Chapter II

ASYMPTOTICS OF UNIVARIATE EXTREMES

Introduction

This chapter can be divided into two different parts: the first (laws of large numbers) which expresses, in a formal way, the intuitive ideas relative to the behaviour of sample maxima and minima and the way they converge to some values (finite or infinite) and the second (asymptotic distributions of extremes) dealing with the behaviour of extremes in large samples, which is, in fact, the basis for the development of the whole book. Although the first part is of almost entirely theoretical importance only, as its practical application is essentially nil, it helps to pave the way for the following developments. Practically speaking hardly anyone will use the laws of large numbers, but many of the practical applications in which extremes have to be taken into account, such as the design of dams, the study of fatigue fractures, etc, are based on the asymptotic distributions of extremes.

General introduction to the laws of large numbers

We say that a sequence of random variables $\{X_k, k = 1,2,\ldots\}$ verifies *a* law of large numbers if there exists a sequence of functions $\{\varphi_k\}$ (possibly one function φ), a sequence of constants $\{c_k\}$, and a finite constant c such that $\varphi_k(X_k, c_k) \overset{P}{\to} c$ as $k \to \infty$; c can evidently be taken as zero.

The interpretation of this definition is easy : it means that for large k, $\varphi_k(X_k, c_k)$ is close in probability to c, i.e. , given $\in (>0)$ and $\delta (0 < \delta < 1)$ we know that for $k > N(\in, \delta)$ we have

$$\text{Prob}\,\{c + \in \leq \varphi_k(X_k, c_k) \leq c + \in\} \geq 1 - \delta\,,$$

thus obtaining an approximate evaluation of the values of X_k for large k. How good the approximation is depends on how large k must be, for small $\in$ and δ; sometimes the approximation is bad even for very large k, in real–life conditions. Be that as it may, it gives a *deterministic* approximation that may be useful for rough calculations.

Such a large framework, depending also on the choice of $\{\varphi_k\}$, has not aroused interest in itself and, until now, studies have specialized in two forms of φ_k: the additive and the multiplicative laws of large numbers, denoted ALLN and MLLN respectively.

The sequence $\{X_k\}$ satisfies the ALLN if there exist $\{c_k\}$ such that $X_k - c_k \xrightarrow{P} 0$. Evidently the c_k are not uniquely defined because if $\{\in_k\} \to 0$ *also* $X_k - (c_k + \in_k) \xrightarrow{P} 0$.

If $h(x) = \alpha + \beta x \ (\beta \neq 0)$ is a linear transformation then $h(X_k) - (\alpha + \beta\, a_k) \xrightarrow{P} 0$, i.e., $\tilde{X}_k = h(X_k)$ also satisfies the ALLN if with coefficients $\tilde{a}_k = \alpha + \beta\, a_k$.

The sequence $\{X_k\}$ satisfies the MLLN if there exist $\{c_k \neq 0\}$ such that $X_k / c_k \xrightarrow{P} 1$. It is evident, also, that c_k is not uniquely defined, as $X_k/ (c_k / \eta_k) \xrightarrow{P} 1$ if $\eta_k(> 0) \to 1$. Evidently $\{\beta X_k\}$ $(\beta \neq 0)$ satisfies the MLLN with $\{\beta c_k\}$.

Putting aside the case where $\{c_k\}$ oscillates permanently in sign, we can suppose that we have c_k ultimately (i.e., for all $k > k_0$) positive or negative. If c_k is ultimately positive we can define $\tilde{c}_k = 1$ if $c_k \leq 0$ and $\tilde{c}_k = c_k$ if $c_k > 0$ which is such that $X_k / \tilde{c}_k \xrightarrow{P} > 1$; also if c_k is ultimately negative we can define $\tilde{c}_k = -1$ if $c_k > 0$, $\tilde{c}_k = c_k$ if $c_k < 0$ which is such that $X_k/ \tilde{c}_k \xrightarrow{P} 1$; we can take it that we always have $c_k > 0$ or $c_k < 0$. But if $X_k / c_k \xrightarrow{P} 1$ and $c_k < 0$ then $(-X_k)/ (-c_k) \xrightarrow{P} 1$, and so we can — with a change of sign for X_k if needed — always take $c_k > 0$; an alternative was always to take $c_k > 0$ and have $X_k / c_k \xrightarrow{P} -1$ or $X_k / c_k \xrightarrow{P} 1$. Note that the choice between $c_k > 0$ and $c_k < 0$ is easy. If MLLN is valid with $c_k > 0$ we have, $F_k(x)$ denoting $\text{Prob}\,\{X_k \leq x\}$, $F_k(0) = \text{Prob}\,\{X_k \leq 0\}$ $\leq \text{Prob}\,\{X_k \leq c_k\,(1 - \in)\} = \text{Prob}\,\{X_k / c_k - 1 < - \in\} \leq \text{Prob}\,\{|X_k / c_k - 1| > \in\} \to 0$ and $F_k(0) \to 0$; in reverse, if the MLLN is valid with $c_k < 0$ we have $F_k(0) = \text{Prob}\,\{X_k \leq 0\} \geq \text{Prob}\{X_k \leq c_k\,(1 - \in)\} \geq \text{Prob}\,\{|X_k / c_k - 1| < \in\} \to 1$. $F_k(0) \to 0$ and $F_k(0) \to 1$ are the indicators for the choice of the sign of c_k. $F_k(0) \to 0$ means that $\{X_k > 0\}$ has a probability close to 1 when k is large and so X_k can, then, be considered practically positive; dually, if

$F_k(0) \to 1$, X_k, for large k, is practically negative. This frame is sufficient for the study of maxima, minima being dealt with by decreasing transformations.

Let us obtain some simple consequences :

1) if $X_k \geq 0$ (i.e., $F_k(0) = 0$) verifies the MLLN, with coefficients $c_k > 0$, then $\log X_k$ verifies the ALLN with coefficients $\log c_k$; more generally if $\{X_k\}$ verifies the MLLN with $c_k > 0$ then $\log \max(c_k, X_k)$ verifies the ALLN with $\log c_k$ as coefficients; conversely, if X_k verifies the ALLN with coefficients c_k then $\exp X_k$ (>0) verifies the MLLN with coefficients $\exp c_k$(>0);

2) if $\{X_k\}$ verifies the ALLN, with $c_k > 0$, then $\{X_k\}$ verifies the MLLN with the same coefficients c_k : it is a simple consequence of $X_k - c_k = c_k (X_k / c_k - 1) \xrightarrow{P} 0$

if $\lim c_k > 0$; also if $c_k < 0$ and $\lim c_k < 0$ the same happens, $(-X_k)$ and $(-c_k)$ are in the previous conditions and so $(-X_k / -c_k) \xrightarrow{P} 1$.

Let us now obtain the necessary and sufficient conditions to have ALLN for $\{X_k\}$ with coefficients $\{c_k\}$ and the MLLN for $\{X_k\}$ with coefficients $\{c_k > 0\}$.

The condition $X_k - c_k \xrightarrow{P} 0$ is equivalent to $\text{Prob}\{|X_k - c_k| \leq \epsilon\} \to 1$, $\forall \epsilon > 0$, i.e., $F_k(c_k + \epsilon) - F_k((c_k - \epsilon)^-) \to 1$ and so $F_k(c_k + \epsilon) \to 1$ and $F_k((c_k - \epsilon)^-) \to 0$; as $F_k((c_k + \epsilon)^-) \leq F_k(c_k - \epsilon) \leq F_k((c_k - \epsilon')^-) \to 0$ for $\epsilon' < \epsilon$ we get that $X_k - c_k \xrightarrow{P} 0$ implies $F_k(c_k + \epsilon) \to 1$ and $F_k(c_k - \epsilon) \to 0$; the converse is immediate and so ALLN is valid iff $F_k(c_k + \epsilon) \to 1$ and $F_k(c_k - \epsilon) \to 0, \forall \epsilon > 0$. By the same technique we see that the MLLN is valid (with $c_k > 0$) iff $F_k((1 + \epsilon) c_k) \to 1$ and $F_k((1 - \epsilon) c_k) \to 0$, $\forall \epsilon > 0$.

The laws of large numbers (LLN) for extremes

It is intuitive, at least in the i.i.d. case, that the maxima and minima of a sequence of samples should converge to the right–end point $\bar{w}$ (i.e., $F(x) < 1$ if $x < \bar{w}$, $F(x) = 1$ if $x \geq \bar{w}$) and to the left-end point $\underline{w}$ (i.e., $F(x) = 0$ if $x < \underline{w}$, $F(x) > 0$ if $x > \underline{w}$).

Let us suppose $\bar{w} < +\infty$. Then $\text{Prob}\,\{|\max_1^k \{X_i\} - \bar{w}| \le \epsilon\} = 1 - F^k((w-\epsilon)^-)$ by the definition of $\bar{w}$, which also implies $F^k((\bar{w}-\epsilon)^-) \to 0$ and so $\max\{X_i\} \xrightarrow{P} \bar{w}$. Suppose, now, that $\bar{w} = +\infty$; then $\text{Prob}\,\{\max_1^k \{X_i\} > M\} = 1 - F^k(M) \to 1$. In both cases we have shown that $\max_1^k \{X_i\} \xrightarrow{P} \bar{w}$. In the same way we can show that $\min_1^k \{X_i\} \xrightarrow{P} \underline{w}$ (finite or infinite). These results are independent of the knowledge of the distribution function $F(x)$, whether it is continuous, discrete or a mixture of both. They were shown in the proper (i.e., with zero probability at the points $\pm\infty$ if we are dealing with the complete real line) and non–degenerate (i.e., a random variable not taking some finite value with probability 1) cases; they can be extended to these situations but there seems little point in doing so.

But, with some knowledge of $F(x)$, we can seek some *deterministic* sequences, essentially for $\underline{w} = -\infty$ and $\bar{w} = +\infty$, that in some way 'measure' the type of increase (for maxima) and of decrease (for minima) of i.i.d. samples.

Although we could try to use a convenient sequence of increasing or decreasing functions, we will only consider the situations leading to an ALLN or MLLN. As stated, we stress that we are dealing only with an i.i.d. situation with distribution function $F(x)$, the increasing sample being denoted by $\{X_k, k = 1,2,\ldots\}$.

We say that $\{X_k\}$ verifies the ALLN for maxima, or that $\max_1^k \{X_i\}$ verifies the ALLN, if there exist constants c_k (not uniquely defined, as said before) such that $\max_1^k \{X_i\} - c_k \xrightarrow{P} 0$. From the above we see that if $\bar{w} < +\infty$ we can take $c_k = \bar{w}$. Let us consider the case $\bar{w} = +\infty$. We have ALLN iff $F^k(c_k + \epsilon) \to 1$ and $F^k(c_k - \epsilon) \to 0$: this will be shown to be equivalent, when $\bar{w} = +\infty$, to $\dfrac{1-F(y+x)}{1-F(y)} \to 0$ if $x > 0$. From $F^k(c_k + \epsilon) \to 1$ or $k \log F(c_k + \epsilon) \to 0$ as $\dfrac{\log u}{u-1} \to 1$ when $u \to 1$ we get $k(1 - F(c_k + \epsilon)) \to 0$ and $F^k(c_k - \epsilon) \to 0$ is equivalent to $k(1 - F(c_k - \epsilon)) \to +\infty$. Note that we can always take c_k as non–decreasing, substituting c_k by $\max(c_1, \ldots, c_k)$: it is evident that if $F^k(c_k + \epsilon) \to 1$ also $F^k(\max(c_1, \ldots, c_k) + \epsilon) \to 1$. We need to prove, now, that $F^k(\max(c_1, \ldots, c_k) - \epsilon) \to 0$ also : as $F^k(c_k - \epsilon) \to 0$, for $k > N(\delta)$ ($\delta > 0$, fixed) we have $F^k(c_k - \epsilon) < \delta$. Take then $k > N(\delta)$ and we have

$$F^k(\max(c_1, \dots, c_k) - \epsilon) = \max_{1}^{k} (F^k(c_i - \epsilon)) = \max (\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon), \max_{i=N(\delta)+1}^{k} F^k(c_i - \epsilon)).$$

But $F^k(c_i - \epsilon) \leq F^i(c_i - \epsilon) < \delta$ if $i > N(\delta)$ and as $N(\delta)$ if fixed $\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon) \to 0$ and so for $k > N'(\delta)$ we have $\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon) < \delta$.

Consequently for $k \geq \max(N(\delta), N'(\delta))$ we have $F^k(\max(c_1, \dots, c_k) - \epsilon) < \delta$ and $F^k(\max(c_1, \dots, c_k) - \epsilon) \to 0$. We can thus suppose $c_k \leq c_{k+1}$ and $c_k \to +\infty$. Situations like guaranteeing that for $k > N(\delta)$ some event or property happens will be described "as for large k" – its formalization is made as previously.

Then for y large we can obtain k such that $c_k \leq y \leq c_{k+1}$ and so we get $1 - F(c_k + \eta) \geq 1 - F(y + \eta)$ and $1 - F(c_{k+1} - \eta) \leq 1 - F(y - \eta)$ so that

$$\frac{1 - F(y+\eta)}{1 - F(y - \eta)} \leq \frac{1 - F(c_k + \eta)}{1 - F(c_{k+1} - \eta)}$$ and as $k(1 - F(c_k + \epsilon)) \to 0$ and $k(1 - F(c_k - \epsilon)) \to \infty$ we see that $\frac{1 - F(y + \eta)}{1 - F(y - \eta)} \to 0$ as $y \to \infty$.

Let us now prove the converse, also introducing one way of calculating the $\{c_k\}$. Let us define c_k such that $F(c_k^-) \leq 1 - 1/k \leq F(c_k)$ and so $c_k \to \infty$; $\{c_k\}$ is non–decreasing. From the condition $\frac{1 - F(y + x)}{1 - F(y)} \to 0$ as $y \to \infty$ or under the form $\frac{1 - F(y + \eta)}{1 - F(y - \eta)} \to 0$ as $y \to \infty$ (used above) we see that for $\epsilon > \eta > 0$ we have $k(1 - F(c_k + \epsilon)) \leq \frac{1 - F(c_k + \epsilon)}{1 - F(c_k)} \to 0$ and so $k(1 - F(c_k + \epsilon)) \to 0$; also $k(1 - F(c_k - \epsilon)) \geq \frac{1 - F(c_k - \epsilon)}{1 - F(c_k^-)} \to +\infty$. Thus we have shown that when $\bar{w} = +\infty$ the ALLN is valid iff $\frac{1 - F(y + x)}{1 - F(y)} \to 0$ as $y \to \infty$ and that *one* system of coefficients $\{c_k\}$ is given by $F(c_k^-) \leq 1 - 1/k \leq F(c_k)$.

As a summary we have seen that :

The ALLN for i.i.d. maxima is valid either when $\bar{w} < +\infty$ *or* $\bar{w} = +\infty$ *iff*

$$\frac{1 - F(y + x)}{1 - F(y)} \rightarrow 0 \text{ if } y \rightarrow +\infty \text{ for any } x > 0.$$

It is now easy to obtain the MLLN. If $0 < \bar{w} < +\infty$ it is obvious that $\max_1^k X_i / \bar{w} \xrightarrow{P} 1$. Consider then the case $\bar{w} = +\infty$. Then $\max_1^k X_i / c_k \xrightarrow{P} 1$, $(c_k > 0)$, as $\bar{w} = +\infty$, is equivalent to $\max_1^k X_i^+/c_k \rightarrow +\infty$ where $X^+ = \max(0,X)$ and so equivalent to the ALLN for $\max_1^k \{\log X_i^+\} - \log c_k \xrightarrow{P} 0$.

Denoting by $F^+(x) = \text{Prob}\{\log X_i^+ \leq x\}$, i.e., $F^+(x) = 0$ if $x < 0$, $F^+(x) = F(e^x)$ if $x \geq 0$ the conditions for the ALLN and for obtaining the $\{c_k\}$ for $\max_1^k \{\log X_i^+\}$ are $\frac{1 - F^+(y + x)}{1 - F^+(y)}$ $\rightarrow 0$ as $y \rightarrow \infty$ and $F^+(\log c_k^-) \leq 1 - 1/k \leq F^+(\log c_k)$ or, returning to $F(.)$, and substituting e^y and e^x by y and $x(>1)$, $\frac{1 - F(y\,.\,x)}{1 - F(y)} \rightarrow 0$ as $y \rightarrow \infty$ and $F(c_k^-) \leq 1 - 1/k \leq F(c_k)$.

As a summary we have seen that :

The MLLN for i.i.d. maxima is valid either when $0 < \bar{w} < +\infty$ *or when* $\bar{w} = +\infty$ *iff* $\frac{1 - F(y\,.\,x)}{1 - F(y)} \rightarrow 0$ *if* $y \rightarrow +\infty$ *for any* $x > 1$.

The conditions on $F(x)$ can be written as conditions on $S(x)$.

For minima we can obtain the ALLN and MLLN from the previous results using the fact that $\min_1^k \{X_i\} = -\max_1^k \{-X_i\}$, or directly.

Then if $\underline{w} > -\infty$ the ALLN is valid and also if $\underline{w} > 0$ the MLLN is valid. Consider now the case for $\underline{w} = -\infty$; the ALLN is valid iff $\frac{F(y)}{F(y+x)} \rightarrow 0$ as $y \rightarrow -\infty$ with $x > 0$; *one* system of coefficients is $\{c_k\}$ (non–increasing) given by $F(c_k^-) \leq 1/k \leq F(c_k)$; the MLLN has no meaning if we continue to use $c_k > 0$; if we accept it as $c_k < 0$ then it is valid iff $\frac{F(y)}{F(y+x)} \to 0$ as $y \to -\infty$, $x > 1$; *one* system of $\{c_k\}$ $(c_k < 0)$ is given by $F(c_k^-) \le 1/k \le F(c_k)$.

The statements for the ALLN and MLLN for i.i.d. minima are the conversion of the ones for i.i.d. maxima with the substitution of $\bar{w}$ *by* $\underline{w}$ *and* $1 - F(x)$ *by* $F(-x)$.

The initial study connected with LLN is the one by Dodd (1923), using different terminology; the modern aspect appears, with unnecessary restrictions, in de Finneti (1932) and Gnedenko (1943); the essential aspects are contained in Galambos (1978) and Tiago de Oliveira ed. (1984).

Some examples concerning the LLN for extremes

Let us now give some examples that can clarify for maxima the border between the non–validity of the LLN, the validity of ALLN (and consequently of the MLLN if $\bar{w} > 0$) and the validity of MLLN.

1. Consider, as the first example, the uniform distribution in $[a,b]$. As $\bar{w} = b$ the ALLN for maxima is valid but the MLLN is valid only if $b > 0$; for minima as $\bar{w} = a$ the ALLN is valid and the MLLN is valid with $c_k = a$ if $a \neq 0$, if we accept the possibility of $c_k < 0$.

We will now consider cases where $\bar{w} = +\infty$ (and $\underline{w} = -\infty$).

2. Take as the second example the standard exponential distribution $E(x) = 0$ if $x \le 0$, $E(x) = 1 - e^{-x}$ if $x \ge 0$ ($\underline{w} = 0$, $\bar{w} = +\infty$). As $\frac{1 - E(y+x)}{1 - E(y)} = e^{-x} \neq 0$ the ALLN is not valid but the MLLN is valid because $\frac{1 - E(y\,.\,x)}{1 - E(y)} = e^{-(x-1)y} \to 0$ when $y \to +\infty$ with $x > 1$ (with $c_k = \log k$); for minima the ALLN is trivially valid as $\underline{w} = 0$ but the MLLN is not.

3. Consider now the standard normal distribution with $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt$ ($\underline{w} = -\infty$, $\bar{w} = +\infty$), which is symmetrical and so we need to deal only with maxima.

We have

$$\frac{1-N(y+x)}{1-N(y)} = \frac{\int_{y+x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt}{\int_{y}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt} \sim \frac{e^{-(y+x)^2/2}}{e^{-y^2/2}} \rightarrow 0$$

and the ALLN is valid; also MLLN is valid because

$$\frac{1-N(y\,.\,x)}{1-N(y)} = \frac{\int_{y.x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt}{\int_{y}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt} \sim x\,\frac{e^{-y^2x^2/2}}{e^{-y^2/2}} \rightarrow 0$$

because $x > 1$; for $c_k = N^{-1}(1 - 1/k)$ we can use the easily obtained approximation $\tilde{c}_k = \sqrt{2 \log k}$ using $k(1 - N(\tilde{c}_k + \in)) \rightarrow 0$, $k(1 - N(\tilde{c}_k - \in)) \rightarrow +\infty$ and l'Hôpital's rule.

4. The geometric or Pascal distribution $F(x) = 0$ if $x < 0$, $F(x) = 1 - e^{-\theta[x]}$ for $x \geq 0$ (with $[x]$ denoting the integer part of x) is a discrete distribution with jumps at the integers $x = 0,1,2,\ldots$, $\underline{w} = 0$, $\bar{w} = +\infty$ (sometimes written as Prob $\{X = m\} = (1-p)\,p^{m-1}$, with $\theta = -\log p$).

Then $\dfrac{1 - F(y + x)}{1 - F(y)} = \dfrac{e^{-\theta[y+x]}}{e^{-\theta[y]}}$ which for $x = 1$ takes the constant value $e^{-\theta}$ and so the ALLN is not valid but $\dfrac{1 - F(y\,.\,x)}{1 - F(y)} = \dfrac{e^{-\theta[y\,.\,x]}}{e^{-\theta[y]}} \leq e^{-\theta\{(x-1)\,y-1\}} \rightarrow 0$ as $y \rightarrow \infty$ and so the MLLN is valid.

Let us consider three examples connected with the asymptotic distributions of extremes, to be dealt with in the next sections: the Gumbel distribution and the Fréchet distribution for maxima, and the Weibull distribution for minima. They seem to be, in applications, the most important ones.

5. For the reduced Gumbel distribution we have $\Lambda(x) = \exp(-e^{-x})$, $\underline{w} = -\infty$, $\bar{w} = +\infty$.

Then for maxima as $\dfrac{1 - \Lambda(y + x)}{1 - \Lambda(y)} \sim e^{-x}\, e^{-e^{-y}\,(e^{-x}-1)} \rightarrow e^{-x}$ and the ALLN is not valid; but the MLLN is valid because $\dfrac{1 - \Lambda(y\,x)}{1 - \Lambda(y)} \rightarrow 0$ when $y \rightarrow +\infty$ and $x > 1$. The coefficients are $c_k = -\log\log\frac{k}{k-1}$ or $\tilde{c}_k = \log k$; this result could be expected because, for large x, $\Lambda(x)$ and $E(x)$ behave similarly. For minima we have $\frac{\Lambda(y)}{\Lambda(y+x)} \to 0$ as $y \to -\infty$, $x > 0$ and so ALLN is valid, a set of coefficients being $c_k = -\log\log k$; accepting $c_k < 0$ the MLLN is also true as $\frac{\Lambda(y)}{\Lambda(y\,.\,x)} \to 0$ as $y \to -\infty$ if $x > 1$: the coefficients $c_k\ (\le 0)$ may be the same.

6. Consider now the reduced Fréchet distribution for maxima $\Phi_\alpha(x) = 0$ if $x \le 0$, $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $\alpha > 0$, if $x \ge 0$ The ALLN is not valid as $\frac{1-\Phi_\alpha(y+x)}{1-\Phi_\alpha(y)} \to 1$ as $y \to +\infty$ $(x > 1)$, nor is the MLLN as $\frac{1-\Phi_\alpha(y\,.\,x)}{1-\Phi_\alpha(y)} \to x^{-\alpha}$ as $y \to +\infty$ $(x > 1)$. For minima we have $\frac{\Phi_\alpha(y)}{\Phi_\alpha(y+x)} \to 0$ for $y \to 0$ and so the ALLN is valid, a system of coefficients being $c_k = (\log k)^{-1/\alpha}$; here the MLLN has direct meaning and is valid because we have $\frac{\Phi_\alpha(y)}{\Phi_\alpha(y\,.\,x)} \to 0$ as $y \to 0^+$, $x > 1$, the coefficients being the same.

7. For the reduced Weibull distribution for minima $W_\alpha(x) = 0$ if $x \le 0$, $W_\alpha(x) = 1 - \exp(-x^\alpha)$ if $x \ge 0$, $\alpha > 0$, we have $\frac{1-W_\alpha(y+x)}{1-W_\alpha(y)} = \exp\{y^\alpha - (y+x)^\alpha\} \to 0$ iff $\alpha > 1$, and so only if $\alpha > 1$ the ALLN is valid with coefficients $c_k = (\log k)^{1/\alpha}$; the MLLN is valid as $\frac{1-W_\alpha(y)}{1-W_\alpha(y\,.\,x)} \to 0$ as $y \to \infty$ with $x > 1$, the coefficients being the same. For minima we get $\frac{W_\alpha(y)}{W_\alpha(y+x)} \to 0$ $(x > 0)$, and so the ALLN is valid with coefficients $c_k = (-\log(1-1/k))^{1/\alpha}$ or more simply $c_k = k^{-1/\alpha}$; the MLLN is not valid as $\frac{W_\alpha(y)}{W_\alpha(y\,.\,x)} \to x^{-\alpha}$ as $y \to 0^+$ $(x > 1)$. Notice that $W_1(x) = E(x)$, the standard exponential distribution, and so the results could be expected.

Note that the non–validity of the MLLN in the last two cases is associated with the attraction conditions for the Fréchet and Weibull distribution of extremes, as will be seen in one of the next sections.

Although we have analysed — as an exercise — the ALLN and MLLN in each case, part of the conclusions could be reached if we used the relations between the LLN given in a previous section.

8. Consider now, finally, as a counter-example, the distribution function $F(x) = 0$ if $x \le e$, $F(x) = 1 - 1/\log x$ if $x \ge e$; it has $\underline{w} = e$ and $\bar{w} = +\infty$.

For maxima as $\frac{1 - F(y + x)}{1 - F(y)} \to 1$ when $y \to +\infty$ $(x > 0)$ the ALLN is not valid as well as the MLLN because $\frac{1 - F(y \,.\, x)}{1 - F(y)} \to 1$ when $y \to +\infty$ $(x > 1)$. For minima as $\underline{w} = e$ the ALLN is trivially valid as well as the MLLN.

The asymptotic distributions of extremes - some examples

The results that are contained in this and the following sections constitute the initial core of *Statistical Extremes Theory* and are the basis for many applications. So we will proceed at a slower pace, with considerable independence from what was said about the LLN.

As has been said many times, the distribution of maxima and minima in a i.i.d. univariate sample $(X_1,\dots, X_k)$ is given by

$$\text{Prob}\{\max_1^k \{X_i\} \le x\} = F^k(x) \text{ and } \text{Prob}\{\min_1^k \{X_i\} \le x\} = 1 - (1 - F(x))^k \,;$$

for survival functions (which satisfy $F(x) + S(x) = 1$) we have

$$\text{Prob}\{\max_1^k \{X_i\} > x\} = 1 - (1 - S(x))^k \text{ and } \text{Prob}\{\min_1^k \{X_i\} > x\} = S^k(x).$$

From these formulae we see that $F(x)$ and $S(x)$ play symmetrical roles for maxima and minima.

In the case where $F(x)$ or $S(x)$ are known, we could proceed to a classical probabilistic and statistical analysis or, if $F(x)$ or $S(x)$ are known to be continuous, symmetrical, etc., in some cases we could have recourse to non–parametric methods. But, in many cases we do not even know if the observations come from some known parametric family of distributions and, in general, we are not under the i.i.d. hypothesis. So we have to resort, in practical applications, to asymptotic results of the type: if for some $\{F_k(x)\}$ *attraction coefficients* $\{\lambda_k, \delta_k > 0\}$ exist such that $F_k(\lambda_k + \delta_k x) \xrightarrow{w} L(x)$, *with* $L(x)$ *continuous (and so uniform convergence)*, as $F_k(\lambda_k + \delta_k x)$ is close to $L(x)$ we will take $L(\frac{y-\lambda'}{\delta'})$ as an approximation to $F_k(y)$, (λ', δ') not being necessarily (λ_k, δ_k) to allow for a better fit in statistical analysis. It must be said, at this preliminary stage, that the i.i.d conditions are not essential and can be weakened — as a rule the margins of the sequence $\{X_1,\ldots,X_k,\ldots\}$ should not be very different and the correlation/association between X_i and X_j must wane out as the distance between i and j increases; two examples will shed some light on this question.

As the relation $\min_1^k\{X_i\} = -\max_1^k\{-X_i\}$ is true we can deal only with maxima or with minima and translate the results if necessary; we will deal, almost always, with maxima.

Thus, our purpose now is to obtain limiting (proper and non–degenerate) distributions of $F^k(\lambda_k + \delta_k x)$. When the asymptotic or limiting distributions were degenerate — Laws of Large Numbers — the three previous examples (5., 6. and 7.) correspond to the three possible limiting distributions and the last one to a case where there does not exist a limiting distribution, as we shall see later. Let us say, as will be shown in the next section, that if $\tilde{L}(x)$ is a possible limiting distribution for maxima then $\underset{\sim}{L}(x) = 1 - \tilde{L}(-x)$ is a possible limiting distribution for minima; but that the same distribution can't have the corresponding distributions $\underset{\sim}{L}$ and $\tilde{L}$, as seen in some examples below; also it can happen that one of $\underset{\sim}{L}$ or $\tilde{L}$ does not exist but the other one exists or, even, that both do not exist. Let us finally recall that (λ_k, δ_k) are not uniquely defined: by Khintchine's convergence of types theorem, (λ_k, δ_k) and $(\lambda_k', \delta_k'$, such that $(\lambda_k' - \lambda_k)/\delta_k \to A$ and $\delta_k'/\delta_k \to B$ $(-\infty < A < +\infty, 0 < B < +\infty)$ as $k \to \infty$, lead to limiting distributions of the same type, respectively $L(x)$ and $L(A + Bx)$ (and conversely), the sets $\{(\lambda_k', \delta_k')\}$ being thus equivalent for limiting purposes; in many cases we have $A = 0$ and $B = 1$, the *total* equivalence. For Khintchine's convergence of types theorem see the Annex II to Part I.

A calculation facility is convenient: as for maxima $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$ is equivalent to $k \log F(\lambda_k + \delta_k x) \to \log \tilde{L}(x)$, as $F(\lambda_k + \delta_k x) \to 1$ if $0 < \tilde{L}(x) < 1$, and $\frac{\log u}{u-1} \to 1$ as $u \to 1$, the convergence relation $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$ is equivalent to $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$; also $1-(1-F(\lambda_k + \delta_k x))^k \xrightarrow{w} \underset{\sim}{L}(x)$ is equivalent, as $F(\lambda_k + \delta_k x)) \to 0$, to $k\,F(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ in the region $0 < \underset{\sim}{L}(x) < 1$. Those equivalent convergence relations (which are not uniform) $k(1-F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$ and $k\,F(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ are very convenient for calculations; corresponding relations can be written for survival functions.

The three reduced limiting distributions for maxima $(L(x))$ are $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$ if $x \le 0$ with $\alpha > 0$, $\Psi_\alpha(x) = 1$ if $x \ge 0$ (Weibull distribution), $\Lambda(x) = \exp(-e^{-x})$ (Gumbel distribution) and $\Phi_\alpha(x) = 0$ if $x < 0$, $\Psi_\alpha(x) = \exp\{-x^{-\alpha}\}$ if $x \ge 0$ with $\alpha > 0$) (Fréchet distribution); the corresponding distributions for minima $\underset{\sim}{L}(x))$ are $W_\alpha(x) = 1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x) = 1 - \exp\{-e^x\}$ and $1 - \Phi_\alpha(-x)$, with the same denominations. Notice that the following *stability equations* will be shown: $\tilde{L}^k(x) = L(\alpha_k + \beta_k\, x)$ and $\underset{\sim}{S}^k(x) = \underset{\sim}{S}(\alpha_k + \beta_k\, x)$ for convenient (α_k, β_k). Note that $(\lambda_{k+1} + \delta_{k+1})$ are, in both cases, also attraction coefficients by Khintchine's convergence of types theorem; this is left as exercise.

We will now give some examples that lead to some of the distributions above : note that each of $\tilde{L}(x)$ or of $\underset{\sim}{L}(x)$ is a limiting distribution for itself, as will be seen in some examples. All these examples will have a full justification in the next sections.

1. Take $F(x) = \Lambda(x)$. Then $F^k(\lambda_k + \delta_k x) = \Lambda^k(\lambda_k + \delta_k x) = \Lambda(\lambda_k + \delta_k x - \log k) = \Lambda(x)$ when $\lambda_k = \log k$ and $\delta_k = 1$ and so $\tilde{L}(x) = \Lambda(x)$; $\Lambda(x)$ is thus stable for maxima. Recall, once more, that equivalent coefficients could be used; in general, we will try the simpler and more manageable ones.

Suppose we are now dealing with minima of Gumbel distribution, we should have $k\,\Lambda(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ and it is easy to see that if $\lambda_k = -\log\log k$, $\delta_k = 1/\log k$ then $1-(1-F(\lambda_k + \delta_k x))^k \to 1 - \Lambda(-x)$. What is important is the behaviour of the right tail for maxima and the left tail for minima, and they can be very different which is not the case: note that for $x > 0$ the ratio of the right tail to the left one is $(1 - \Lambda(x))/\Lambda(-x) \to 1$ as $x \to +\infty$; only when this ratio is 1 the distribution would be the pair $\tilde{L}(x)$ and $\underset{\sim}{L}(x) = 1 - \tilde{L}(x)$; if the limit of the ratio is $c(0 < c < +\infty)$ then $\tilde{L}$ and $\underset{\sim}{L}$ would be connected in the same way apart from a power transformation.

2. Consider the exponential distribution: $F(x) = E(x) = W_1(x)$ where $E(x) = 0$ if $x < 0$, $E(x) = 1 - e^{-x}$ if $x \geq 0$. As $F^k(\lambda_k + \delta_k x) \overset{w}{\to} \tilde{L}(x)$ is equivalent to $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$, we must have $k\, e^{-\lambda_k - \delta_k x} \to -\log \tilde{L}(x)$ when $\lambda_k + \delta_k x > 0$; a solution is $\lambda_k = \log k$, $\delta_k = 1$ and $\tilde{L}(x) = \Lambda(x)$.

As far as minima are concerned, we should have $k(1 - e^{-\lambda_k - \delta_k x}) \to -\log(1 - \tilde{L}(x))$ for $\lambda_k + \delta_k x > 0$; taking $\lambda_k = 0$, $\delta_k = 1/k$ we get $\tilde{L}(x) = E(x) = W_1(x)$ and the exponential distribution is stable for minima.

3. Take now the Pareto distribution: $F(x) = 0$ if $x \leq 1$, $F(x) = 1 - x^{-\alpha}$ if $x \geq 1$, with $\alpha > 0$. To have $F^k(\lambda_k + \delta_k x) \to \tilde{L}(x)$, we must have $k(1 - F(\lambda_k + \delta_k x)) = k(\lambda_k + \delta_k x)^{-\alpha} \to -\log \tilde{L}(x)$ when $\lambda_k + \delta_k x \geq 1$; a solution is $\lambda_k = 0$, $\delta_k = k^{1/\alpha}$ and the limiting distribution is $\tilde{L}(x) = \Phi_\alpha(x)$.

For minima we must have $k\left(1 - \dfrac{1}{(\lambda_k + \delta_k x)^\alpha}\right) \to -\log(1 - \underset{\sim}{L}(x))$ when $\lambda_k + \delta_k x \geq 1$; if we take $\lambda_k = 1$, $\delta_k = 1/\alpha k$ we get $\underset{\sim}{L}(x) = E(x) = W_1(x)$, the exponential distribution, which is a Weibull form for minima and not a Fréchet one.

4. When we consider the uniform distribution in $[0, 1]$, $F(x) = 0$ if $x \leq 0$, $F(x) = x$ if $0 \leq x \leq 1$ and $F(x) = 1$ if $x \geq 1$, the condition is $k(1 - \lambda_k - \delta_k x) \to -\log(1 - \tilde{L}(x))$ when $0 \leq \lambda_k + \delta_k x \leq 1$; taking $\lambda_k = 1$ and $\delta_k = 1/k$ we get $\tilde{L}(x) = \Psi_1(x) = e^x$ if $x \leq 0$, $\Psi_1(x) = 1$ if $x \geq 0$, i.e., the Weibull distribution for maxima.

For minima we should have $k(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ when $0 \leq \lambda_k + \delta_k x \leq 1$ which is obtained with $\lambda_k = 0$, $\delta_k = 1/k$ and $\underset{\sim}{L}(x) = W_1(x)$, the Weibull distribution for minima. Here situations are in correspondence because the uniform distribution is symmetrical about the mid–point 1/2.

We could study more cases of classical distributions but it does not seem necessary except for the normal, the logistic being a simple exercise. As seen, (λ_k, δ_k) were given with a hint about the way we could have arrived at them; the standard techniques will be explained in the next section.

5. Consider now the standard normal distribution function $N(x)$. It can be shown very easily, using the attraction conditions to be given later, that the possible limits of $N^k(\lambda_k + \delta_k x)$ cannot be either $\Psi_\alpha(x)$ or $\Phi_\alpha(x)$, Weibull and Fréchet distributions. We should thus have, if the limit exists, $N^k(\lambda_k + \delta_k x) \to \Lambda(x)$ or $k(1 - N(\lambda_k + \delta_k x)) \to e^{-x}$, where as known $N(\lambda_k + \delta_k x) \to 1$ and thus $\lambda_k + \delta_k x \to +\infty$.

A very well known result is that $1 - N(x) \sim \dfrac{N'(x)}{x}$ when $x \to +\infty$ and thus the previous condition is equivalent to

$$\log k + \log N'(\lambda_k + \delta_k x) - \log(\lambda_k + \delta_k x) + x \to 0 \quad \text{as} \quad k \to \infty$$

or

$$\log k - \log\sqrt{2\pi} - \frac{1}{2}(\lambda_k + \delta_k x)^2 - \log(\lambda_k + \delta_k x) + x = o(1);$$

dividing by $\log k$, we also see that $\dfrac{(\lambda_k + \delta_k x)^2}{2\log k} \to 1$ as $\dfrac{\log(\lambda_k + \delta_k x)}{(\lambda_k + \delta_k x)^2} \to 0$ and thus $2\log(\lambda_k + \delta_k x) - \log 2 - \log\log k \to 0$ and so $\log(\lambda_k + \delta_k x) = \dfrac{1}{2}(\log 2 + \log\log k) + o(1)$. Substituting above $\log(\lambda_k + \delta_k x)$ we get

$$(\lambda_k + \delta_k x)^2 = 2\{\log k + x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k\} + o(1) =$$

$$= 2\log k \times \{1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k}{\log k} + o(\frac{1}{\log k})\},$$

$$\lambda_k + \delta_k x = \sqrt{2\log k}\,(1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k}{2\log k} + o(\frac{1}{\log k}))$$

and thus we can take

$$\lambda_k = \sqrt{2\log k}\,(1 - \frac{\log 4\pi + \log\log k}{4\log k}) = \sqrt{2\log k} - \frac{\log 4\pi + \log\log k}{2\sqrt{2\log k}}$$

and $\delta_k = \dfrac{1}{\sqrt{2 \log k}}$.

The computation is easy, although tedious. But the verification is simple using $\sqrt{2 \log k} = t \to +\infty$.

6. Let us study the last distribution function considered in the previous section : $F(x) = 0$ if $x \le e$, $F(x) = 1 - \dfrac{1}{\log k}$ if $x \ge e$.

By the attraction conditions to be given later, it is immediately excluded that, for some (λ_k, δ_k), $F^k(\lambda_k + \delta_k x)$ would converge to $\Psi_\alpha(x)$ or $\Phi_\alpha(x)$; the exclusion of $\Psi_\alpha(x)$ as a possible limit comes from the fact $\bar{w} = +\infty$; the other exclusion is almost as easy.

Let us show, directly, that $F^k(\lambda_k + \delta_k x)$ cannot converge (weakly) to $\Lambda(x)$ or, equivalently, that $k(1 - F(\lambda_k + \delta_k x))$ does not converge to $-\log \Lambda(x) = e^{-x}$. If it did converge we would have $F(\lambda_k + \delta_k x) \to 1$ and so it is sufficient to study the function in the right tail: we should thus have $\dfrac{k}{\log(\lambda_k + \delta_k x)} \to e^{-x}$ $(\lambda_k + \delta_k x > 1)$ or $\dfrac{\log(\lambda_k + \delta_k x)}{k} \to e^x$.

For $x = 0$ we have $\dfrac{\log \lambda_k}{k} \to 1$ and so $\lambda_k \to +\infty$ and $\lambda_k > 0$ ultimately.

Then we would have $\dfrac{\log(\lambda_k + \delta_k x)}{k} = \dfrac{\log \lambda_k + \log(1 + \delta_k/\lambda_k \,.\, x)}{k} \to e^x$ or $\dfrac{1}{k} \log(1 + \delta_k/\lambda_k \,.\, x) \to e^{-x} - 1$. Let a (≥ 0) be one of the possible limits of δ_k/λ_k : using the subsequence of $\{k\}$ such that $\delta_k/\lambda_k \to a$ if $a < +\infty$ we get $\dfrac{1}{k} \log(1 + \delta_k/\lambda_k \,.\, x) \to 0 \ne e^x - 1$; suppose now $a = +\infty$ and $x > 0$: then $\dfrac{\log(1 + \delta_k/\lambda_k \,.\, x)}{\log(\delta_k/\lambda_k \,.\, x)} \dfrac{\log(\delta_k/\lambda_k \,.\, x)}{k} \to e^x - 1$ is impossible. Thus if $F(x) = 0$ if $x \le e$, $F(x) = 1 - 1/\log x$ if $x \ge e$, it does not exist (λ_k, δ_k) such that $F^k(\lambda_k + \delta_k x)$ has a proper and nondegenerate limiting distribution. For minima we should have $k\,F(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ or equivalently $k\left(1 - \dfrac{1}{\log(\lambda_k + \delta_k x)}\right) \to -\log(1 - \underset{\sim}{L}(x))$ and so $\dfrac{1}{\log(\lambda_k + \delta_k x)} \to 1$ or $\lambda_k + \delta_k x \to e$.
Let us take $\lambda_k = e$ and $\delta_k = e/k$: we get $k\left(1 - \dfrac{1}{(\log e + ex/k)}\right) = \dfrac{k \log(1 + x/k)}{1 + \log(1 + x/k)} \to x = -\log(1 - \underset{\sim}{L}(x))$.

This shows that we have $\underset{\sim}{L}(x) = 1 - W_1(x) = E(x)$, which is the exponential distribution, an Weibull distribution for minima with $\alpha = 1$.

7. Let us now consider, finally, an example that shows how misleading some intuitive approximations can be. Consider the geometric or Pascal distribution $F(x) = 0$ if $x \leq 0$, $F(x) = 1 - e^{\theta[x]}$ if $x \geq 0$. We could expect that F(x) would behave practically like the exponential distribution to which it is very similar, say a discretized version. As $\overline{w} = +\infty$ we can exclude the Weibull distribution as a limit; by other conditions below we can exclude the Fréchet distribution as a limit. We still have the Gumbel distribution as a possible limit which we could expect to be the one as happens with exponential distribution.

The condition $k(1 - F(\lambda_k + \delta_k x)) \to e^{-x}$ is equivalent, to $k\, e^{-\theta[\lambda_k + \delta_k x]} \to e^{-x}$ or $\theta[\lambda_k + \delta_k x] - \log k \to x$. If this happens we will show that $\theta\, \delta_k = 1$, defining uniquely δ_k, which is impossible by Khintchine's theorem. In fact, for $x = 0$ we also get $[\lambda_k] - \log k \to 0$ and so $\lambda_k \to +\infty$, as could be expected. By subtraction we get $\theta\{[\lambda_k + \delta_k x] - [\lambda_k]\} \to x$; denoting by $r(\alpha) = \alpha - [\alpha]$ the fractional part of α, the previous relation can be written $\theta\{\delta_k x - (r(\lambda_k + \delta_k x) - r(\lambda_k))\} \to x$ and with $x > 0$ (to simplify) we get $\theta\, \delta_k \frac{r(\lambda_k + \delta_k x) - r(\lambda_k)}{x} \to 1$. As $|r(\alpha) - r(\beta)| < 1$, for k fixed, $\in (> 0)$ fixed and $k > N(\in)$ we have $1 - \in - \frac{1}{x} \leq \sigma\, \delta_k \leq 1 + \in + \frac{1}{x}$ and letting $x \to \infty$, we obtain finally $\delta_k = 1/\theta$, a *unique* result that is impossible so rejecting the Gumbel limit. But let us go further in the analysis. Accepting $\delta_k = 1/\theta$ we should have then $r(\lambda_k + x/\theta) - r(\lambda_k) \to 0$ as $k \to \infty$, $\forall\, x$. Suppose we fix x such that $0 < \varphi = x/\theta < 1$: we have $r(\lambda_k) = \tau_k$ $(0 \leq \tau < 1)$ and $r(\lambda_k + \varphi) = \tau_k + \varphi$ when $\tau_k + \varphi < 1$ and $\tau_k + \varphi - 1$ when $1 \leq \tau_k + \varphi\ (< 2)$ and so $r(\lambda_k + \varphi) - r(\lambda_k)$ has as possible limits φ and $\varphi - 1$, showing, once more, that a Pascal distribution cannot have, for maxima, a Gumbel distribution as a limit.

8. Also a Poisson distribution does not have a limiting distribution for maxima.

9. In principle all lattice distributions, i.e., discrete distributions with successive jumps at equal steps, can be dealt with by the use of a theorem of Gnedenko (1943) and in a simpler way. We have used those elementary proofs to clarify the issue from the start.

Let us now give two simple examples showing that i.i.d. conditions can be weakened.

Take a random i.i.d. sequence $\{X_i, i = 1,2, \dots\}$ where each random variable has the distribution function $F(x)$ and that $\max_1^k \{X_i\}$ has a limiting distribution $\underset{\sim}{L}(x)$ (as said before $\underset{\sim}{L}$ can only be Ψ_α, Λ, Φ_α apart from location-dispersion parameters), i.e., there exist attraction coefficients $(\lambda_k, \delta_k > 0)$ such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$. Now define the new dependent sequence $\{X_i^* = \max(X_i, X_{i+1}), i = 1,2, \dots\}$. We have

$$F_k^*(x) = \text{Prob}\{\max_1^k \{X_i^*\} \le x\} = \text{Prob}\{\max_1^{k+1} \{X_i\} \le x\} = F^{k+1}(x) \text{ and thus } \lambda_k^* = \lambda_{k+1}$$

$\delta_k^* = \delta_{k+1}$ is a system of attraction coefficients for the new (dependent) sequence leading to the same limiting distribution $\tilde{L}(x)$. We could also use the coefficients (λ_k, δ_k), by Khintchine's theorem as said before.

As concrete examples for $\{\max_1^k X_i^*\}$, consider the cases where $F(x)$ is one of the three limiting distributions:

1. if $F(x) = \Psi_\alpha$ we have the stability relation $\Psi_\alpha^k(x) = \Psi_\alpha(k^{1/\alpha} x)$, $\Psi_\alpha^k(k^{-1/\alpha} x) = \Psi_\alpha(x)$; so we can take $\lambda_k = 0$ and $\delta_k = k^{-1/\alpha}$ and thus we can use $\lambda_k^* = \lambda_k = 0$ and $\delta_k^* = \delta_k = k^{-1/\alpha}$;

2. if $F(x) = \Lambda(x)$ we have the stability equation $\Lambda^k = \Lambda(x - \log k)$ or $\Lambda^k(\log k + x) = \Lambda(x)$; so we can take $\lambda_k = \log k$, $\delta_k = 1$ and thus we can use $\lambda_k^* = \lambda_k = \log k$ and $\delta_k^* = \delta_k = 1$;

3. if $F(x) = \Phi_\alpha(x)$ we have the stability equation $\Phi_\alpha^k(x) = \Phi_\alpha^k(k^{-1/\alpha} x)$ and so $\Phi_\alpha^k(k^{1/\alpha} x) = \Phi_\alpha^k(x)$; so we can take $\lambda_k = 0$, $\delta_k = k^{1/\alpha}$; and thus we can also use $\lambda_k^* = \lambda_k = 0$, $\delta_k^* = \delta_k$.

Let us note that this is a special case of 1–dependence as $X_i^* = \max(X_i, X_{i+1})$ and $X_{i+1}^* = \max(X_{i+1}, X_{i+1})$ are dependent but $X_i^* = \max(X_i, X_{i+1})$ and $X_j^* = \max(X_j^*, X_{j+1})$ are independent if $|j - i| > 1$; all the cases of m–dependence (X_i and X_j independent if $|i - j| > m$) lead to the same situation.

Let us now consider the weakening of the identical distributions condition. Consider a sequence $\{X_i\}$ of i.i.d. random variables with distribution function F(x) and such that

$$\text{Prob}\ \{(\max_1^k\{X_i\} - \lambda_k)/\delta_k \le x\} = F^k(\lambda_k + \delta_k\, x) \to \tilde{L}(x) \text{ when } k \to \infty\ .$$

Define the new independent, but not identically distributed, sequence $\{X_i^*\}$ by the equation

$$X_i^* = X_i + (-1)^i.$$

Let $F_k^*(x) = \text{Prob}\ \{(\max_1^k\{X_i^*\} \le x)$. We have for the distribution of $\max_1^k\{X_i^*\}$:

for $k = 2\,p$, $F_k^*(x) = \text{Prob}\{X_1 - 1 \le x, X_3 - 1 \le x, \ldots, X_{2p-1} - 1 \le x\}\ \text{x}$

$$\text{x}\ \text{Prob}\{X_2 + 1 \le x, \ldots, X_{2p} + \le x\} = F^p(x+1)\, F^p(x-1);$$

for $k = 2\,p+1$ $\qquad F_{2p+1}^*(x) = F^{p+1}(x+1)\ F^p(x-1)\,.$

Do we have coefficients $(\lambda_k^*, \delta_k^*)$ such that $F_k^*(\lambda_k^* + \delta_k^*\, x)$ converges to a limiting distribution function, possibly to $\tilde{L}(x)$? This is a special case of a problem (periodic disturbances, dealt with in the last section) where we alternately add or subtract 1, and can be studied directly. We leave it to the reader.

For simplicity of the example — but where all the ingredients are present — we will study only the cases where $F(x) = \tilde{L}(x)$, i.e., where F(x) is $\Psi_\alpha(x)$, $\Lambda(x)$ or $\Phi_\alpha(x)$ and we intend to obtain $\tilde{L}(x)$ as a limit.

Consider the case where $F(x) = \Psi_\alpha(x)$. As is known, $\lambda_k = 0$ and $\delta_k = k^{1/\alpha}$, i.e., the stability equation takes the form $\Psi_\alpha^k(x) = \Psi_\alpha(k^{1/\alpha} x)$. We will compute, separately, $(\lambda_{2p}^*, \delta_{2p}^*)$ and $(\lambda_{2p+1}^*, \delta_{2p+1}^*)$ such that $F_k^*(\lambda_k^* + \delta_k^* x) \to \Psi_\alpha(x) = \exp\{(-\min(x,0))^\alpha\}$. Let us note that $\bar{w}_{2p}^* = 1$, $\bar{w}_{2p+1}^* = -1$ and so the right-end point of $\max_1^k X_i^*$ is 1 for $k > 1$. For $k = 2p$ if we take $\lambda_k^* = 1$ and $\delta_k^* = (2p)^{-1/\alpha}$ and if for $k = 2p+1$ we take the same expressions, we get $F_k^*(\lambda_k^* + \delta_k^* x) \to \Psi_\alpha(x)$. In general we have $\lambda_k^* = 1$, $\delta_k^* = [k/2]^{-1/\alpha}$. Another proof of the result can be given by showing that $F_{2p}^k(x)$ and $F_{2p+1}^k(x)$ are both equivalent for maxima to $W_\alpha^p(x)$; see the next chapter. This is left as an exercise as well as for the immediate examples.

When $F(x) = \Lambda(x)$ we have $\lambda_k = \log k$ and $\delta_k = 1$ as seen. The relation $F_k^*(\lambda_k^* + \delta_k^* x) \to \Lambda(x)$ gives a solution $\lambda_k^* = \log\frac{e + e^{-1}}{2} + \log k$, $\delta_k^* = 1$.

For Fréchet distribution , $F(x) = \Phi_\alpha(x)$ we have $\lambda_k^* = 0$ and $\delta_k^* = k^{1/\alpha}$ and also $\lambda_k^* = 0$ and $\delta_k^* = k^{1/\alpha}$.

In all cases, the coefficients can be the same owing to the special (limiting) form of $F(x)$ $(= \tilde{L}(x))$.

The asymptotic distributions of extremes

After these examples — where, for instance, logistic distribution, Cauchy distribution and other 'textbook' distribution have not been dealt with here for reasons to be seen shortly — we will proceed to obtain the limiting distributions of maxima from i.i.d. samples $\tilde{L}(x)$ which, as was said before, are $\Psi_\alpha(x)$, $\Lambda(x)$ and $\Phi_\alpha(x)$.

We say that a random sequence $\{X_i\}$ of i.i.d. random variables with distribution function *$F(x)$ is attracted for maxima to $\tilde{L}(x)$* if there exist coefficients $(\lambda_k, \delta_k > 0)$ — already called *attraction coefficients* (for maxima) and not uniquely defined — such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$. As the limiting distributions are continuous the convergence is uniform, as is well known. We will say also that *F(x) is in the domain of attraction of $\tilde{L}(x)$* which will be denoted by $F \in \mathcal{D}(\tilde{L})$. Correspondingly we will say that *F(x) is stable for maxima* (or *max–stable*) if $F^k(x) = F(\alpha_k + \beta_k x)$ for convenient α_k and $\beta_k > 0$. Evidently $\Psi_\alpha(x)$, $\Lambda(x)$ and $\Phi_\alpha(x)$ are max-stable with $\alpha_k = 0, \beta_k = k^{1/\alpha}$, $\alpha_k = -\log k, \beta_k = 1$ and $\alpha_k = 0, \beta_k = k^{-1/\alpha}$. It is immediate that if F(x) is max–stable, F(x) is a limiting distribution or is attracted to F(x): in fact if we take $\lambda_k = -\alpha_k/\beta_k$ and $\delta_k = 1/\beta_k$ we obtain $F^k(\lambda_k + \delta_k x) = F(x)$ and the convergence is verified. The crux of the proof is to show, in reverse, that the limiting distributions are max–stable, to solve the corresponding functional equation, and to obtain one (or more) systems of attraction coefficients.

The conversion of these definitions for minima is immediate and can be done in terms of either F(x) or of the survival function S(x).

We will continue, out of habit, to use F(x), the conversion to S(x) being the reproduction of the previous definitions, but with the introduction of some additional notation which would not be used in the sequel; but this is an interesting exercise because it stresses the duality.

We say that *F(x) is attracted for minima to* $\underset{\sim}{L}(x)$ or *is in the domain of a attraction of $\underset{\sim}{L}(x)$*, denoted by $F \in \mathcal{D}(\underset{\sim}{L})$ (the symbol $\mathcal{D}$ does not introduce ambiguity because of the context and of the pair of dual symbols $\tilde{L}$ and $\underset{\sim}{L}$) if there exist *attraction coefficients* (for minima) $\{(\lambda_k, \delta_k), \delta_k > 0\}$ not uniquely defined, such that $1 - (1 - F(\lambda_k + \delta_k x))^k \xrightarrow{w} \underset{\sim}{L}(x)$; as said, and to be shown, the limiting $\underset{\sim}{L}(x)$ are $1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x)$ and $1 - \Phi_\alpha(-x)$. The stability relation is either $(1 - F(x))^k = 1 - F(\alpha_k + \beta_k x)$ or, using survival functions, $S^k(x) = S(\alpha_k + \beta_k x)$. We say that F (or S) is *stable for minima* (or *min–stable*) if the last relation hold. There is no difference in the definitions, arising from $F(x) + S(x) = 1$. An exercise is to compute the expressions of (α_k, β_k) for min–stability. It is also evident that min–stable distribution (or survival functions) are also limiting ones. The result comes from the solution obtained for maxima.

Let us, then, study maxima. We will show that:

The limiting distributions of maxima are max–stable.

In fact, suppose that there exist $\{(\lambda_k, \delta_k), \delta_k > 0\}$ such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$. Then we have $F^{mk}(\lambda_{mk} + \delta_{mk} x) \xrightarrow{w} \tilde{L}(x)$ also, and thus $F^k(\lambda_{mk} + \delta_{mk} x) \to \tilde{L}^{1/m}(x)$. Let

$\alpha'_m, \beta'_m > 0$ be a limit of a subsequence of $(\lambda_{mk} - \lambda_k)/\delta_k \to \alpha'_m$, $\delta_{mk}/\delta_k \to \beta'_m$, as $k \to \infty$ through the subsequence; then by Khintchine's theorem we have $F^k(\lambda_{mk} + \delta_{mk}\, x) \to \tilde{L}(\alpha'_m + \beta'_m\, x) = \tilde{L}^{1/m}(x)$ or putting $x' = \alpha'_m + \beta'_m\, x$ we get the stability equation

$$\tilde{L}^m(x') = \tilde{L}\left(-\frac{\alpha'_m}{\beta'_m} + \frac{1}{\beta'_m}\, x\right) = \tilde{L}(\alpha_m + \beta_m\, x)$$

and the result is independent of the subsequence. Notice that (α_m, β_m) can be defined by $\alpha_m = -\lim (\lambda_{mk} - \lambda_k)/\delta_{mk}$ and $\beta_m = \lim \delta_k/\delta_{mk}$ when $k \to +\infty$.

Thus we now have to solve the stability functional equation. The first step is to extend for any positive real t the stability equation, i.e., that there exist $\{(\alpha(t), \beta(t)), \beta(t) > 0\}$ such that $\tilde{L}^t(x) = \tilde{L}(\alpha(t) + \beta(t)\, x)$.

When $k \to +\infty$ we have $F^{[kt]}(\lambda_{[kt]} + \delta_{[kt]}\, x) \overset{w}{\to} \tilde{L}(x)$ and as $[kt]/k \to t$ also, we have

$$[F^k(\lambda_{[kt]} + \delta_{[kt]}\, x)]^t \to \tilde{L}(x);$$

consequently $(\lambda_{[kt]} - \lambda_k)/\delta_k$ and $\delta_{[kt]}/\delta_k$ converge as $k \to +\infty$. Let us define $\alpha(t)$ and $\beta(t)$ as

$$\alpha(t) = -\lim (\lambda_{[kt]} - \lambda_k)/\delta_{[kt]} \quad \text{and} \quad \beta(t) = \lim \delta_k/\delta_{[kt]}:$$

we get, as desired, $\tilde{L}^t(x) = \tilde{L}(\alpha(t) + \beta(t)\, x)$, which shows the continuity of $\tilde{L}(x)$.

To solve this functional equation we must know that the functional equation $u(ts) = u(t)\, u(s)$ $(t, s > 0)$ has the solution $u(t) = 0$ or $u(t) = t^{\rho}$ (*).

Then the relation

$$\tilde{L}^{st}(x) = \tilde{L}(\alpha(st) + \beta(st)\, x) = (\tilde{L}^{s}(x))^{t} = \tilde{L}^{t}(\alpha(s) + \beta(s)\, x))$$

$$= \tilde{L}(\alpha(t) + \beta(t)\, (\alpha(s) + \beta(s)\, x))$$

which exchanging s and t is equal to $\tilde{L}(\alpha(s) + \beta(s)\, (\alpha(t) + \beta(t)\, x))$. This leads to the equation

$$\alpha(st) = \alpha(t) + \alpha(s)\, \beta(t) = \alpha(s) + \alpha(t)\, \beta(s)$$

$$\beta(st) = \beta(s)\, \beta(t)\, .$$

The second equation can have the solution $\beta(t) = 0$ or $\beta(t) = t^{\rho}$. But $\beta(t) = 0$ leads to $\tilde{L}^{t}(x) = \tilde{L}\, (\alpha(t))$ and should be disregarded.

We have, then, as $\beta(t) = t^{\rho}$ the relations $\alpha(t)\, (1 - s^{\rho}) = \alpha(s)\, (1 - t^{\rho})$.

(*) — This is true for $u(t)$ continuous, monotonic or measurable; we will prove it for the continuous case. If $u(t) = 0$ the theorem is proved. Suppose that there exists $t_o (> 0)$ such that $u(t_o) \neq 0$ and, thus, $u(1) = 1$. Then for $u(t_o^m) = u(t_o)\, u(t_o^{m-1})$ we get $u(t_o^m) = u(t_o)^m$ and as $u(t_o^{m/n}) = u(t_o^{1/n})^m = u(t_o)^{m/n}$ we see that for every rational $r > 0$ we have $u(t_o^r) = u(t_o)^r$. But the positive rationals are dense in the positive reals and as $u(t)$ is continuous, we get $u(t_o^s) = u(t_o)^s$. Putting $t_o^s = t$ we get $u(t) = u(t_o)^{\log t/\log t_o}$ and putting $u(t_o)^{1/\log t_o} = e^{\rho}$ we get $u(t) = t^{\rho}$ as could be expected.

And here, we can have two cases : $\rho = 0$ or $\rho \neq 0$.

If $\rho = 0$ we get $\alpha(st) = \alpha(s) + \alpha(t)$ and so $e^{\alpha(t)}$ satisfies the equation for $u(t)$; then $\alpha(t) = -\theta \log t$ and the functional equation reads as $\tilde{L}^t(x) = \tilde{L}(-\theta \log t + x)$ whose solution is $\tilde{L}(y) = \exp\{-\exp-\{y/\theta - \log(-\log \tilde{L}(0))\}\}$ which is a Gumbel distribution when $\theta > 0$. For $\theta \leq 0$ the solution is not a distribution function.

Suppose now that $\rho \neq 0$. Then we have (as $t^\rho \neq 1$ for $t \neq 1$) we have (for $t, s \neq 1$) $\frac{\alpha(s)}{1-s^\rho} = \frac{\alpha(t)}{1-s^t} = c$ and so we have the equation $\tilde{L}^t(x) = \tilde{L}(c(1-t^\rho) + t^\rho x)$.

Consider that $\rho > 0$. For $t \to 0$ we get $1 = \tilde{L}(c)$ but for no $x_0 < c$ is $\tilde{L}(x_0) = 1$ because otherwise we would have $\tilde{L}(x_0) = 1 = \tilde{L}^t(x_0) = \tilde{L}(c + t^\rho(x_0 - c))$ which for $t^\rho = \frac{y-c}{x_0-c} > 0$ (implying $y < 0$) would lead to $\tilde{L}(x_0) = \tilde{L}(y) = 1$ for any y; this shows that c is the right–end point. Let us put $x - c = \varphi\ (\leq 0)$. We get $\tilde{L}^t(c+\varphi) = \tilde{L}(c+t^\rho \varphi)$. Put $\tilde{L}(c+\varphi) = e^{-h(\varphi)}$ where $h(\varphi)$ is continuous and non-increasing from $+\infty$ to 0 in $[-\infty, 0]$. We get, for $\varphi \leq 0$,

$$t\, h(\varphi) = h(t^\rho \varphi) \text{ or } h(\varphi) = h(-1)(-\varphi)^{1/\rho} \quad \text{and thus}$$

$$\tilde{L}(x) = 1 \quad \text{for} \quad x \geq c$$

$$\tilde{L}(x) = \tilde{L}(c + (x-c)) = e^{-h(x-c)} = e^{-h(-1)(c-x)^{1/\rho}}$$

which is a Weibull distribution with location parameter c($X \leq c$ with probability 1), dispersion parameter $h(-1)^{-\rho}$, and shape parameter $1/\rho$.

If we are dealing with the case $\rho < 0$, when we take $t \to +\infty$ we get $0 = \tilde{L}(c)$ and so c is the left–end point of $\tilde{L}(x)$.

Put also $x - c = \varphi\ (\geq 0)$: the equation reads as $\tilde{L}(c+\varphi) = \tilde{L}(c+t^\rho \varphi)$. Now put $\tilde{L}(c+\varphi) = e^{-h(\varphi)}$. The equation gives $t\, h(\varphi) = h(t^\rho \varphi)$ and analogously $h(\varphi) = h(1)\, \varphi^{1/\rho}$ and thus

$$\tilde{L}(x) = 0 \quad \text{if} \quad x < c$$

$$\tilde{L}(x) = \tilde{L}(c + (x-c)) = e^{-h(x-c)} = e^{-h(1)(x-c)^{1/\rho}}$$

which is a Fréchet distribution with left–end point c, dispersion parameter $h(1)^{-\rho}$, and shape coefficient $1/\rho$. We have thus proved that the classes of max–stable and limiting distributions for maxima coincide; the changes of the location–dispersion parameters are integrated in the type (or equivalence) changes.

Note that, by the proof given above connected with the stability equation, we have shown that $F^k(\lambda_k + \delta_k x) \overset{w}{\to} \tilde{L}(x)$, k (>0) integer, and $F^t(\lambda_t + \delta_t x) \overset{w}{\to} \tilde{L}(x)$, t (>0) real, are equivalent and this will be used sometimes later.

Notice that the $\tilde{L}$ are continuous and, consequently, *the convergence is uniform.*

For minima, let $F^*(x) = \text{Prob}\{-X \le x\} = \text{Prob}\{X \ge -x\} = S(-x)$ in the continuity set of S.

We have $\text{Prob}\{(\overset{k}{\underset{1}{\min}}\{X_i\} - \lambda_k)/\delta_k < x\} = \text{Prob}\{\overset{k}{\underset{1}{\min}}\{X_i\} < \lambda_k + \delta_k x\} =$

$$= \text{Prob}\{\overset{k}{\underset{1}{\max}}\{-X_i\} > -(\lambda_k + \delta_k x)\} = 1 - \text{Prob}\{\overset{k}{\underset{1}{\max}}\{-X_i\} \le -(\lambda_k + \delta_k x)\} =$$

$$= 1 - F^{*k}(-\lambda_k - \delta_k x) \text{ and if } F^{*k}(-\lambda_k + \delta_k y) \to \tilde{L}(y) \text{ we get}$$

$$\text{Prob}\{(\overset{k}{\underset{1}{\min}}\{X_i\} - \lambda_k)/\delta_k < x\} \to 1 - \tilde{L}(-x) \text{ as } \tilde{L}(y) \text{ is continuous.}$$

$\tilde{L}$ is thus associated with the limiting distribution, if it exists, of $F^*(x)$ and not of $F(x)$. The set of limiting distribution functions for minima is the one already described.

Another proof could be obtained using the survival functions, which is left as an exercise.

We have then proved the *Extremal Limit Theorem*:

The reduced asymptotic distributions of maxima $\tilde{L}(x)$, *when they exist, are* $\Psi_\alpha(x)$, $\Lambda(x)$ *and* $\Psi_\alpha(x)$ *and the reduced asymptotic distribution of minima* $\underset{\sim}{L}(x)$, *when they exist are* $1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x)$ *and* $1 - \Psi_\alpha(-x)$.

Notice that the right–tail behaviour of $F^*(x)$ defining the maxima of $\{-X_i\}$ corresponds to the left–tail behaviour of $F(x)$ corresponding to the minima.

Consider a distribution function (proper and non–degenerate) and a sequence of real numbers $\{u_k\}$ such that $k(1 - F(u_k)) \to \tau$ or $k(1 - F(u_k)) = \tau + o(1) = \tau(1 + o(1))$; the u_k are, evidently, a function of τ and a functional of F.

It is immediate that, if $\{X_i\}$ are i.i.d. with distribution function F, then $= \text{Prob}\{\max_1^k\{X_i\} \le u_k\} = F^k(u_k) = (1 - \frac{\tau(1+o(1))}{k})^k \to e^{-\tau}$ as $k \to \infty$; the inverse is also valid as seen by taking logarithms; the case $\tau = +\infty$ is dealt with analogously. The connection between the changing (non–decreasing in general) levels for overpassing and the previous results corresponds for instance to taking $u_k(\tau, x) = \lambda_k + \delta_k x$ and $\tau = e^{-x}$, etc.. Note that u_k does not necessarily exist; for $F(x) = 0$ if $x \le 0$, $F(x) = x/2$ if $0 \le x \le 1$ and $F(x) = 1$ for $x \ge 1$ (a jump of 1/2 at $x = 1$) we see that for $\tau = 1$ we can take neither $u_k \ge 1$, because we would get $k(1 - F(u_k)) = 0$, nor $0 \le u_k < 1$, because for $k(1 - u_k/2) \to 1$ we should have taken $u_k = 2 - \frac{2 + o(1)}{k}$. Also if $F(x) = 1 - e^{-[x]}$ for $x \ge 0$, $F(x) = 0$ for $x < 0$, it is easy to see that for no $\{u_k\}$ we can obtain $k(1 - F(u_k)) \to \tau$, as $\log k - [u_k]$ is oscillating. But for $F(x) = 0$ if $x \le e$, $1 - 1/\log x$ if $x \ge e$, we get $u_k(\tau) = e^{k/\tau}$.

The classical theory already developed used $u_k = \lambda_k + \delta_k x$ with τ depending on x, giving the desired results when $\tau = (-x)^\alpha$ $(x < 0)$, $\tau = e^{-x}$ or $\tau = x^{-\alpha}$ $(x > 0)$. In some cases we can transform $u_k(\tau)$ into an equivalent linear function of x (x and τ related as above), but other cases exist, like the last example in the previous paragraph, where such a transformation for a linear function cannot be made.

Finally let us speak of a general form that integrates in one expression the three limiting distributions of maxima, called von Mises-Jenkinson formula

$$G(z|\theta) = \exp\{-(1 + \theta z)_+^{-1/\theta}\}, \quad \theta \text{ real, i.e.,}$$

$$G(z \mid \theta) = \exp\{-(1 + \theta z)-1/\theta\}, \quad \theta \text{ real,}$$

where $1 + \theta z \ge 0$ with the natural truncation when $1 + \theta z < 0$: if $\theta > 0$, for $z < -1/\theta$ we have $G(z|\theta) = 0$ and if $\theta < 0$ for $z > -1/\theta$ we have $G(z|\theta) = 1$.

If $\theta < 0$ we have immediately $G(-(1 + x)/\theta|\theta) = \Psi_{-1/\theta}(x)$ where $\alpha = -1/\theta > 0$; if $\theta > 0$ we have $G(-(1 - x)/\theta|\theta) = \Phi_{-1/\theta}(x)$ where $\alpha = 1/\theta > 0$ and finally we have $G(z|0^+) = G(z|0^-) = 1(z)$.

The graphics of the densities of $G'(z|\theta)$, for maxima, are given in Fig.II.1. They will be given for each form in the chapters V, VI and VII (for minima)

Fig. II.1 — Graphs of the densities $G'(z|\theta)$

The graphs of the densities for minima $G'(-z|\theta)$ are the miror images of the previous ones.

It should be noted that when $\theta \to 0$ the graphs of $G'(z|\theta)$ converge to $G'(z|0) = \Lambda'(z)$; this corresponds to the fact that for large $\alpha = 1/\theta$ Fréchet and Weibull distributions are very close to the Gumbel distribution.

This integrated formula will be useful for the statistical choice of models in chapter VIII.

The basic texts are Fisher and Tippet (1928), Fréchet (1927), de Finetti (1932), Gumbel (1935), von Mises (1936) and Gnedenko (1943); exposés can be found in Galambos (1978) and de Hann (1976).

It is of interest to consider the joint behaviour of $(\overset{k}{\underset{1}{\min}}\{X_i\}, \overset{k}{\underset{1}{\max}}\{X_i\})$. Its joint distribution function is

$$\mathrm{Prob}\{\overset{k}{\underset{1}{\min}}\ \{X_i\} \leq x,\ \overset{k}{\underset{1}{\max}}\ \{X_i\} \leq y) = \mathrm{Prob}\{\overset{k}{\underset{1}{\min}}\ \{X_i\} \leq \min(x, y)\ ,$$

$$\overset{k}{\underset{1}{\max}}\ \{X_i\} \leq y\} = F^k(y) - (F(y) - F(\min(x, y))^k\ .$$

Supposing that

$$\mathrm{Prob}\{\overset{k}{\underset{1}{\min}}\ \{X_i\} \leq \lambda'_k + \delta'_k\ x\} = 1 - (1 - F(\lambda'_k + \delta'_k\ x))^k \to \tilde{L}\ (x)$$

and

$$\text{Prob}\{\max_1^k \{X_i\} \le \lambda_k + \delta_k\, y\} = F^k(\lambda_k + \delta_k\, y)) \to \tilde{L}(y)$$

we get

$$\text{Prob}\{\min_1^k \{X_i\} \le \lambda_k' + \delta_k'\, x,\ \max_1^k \{X_i\} \le \lambda_k + \delta_k\, y\} =$$

$$= F^k(\lambda_k + \delta_k\, y) - \{F(\lambda_k + \delta_k\, y) - F(\min(\lambda_k' + \delta_k\, x, \lambda_k + \delta_k\, y))\}^k .$$

The first summand converges to $\tilde{L}$ (y). Let us thus study the second one which can be written as

$$F^k(\lambda_k + \delta_k\, y)\ \{1 - \frac{F(\min(\lambda_k' + \delta_k'\, x, \lambda_k + \delta_k\, y))}{F(\lambda_k + \delta_k\, y)}\}^k$$

As the first factor converges to $\tilde{L}(y)$, denote by $\varphi(x, y)$ the limit of the second factor. We have, if $\varphi(x, y)$ exists,

$$\text{Prob}\{\min_1^k \{X_i\} \le \lambda_k' + \delta_k'\, x,\ \min_1^k \{X_i\} \le \lambda_k + \delta_k\, y\} \to \tilde{L}(y)$$

But $$\varphi(x, y) = \lim\ (1 - \frac{F(\min(\lambda_k' + \delta_k'\, x, \lambda_k + \delta_k\, y))}{F(\lambda_k + \delta_k\, y)})^k$$

$$= \exp\{-\lim \frac{k\, F(\min(\lambda_k' + \delta_k'\, x, \lambda_k + \delta_k\, y))}{F(\lambda_k + \delta_k\, y)}\}$$

As said before $F(\lambda_k + \delta_k\, y) \to 1$ for y in the support of $\tilde{L}(y)$; on the other hand we know that for x and y in the supports of $\underset{\sim}{L}(x)$ and $\tilde{L}(y)$, $\lambda_k' + \delta_k'\, x \to \underline{w}$ and $\lambda_k + \delta_k\, y \to \bar{w}$ and for large k we have $\lambda_k' + \delta_k'\, x < \lambda_k + \delta_k\, y$ and, thus, we get

$$\varphi(x, y) = \exp\{-\lim k\, F(\lambda_k' + \delta_k' x)\} = \exp\{-[-\log(1 - \underset{\sim}{L}(x))]\} = 1 - \underset{\sim}{L}(x)$$

and so

$$\text{Prob}\{\min_1^k \{X_i\} \leq \lambda_k' + \delta_k' x,\ \max_1^k \{X_i\} \leq \lambda_k + \delta_k y\} \to \underset{\sim}{L}(x)\ \tilde{L}(y).$$

Consequently the reduced extremes are asymptotically independent (in the sense defined in the proof). For more details on asymptotic independence of sample extremes see Geffroy (1958/59).

From asymptotic independence, it is evident that the distributions of the reduced range, centre, and absolute maximum depend only on the prevailing reduced extremes. If they are of the corresponding asymptotic form their distributions are, for large samples, those of $Y - X$, $\frac{1}{2}(Y + X)$ and $\max(-X, Y)$.

Similarly it was proved by Rosengard (1962, 1966), Tiago de Oliveira (1961), and Rossberg (1963), (1965), that the sample extremes and mean, under very general conditions, are asymptotically independent.

The asymptotic distributions of the m–th extremes

The asymptotic theory of m–th extremes was, essentially, initiated in a paper by Gumbel (1935).

Consider a sequence of i.i.d. random variables $\{X_i\}$ with distribution function $F(x)$ and suppose that

$$\text{Prob}\{\max_1^k (X_i - \lambda_k)/\delta_k \leq x\} = F^k(\lambda_k + \delta_k x) \to \tilde{L}(x).$$

Let us see what happens to the m–th order maximum $X'_{k-m+1} = X''_m$ in the notation of chapter I.

We have $F_{m,k}(x) = \text{Prob}\{X''_m \leq x\} = \text{Prob}\,\{\text{at least } k\text{–}m\text{+}1 \text{ of the } X_i \leq x\} =$

$$\sum_{k-m+1}^{k} \binom{k}{j} F(x)^j (1-F(x))^{k-j} = \sum_{i=o}^{m-1} \binom{k}{k-i} F^{k-i}(x)(1-F(x))^i$$

and so

$$F_{m,k}(\lambda_k + \delta_k\ x) \to \tilde{L}_m(x) = \lim \sum_{i=o}^{m-1} \binom{k}{k-i} F^{k-i}(\lambda_k + \delta_k\ x)(1 - F(\lambda_k + \delta_k\ x))^i$$

$$= \tilde{L}(x) \sum_{i=o}^{m-1} \frac{(-\log \tilde{L}(x))^i}{i\,!} \quad \text{as } F^{k-i}(\lambda_k + \delta_k\, x) \to \tilde{L}(x) \qquad \text{and}$$

$$\binom{k}{k-i}(1 - F(\lambda_k + \delta_k\ x))^i = \binom{k}{i}(1 - F(\lambda_k + \delta_k\ x))^i$$

$$= \frac{k(k-1)\ldots(k-i+1)\,(1 - F(\lambda_k + \delta_k\, x))^i}{i\,!} \sim [k(1 - F(\lambda_k + \delta_k\ x))]^i/i\,!$$

and as seen before $k(1 - F(\lambda_k + \delta_k\ x)) \to -\log \tilde{L}(x)$ as $k \to +\infty$. Evidently $\tilde{L}_1(x) = \tilde{L}(x)$.

Analogously we can obtain the distribution of the m–th minimum: suppose that $1-(1 - F(\lambda_k + \delta_k\ x))^k \to -\log \underset{\sim}{L}(x)$ or equivalently that $k\,F(\lambda_k + \delta_k\ x)) \to -\log(1 - \underset{\sim}{L}(x))$.

Then

$$F^*_{m,k}(x) = \text{Prob}\{X'_m \le x\} = 1 - \text{Prob}\{X'_m > x\} = 1 - \sum_{j=o}^{m-1} \binom{k}{j} F^j(x)(1-F(x))^{k-j}$$

and consequently

$$F^*_{m,k}(\lambda_k + \delta_k\ x) = 1 - \sum_{j=o}^{m-1} \binom{k}{j} F^j(\lambda_k + \delta_k\ x)\ (1 - F(\lambda_k + \delta_k\ x))^{k-j}\ .$$

But $(1 - F(\lambda_k + \delta_k\ x))^{k-j} \to 1 - \underset{\sim}{L}(x)$ and

$$\binom{k}{j} F^j(\lambda_k + \delta_k\ x) = \frac{k(k-1)\ \ldots\ (k-i+1)}{j\,!} F^j(\lambda_k + \delta_k\ x) \sim \frac{(k\,F(\lambda_k + \delta_k\, x))^j}{j\,!} \to$$

$$\to \left(\frac{-\log(1-\underset{\sim}{L}(x))}{j!}\right)^j \qquad \text{and we get}$$

$$\underset{\sim}{L}_m(x) = 1-(1-\underset{\sim}{L}_m(x)) \sum_{j=o}^{m-1} \left(\frac{-\log(1-\underset{\sim}{L}(x))}{j!}\right)^j .$$

Using the survival function $\underset{\sim}{S}(x) = 1 - \underset{\sim}{L}(x)$ we get

$$\underset{\sim}{S}_m(x) = \underset{\sim}{S}(x) \sum_{j=o}^{m-1} \left(\frac{-\log \underset{\sim}{S}(x)}{j!}\right)^j ,$$

dual of the expression for $\underset{\sim}{L}_m(x)$; evidently $\underset{\sim}{S}_1(x) = \underset{\sim}{S}(x)$.

To compute $\underset{\sim}{L}_m(x)$ and $\underset{\sim}{S}_m(x)$ from the three possible expression of $\underset{\sim}{L}_1(x)$ and $\underset{\sim}{S}_1(x)$ is a simple exercise.

In what has been said we have always supposed that we were dealing with the m–th maximum or m–th minimum, with m fixed. We can even allow m to be a function m(k) such that $m(k)/k \to 0$ as $k \to +\infty$ and obtain similar results: see Smirnov (1952) as the first step, and Mejzler (1984) for the general overview.

We will not develop this points further because it does not seem to be particularly interesting.

Let us now recall that maxima in between themselves and minima in between themselves are associated; but, as shown the maximum and the minimum of a sample are asymptotically independent; a m_1–th maximum and a m_2–minimum are also asymptotically independent by a simple extension of the proof.

An overview of some extensions

Here we will sketch, briefly, without proof, results extending the three limiting distributions under conditions which are not i.i.d., for which we have given already some examples.

If the identically distributed condition is relaxed, but independence is maintained, the initial papers of Juncosa (1948) and Mejzler (1949), detailed in Mejzler (1984), have shown that under general reasonable conditions, the three limiting distributions are still valid. But this is not completely general: as a counter-example, take any distribution function F(x) and consider the sequence of independent random variables $\{X_i\}$ with distribution functions $F_i(x) = (F(x))^{2^{-i}}$. It is immediate that

$$\text{Prob}\{\max(X_1, \dots, X_k) \le x\} = \prod_1^k F_i(x) = (F(x))^{1-2^{-k}} \to F(x)$$

as $k \to \infty$, with a system of attraction coefficients $\lambda_k = 0, \delta_k = 1$. Thus any distribution function can be the limiting distribution function of max $(X_1, \dots, X_k)$ and so some moderate condition about the set of $\{F_i(.)\}$ must be introduced to maintain the classical limiting distribution of maxima (and minima, obviously). See the above–mentioned papers for more details, and in the case where $F_i(x) = F(x + \theta_i)$ with $\theta_i = \theta_{i+P}$ (periodic disturbances of period P), see Tiago de Oliveira (1976).

Also if the independence condition is lifted, but i.d. margins are maintained, the asymptotic distribution of maxima may be different from the three given forms. As a counter example take the sequence $Y_o, Y_1, \dots, Y_i, \dots$ i.i.d. with a standard normal distribution function and define

$$X_i = \sqrt{\rho}\, Y_o + \sqrt{1-\rho}\; Y_i \quad (i = 1,2, \dots)\ ,\ \rho > 0.$$

The X_i are multinormal with standard margins and (constant) correlation coefficient ρ (> 0). Then

$$\max_1^k X_i = \sqrt{\rho}\, Y_o + \sqrt{1-\rho}\; \max_1^k Y_i$$

and, as by the ALLN we have $\max_1^k X_i - \sqrt{2 \log k} \xrightarrow{P} 0$, we see that

$$\frac{\max_1^k X_i - \sqrt{1-\rho}\;\sqrt{2 \log k}}{\sqrt{\rho}} - Y_o \xrightarrow{P} 0$$

and so

$$\text{Prob}\ \{\max_1^k X_i \le \sqrt{1-\rho}\; \sqrt{2 \log k} + \sqrt{\rho}\; x\} \to N(x).$$

This example suggests that the correlation in dependent sequences must be waning out at a reasonable rate to obtain the classical limiting forms for maxima.

The case where we deal with a random (dependent) sequence with the same marginal distributions (F(.)) — i.e., stationarity of order 1 — and where the sequence of maxima is attracted by the same distribution as the i.i.d. sequence with the same distributions F(.), has been initially considered by Watson (1954) and Newell (1964) for the case of m–dependence, i.e., (X_i, X_j) independent if $|i-j| > m$, and later by others — this point will be developed in part IV; see references therein.

The application to Meteorology and Oceanography of m–dependence is obvious as after a few days (2 or 3) there is practical independence for the observations.

For a relaxing of both i. and i.d. conditions, but having yet one of the limiting distributions, see Tiago de Oliveira (1978).