamma\frac{y-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right)\left(1+\gamma\frac{y-\mu}{\sigma}\right)^{-\frac{1}{\gamma}-1}\frac{1}{\sigma}\,dy.$$

Performing the change of variables $z = y - \mu + \frac{\sigma}{\gamma}$ and $t = \frac{\sigma}{\gamma}$, we get:

$$M_{1,r,0} = \int_0^{\infty} (\mu - \frac{\sigma}{\gamma}(1-t)) \exp\left(-(r+1)\,(t)^{-\frac{1}{\gamma}}\right) t^{-\frac{1}{\gamma}-1}\frac{1}{\gamma}\,dy;$$

using the property of Weibull distribution, it is not difficult to show that :

$$M_{1,r,0} = \frac{\mu + \frac{\sigma}{\gamma}[(r+1)^{\gamma}\Gamma(1-\gamma) - 1]}{r+1}, \quad \gamma < 1. \tag{18}$$

In order to derive the estimators, it is necessary to solve equation (18) for $r = 0, 1, 2$, getting the following system:

$$\begin{cases} M_{1,0,0} = \mu - \frac{\sigma}{\gamma}(1 - \Gamma(1-\gamma)) \\ 2M_{1,1,0} - M_{1,0,0} = \frac{\sigma}{\gamma}\Gamma(1-\gamma)(2^{\gamma} - 1) \\ \frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^{\gamma}-1}{2^{\gamma}-1}, \end{cases} \tag{19}$$

where $M_{1,r,0}$ is then replaced by its empirical version given by :

$$\bar{M}_{1,r,0} = \frac{\sum_{j=1}^{m}\left(\frac{j}{m+1}\right)^r Y_{j,m}}{m}.$$

The last equation in order to find γ must be solved numerically, while the other two equations can be solved also analytically given the value of γ. Defining $\bar{\gamma}$ as the value that satisfies the last equation in the system (19), we have:

$$\bar{\sigma} = \frac{\bar{\gamma}(2\bar{M}_{1,1,0} - \bar{M}_{1,0,0})}{\Gamma(1-\bar{\gamma})(2^{\bar{\gamma}} - 1)}$$

and

$$\bar{\mu} = \bar{M}_{1,0,0} + \frac{\bar{\sigma}}{\bar{\gamma}}(1 - \Gamma(1-\bar{\gamma})).$$

Now we want to derive the limiting distributions of $(\bar{\sigma}, \bar{\gamma}, \bar{\mu})$. We Define $M = (M_{1,0,0}, M_{1,1,0}, M_{1,2,0})$ and $\bar{M} = (\bar{M}_{1,0,0}, \bar{M}_{1,1,0}, \bar{M}_{1,2,0})$;
then it is possible to show that, if $\gamma < 0.5$:

$$\sqrt{m}(\bar{M} - M) \xrightarrow{m\to\infty} N(0, V), \tag{20}$$

where V is the inverse of Fisher Information Matrix whose elements depend on r, s, γ and σ.
Now we apply delta method, according to which, if $\sqrt{n}(X_n - \theta) \to N(0, \sigma^2)$, then

$$\sqrt{n}(g(X_n) - g(\theta)) \to N(0, \sigma^2 (g^{'})^2(\theta)),$$

with $\theta \in \mathbb{R}$. Let us call $f : \mathbb{R}^3 \to \mathbb{R}^3$ the function such that $f(M) = (\sigma, \gamma, \mu)$; if (20) holds, then:

$$\sqrt{m}(f(\bar{M}) - f(M)) \to N(0, gVg^T),$$

where g is obviously the 3×3 matrix of partial derivatives of f with respect to the vector $M = (M_{1,0,0}, M_{1,1,0}, M_{1,2,0})$. So:

$$\sqrt{m}((\bar{\sigma}, \bar{\gamma}, \bar{\mu}) - (\sigma, \gamma, \mu)) \to N(0, gVg^T). \tag{21}$$

To sum up, the choice of MLE or Probability-Weighted Moments Method depends on the initial starting guess for γ: if γ is deeply negative, the latter method is recommended, if γ is strongly positive, the maximum likelihood estimation is definitely better, while if $-0.5 < \gamma < 0.5$, both methods can be implemented.
Block Maxima method can be used also to construct confidence intervals for the parameters: in particular, we focus on the role of γ, since, in section 5, a confidence interval for the shape parameter of the distribution of the minimum of two random variables with extreme-value distributions will be built up.
Let us call the asymptotic variance in (21) V_2, then it is not difficult to understand that the second element on the main diagonal of V_2, $v_{2,2}$, is just the asymptotic variance of γ, so:

$$\sqrt{m}(\bar{\gamma} - \gamma) \to N(0, v_{2,2}).$$

So a confidence interval with level of confidence $1 - \alpha$ for γ is given by:

$$\left[\bar{\gamma} - \sqrt{\frac{v_{2,2}}{m}}\Phi^{-1}\left(1 - \frac{1}{2}\alpha\right), \bar{\gamma} + \sqrt{\frac{v_{2,2}}{m}}\Phi^{-1}\left(1 - \frac{1}{2}\alpha\right)\right],$$

where $\Phi^{-1}(x)$ is the quantile function of a standard normal distribution evaluated in x. However, the rate of convergence of equation (21) is quite slow, so, for small samples, normal approximation can be quite bad.
This is the reason why we usually prefer confidence intervals based on profile likelihood, that is defined as:

$$L_p(\gamma) = \max_{\sigma,\mu|\gamma} L(\sigma, \gamma, \mu)$$

So it is possible to test $H_0 : \gamma = \gamma_0$ and $H_1 : \gamma \neq \gamma_0$ by considering the profile likelihood ratio, $\Lambda = \frac{L_p(\gamma_0)}{L_p(\bar{\gamma})}$, which, according to Wilks Theorem , follows, asymptotically, a Chi-Square Distribution with one degree of freedom. In general, it is quite clear that the null hypothesis will be rejected if Λ is small, or, equivalently, when $-2\log(\Lambda)$ is large: so, given a significance level α, the critical region will be given by:

$$R = \left(\gamma : \log(L_p(\gamma)) < \log(L_p(\bar{\gamma})) - \frac{Q_{X_1^2}(1-\alpha)}{2}\right),$$

where $Q_{X_1^2}(1-\alpha)$ is the quantile function of a Chi-Square Distribution with one degree of freedom. The log-likelihood ratio is much more robust than normal approximation and it is strongly preferable in empirical applications.
However, one of the problems of Block Maxima Method is that only the maximum of the sample is used, implying waste of data: this is the reason why we present another method in the following subsection.

2.2.3 Peaks Over Threshold Method

In this section, we'll present another way to estimate the parameter γ from the data, the Peak Over Threshold (POT) method, by analysing the distribution of the exceedances of a random sample over a given threshold t: so given a sample $\{X_1, X_2, \dots X_N\}$, we'll focus on the sub-sample $\{Y_1, Y_2 \dots Y_{N_t}\}$, where $Y_i = X_i - t$ if $X_i > t$ and $N_t = \sum_{i=1}^{N} 1_{X_i > t}$. First of all, we can notice that:

$$P[Y > yb(t)|Y > 0] = P\left[\frac{X-t}{b(t)} > y|X > t\right] = \frac{\bar{F}_X(t + yb(t))}{\bar{F}_X(t)}.$$

If $\gamma = 0$, applying limit (14),

$$\lim_{t\to\infty} \frac{\bar{F}_X(t + yb(t))}{\bar{F}_X(t)} = \exp(-y).$$

If $\gamma \neq 0$, it is possible to prove, using a technique very similar to the one used to prove (14), that, when t is sufficiently large,

$$\frac{\bar{F}_X(t + yb(t))}{\bar{F}_X(t)} = (1 + \gamma y))^{-\frac{1}{\gamma}}.$$

We can now simplify a little bit by assuming that $b(t)$ is constant and equal to σ. Then we get:

$$P[Y < y|Y > 0] = 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{1}{\gamma}} \text{ if } \gamma \neq 0, 1 + \frac{\gamma y}{\sigma} > 0 \tag{22}$$

and

$$P[Y < y|Y > 0] = 1 - \exp\left(-\frac{y}{\sigma}\right) \text{ if } \gamma = 0. \tag{23}$$

So the distribution of the exceedances over t follows a generalised extreme value distribution, motivating the use of POT method.
Another justification of Peak Over Threshold Method can be found in the point process characterisation of the exceedances. Let us state the following proposition:

Proposition 2.1. *Let $\{X_1, X_2 \dots X_n\}$ be an iid sequence of random variables. Suppose that $P[X_{n,n} < u] \to \exp(-\tau)$, where $X_{n,n} = \max\{X_1 \dots X_n\}$, then:*

$$n\bar{F}(u) \to \tau \tag{24}$$

The proof is given in the appendix. Now we state and prove the following theorem.

Theorem 2.11. *Let X_i be a sequence of iid random variables satisfying proposition 2.1 and let $N_n(\cdot) = \sum_{i=1}^{N} 1_{\frac{i}{n}\in\{\cdot\}} 1_{X_i > t}$ be a point process with threshold t and $N(\cdot)$ be a homogeneous Poisson process on $[0, 1]$, with measure given by*

$$\tau|\cdot|. \tag{25}$$

Let A_i be an interval in the form of $[a_i, b_i]$, where $0 < a_i < b_i < 1$. If, for any unions of mutually disjoint intervals $B = \bigcup_{i=1}^{k} A_i = \bigcup_{i=1}^{k} [a_i, b_i] \subseteq [0, 1]$

$$P(N_n(B) = 0) = P(N(B) = 0), \tag{26}$$

and if, for any A_i,

$$E[N_n(A_i)] = E[N(A_i)], \tag{27}$$

then $N_n(\cdot) \to N(\cdot)$.

Proof. First of all, It is not difficult to understand that $N_n(A) = \sum_{i=\lfloor na+1 \rfloor}^{\lfloor nb \rfloor} 1_{X_i > t}$, where $A = [a, b]$. Moreover, we notice that $N_n(A) = \text{Bin}(\lfloor nb \rfloor - \lfloor na \rfloor, \bar{F}(t))$. Moreover, by equation (24),

$$E[N_n(A)] \to \frac{\tau}{n}(\lfloor nb \rfloor - \lfloor na \rfloor) = \tau(b - a) = \tau|A|,$$

that is exactly the measure of the Poisson Process: this proves condition (27). Now we need to prove condition (26). We take a subset of B, A_j, then:

$$P(N_n(A_j) = 0) = F^{\lfloor b_j n \rfloor - \lfloor a_j n \rfloor}(t) = \exp\left\{(\lfloor nb_j \rfloor - \lfloor na_j \rfloor)(\log(1 - \bar{F}(t))\right\},$$

which, by equation (24) and by Mclaurin expansion of $\log(1 - x)$, converges to:

$$\exp\left(-(\lfloor nb_j \rfloor - \lfloor na_j \rfloor)\left(\frac{\tau}{n}\right)\right) = \exp\left(-\tau(b_j - a_j)\right).$$

Using independence of the sequence $\{X_i\}$ and disjointness of the sets A_j, we get: $P[N_n(B) = 0] = \prod_{j=1}^{k} \exp\left(-\tau(b_j - a_j)\right) = \exp\left(-\tau \sum_{j=1}^{k} b_j - a_j\right)$. But $P[N(B) = 0] = \exp\left(-\tau|B|\right) = \exp\left(-\tau \sum_{j=1}^{k} b_j - a_j\right)$, so we have proved condition (26). □

So we can define the two-dimensional process:

$$P_N = \left\{\left(\frac{i}{N+1}; \frac{X_i - b_n}{a_n}\right) i = 1 \dots N\right\},$$

where a_n and b_n are the normalising sequences of $X_{n,n}$, which converges to a two-dimensional Poisson Process by theorem 2.11. Since $P[\frac{X_{n,n} - b_n}{a_n} \leq x] \to \exp\left(-(1 + \gamma x)^{-\frac{1}{\gamma}}\right)$, then the intensity measure of the limiting Poisson process on the set $A_i = (a_i, b_i)$ is given by:

$$\Lambda(A_i) = (b_i - a_i)(1 + \gamma x)^{-\frac{1}{\gamma}}. \tag{28}$$

Setting $A_{x+t} = (0, 1) \times (t + x, \infty)$ and $A_t = (0, 1) \times (t, +\infty)$, we have, for a sufficiently large t,

$$P\left[\frac{X_i - b_n}{a_n} > t + x | \frac{X_i - b_n}{a_n} > t\right] \sim \frac{\Lambda(A_{x+t})}{\Lambda(A_t)} = \left(1 + \frac{\gamma x}{1 + \gamma t}\right)^{-\frac{1}{\gamma}},$$

ending up again with a GP distribution as before, where $b(t) = 1 + \gamma t$. So this is another justification of POT method.
We now present different ways of estimation within POT methodology. As usual, we start with the simplest one, maximum likelihood. From equation (22), it is not difficult to see that log-likelihood function, if γ is non-zero, is given by:

$$\log\left(L(\sigma, \gamma)\right) = -N_t \log \sigma - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{N_t} \log\left(1 + \frac{\gamma Y_i}{\sigma}\right),$$

while, if $\gamma = 0$, from equation (23), it is equal to:

$$\log L(\sigma, 0) = -N_t \log(\sigma) - \frac{1}{\sigma} \sum_{i=1}^{N_t} Y_i.$$

We focus only on the case $\gamma \neq 0$, since it is quite trivial to recover maximum likelihood estimator for σ in the case $\gamma = 0$. Setting $\tau = \frac{\gamma}{\sigma}$, it is possible to find estimates of γ and σ by setting partial derivatives equal to 0 as usual: after determining, numerically, the value of τ, say $\bar{\tau}$, that satisfies

$$\frac{1}{\tau} - \left(\frac{1}{\tau} + 1\right) \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{Y_i}{1 + \tau Y_i} = 0,$$

it is not difficult to show that:

$$\bar{\gamma} = \frac{\sum_{i=1}^{N_t} \log\left(1 + \bar{\tau} Y_i\right)}{N_t}.$$

Another way to implement POT method is probability-weighted moments technique. We recall from the previous subsection that the quantity $M_{p,r,s}$ is defined as:

$$M_{p,r,s} = E[Y^p F^r(Y)(1 - F(Y))^s].$$

If we set $p = 1, 2$ and $r = s = 0$, we recover a system of equations whose solutions correspond to method of moments estimators, which are:

$$\bar{\gamma} = \frac{1}{2}\left(1 - \frac{\bar{Y}}{\bar{S}^2(Y)}\right)$$

and

$$\bar{\sigma} = \frac{\bar{Y}}{2}\left(1 + \frac{\bar{Y}^2}{\bar{S}^2(Y)}\right),$$

where $\bar{Y}$ and $\bar{S}^2(Y)$ are unbiased empirical estimators of the mean and the variance of Y. We now set $p = 1$ and $r = 0$, with $s \in N$. Then:

$$M_{1,0,s} = E\left[y\left(1 + \frac{y\gamma}{\sigma}\right)^{-\frac{s}{\gamma}}\right] = \int_0^\infty y\left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{s+1}{\gamma} - 1} \frac{1}{\sigma}\, dy.$$

By setting $t = y(s + 1)$ and by using the property of Pareto Distribution, it is not difficult to show that:

$$M_{1,0,s} = \frac{\sigma}{(s + 1)(s + 1 - \gamma)}, \ \gamma < 1.$$

Estimating $M_{1,0,s}$ by its empirical version $\bar{M}_{1,0,s}$

$$\frac{\sum_{i=1}^{N_t} \left(1 - \frac{j}{n+1}\right)^s Y_{j,N_t}}{N_t},$$

and solving for $s = 0$ and $s = 1$, we end up with a system of equations in two unknowns, whose solutions are given by:

$$\hat{\gamma} = 2 - \frac{\bar{M}_{1,0,0}}{\bar{M}_{1,0,0} - 2\bar{M}_{1,0,1}}$$

$$\hat{\sigma} = \frac{2\bar{M}_{1,0,0}\bar{M}_{1,0,1}}{\bar{M}_{1,0,0} - 2\bar{M}_{1,0,1}}.$$

However, probability-weighted moments method presents an important problem: it can be used only when $\gamma < 1$, otherwise the integral $M_{1,0,s}$ does not converge; moreover, if $\gamma < 0$, some observations could be larger than $-\frac{\sigma}{\gamma}$ and so outside of the support of the GP Pareto Distribution fitted.

In order to overcome these issues, we now present element percentile method. If $\gamma = 0$, we can use MLE estimation efficiently, so we focus only on the case in which $\gamma \neq 0$. We set $p_{i,n} = \frac{i}{n+1}$, then, given order statistics $\{Y_{1,N_t}, \dots Y_{j,N_t} \dots Y_{N_t,N_t}\}$, we get the following system of equations:

$$\begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma} Y_{i,N_t}\right)^{-\frac{1}{\gamma}} = p_{i,n} \\ 1 - \left(1 + \frac{\gamma}{\sigma} Y_{j,N_t}\right)^{-\frac{1}{\gamma}} = p_{j,n}. \end{cases} \tag{29}$$

Setting $\tau = \frac{\gamma}{\sigma}$, and taking logarithms, we obtain:

$$C_j \log\left(1 + \tau Y_{i,N_t}\right) = C_i \log\left(1 + \tau Y_{j,N_t}\right),$$

where $C_j = -\log\left(1 - p_{j,N_t}\right)$. The solution of the equation above, say $\bar{\tau}$, must be found numerically, then:

$$\bar{\gamma} = \frac{\log\left(1 + \bar{\tau} Y_{i,N_t}\right)}{C_i}.$$

The advantage with respect to probability weighted moments method is that it works for any $\gamma \neq 0$. Finally, we discuss briefly about the choice of the threshold t, that is crucial in order to estimate γ: a possible solution takes into consideration the mean excess function. In fact, the mean excess function of a Pareto Distribution is given by:

$$e(t) = \frac{\sigma + t\gamma}{1 - \gamma},$$

so it is linear in t: a possible solution is to choose t such that the empirical mean excess function, $\hat{e}(t)$, is linear to the right of the point $(t, \hat{e}(t))$.
However, the POT method is far more stable than Block Maxima one: in fact, in empirical applications, we can see that the estimation of γ does not fluctuate too much as t varies, and this is the main advantage with respect to Block Maxima method. Unfortunately, a bias is still present.

3 Bivariate Extreme Value Theory

After presenting the most important elements in univariate case, we now move to the most interesting part of this thesis, that is bivariate extreme value theory: after a gentle introduction, we'll introduce two new measures, the exponent and the spectral measures. We denote by $\boldsymbol{x}$ the vector (x_1, x_2).
In order to start, let us define the maximum $\boldsymbol{M} = (M_1, M_2)$ as :

$$M_j = \max\{X_{1,j} \dots X_{n,j}\} \; j = 1, 2 \ ,$$

where $\{X_1 \dots X_n\}$ is the bivariate random sample available. Then it is not difficult to understand that:

$$P(M_1 < x_1, M_2 < x_2) = P^n[X_1 < x_1, X_2 < x_2] = F^n(x_1, x_2),$$

where F is the bivariate distribution of $\boldsymbol{X} = (X_1, X_2)$. In bivariate setting, F is in the domain of attraction of G, $F \in D(G)$, if and only if:

$$F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}), \tag{30}$$

where G is the bivariate extreme value distribution appearing in the limit. We now consider the convergent sequences $\{\boldsymbol{a}_n\}, \{\boldsymbol{b}_n\}, \{\boldsymbol{a}_{nk}\}, \{\boldsymbol{b}_{nk}\}$, where k is a fixed integer: it is then possible to prove that there exist vectors $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ such that:

$$\frac{\boldsymbol{a}_{nk}}{\boldsymbol{a}_n} \xrightarrow{n\to\infty} \boldsymbol{\alpha}_k$$

and

$$\frac{\boldsymbol{b}_{nk} - \boldsymbol{b}_n}{\boldsymbol{a}_n} \xrightarrow{n\to\infty} \boldsymbol{\beta}_k.$$

These equations, together with condition (30), implies that:

$$F^{nk}(\boldsymbol{x}\boldsymbol{a}_{\boldsymbol{nk}} + \boldsymbol{b}_{nk}) \xrightarrow{n\to\infty} G^k(\boldsymbol{x}\boldsymbol{\alpha}_k + \boldsymbol{\beta}_k). \tag{31}$$

But combining equations (31) and (30), we obtain:

$$G^k(\boldsymbol{x}\boldsymbol{\alpha}_k + \boldsymbol{\beta}_k) = G(\boldsymbol{x}), \forall k \in \mathbb{N}. \tag{32}$$

In other words, we have discovered that the class of bivariate extreme value distributions and the class of bivariate max-stable [1] distributions actually coincide.

Theorem 3.1. *Let us define the distribution function $\tilde{G}$ in the following way:*

$$\begin{aligned}\tilde{G}(z_1, z_2) &= P\left[-\frac{1}{\log G_1(Y_1)} < z_1, -\frac{1}{\log G_2(Y_2)} < z_2\right] = \\ &= G\left(G_1^{-1}\left\{\exp\left(-\frac{1}{z_1}\right)\right\}, G_2^{-1}\left\{\exp\left(-\frac{1}{z_2}\right)\right\}\right) = \\ &= C\left(\exp\left(-\frac{1}{z_1}\right), \exp\left(-\frac{1}{z_2}\right)\right),\end{aligned} \tag{33}$$

where G is the distribution function of the bivariate random vector Y with margins G_1 and G_2; then, if G is max-stable, $\tilde{G}$ is max stable too.

[1] A bivariate distribution function F is max-stable if and only if there exist sequences a_k, b_k such that:

$$F(x_1, x_2) = F^k(\alpha_k x_1 + \beta_k, \alpha_k x_2 + \beta_k) \; \forall k \in \mathbb{N}$$

Proof. Setting $G_j^k(\alpha_{jk}x_j + \beta_{jk}) = G_j(x_j) = t_j \ \forall j$, with $k \in \mathbb{N}$,implying

$$x_j = \frac{G_j^{-1}(t^{\frac{1}{k}}) - \beta_{kj}}{\alpha_{kj}},$$

we perform a change of variables $G_j(x_j) = t_j = \exp\left(-\frac{1}{y_j}\right)$. So we obtain

$$G_j^{-1}\left(\exp\left(-\frac{1}{y_j k}\right)\right) = \alpha_{kj}x_j + \beta_{kj}$$

that we plug into the definition of $\tilde{G}(x)$ in order to get:

$$\begin{aligned}
\tilde{G}(k\boldsymbol{y})^k &= G^k\left(G_1^{-1}\left\{\exp\left(-\frac{1}{y_1 k}\right)\right\}, G_2^{-1}\left(\exp\left\{-\frac{1}{y_2 k}\right)\right\}\right) \\
&= G^k\left(\alpha_{k,1}x_1 + \beta_{k,1}, \alpha_{k,2}x_2 + \beta_{k,2}\right) = G(x_1, x_2) = \\
&= G\left(G_1^{-1}\left\{\exp\left(-\frac{1}{y_1}\right)\right\}, G_2^{-1}\left\{\exp\left(-\frac{1}{y_2}\right)\right\}\right) = \\
&= \tilde{G}(\boldsymbol{y}).
\end{aligned} \tag{34}$$

To sum up,

$$\tilde{G}^k(k\boldsymbol{x}) = \tilde{G}(\boldsymbol{x}) \ \forall k \in \mathbb{N}, \tag{35}$$

so we prove the max-stability of $\tilde{G}(\boldsymbol{x})$, with $\alpha_k = k$ and $\beta_k = 0$. □

Using continuity and the equation above, it is not difficult to show that:

$$\tilde{G}^s(s\boldsymbol{z}) = \tilde{G}(\boldsymbol{z}), \forall s \text{ in } \mathbb{R}. \tag{36}$$

Remark 3.1. *$\tilde{G}$ has the same dependence structure of G, but with standard Frechet margins.*

We now present two fundamental measures, the Exponent and the Spectral Measures.

3.1 Exponent Measure

From now on, we assume that $\boldsymbol{X} = (X_1, X_2)$ is a bivariate random vector with distribution G: for sake of simplicity, we assume in this section that the random variables X_1 and X_2 are not negatively correlated. The Exponent Measure $\mu(\cdot)$ is defined as the measure satisfying

$$G(x) = \exp\left(-\mu\{[\mathbf{0}, +\infty) \setminus [\mathbf{0}, \boldsymbol{x}]\} = \exp\left(-V(\boldsymbol{x})\right)\right.$$

or, equivalently, the exponent measure $\tilde{\mu}(\cdot)$ is such that

$$\tilde{G}(\boldsymbol{x}) = \exp\left(-\tilde{\mu}\{[\mathbf{0}, +\infty) \setminus [\mathbf{0}, \boldsymbol{x}]\}\right) = \exp\left(-\tilde{V}(\boldsymbol{x})\right), \tag{37}$$

with G is the distribution function of a random variable with support $[\boldsymbol{q}, +\infty]$, where $q_j = \inf\{y : G_j(y) > 0\} < x_j \ j = 1, 2$. By combining equations (37) and (36), we can prove the homogeneity of $\tilde{\mu}$:

$$\begin{aligned}
\tilde{\mu}\{[\boldsymbol{z}, +\infty)\} &= -\log\left(\tilde{G}(\boldsymbol{z})\right) = \\
&= -s\log\tilde{G}(s\boldsymbol{z}) = s\tilde{\mu}\{[s\boldsymbol{z}, +\infty)\},
\end{aligned} \tag{38}$$

which can be expressed in a more compact way as:

$$\tilde{\mu}(\cdot) = s\tilde{\mu}(s\cdot). \tag{39}$$

Now we define the stable tail dependence function as:

$$l(v_1, v_2) = \tilde{V}\left(\frac{1}{v_1}, \frac{1}{v_2}\right); \tag{40}$$

then it is not difficult to understand that:

$$l(v_1, v_2) = -\log\{\left(G\left(G_1^{-1}\{\exp(-v_1)\}, G_2^{-1}\{\exp(-v_2)\}\right)\right)\} = -\log\tilde{G}\left(\frac{1}{v_1}, \frac{1}{v_2}\right). \tag{41}$$

Equivalently,

$$-\log G(\boldsymbol{v}) = l\{-\log(G_1(v_1)), -\log(G_2(v_2))\}. \tag{42}$$

Theorem 3.2. *Let $l(v_1, v_2)$ as defined in equations (40) and (41). Then the following properties hold:*

$$l(s\cdot) = sl(\cdot) \tag{43}$$

and

$$\max\{v_1, v_2\} \leq l(v_1, v_2) \leq v_1 + v_2. \tag{44}$$

Proof. Proof of (43) is trivial: in fact, $l(s\boldsymbol{x}) = -\log\left(\tilde{G}\left(\frac{1}{sx_1}, \frac{1}{sx_2}\right)\right)$, and, by equation (36),

$$l(s\boldsymbol{x}) = -\log\left(\tilde{G}\left(\frac{1}{sx_1}, \frac{1}{sx_2}\right)\right) = -\log\left(\tilde{G}^s\left(\frac{s}{sx_1}, \frac{s}{sx_2}\right)\right) = -s\log\left(\tilde{G}\left(\frac{1}{x_1}, \frac{1}{x_2}\right)\right) = sl(\boldsymbol{x}).$$

We now prove (44): in order to do that, we need to express the stable tail dependence function in terms of a copula. By equation (41), we know that:

$$\begin{aligned} l(v_1, v_2) &= -\log\{G(G^{-1}\{\exp(-v_1)), G^{-1}(\exp(-v_2)\})\} = \\ &= -\log P[G(X_1) < \exp(-v_1), G(X_2) < \exp(-v_2)] = \\ &= -\log C(\exp(-v_1), \exp(-v_2)), \end{aligned} \tag{45}$$

where C is the distribution function of a bivariate random vector with uniform margins, known as copula function. We define $C^I(u, v)$ the independence copula and $C^M(u, v)$ the co-monotonic copula as the distribution functions of bivariate random vectors with uniform margins and,respectively, with independent and perfectly dependent components.It is not difficult to show that:

$$C^I(u, v) = uv$$

and

$$C^M(u, v) = \min\{u, v\}.$$

By Copula Theory, see Nelsen [1], we know that any bivariate copula with positive association between the two random variables satisfy the following inequality:

$$uv = C^I(u, v) \leq C(u, v) \leq C^M(u, v) = \min\{u, v\}.$$

Then, by using (45), we can see that:

$$-\log(\exp(-v_1 - v_2)) \leq l(v_1, v_2) \leq -\log(\min\{\exp(-v_1), \exp(-v_2)\}),$$

which implies:

$$\max(v_1, v_2) \leq l(v_1, v_2) \leq v_1 + v_2. \tag{46}$$

□

In particular,

$$l(v_1, v_2) = v_1 + v_2 \tag{47}$$

corresponds to independence, while

$$l(v_1, v_2) = \max\{v_1, v_2\} \tag{48}$$

is associated with perfect positive dependence.

3.2 Spectral Measure

In this section, we'll describe $\mu(\cdot)$ in terms of pseudo-polar coordinates and we'll define the Spectral Measure. Assume, as before, that $\boldsymbol{X} = (X_1, X_2)$ is distributed as G and that $\boldsymbol{Z} = (Z_1, Z_2) = \left(-\frac{1}{\log(G_1(X_1))}, -\frac{1}{\log(G_2(X_2))}\right)$ follows the bivariate distribution $\tilde{G}$, to which the measure $\tilde{\mu}$ is associated. So let $T(\boldsymbol{z}) = (r, \boldsymbol{w})$ be a mapping from $\mathbb{R}^2$ to $\{(0, \infty) \times S_2\}$, where $\boldsymbol{w} = \frac{\boldsymbol{z}}{||\boldsymbol{z}||_2}$, $r = ||\boldsymbol{z}||_1$ and $S_2 = \{\boldsymbol{w} : ||\boldsymbol{w}||_2 = 1\}$: in other terms, $\boldsymbol{w}$ can be considered as the angular component of $\boldsymbol{z}$, while r can be seen as the radial part.
The spectral measure is then defined as:

$$S(B) = \tilde{\mu}\{\boldsymbol{z} \in \mathbb{R}^2 : ||\boldsymbol{z}|| \geq 1, \boldsymbol{w} \in B\}, \tag{49}$$

where B is a Borel subset of $E = S_2 \cap [0, +\infty)^2$. We want now to prove the following theorem that describes the link between the stable tail dependence function and the spectral measure.

Theorem 3.3. *Let $l(v_1, v_2)$ be the stable tail dependence function of the distribution $\tilde{G}$ associated to the measure $\tilde{\mu}$. Then:*

$$l(\boldsymbol{v}) = \int_E \max\left\{\frac{w_1 v_1}{||\boldsymbol{w}||_1}, \frac{w_2 v_2}{||\boldsymbol{w}||_1}\right\} \; S(d\boldsymbol{w}), \tag{50}$$

where $S(B) = \tilde{\mu}\{\boldsymbol{z} \in \mathbb{R}^2 : ||\boldsymbol{z}|| \geq 1, \boldsymbol{w} \in B\}$ as above.

Proof. By equation (38) and by the definition of spectral measure above, we can write

$$r\tilde{\mu}\{\boldsymbol{z} \in \mathbb{R}^2 : ||\boldsymbol{z}|| \geq r, \boldsymbol{w} \in B\} = \tilde{\mu}\{\boldsymbol{z} \in \mathbb{R}^2 : ||\boldsymbol{z}|| \geq 1, \boldsymbol{w} \in B\} = S(B) :$$

dividing both terms by r, we can see the decomposition of the spectral measure $\mu(\cdot)$ into the product of two components in pseudo-polar coordinates, the radial part equal to $r^{-2}dr$ and the angular part $S(d\boldsymbol{w})$. In formal terms,

$$\tilde{\mu}(d\boldsymbol{z}) = \tilde{\mu}(T^{-1}(dr, d\boldsymbol{w})) = r^{-2} dr S(d\boldsymbol{w}), \tag{51}$$

that can be used in order to compute integrals with respect to exponent measure. In fact, by using equation above and the definition of the function $T(\boldsymbol{z})$, we can easily write:

$$\int_{(0,\infty)^2} g(\boldsymbol{z})\tilde{\mu}(d\boldsymbol{z}) = \int_E \int_0^\infty g\left(\frac{r\boldsymbol{w}}{||\boldsymbol{w}||_1}\right) r^{-2} \; dr \; S(d\boldsymbol{w}),$$

where the set E has been defined above. Moreover, we can notice that:

$$\begin{aligned}\tilde{V}(\boldsymbol{z}) &= \tilde{\mu}\{[\boldsymbol{0}, +\boldsymbol{\infty}) \setminus [\boldsymbol{0}, \boldsymbol{z}]\} = \tilde{\mu}\{(Z_1, Z_2) : \exists j \in \{1, 2\} \; Z_j > z_j\} \\ &= \tilde{\mu}\left\{(Z_1, Z_2) : \max\left\{\frac{Z_1}{z_1}, \frac{Z_2}{z_2}\right\} > 1\right\} = \\ &= \int_{\mathbb{R}^2} 1\left\{\max\left\{\frac{Z_1}{z_1}, \frac{Z_2}{z_2}\right\} > 1\right\} \; \tilde{\mu}(d\boldsymbol{z});\end{aligned} \tag{52}$$

then, by using the change of variables presented above, we can rewrite the integral in this way:

$$
\begin{aligned}
\tilde{V}(\boldsymbol{z}) &= \int_{\mathbb{R}^2} 1\left(\max\left\{\frac{Z_1}{z_1}, \frac{Z_2}{z_2}\right\} > 1\right) \tilde{\mu}(d\boldsymbol{z}) = \\
&= \int_E \int_0^\infty 1\left(\max\left\{\frac{rw_1}{||\boldsymbol{w}||_1 z_1}, \frac{rw_2}{||\boldsymbol{w}||_1 z_2}\right\}\right) r^{-2}\, dr\, S(d\boldsymbol{w}) = \\
&= \int_E \max\left\{\frac{w_1}{z_1||\boldsymbol{w}||_1}, \frac{w_2}{||\boldsymbol{w}||_1 z_2}\right\} S(d\boldsymbol{w}).
\end{aligned} \tag{53}
$$

By applying the definition of l in terms of $\tilde{V}$ described by equation(40), we have proved relation (50). □

Moreover, since each margin of $\boldsymbol{Z}$ follows a standard Frechet distribution, we can write:

$$
\begin{aligned}
&\lim_{z_2\to\infty} \tilde{V}(z_1, z_2) = \\
&= \lim_{z_2\to\infty} -\log\left\{G\left(G_1^{-1}\left\{\exp\left(-\frac{1}{z_1}\right)\right\}, G_2^{-1}\left\{\exp\left(-\frac{1}{z_2}\right)\right\}\right)\right\} = \\
&= -\log\left(P\left[G_1(X_1) \le \exp\left(-\frac{1}{z_1}\right)\right]\right) = \\
&= \frac{1}{z_1}.
\end{aligned} \tag{54}
$$

Combining equations (54) and (53), we get:

$$
\begin{aligned}
&\lim_{z_2\to\infty} \tilde{V}(z_1, z_2) = \\
&= \lim_{z_2\to\infty} \int_E \max\left\{\frac{w_1}{||\boldsymbol{w}||_1 z_1}, \frac{w_2}{||\boldsymbol{w}||_1 z_2}\right\} S(dw) = \\
&= \int_E \frac{w_1}{||\boldsymbol{w}||_1 z_1} S(dw) = \frac{1}{z_1},
\end{aligned} \tag{55}
$$

which implies, by multiplying both terms by z_1, that:

$$
\int_E \frac{w_1}{||\boldsymbol{w}||_1} S(d\boldsymbol{w}) = 1. \tag{56}
$$

By the same way of reasoning, we can easily see that:

$$
\int_E \frac{w_2}{||\boldsymbol{w}||_1} S(d\boldsymbol{w}) = 1. \tag{57}
$$

Proposition 3.1. *Let $l(\boldsymbol{v})$ be the stable tail dependence function defined in (40), then $l(\boldsymbol{v})$ is convex.*

Proof. By equation (50), we can write:

$$
\begin{aligned}
&l(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) = \int_E \max\left\{\frac{w_1(\lambda x_1 + (1-\lambda)y_1)}{||\boldsymbol{w}||_1}, \frac{w_2(\lambda x_2 + (1-\lambda)y_2)}{||\boldsymbol{w}||_1}\right\} S(d\boldsymbol{w}) \le \\
&\int_E \max\left\{\frac{w_1\lambda x_1}{||\boldsymbol{w}||_1}, \frac{w_2\lambda x_2}{||\boldsymbol{w}||_1}\right\} S(d\boldsymbol{w}) + \int_E \max\left\{\frac{w_1(1-\lambda)y_1}{||\boldsymbol{w}||_1}, \frac{w_2(1-\lambda)y_2}{||\boldsymbol{w}||_1}\right\} S(d\boldsymbol{w}) = \\
&= l(\lambda\boldsymbol{x}) + l((1-\lambda)\boldsymbol{y}),
\end{aligned} \tag{58}
$$

by using the fact that $\max\{a+b, c+d\} \le \max\{a, c\} + \max\{b, d\}$. □

The most popular choice for the norms $||\boldsymbol{x}||_1$ and $||\boldsymbol{x}||_2$ is the sum-norm, $|\boldsymbol{x}| = x_1 + x_2$: in this case, the spectral measure S is often denoted by H and the space S_2 on which it is defined becomes

$$S_2 = \{\boldsymbol{w} : w_1 + w_2 = 1\}.$$

Trivially, equation (50) becomes:

$$l(v_1, v_2) = \int_0^1 \max\{wv_1, (1-w)v_2\} H(dw), \tag{59}$$

while equations (56) and (57) can be written as:

$$\int_0^1 w \ H(dw) = \int_0^1 (1-w) \ H(dw) = 1. \tag{60}$$

In particular, we can easily notice that $H[0,1] = 2$, in fact:

$$H([0,1]) = \int_0^1 H(dw) = \int_0^1 wH(dw) + \int_0^1 (1-w)H(dw) = 2. \tag{61}$$

3.3 Pickands Dependence Function

We define the Pickands Dependence Function as the function A such that

$$l(v_1, v_2) = (v_1 + v_2)A\left(\frac{v_2}{v_1 + v_2}\right). \tag{62}$$

By combining equations (42) and (62),we get:

$$G(x_1, x_2) = \exp\left(\log(G_1(x_1)G_2(x_2))A\left(\frac{\log(G_2(X_2))}{\log(G_1(x_1)G_2(x_2))}\right)\right). \tag{63}$$

Then, setting $G_1(x_1) = u$ and $G_2(x_2) = v$, we get the canonical representation of an extreme value copula:

$$C(u, v) = \exp\left(\log(uv)A\left(\frac{\log(v)}{\log(uv))}\right)\right).$$

Focusing on the unit simplex of (v_1, v_2), id est the set $V_2 = \{(v_1, v_2) : v_1 + v_2 = 1\}$, we may rewrite the equation (62) also as:

$$l(1-t, t) = A(t), \ t \in [0,1]. \tag{64}$$

We now describe some properties of $A(t)$ using those of the function $l(1-t, t)$ presented in the previous section:
a) $A(t)$ is obviously convex as well as $l(1-t, t)$, see Proposition 3.1;
b) By equation (46), $\max\{1-t, t\} \leq A(t) \leq 1$. The lower bound $A(t) = \max\{1-t, t\}$ corresponds to complete dependence, while the upper one, $A(t) = 1$, to perfect independence.
There is also a connection between Pickands Dependence Function and the spectral measure $H[0, w]$ on the unit simplex: in fact, it is possible to show, see Berlaint and alias [12], that:

$$H[0, w] = \begin{cases} 1 + A^{'}(w) \text{ if } 0 \leq w < 1 \\ 2 \text{ if } w = 1 \end{cases}, \tag{65}$$

using the fact that :

$$A(t) = \int_0^1 \max\{w(1-t), (1-w)t\} H(dw). \tag{66}$$

3.4 Measures of Extremal Dependence

In this subsection, we'll introduce some measures in order to quantify the level of extreme dependence between two random variables at finite level: we include the extremal coefficient and some copula-based indicators, such as Kendall's Tau and Spearman's Rho.
Let us start with the former one.

Definition 3.1. *Let us consider a non-empty set $V \subseteq \{1,2\}$ and the related set $\boldsymbol{e}_V = \{1_{1\in V}, 1_{2\in V}\}$. We define the extremal coefficient θ_V as*

$$\theta_V = l(\boldsymbol{e}_V) = \int_E \max_{j\in V} \left\{ \frac{w_j}{||\boldsymbol{w}||_1} \right\} S(d\boldsymbol{w}), \tag{67}$$

where l denotes the stable tail dependence function and S the Spectral measure as usual.

Theorem 3.4. *Let θ_V be the extremal coefficient associated to the extreme value distribution G with Pickands Dependence Function A and stable tail dependence l.Then:*

$$P[G_j(Y_j) \leq p \ \forall j \in V] = p^{\theta_V}, \tag{68}$$

where Y is a bivariate random vector with distribution G and margins G_1 and G_2.

Proof. We start by considering $V = \{1,2\}$, so:

$$l(1,1) = 2A\left(\frac{1}{2}\right) = \theta_{\{1,1\}}. \tag{69}$$

So:

$$\begin{aligned} P[G_1(Y_1) < p, G_2(Y_2) < p] &= \exp\left(\log(p^2) A\left(\frac{\log(p)}{\log(p^2)}\right)\right) \\ &= \exp\left(\frac{1}{2}\theta_V \log(p^2)\right) = p^{\theta_V}, \end{aligned} \tag{70}$$

that concludes the proof for this case. If $V = \{1\}$ or $V = \{2\}$, it is sufficient to notice that $\theta_V = l(1,0) = l(0,1) = 1$ by property (60). □

We now move to copula-based measures of extreme dependence, starting with Kendall's Tau.

Definition 3.2. *Let $\boldsymbol{Y} \sim G$ with margins G_1 and G_2 and with copula C to which the Pickands Dependence Function A is associated. We define Kendall's Tau as the difference between the probability of concordance of two realizations and the probability of discordance:*

$$\tau = P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(2)}) > 0] - P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(2)}) < 0], \tag{71}$$

where $(Y_1^{(1)}, Y_2^{(1)})$ and $(Y_1^{(2)}, Y_2^{(2)})$ are randomly generated from G.

Proposition 3.2. *Let τ and Y be defined as in definition 3.2. Then the following relation holds:*

$$\tau = 4E[C(G_1(Y_1), G_2(Y_2)] - 1. \tag{72}$$

Proof. First of all, we note that:

$$P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(2)}) > 0] = 2P[Y_1^{(1)} > Y_1^{(2)}, Y_2^{(1)} > Y_2^{(2)}];$$

that probability can be evaluated by integration with respect to one of the two realizations, say $(Y_1^{(1)}, Y_2^{(1)})$. Then we have:

$$\begin{aligned} P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(2)}) > 0] &= 2\int_0^{\infty}\int_0^{\infty} P[Y_1^{(2)} < y_1, Y_2^{(2)} < y_2]\; dC(G_1(y_1), G_2(y_2)) = \\ &= 2\int_0^{\infty}\int_0^{\infty} C(G_1(y_1), G_2(y_2))\; dC(G_1(y_1), G_2(y_2)) = \\ &= 2E[C(G_1(Y_1), G_2(Y_2))]. \end{aligned} \tag{73}$$

Writing definition (71) as

$$\tau = 2P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(2)}) > 0] - 1$$

and plugging into it equation (73), we get the result desired. □

Moreover, it is possible also to write τ as an integral of the Pickands Dependence Function, as we state in the following proposition.

Proposition 3.3. *Let τ be as in definition 3.2, then τ can be equivalently written as:*

$$\tau = \int_0^1 \frac{t(1-t)dA^{(1)}(t)}{A(t)}.$$

The proof is provided in the appendix.
The last measure of dependence we present in this thesis is Spearman's Rho.

Definition 3.3. *Given a random vector (Y_1, Y_2) as in definition 3.2, we define Spearman's Rho as*

$$\rho = 3\{P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(3)}) > 0] - P[(Y_1^{(1)} - Y_1^{(2)})(Y_2^{(1)} - Y_2^{(3)}) < 0]\}, \tag{74}$$

where $(Y_1^{(1)}, Y_2^{(1)})$, $(Y_1^{(2)}, Y_2^{(2)})$ and $(Y_1^{(3)}, Y_2^{(3)})$ are independent realizations from G.

Proposition 3.4. *Let the random vector (Y_1, Y_2) be distributed with distribution G, with extreme-value copula C and Pickands Dependence function A. Moreover let ρ be as in definition 3.3, then ρ can be equivalently written as*

$$\rho = 12\int_0^1\int_0^1 uv\; dC(u, v) - 3;$$

and also as

$$\rho = 12\int_0^1 \frac{1}{(1 + A(t))^2}dt - 3.$$

Proof. The proof is very similar to the proof of Proposition 3.2 and 3.3. □

3.5 Dependence Structure

We recall that $F \in D(G)$ if and only if the following limit holds:

$$\lim_{n\to\infty} F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}) :$$

so, in a bivariate context, it is quite intuitive to understand that the convergence of that limit implies the convergence of both margins and, also, of the dependence structure. Clearly, the convergence of the former is satisfied when:

$$\lim_{n\to\infty} F_j^n(a_{nj}x_j + b_{nj}) \to G_j(x_j),\; j \in \{1, 2\}, \tag{75}$$

so let us focus on the dependence structure that we model through copula functions. So, given a sequence of independent bivariate random vectors $\{X_1 \dots X_n\}$ with distribution F, it is not difficult to understand that the distribution of the sample maximum of each component evaluated in x_j is just $F_j^n(x_j)$, so that the copula of the sample maximum will be just:

$$\begin{aligned}
C_{F^n}(\boldsymbol{u}) &= P[\max_{i=1\dots n} F_1^n(X_{1i}) < u_1, \max_{i=1\dots n} F_2^n(X_{2i}) < u_2] = \\
&= P^n[F_1^n(X_{1i}) < u_1, F_2^n(X_{2i}) < u_2] = P^n[F_1(X_{1i}) < u_1^{\frac{1}{n}}, F_2(X_{2i}) < u_2^{\frac{1}{n}}] = \\
&= C_F^n(u_1^{\frac{1}{n}}, u_2^{\frac{1}{n}}).
\end{aligned} \tag{76}$$

The convergence of the copula structure then holds if:

$$\lim_{n\to\infty} C_{F^n}(\boldsymbol{u}) = \lim_{n\to\infty} C_F^n(u_1^{\frac{1}{n}}, u_2^{\frac{1}{n}}) = C_G(\boldsymbol{u}). \tag{77}$$

Now we propose the following approximation that will be used widely in the empirical part in order to apply Excesses over Threshold method.

Theorem 3.5. *Let $F \in D(G)$, where F denotes the distribution function of a bivariate random vector X: then the following approximation for the distribution function F holds:*

$$F(\boldsymbol{x}) \approx \exp(-l\{-\log(F_1(x_1)), -\log(F_2(x_2))\}), \tag{78}$$

when $F_1(x_1)$ and $F_2(x_2)$ are sufficiently close to 1 and l is clearly the stable tail dependence function of G.

Proof. Since $F \in D(G)$, both limits (75) and (77) hold; moreover, equation (77) holds uniformly, so it is possible to replace the integer variable n by a real one t, obtaining:

$$\lim_{t\to\infty} C_{F^t}(\boldsymbol{u}) = \lim_{t\to\infty} C_F^t(u_1^{\frac{1}{t}}, u_2^{\frac{1}{t}}) = C_G(\boldsymbol{u}). \tag{79}$$

But, by using equation (45) and the homogeneity relation of $l(\cdot)$, we have:

$$\begin{aligned}
C_G^s(\boldsymbol{u}^{\frac{1}{s}}) &= \exp\left(-sl\{-\log(u_1^{\frac{1}{s}}), -\log(u_2^{\frac{1}{s}})\}\right) = \\
&= \exp\left(-l\{-\log(u_1), -\log(u_2)\}\right) = C_G(u_1, u_2),
\end{aligned} \tag{80}$$

that implies that $C_F(\boldsymbol{u}) \approx C_G(\boldsymbol{u})$ for $\boldsymbol{u}$ sufficiently close to $\mathbf{1}$. So, using again equation (45) and the approximation just proved,

$$\begin{aligned}
F(\boldsymbol{x}) &= C_F(F_1(x_1), F_2(x_2)) \approx C_G(F_1(x_1), F_2(x_2)) = \\
&= \exp(-l\{-\log(F_1(x_1)), -\log(F_2(x_2))\}),
\end{aligned} \tag{81}$$

proving the theorem. □

This approximation shows the possibility to connect the distribution function F to the stable tail dependence function of the distribution G to which F is attracted.
We want now to introduce a new useful function, the tail dependence function, $D_F(\boldsymbol{u})$, but, before doing that, we need to state the following theorem.

Theorem 3.6. *Let $\boldsymbol{X} = (X_1, X_2) \sim F$, with $F \in D(G)$, and let $\hat{\boldsymbol{X}} = (\hat{X}_1, \hat{X}_2) \sim \hat{F}$, where*

$$\hat{X}_1 = -\frac{1}{\log(F_1(X_1))}$$

and

$$\hat{X}_2 = -\frac{1}{\log(F_2(X_2))}.$$

Then the following limit holds:

$$\lim_{t\to\infty} t(1-\hat{F}(t\boldsymbol{z})) = -\log(\hat{G}(\boldsymbol{z})) = l\left(\frac{1}{z_1}, \frac{1}{z_2}\right),$$

where $\hat{G}$ has been obtained after transformation of G into standard Frechet Margins and l is the stable tail dependence function of G as usual.

Proof. It is not difficult to check that the distribution of $\hat{X}$ is given by:

$$\hat{F}(\boldsymbol{z}) = P[\hat{X}_1 < z_1, \hat{X}_2 < z_2] = F\left(F_1^{-1}\left\{\exp\left(-\frac{1}{z_1}\right)\right\}, F_2^{-1}\left\{\exp\left(-\frac{1}{z_2}\right)\right\}\right).$$

So, by applying limit (79), the following result holds:

$$\begin{aligned}
\lim_{t\to\infty} \hat{F}^t(t\boldsymbol{z}) &= \lim_{z\to\infty} F^t\left(F_1^{-1}\left\{\exp\left(-\frac{1}{z_1 t}\right)\right\}, F_2^{-1}\left\{\exp\left(-\frac{1}{z_2 t}\right)\right\}\right) = \\
&= \lim_{t\to\infty} C_F^t\left(\exp\left(-\frac{1}{z_1 t}\right), \exp\left(-\frac{1}{z_2 t}\right)\right) = \\
&= C_G\left(\exp\left(-\frac{1}{z_1}\right), \exp\left(-\frac{1}{z_2}\right)\right) = \\
&= G\left(G_1^{-1}\left\{\exp\left(-\frac{1}{z_1}\right)\right\}, G_1^{-1}\left\{\exp\left(-\frac{1}{z_2}\right)\right\}\right) = \\
&= \hat{G}(\boldsymbol{z}),
\end{aligned} \tag{82}$$

where $\hat{G}(\boldsymbol{z})$ has been obtained after transformation of $G(\boldsymbol{z})$ into standard Frechet margins. Taking logarithms and a first order Taylor expansion of the above limit, we get:

$$\lim_{t\to\infty} t(1-\hat{F}(t\boldsymbol{z})) = -\log(\hat{G}(\boldsymbol{z})) = l\left(\frac{1}{z_1}, \frac{1}{z_2}\right), \tag{83}$$

where the stable tail dependence function is the same for both $G(\boldsymbol{z})$ and $\hat{G}(\boldsymbol{z})$, as a consequence of invariance Copula property. □

Definition 3.4. *Given a random vector $\boldsymbol{X} = (X_1, X_2) \sim F$, we define the tail dependence function $D_F(\boldsymbol{z})$ as:*

$$D_F(z_1, z_2) = 1 - F(F_1^{-1}(1-z_1), F_2^{-1}(1-z_2)). \tag{84}$$

If we apply the transformation $\dot{X}_j = \frac{1}{1-F_j(X_j)}$ $j = 1, 2$, then it is easy to show that the distribution of $\dot{X}$ is given by:

$$\dot{F}(z_1, z_2) = F\left(F_1^{-1}\left(1-\frac{1}{z_1}\right), F_2^{-1}\left(1-\frac{1}{z_2}\right)\right),$$

which satisfies, as well as $\hat{F}(z)$, limit (83), with limiting distribution $\dot{G}$, which is obtained after transformation of G into standard Pareto margins, since Pareto Distribution is in its own domain of attraction. Moreover,

$$D_F(z_1, z_2) = 1 - \dot{F}\left(\frac{1}{z_1}, \frac{1}{z_2}\right).$$

Applying now limit (83) and the new definition of $D_F(\boldsymbol{z})$ in terms of $\dot{F}(\boldsymbol{z})$, we get the following fundamental result:

$$\begin{aligned}&\lim_{t\to\infty} t\left(1-\dot{F}\left(\frac{t}{z_1},\frac{t}{z_2}\right)\right)=\lim_{t\to\infty} tD_F\left(\frac{\boldsymbol{z}}{t}\right)=\\&=\lim_{s\to 0} s^{-1}D_F(s\boldsymbol{z})=l(\boldsymbol{z}),\end{aligned}\tag{85}$$

where $l(z)$ is the stable tail dependence function of $\dot{G}$. Moreover, it is easy to show that

$$l(\boldsymbol{z})=\lim_{s\to 0} s^{-1}P[\exists\ j=1,2: F_j(X_j)>1-sz_j]=\lim_{t\to\infty} tP\left[\max_{j=1,2}\left\{\frac{z_j}{1-F_j(X_j)}\right\}>t\right]\tag{86}$$

and

$$l(\boldsymbol{z})=\lim_{s\to 0} s^{-1}(1-C_F(1-sz_1,1-sz_2)).\tag{87}$$

By the last limit, we can even easily find the Pickands Dependence Function and the extremal coefficient as a limit of the tail dependence function.
We now state the following fundamental theorem that provides an approximation for F in terms of l that will be widely used in Block Maxima approach in the empirical application.

Theorem 3.7. *Let $F\in D(G)$ be the distribution of (X_1,X_2) and l the stable tail dependence function of the limit distribution G. Then the following approximation holds for values of $F_j(x_j)$ $j=1,2$ sufficiently close to 1:*

$$1-F(x_1,x_2)\approx l(1-F_1(x_1),1-F_2(x_2)).\tag{88}$$

Proof. We consider limit (85) as an exact equality for small values of s. Setting $sz_j=1-F_j(x_j)$ $j=1,2$, with $F_j(x_j)$ as close as possible to 1, we have, on the left hand side,

$$D_F(s\boldsymbol{z})=1-F(F_1^{-1}(1-1+F_1(x_1)),F_2^{-1}(1-1+F_2(x_2)))=1-F(x_1,x_2)$$

and, on the right hand side,

$$sl(\boldsymbol{z})=l(s\boldsymbol{z})=l(1-F_1(x_1),1-F_2(x_2)).$$

Combining both sides, we get:

$$1-F(x_1,x_2)\approx l(1-F_1(x_1),1-F_2(x_2)),\tag{89}$$

for values of $F_j(x_j)$ $\forall j$ sufficiently close to 1, as we wish. □

Now I want to make a comparison between approximations (81) and (88) under exact independence.

Remark 3.2. *Under independence, using approximation (81) and property (47) of l, for values of $F_1(x_1)$ and $F_2(x_2)$ sufficiently close to 1, we have:*

$$P(X_1>x_1,X_2>x_2)\approx 1-F_1(x_1)-F_2(x_2)+\exp\{-\log(F_1(x_1))-\log(F_2(x_2))\}=\bar{F}_1(x_1)\bar{F}_2(x_2),$$

which is exactly what we expect under independence. By approximation (88),under independence, when $F_1(x_1)$ and $F_2(x_2)$ are close to 1,

$$P[X_1<x_1,X_2<x_2]=F(x_1,x_2)\approx 1-l(1-F_1(x_1),1-F_2(x_2))=1-(1-F_1(x_1)+1-F_2(x_2)),$$

so the following equation holds:

$$P[X_1>x_1,X_2>x_2]=1-F_1(x_1)-F_2(x_2)+P[X_1<x_1,X_2<x_2]\approx 0.$$

So the probability of joint extremes is almost 0 under approximation (88): so any models based on this approximation will undervalue the probability, under independence, that simultaneous extreme realizations occur, pointing to stronger dependence than actual one. In other words, estimations will be biased towards stronger dependence. This is the reason why we prefer approximation (81).

3.6 Asymptotic Theory

In this subsection, we'll discuss about the strength of dependence of two random variables at asymptotic level, focusing on the limit cases of independence and complete dependence. We start by complete dependence.

Definition 3.5. *Given (X_1, X_2) a bivariate random vector with distribution $F \in D(G)$ and margins F_1 and F_2 and with copula C_F, we say that X_1 and X_2 are asymptotically completely dependent if and only if*

$$\lim_{n\to\infty} C_F^n(u_1^{\frac{1}{n}}, u_2^{\frac{1}{n}}) = C_G(u_1, u_2) = \min\{u_1, u_2\}$$

or

$$\lim_{s\to 0} s^{-1} D(s\boldsymbol{u}) = \max\{u_1, u_2\}.$$

It is not difficult to see that the two conditions are equivalent. In fact,

$$l(u_1, u_2) = \lim_{s\to 0} s^{-1} D(s\boldsymbol{u}) = \lim_{s\to 0} s^{-1}(1 - C(1 - su_1, 1 - su_2)) = \max\{u_1, u_2\},$$

under asymptotic dependence. We now state the following theorem.

Theorem 3.8. *Let us consider a bivariate random vector X with distribution function F in the domain of attraction of G. A sufficient condition in order to have asymptotic complete dependence is given by:*

$$\exists w \in [0, +\infty) \ \lim_{s\to 0} s^{-1} D(sw, sw) = w. \tag{90}$$

Proof. We set $v = \max\{v_1, v_2\}$, so we have

$$v \le s^{-1} D(sv_1, sv_2) \le s^{-1} D(sv, sv) = \frac{v}{w}\frac{w}{sv} D\left(\frac{sv}{w} w, \frac{sv}{w} w\right). \tag{91}$$

But, if (90) is true, then the right-hand side converge to v as s approaches 0. This implies that, by squeeze theorem, $\lim_{s\to 0} s^{-1} D(sv_1, sv_2)$ is equal to v too. Then the two variables are asymptotic dependent by definition 3.5. □

We now move to the independence case.

Definition 3.6. *Let $\boldsymbol{X} = (X_1, X_2)$ be a bivariate random vector with distribution function F in the domain of attraction of G, copula C_F and margins F_X and F_Y. Then X_1 and X_2 are asymptotically independent if and only if:*

$$\lim_{n\to\infty} C_F^n(u_1^{\frac{1}{n}}, u_2^{\frac{1}{n}}) = C_G(u_1, u_2) = u_1 u_2$$

or

$$\lim_{s\to 0} s^{-1} D(sv_1, sv_2) = v_1 + v_2.$$

Theorem 3.9. *Let $\boldsymbol{X} = (X_1, X_2)$ be again a bivariate random vector with distribution function F in the domain of attraction of G and copula C_F. A necessary condition between the two components of the random vector in order to have asymptotic independence is the following:*

$$\exists v \in (\boldsymbol{0}, \infty) : \lim_{s\to 0} s^{-1} D(sv_1, sv_2) = v_1 + v_2 \ , \tag{92}$$

where $D_F(v_1, v_2)$ is clearly the tail dependence function of F.

Proof. The proof is very similar to the previous one, so we omit it. □

We now want to determine some coefficients of dependence in order to identify the class of asymptotically dependent and independent random variables and describe the strength of their dependence.

Definition 3.7. *Given a random vector (X, Y) with distribution F, we define the function $\chi(u)$ as:*

$$\chi(u) = 2 - \frac{\log(C_F(u, u))}{\log(u)}. \tag{93}$$

Since $\max\{2u - 1, 0\} \leq C(u, u) \leq u$, see Nelsen [1] for the proof, then:

$$2 - \frac{\log(\max\{2u - 1, 0\})}{\log(u)} \leq \chi(u) \leq 1.$$

The right hand side corresponds to perfect positive dependence, while the left one to perfect negative correlation. The function $\chi(u)$ is useful in the investigation of tail dependence: in fact, when u is close to 1,

$$\chi(u) = 2 - \frac{\log(C_F(u, u))}{\log(u)} \approx 2 - \frac{1 - C_F(u, u)}{1 - u} = \frac{1 - 2u + C_F(u, u)}{1 - u} = P[F_X(X) > u | F_Y(Y) > u]. \tag{94}$$

Moreover, it is not difficult to check that $\chi(u) > 0 \leftrightarrow C(u, u) > u^2$, meaning that $\chi(u)$ can be considered as a way to establish the kind of relationship among the two random variables: in fact they are positively dependent if $\chi(u) > 0$, negatively dependent when $\chi(u) < 0$ and independent otherwise. We now define the coefficient χ as :

$$\chi = \lim_{u \to 1} \chi(u). \tag{95}$$

It lies in the region $[0, 1]$ and it increases as we move away from independence. The coefficient we have just described can be also expressed in terms of the extremal coefficient θ of the extreme value distribution G to which F is attracted:

$$\chi = 2 - \theta.$$

By using equation (69),

$$C_F(u, u) = \exp\left(\log(u^2) A\left(\frac{\log(u)}{\log(u^2)}\right)\right) = u^\theta,$$

so we can write, by definition 3.7,

$$\begin{aligned} \chi &= 2 - \lim_{u \to 1} \frac{\log(C_F(u, u))}{\log(u)} = \\ &= 2 - \lim_{u \to 1} \frac{\log(u^\theta)}{\log(u)} = \\ &= 2 - \theta. \end{aligned} \tag{96}$$

To sum up, $F \in D(G)$ implies that $\chi(u) = 2 - \theta$.
We now focus on the class of asymptotically independent variables with $\chi = 0$: we want, in particular, to look for some measures of the strength of dependence within this group, and a possible source of information is given by the analysis of the function $\bar{\chi}(u)$.

Definition 3.8. *Given a random vector (X, Y) with distribution F, margins F_X, F_Y and copula C_F, we define the function $\bar{\chi}(u)$ as*

$$\bar{\chi}(u) = \frac{2\log(1-u)}{\log(\bar{C}_F(u,u))} - 1, \tag{97}$$

where $\bar{C}_F(u,u)$, known as survival copula, is given by

$$\bar{C}_F(u,u) = P[U_1 > u, U_2 > u] = 1 - 2u + C(u,u).$$

We now set $\bar{\chi}$ equal to:

$$\bar{\chi} = \lim_{u\to 1} \bar{\chi}(u),$$

as done for χ. Using again the Frechet-Hoeffding bounds for copulas, it is not difficult to show that:

$$\frac{\log(1-u)}{\log(\max\{1-2u; 0\})} - 1 \leq \bar{\chi}(u) \leq 1,$$

which implies that

$$-1 \leq \bar{\chi} \leq 1.$$

So, if two variables are asymptotically positive dependent, then $\bar{\chi} = 1$, with $0 < \chi < 1$ as a measure of the strength of dependence; if they are asymptotically independent, $\chi = 0$, with a strength of dependence given by $-1 \leq \bar{\chi} < 1$. In particular, if the random variables are independent not only asymptotically but also at finite level, the coefficient $\bar{\chi} = 0$ too. Moreover, looking at the sign of $\bar{\chi}(u)$, it is not difficult to show that $\bar{\chi}(u)$ is positive when there is positive association between the two variables, id est when $u^2 < C(u,u) \leq u$, while it is negative in the opposite case.
In order to introduce another measure of tail dependence, the coefficient of tail dependence, we need to transform the margins of F into standard Frechet Margins, by setting as usual

$$Z_j = -\frac{1}{\log(U_j)},\ \ j = 1, 2,$$

where $U_1 = F_X(X)$ and $U_2 = F_Y(Y)$. Then we can write:

$$\begin{aligned} P[Z_1 > z_1, Z_2 > z_2] &= P\left[-\frac{1}{\log(U_1)} > z_1, -\frac{1}{\log(U_2)} > z_2\right] = \\ &= P\left[U_1 > \exp\left(-\frac{1}{z_1}\right), U_2 > \exp\left(-\frac{1}{z_2}\right)\right] = \\ &= \bar{C}\left(\exp\left(-\frac{1}{z_1}\right), \exp\left(-\frac{1}{z_2}\right)\right). \end{aligned} \tag{98}$$

Definition 3.9. *The coefficient of tail dependence is the coefficient η such that, for values of z sufficiently large,*

$$P[Z_1 > z, Z_2 > z] \sim L(z) z^{-\frac{1}{\eta}}, \tag{99}$$

where $L(z)$ is s.v..

Since $P[Z_1 > z, Z_2 > z] \leq P[Z_1 > z] = 1 - \exp\left(-\frac{1}{z}\right) \approx \frac{1}{z}$, then $\eta \leq 1$.

Theorem 3.10. *Let Z be a bivariate random vector and suppose that*

$$P[Z_1 > z, Z_2 > z] \sim L(z) z^{-\frac{1}{\eta}},$$

for large values of z. Then the following relation holds:

$$\bar{\chi} = 2\eta - 1.$$

Proof. Combining equations (98) and (99), we have:

$$\lim_{u\to 1}\bar{C}(u,u)=\lim_{u\to 1}L\left(-\frac{1}{\log(u)}\right)(-\log(u))^{\frac{1}{\eta}}.$$

So we can write the coefficient $\bar{\chi}$ as a function of η:

$$\begin{aligned}\bar{\chi}&=\lim_{u\to 1}\frac{2\log(1-u)}{\log(L(-\frac{1}{\log(u)}))+\frac{1}{\eta}\log(-\log(u))}-1=\\&=2\eta-1,\end{aligned}\tag{100}$$

by using the fact that $\log(u)\approx u-1$ as u approaches 1 and that

$$\lim_{u\to 1}\frac{\log(L(-\frac{1}{\log(u)}))}{\log(-\frac{1}{\log(u)})}=0,$$

due to the application of the corollary 2.3.1. □

Moreover, by application of equation (99) and definition (95), we can see that:

$$\lim_{z\to\infty}L(z)z^{1-\frac{1}{\eta}}=\lim_{z\to\infty}P[Z_1>z|Z_2>z]=\lim_{u\to 1}P[U_1>u|U_2>u]=\chi,\tag{101}$$

where $U_i=F_X(X_i)$ and $V_i=F_Y(Y_i)$. So, if $\eta=1$ and $\lim_{z\to\infty}L(z)\to c$, then, by equation (100), the two variables are asymptotically dependent and, by (101), their strength of dependence is equal to c. By contrast, if $0<\eta<1$, then $\chi=0$, so the variables are asymptotically independent, with a strength of dependence measured by $\bar{\chi}=2\eta-1$. Within the class of asymptotically independent variables, positive association happens when $0.5<\eta<1$, while negative correlation arises when $0<\eta<0.5$: near independence occurs when $\eta=0.5$ and perfect independence when, in addition, $\lim_{z\to\infty}L(z)=1$. We conclude this section with an example.

Example 3.1. *Let us consider the following copula function:*

$$C(u_1,u_2)=\exp(-[(-\log(u_1))^{\frac{1}{\alpha}}+(-\log(u_2))^{\frac{1}{\alpha}}]^{\alpha}),\tag{102}$$

which is the distribution of the uniform random vector (U_1,U_2)*. We perform the usual transformation to standard Frechet Margins,*

$$Z_j=-\frac{1}{\log(U_j)},$$

so we have:

$$P[Z_1<u,Z_2<u]=\exp\left(-\frac{2^{\alpha}}{u}\right).$$

Writing

$$P[Z_1>u,Z_2>u]=1-2\exp\left(-\frac{1}{u}\right)+\exp\left(-\frac{2^{\alpha}}{u}\right)$$

and using the fact that, as z *approaches* $+\infty$,

$$\exp\left(-\frac{1}{z}\right)\approx 1-\frac{1}{z},$$

we have :

$$\lim_{z\to\infty}P[Z_1>z,Z_2>z]=\lim_{z\to\infty}(2-2^{\alpha})z^{-1}+o(z^{-1}).$$

So, by application of equation (99), we can see that, if $0 < \alpha < 1$, $\eta = 1$ *and* $L(z) \to 2 - 2^{\alpha}$*; so a value of* α *smaller than 1 implies asymptotic dependence with strength of dependence equal to* $2 - 2^{\alpha}$*. If* $\alpha = 1$*, by direct substitution into equation (102),*

$$C(u_1, u_2) = u_1 u_2,$$

implying independence.

We'll come back to this copula model in chapter 4.

References

[1] B.Nelsen. *An Introduction to Copulas*. Springer, 2006, p. 270.

[2] 0.Scaillet B.Remillard. “Testing for equality between two copulas”. In: *Journal of Multivariate Analysis* (2009), pp. 377–386.

[3] M.P.Vecchi C.D.Gelatt S.Kirkpatrick. “Optimization by Simulated Annealing”. In: *Science, Volume 220* (1983), pp. 671–680.

[4] B.Rémillard C.Genest J.F.Quessy. “Asymptotic local efficiency of Cramér-Von Mises tests for multivariate independence”. In: *The Annals of Statistics 35* (2007), pp. 166–191.

[5] J.Neslehova C.Genest I.Kojadinovic et al. “A goodness-of-fit test for bivariate extreme-value copulas”. In: *Bernoulli, Volume 17* (2011), pp. 253–275.

[6] J.Segers C.Genest. “Rank-Based Inference for Bivariate Extreme-Value Copulas”. In: *The Annals of Statistics* (2009), p. 34.

[7] P.Caperàa A.L.Fougeres C.Genest. “A non-parametric estimation procedure for bivariate extreme value copulas”. In: *Biometrika Volume 84* (1997), pp. 567–577.

[8] M.Tamraz E.Hashorva G.Ratovomirja. “On some new dependence models derived from multivariate collective models in insurance applications”. In: *Scandinavian Actuarial Journal,Volume 2017* (2017), pp. 730–750.

[9] R. Schmidt G. Frahm M. Junker. “Estimating the tail-dependence coefficient: Properties and pitfalls”. In: *Insurance: Mathematics and Economics, Volume 37* (2005), pp. 80–100.

[10] J.Yan I. Kojadinovic. “Non-parametric rank-based tests of bivariate extreme-value dependence”. In: *Journal of Multivariate Analysis* (2010), pp. 2235–2249.

[11] J.A.Tawn. “Bivariate Extreme Value Theory: Model and Estimation”. In: *Biometrika Volume 75* (1988), pp. 397–415.

[12] J.Segers J.Berlaint Y.Goegebeur et al. *Statistics of Extremes: Theory and Applications*. Wiley, 2004, p. 514.

[13] J.Pickands. “Multivariate extreme value distributions”. In: *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session* (1981), pp. 859–878.

[14] J.R.M.Hosking. “Maximum likelihood estimation of the parameters of the generalized extreme-value distribution”. In: *Journal of the Royal Statistical Society* (1985), pp. 301–310.

[15] S.G.Self K.Y. Liang. “Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions”. In: *Journal of the American Statistical Association, Volume 82* (1987), pp. 605–610.

[16] M.Rytgaard. “Simulation experiments on the mean residual life function m(x)”. In: *Proceedings of the XXVII Astin Colloquium* (1996), pp. 59–81.

[17] N.Tajvidi. *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions*. 1996. URL: `http://www.maths.lth.se/matstat/staff/nader/thesis.pdf`.

[18] P.Billingsley. *Probability and Measure*. Wiley, 1995, p. 587.

[19] T.Mikosch P.Embrechts C.Kluppelberg. *Modelling Extremal Events*. Springer, 1997, p. 645.

[20] S.Haug C. Kluppelberg L. Peng. “Statistical Models and Methods for Dependence in Insurance Data”. In: *Journal of Korean Statistical Society,Volume 40* (2011), pp. 125–139.

[21] P.Hall and N.Tajvidi. "Distribution and dependence-function estimation for bivariate extreme-value distributions". In: *Bernoulli Volume 6* (2000), pp. 835–844.

[22] R.Mason R.Gunst J.Hess. *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*. Wiley, 2003, p. 723.

[23] R.L.Smith. "Maximum likelihood estimation in a class of non-regular cases". In: *Biometrika 72* (1985), pp. 67–90.

[24] J.A.Tawn S.G Coles. "Modelling Extreme Multivariate Events". In: *Journal of the Royal Statistical Society* (1991), pp. 377–392.

[25] S.Resnick. "Discussion of the Danish Data on Large Fire Insurance Losses". In: *Journal of the International Actuarial Association* (1997), pp. 139–151.

[26] S.Resnick. *Heavy Tailed Analysis*. 2005. URL: `https://www.eurandom.tue.nl/EURANDOM_chair/I-Resnick.pdf`.

[27] W.Stute. "The oscillation behaviour of empirical processes: The multivariate case". In: *Annals of Probability* (1984), pp. 361–379.