4

Fluctuations of Upper Order Statistics

After having investigated in Chapter 3 the behaviour of the maximum, i.e. the largest value of a sample, we now consider the joint behaviour of several upper order statistics. They provide information on the right tail of a df.

In Section 4.1, after some basic results on the ordered sample, we present various examples in connection with uniform and exponential order statistics and spacings. Just to mention two items, we touch on the subject of simulation of general upper order statistics (working from uniform random numbers) and prove the order statistics property of a homogeneous Poisson process. Here also Hill's estimator appears for the first time: we prove that it is a consistent estimator for the index of a regularly varying tail. Its importance will be made clear in Chapter 6.

In Section 4.2 we exploit the Poisson approximation, already used to derive limit laws of normalised maxima, in a more advanced way. This leads to the multivariate limit distribution of several upper order statistics. Such results provide the theoretical background when deriving limit properties for tail estimators, as we shall see in Chapter 6. Extensions to randomly indexed samples will be given in Section 4.3.

In Section 4.4 we show under what conditions the previous results can be extended to stationary sequences.

P. Embrechts et al., *Modelling Extremal Events*, Stochastic Modelling and Applied Probability 33, DOI 10.1007/978-3-642-33483-2_5,

4.1 Order Statistics

Let $X, X_1, X_2, \ldots$ denote a sequence of iid non–degenerate rvs with common df F. In this section we summarise some important properties of the upper order statistics of a finite sample $X_1, \ldots, X_n$. To this end we define the *ordered sample*

$$X_{n,n} \le \cdots \le X_{1,n} \,.$$

Hence $X_{n,n} = \min(X_1, \ldots, X_n)$ and $X_{1,n} = M_n = \max(X_1, \ldots, X_n)$. The rv $X_{k,n}$ is called the *kth upper order statistic*. The notation for order statistics varies; some authors denote by $X_{1,n}$ the minimum and by $X_{n,n}$ the maximum of a sample. This leads to different representations of quantities involving order statistics.

The relationship between the order statistics and the empirical df of a sample is immediate: for $x \in \mathbb{R}$ we introduce the *empirical df* or *sample df*

$$F_n(x) = \frac{1}{n} \operatorname{card}\{i : 1 \le i \le n, X_i \le x\} = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \le x\}} \,, \quad x \in \mathbb{R} \,,$$

where I_A denotes the indicator function of the set A. Now

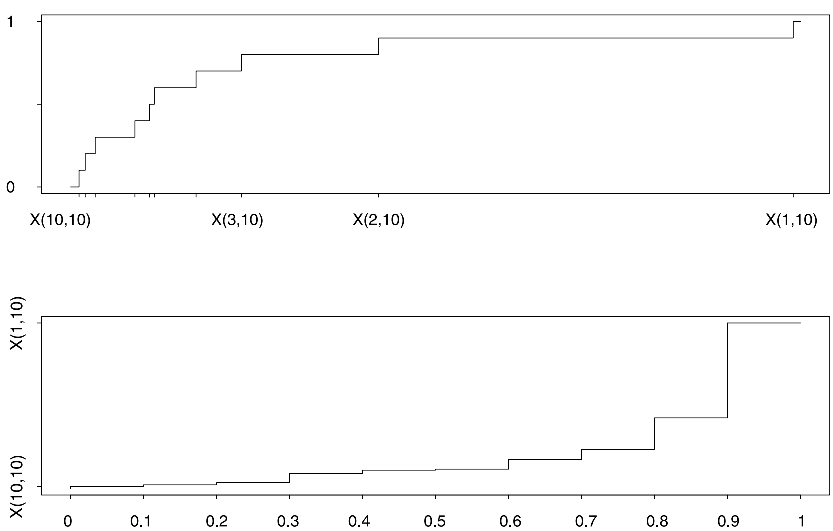


Figure 4.1.1 *Empirical df F_n (top) and empirical quantile function $F_n^{\leftarrow}$ (bottom) of a sample of 10 standard exponential rvs.*

$$X_{k,n} \le x \qquad \text{if and only if} \qquad \sum_{i=1}^{n} I_{\{X_i > x\}} < k\,, \tag{4.1}$$

which implies that

$$P\left(X_{k,n} \le x\right) = P\left(F_n(x) > 1 - \frac{k}{n}\right).$$

Upper order statistics estimate tails and quantiles, and also excess probabilities over certain thresholds. Recall the definition of the quantile function of the df F

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \ge t\}\,, \quad 0 < t < 1\,.$$

For a sample $X_1, \ldots, X_n$ we denote the empirical quantile function by $F_n^{\leftarrow}$. If F is a continuous function, then ties in the sample occur only with probability 0 and may thus be neglected, i.e. we may assume that $X_{n,n} < \cdots < X_{1,n}$. In this case $F_n^{\leftarrow}$ is a simple function of the order statistics, namely

$$F_n^{\leftarrow}(t) = X_{k,n} \quad \text{for } 1 - \frac{k}{n} < t \le 1 - \frac{k-1}{n}\,, \tag{4.2}$$

for $k = 1, \ldots, n$.

Next we calculate the df $F_{k,n}$ of the kth upper order statistic explicitly.

Proposition 4.1.2 (Distribution function of the kth upper order statistic) *For $k = 1, \ldots, n$ let $F_{k,n}$ denote the df of $X_{k,n}$. Then*

(a) $F_{k,n}(x) = \sum_{r=0}^{k-1} \binom{n}{r} \overline{F}^{\,r}(x)\, F^{n-r}(x)$.

(b) If F is continuous, then

$$F_{k,n}(x) = \int_{-\infty}^{x} f_{k,n}(z)\, dF(z)\,,$$

where

$$f_{k,n}(x) = \frac{n!}{(k-1)!\,(n-k)!}\, F^{n-k}(x)\, \overline{F}^{\,k-1}(x)\,;$$

i.e. $f_{k,n}$ is a density of $F_{k,n}$ with respect to F.

Proof. (a) For $n \in \mathbb{N}$ define

$$B_n = \sum_{i=1}^{n} I_{\{X_i > x\}}\,.$$

Then B_n is a sum of n iid Bernoulli variables with success probability

$$EI_{\{X > x\}} = P(X > x) = \overline{F}(x)\,.$$

Hence B_n is a binomial rv with parameters n and $\overline{F}(x)$. An application of (4.1) gives for $x \in \mathbb{R}$

$$\begin{aligned} F_{k,n}(x) &= P(B_n < k) \\ &= \sum_{r=0}^{k-1} P(B_n = r) \\ &= \sum_{r=0}^{k-1} \binom{n}{r} \overline{F}^{\,r}(x)\, F^{n-r}(x)\,. \end{aligned}$$

(b) Using the continuity of F, we calculate

$$\begin{aligned} &\frac{n!}{(k-1)!\,(n-k)!} \int_{-\infty}^{x} F^{n-k}(z)\, \overline{F}^{\,k-1}(z)\, dF(z) \\ &= \frac{n!}{(k-1)!\,(n-k)!} \int_{\overline{F}(x)}^{1} (1-t)^{n-k}\, t^{k-1}\, dt \\ &= \sum_{r=0}^{k-1} \binom{n}{r} \overline{F}^{\,r}(x)\, F^{n-r}(x) = F_{k,n}(x)\,. \end{aligned}$$

The latter follows from a representation of the incomplete beta function; it can be proved by multiple partial integration. See also Abramowitz and Stegun [3], formula 6.6.4. □

Similar arguments lead to the joint distribution of a finite number of different order statistics. If for instance F is absolutely continuous with density f, then the joint density of $(X_1, \ldots, X_n)$ is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)\,, \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n\,.$$

Since the n values of $(X_1, \ldots, X_n)$ can be rearranged in $n!$ ways (by absolute continuity there are a.s. no ties), every specific ordered collection $(X_{k,n})_{k=1,\ldots,n}$ could have come from $n!$ different samples. This heuristic argument can be made precise; see for instance Reiss [526], Theorem 1.4.1, or alternatively use the transformation theorem for densities. The joint density of the ordered sample becomes:

$$f_{X_{1,n},\ldots,X_{n,n}}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i)\,, \qquad x_n < \cdots < x_1\,. \tag{4.3}$$

The following result on marginal densities is an immediate consequence of equation (4.3).

Theorem 4.1.3 (Joint density of k upper order statistics)
If F is absolutely continuous with density f, then

$$f_{X_{1,n},\ldots,X_{k,n}}(x_1,\ldots,x_k) = \frac{n!}{(n-k)!}\, F^{n-k}(x_k) \prod_{i=1}^{k} f(x_i)\,, \quad x_k < \cdots < x_1\,.$$

□

Further quantities which arise in a natural way are the *spacings*, i.e. the differences between successive order statistics. They are for instance the building blocks of Hill's estimator; see Example 4.1.12 and Section 6.4.2.

Definition 4.1.4 (Spacings of a sample)
For a sample $X_1,\ldots,X_n$ the spacings *are defined by*

$$X_{k,n} - X_{k+1,n}\,, \qquad k = 1,\ldots,n-1\,.$$

For rvs with finite left (right) endpoint $\widetilde{x}_F$ (x_F) we define the nth (0th) spacing as $X_{n,n} - X_{n+1,n} = X_{n,n} - \widetilde{x}_F$ ($X_{0,n} - X_{1,n} = x_F - X_{1,n}$). □

Example 4.1.5 (Order statistics and spacings of exponential rvs)
Let (E_n) denote a sequence of iid standard exponential rvs. An immediate consequence of (4.3) is the joint density of an ordered exponential sample $(E_{1,n},\ldots,E_{n,n})$:

$$f_{E_{1,n},\ldots,E_{n,n}}(x_1,\ldots,x_n) = n!\; \exp\left\{-\sum_{i=1}^{n} x_i\right\}, \qquad 0 < x_n < \cdots < x_1\,.$$

From this we derive the joint distribution of exponential spacings by an application of the transformation theorem for densities. Define the transformation

$$T(x_1,\ldots,x_n) = (x_1 - x_2, 2(x_2 - x_3), \ldots, n x_n)\,, \quad 0 < x_n < \cdots < x_1\,.$$

Then $\det(\partial T(\mathbf{x})/\partial \mathbf{x}) = n!$ and

$$T^{-1}(x_1,\ldots,x_n) = \left(\sum_{j=1}^{n} \frac{x_j}{j}, \sum_{j=2}^{n} \frac{x_j}{j}, \ldots, \frac{x_n}{n}\right), \quad x_1, x_2, \ldots, x_n > 0\,.$$

Then the density g of $(E_{1,n} - E_{2,n}, 2(E_{2,n} - E_{3,n}), \ldots, n E_{n,n})$ is of the form

$$\begin{aligned} g(x_1,\ldots,x_n) &= \frac{1}{n!}\, f_{E_{1,n},\ldots,E_{n,n}}\left(\sum_{j=1}^{n} \frac{x_j}{j}, \sum_{j=2}^{n} \frac{x_j}{j}, \ldots, \frac{x_n}{n}\right) \\ &= \exp\Big\{-\sum_{i=1}^{n}\sum_{j=i}^{n} \frac{x_j}{j}\Big\} \\ &= \exp\Big\{-\sum_{i=1}^{n} x_i\Big\}\,. \end{aligned}$$

This gives for $i = 1, \ldots, n$ that the rvs $i(E_{i,n} - E_{i+1,n})$ have joint density

$$g\left(x_1, \ldots, x_n\right) = \exp\left\{-\sum_{i=1}^{n} x_i\right\}, \qquad x_1, \ldots, x_n > 0\,.$$

This implies that the spacings

$$E_{1,n} - E_{2,n}\,,\ E_{2,n} - E_{3,n}\,, \ldots, E_{n,n}$$

are independent and exponentially distributed, and $E_{k,n} - E_{k+1,n}$ has mean $1/k$ for $k = 1, \ldots, n$, where we recall that $E_{n+1,n} = 0$. □

Example 4.1.6 (Markov property of order statistics)
When working with spacings from absolutely continuous dfs one can often make use of the fact that their order statistics form a Markov process, i.e.

$$\begin{aligned} & P(X_{k,n} \leq y \,|\, X_{n,n} = x_n, \ldots, X_{k+1,n} = x_{k+1}) \\ = \quad & P(X_{k,n} \leq y \,|\, X_{k+1,n} = x_{k+1}). \end{aligned}$$

To be precise, $(X_{n,n}, \ldots, X_{1,n})$ is a non-homogeneous, discrete–time Markov process whose initial df is

$$P(X_{n,n} \leq x) = 1 - \overline{F}^{\,n}(x)\,,$$

and whose transition df $P(X_{k,n} \leq y \,|\, X_{k+1,n} = x_{k+1})$ is the df of the minimum of k iid observations from the df F truncated at x_{k+1}. For $k = 1, \ldots, n-1$,

$$P(X_{k,n} > y \,|\, X_{k+1,n} = x_{k+1}) = \left(\frac{\overline{F}(y)}{\overline{F}(x_{k+1})}\right)^k, \quad y > x_{k+1}\,.$$

A proof of the Markov property is straightforward; see for instance Arnold, Balakrishnan and Nagaraja [20], Theorem 2.4.3. They also provide an example showing that the Markov property does not hold for general F; see their Section 3.4. □

Example 4.1.7 (Order statistics property of the Poisson process)
Let $N = (N(t))_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$; for a definition see Example 2.5.2. Then the arrival times T_i of N in $(0, t]$, conditionally on $\{N(t) = n\}$, have the same distribution as the order statistics of a uniform sample on $(0, t)$ of size n; i.e.

$$P\big((T_1, T_2, \ldots, T_{N(t)}) \in A | N(t) = n\big) = P\big((U_{n,n}, \ldots, U_{1,n}) \in A\big)$$

for all Borel sets A in $\mathbb{R}_+$. This property is called the *order statistics property* of the Poisson process. It gives an intuitive description of the distribution of

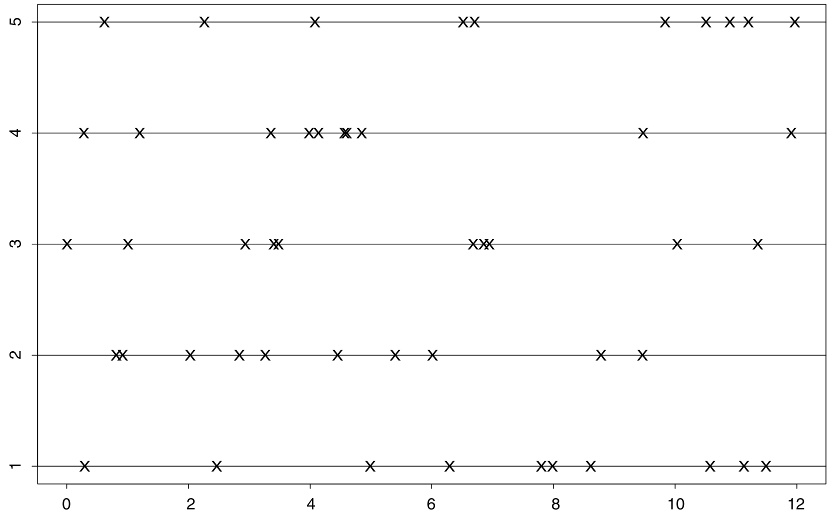


Figure 4.1.8 *Five realisations of the arrival–times of a Poisson process N with intensity 1, conditionally on $\{N(12) = 10\}$. They illustrate the order statistics property (Example 4.1.7).*

the arrival times of a Poisson process.

For a proof we assume that $0 < t_1 < \cdots < t_n < t$ and $h_1, \ldots, h_n$ are all positive, but small enough such that the intervals $J_i = (t_i, t_i + h_i]\,, i = 1, \ldots, n\,,$ are disjoint. Then

$$\begin{aligned} &P(T_1 \in J_1, \ldots, T_n \in J_n \mid N(t) = n) \\ &\quad = \quad P(T_1 \in J_1, \ldots, T_n \in J_n, N(t) = n) \,/\, P(N(t) = n)\,. \end{aligned}$$

Writing $N(J_i) = N(t_i + h_i) - N(t_i)\,, i = 1, \ldots, n$, and using the independence and stationarity of the increments of the Poisson process we obtain for the numerator that

$$\begin{aligned} &P(N(t_1) = 0, N(J_1) = 1, N(t_2) - N(t_1 + h_1) = 0, \\ &\qquad \ldots, N(t_n) - N(t_{n-1} + h_{n-1}) = 0, N(J_n) = 1, N(t) - N(t_n + h_n) = 0) \\ &= \quad P(N(t_1) = 0) P(N(J_1) = 1) P(N(t_2) - N(t_1 + h_1) = 0) \times \\ &\qquad \cdots \times P(N(J_n) = 1) P(N(t) - N(t_n + h_n) = 0) \end{aligned}$$

$$
\begin{aligned}
&= P(N(t_1)=0)P(N(h_1)=1)P(N(t_2-t_1-h_1)=0)\times \\
&\qquad \cdots \times P(N(h_n)=1)P(N(t-(t_n+h_n))=0) \\
&= e^{-\lambda t_1}\times e^{-\lambda h_1}\lambda h_1 \times e^{-\lambda[t_2-(t_1+h_1)]}\times \\
&\qquad \cdots \times e^{-\lambda[t_n-(t_{n-1}+h_{n-1})]}\times e^{-\lambda h_n}\lambda h_n \times e^{-\lambda(t-(t_n+h_n))} \\
&= \lambda^n e^{-\lambda t}\prod_{i=1}^n h_i\,.
\end{aligned}
$$

This implies

$$P(T_1\in J_1,\ldots,T_n\in J_n \mid N(t)=n) = \frac{n!}{t^n}\prod_{i=1}^n h_i\,.$$

The conditional densities are obtained by dividing both sides by $\prod_{i=1}^n h_i$ and taking the limit for $\max_{1\le i\le n} h_i \to 0$, yielding

$$f_{T_1,\ldots,T_n|N(t)}(t_1,\ldots,t_n|n) = \frac{n!}{t^n}\,,\quad 0<t_1<\cdots<t_n<t\,. \tag{4.4}$$

It follows from (4.3) that (4.4) is the density of the order statistics of n iid uniform rvs on $(0,t)$. □

The following concept is called *quantile transformation*. It is extremely useful since it often reduces a problem concerning order statistics to one concerning the corresponding order statistics from a uniform sample. The proof follows immediately from the definition of the uniform distribution.

Lemma 4.1.9 (Quantile transformation)
Let $X_1,\ldots,X_n$ be iid with df F. Furthermore, let $U_1,\ldots,U_n$ be iid rvs uniformly distributed on $(0,1)$ and denote by $U_{n,n}<\cdots<U_{1,n}$ the corresponding order statistics. Then the following results hold:

(*a*) $F^{\leftarrow}(U_1)\stackrel{d}{=}X_1$.

(*b*) *For every $n\in\mathbb{N}$,*

$$(X_{1,n},\ldots,X_{n,n})\stackrel{d}{=}\left(F^{\leftarrow}(U_{1,n}),\ldots,F^{\leftarrow}(U_{n,n})\right)\,.$$

(*c*) *The rv $F(X_1)$ has a uniform distribution on $(0,1)$ if and only if F is a continuous function.* □

Example 4.1.10 (Simulation of upper order statistics)
The quantile transformation above links the uniform distribution to some general distribution F. An immediate application of this result is random number generation. For instance, exponential random numbers can be obtained from uniform random numbers by the transformation $E_1 = -\ln(1-U_1)$. Simulation studies are widely used in an increasing number of applications. A simple algorithm for simulating order statistics of exponentials can be based on Example 4.1.5, which says that

$$(E_{i,n} - E_{i+1,n})_{i=1,\ldots,n} \stackrel{d}{=} (i^{-1}E_i)_{i=1,\ldots,n}\,,$$

with $E_{n+1,n} = 0$. This implies for the order statistics of an exponential sample that

$$(E_{i,n})_{i=1,\ldots,n} \stackrel{d}{=} \left(\sum_{j=i}^{n} j^{-1}E_j\right)_{i=1,\ldots,n}.$$

Order statistics and spacings of iid rvs U_i uniformly distributed on $(0,1)$ and standard exponential rvs E_i are linked by the following representations; see e.g. Reiss [526], Theorem 1.6.7 and Corollary 1.6.9. We write $\Gamma_n = E_1 + \cdots + E_n$, then

$$(U_{1,n}, U_{2,n}, \ldots, U_{n,n}) \stackrel{d}{=} \left(\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, \ldots, \frac{\Gamma_1}{\Gamma_{n+1}}\right),$$

and

$$(1-U_{1,n}\,, U_{1,n} - U_{2,n}\,, \ldots, U_{n,n}) \stackrel{d}{=} \left(\frac{E_{n+1}}{\Gamma_{n+1}}, \ldots, \frac{E_1}{\Gamma_{n+1}}\right).$$

The four distributional identities above provide simple methods for generating upper order statistics or spacings of the exponential or uniform distribution. A statement for general F is given in (4.6) below. For more sophisticated methods based on related ideas we refer to Gerontidis and Smith [259] or Ripley [542], Section 4.1, and references therein. □

Example 4.1.11 (The limit of the ratio of two successive order statistics)
Consider $F \in \mathrm{MDA}(\Phi_\alpha)$, equivalently $\overline{F} \in \mathcal{R}_{-\alpha}$, for some $\alpha > 0$. We want to show that

$$\frac{X_{k,n}}{X_{k+1,n}} \stackrel{P}{\to} 1\,, \quad k = k(n) \to \infty\,, \quad k/n \to 0\,. \tag{4.5}$$

The latter fact will frequently be used in Chapter 6.

For the proof we conclude from Lemma 4.1.9(b) and Example 4.1.10 that

$$\begin{aligned}(X_{1,n}, \ldots, X_{n,n}) &\stackrel{d}{=} \left(F^{\leftarrow}(U_{1,n}), \ldots, F^{\leftarrow}(U_{n,n})\right)\\ &\stackrel{d}{=} \left(F^{\leftarrow}(\Gamma_n/\Gamma_{n+1}), \ldots, F^{\leftarrow}(\Gamma_1/\Gamma_{n+1})\right), \end{aligned} \tag{4.6}$$

where $\Gamma_n = E_1 + \cdots + E_n$ and the E_i are iid standard exponential rvs. Notice that (4.6) holds only for every fixed n. However, we are interested in the weak convergence result (4.5), and therefore it suffices to show (4.5) for one special version of

$$((X_{k,n})_{k=1,\ldots,n})_{n\geq 1}\,.$$

In particular, we may choose this sequence by identifying the lhs and the rhs in (4.6) not only in distribution, but pathwise. Hence we get

$$\frac{X_{k,n}}{X_{k+1,n}} = \frac{F^{\leftarrow}(\Gamma_{n-k+1}/\Gamma_{n+1})}{F^{\leftarrow}(\Gamma_{n-k}/\Gamma_{n+1})}\,. \tag{4.7}$$

Since $\overline{F} \in \mathcal{R}_{-\alpha}$,

$$F^{\leftarrow}(1-t^{-1}) = t^{1/\alpha}L(t)\,, \quad t>0\,, \tag{4.8}$$

for some $L \in \mathcal{R}_0$; see Bingham, Goldie and Teugels [72], Corollary 2.3.4. By the SLLN and since $k/n \to 0$, $\Gamma_{n-k}/\Gamma_{n+1} \stackrel{\text{a.s.}}{\to} 1$. Hence, by (4.7) and (4.8) for sufficiently large n,

$$\frac{X_{k,n}}{X_{k+1,n}} = \left(\frac{\Gamma_{n+1}-\Gamma_{n-k}}{\Gamma_{n+1}-\Gamma_{n-k+1}}\right)^{1/\alpha} \frac{L(\Gamma_{n+1}/(\Gamma_{n+1}-\Gamma_{n-k+1}))}{L(\Gamma_{n+1}/(\Gamma_{n+1}-\Gamma_{n-k}))}\,. \tag{4.9}$$

Again using the SLLN, $k\to\infty$ and $k/n \to 0$,

$$\frac{\Gamma_{n+1}-\Gamma_{n-k}}{\Gamma_{n+1}-\Gamma_{n-k+1}} \stackrel{d}{=} \frac{\Gamma_{k+1}}{\Gamma_k} \stackrel{\text{a.s.}}{\to} 1\,, \tag{4.10}$$

$$\frac{\Gamma_{n+1}-\Gamma_{n-k+1}}{\Gamma_{n+1}} = \frac{\Gamma_{n+1}-\Gamma_{n-k}}{\Gamma_{n+1}}(1+o(1)) \stackrel{\text{a.s.}}{\to} 0 \tag{4.11}$$

Relations (4.9)–(4.11) and the uniform convergence theorem for $L \in \mathcal{R}_0$ (see Theorem A3.2) prove (4.5). □

Example 4.1.12 (Asymptotic properties of the Hill estimator)
Assume X is a positive rv with regularly varying tail $\overline{F}(x) = x^{-\alpha}L_0(x)$ for some $\alpha>0$ and $L_0 \in \mathcal{R}_0$. For applications it is important to know α. In Section 6.4.2 several estimators of α are derived and their statistical properties are studied. The most popular estimator of α was proposed by Hill [326]. It is based on the k upper order statistics of an iid sample:

$$\widehat{\alpha}_n^{-1} = \frac{1}{k-1}\sum_{i=1}^{k-1}\ln\left(\frac{X_{i,n}}{X_{k,n}}\right) = \frac{1}{k-1}\sum_{i=1}^{k-1}\ln X_{i,n} - \ln X_{k,n}\,, \tag{4.12}$$

for $k\geq 2$. We suppress the dependence on k in the notation.

There exist many variations on the theme "Hill" with $k-1$ replaced by k

(and vice versa) at different places in (4.12). By (4.5) all these estimators have the same asymptotic properties provided $k = k(n) \to \infty$ and $k/n \to 0$. We are interested in the asymptotic properties of $\widehat{\alpha}_n^{-1}$ (consistency, asymptotic normality).

By Lemma 4.1.9(b) we may and do assume that $\widehat{\alpha}_n^{-1}$ has representation

$$\widehat{\alpha}_n^{-1} = \frac{1}{k-1} \sum_{i=1}^{k-1} \ln F^{\leftarrow}(U_{i,n}) - \ln F^{\leftarrow}(U_{k,n}) \tag{4.13}$$

for an ordered sample $U_{n,n} < \cdots < U_{i,n}$ from a uniform distribution on $(0,1)$. We are interested only in asymptotic distributional properties of $\widehat{\alpha}_n^{-1}$. For this it suffices to study the distribution of $\widehat{\alpha}_n^{-1}$ at every fixed n. If one wants to study a.s. convergence results one has to consider the distribution of the whole sequence $(\widehat{\alpha}_n)$. Then representation (4.13) is not useful. (Lemma 4.1.9(b) is applicable only for a finite vector of order statistics.) Regular variation of $\overline{F}$ implies that

$$F^{\leftarrow}(y) = (1-y)^{-1/\alpha} L\left((1-y)^{-1}\right) , \quad y \in (0,1) ,$$

for some $L \in \mathcal{R}_0$; see Bingham et al. [72], Corollary 2.3.4. Combining (4.13) with the representation of $(U_{k,n})$ via iid standard exponential rvs E_i (see Example 4.1.10) and writing

$$\Gamma_n = E_1 + \cdots + E_n , \quad n \geq 1 ,$$

we obtain the representation

$$\begin{aligned}
\widehat{\alpha}_n^{-1} &= \frac{1}{k-1} \sum_{i=1}^{k-1} \ln \left[\left(1 - \frac{\Gamma_{n-i+1}}{\Gamma_{n+1}}\right)^{-1/\alpha} L\left(\left(1 - \frac{\Gamma_{n-i+1}}{\Gamma_{n+1}}\right)^{-1}\right)\right] \\
&\qquad - \ln \left[\left(1 - \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}}\right)^{-1/\alpha} L\left(\left(1 - \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}}\right)^{-1}\right)\right] \\
&= \frac{1}{\alpha} \frac{1}{k-1} \sum_{i=1}^{k-1} \ln \frac{\Gamma_{n+1} - \Gamma_{n-k+1}}{\Gamma_{n+1} - \Gamma_{n-i+1}} \\
&\qquad + \frac{1}{k-1} \sum_{i=1}^{k-1} \ln \frac{L(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-i+1}))}{L(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-k+1}))} \\
&= \beta_n^{(1)} + \beta_n^{(2)} .
\end{aligned} \tag{4.14}$$

The leading term in this decomposition is $\beta_n^{(1)}$. It determines the asymptotic properties of the estimator $\widehat{\alpha}_n^{-1}$. Thus we first study $\beta_n^{(1)}$. Again applying Example 4.1.10 we see that for every $k \geq 2$,

$$\left(\frac{\Gamma_{n+1}-\Gamma_{n-i+1}}{\Gamma_{n+1}-\Gamma_{n-k+1}}\right)_{i=1,\ldots,k-1} \stackrel{d}{=} \left(\frac{\Gamma_i}{\Gamma_k}\right)_{i=1,\ldots,k-1} \stackrel{d}{=} (U_{k-i,k-1})_{i=1,\ldots,k-1} \ .$$

Hence, for iid (U_i) uniform on $(0,1)$,

$$\beta_n^{(1)} \stackrel{d}{=} -\frac{1}{\alpha}\frac{1}{k-1}\sum_{i=1}^{k-1}\ln U_i \stackrel{d}{=} \frac{1}{\alpha}\frac{1}{k-1}\sum_{i=1}^{k-1}E_i \, .$$

We immediately conclude from the SLLN and the CLT for iid rvs that

$$\beta_n^{(1)} \stackrel{P}{\rightarrow} \frac{1}{\alpha} \, ,$$

$$\alpha\sqrt{k}\left(\beta_n^{(1)}-\frac{1}{\alpha}\right) \stackrel{d}{\rightarrow} \Phi \, ,$$

where Φ is the standard normal distribution, provided that $k=k(n)\to\infty$. Notice that $\beta_n^{(2)}$ vanishes if the relation $\overline{F}(x)=cx^{-\alpha}$ holds for large x and constant $c>0$, and then the limit theory for $\beta_n^{(1)}$ and for $\widehat{\alpha}_n^{-1}$ is the same. However, for real data one can never assume that the tail $\overline{F}$ has exact power law behaviour. Therefore one also has to understand the limit behaviour of the second term $\beta_n^{(2)}$ in the decomposition (4.14). Recall from the representation theorem (Theorem A3.3) that the slowly varying function L can be written in the form

$$L(x)=c(x)\exp\left\{\int_z^x \frac{\delta(u)}{u}\,du\right\} \, , \quad x\ge z \, , \tag{4.15}$$

for some $z>0$, functions $c(x)\to c_0>0$ and $\delta(x)\to 0$ as $x\to\infty$. With this representation we obtain

$$\begin{aligned}\beta_n^{(2)} &= \frac{1}{k-1}\sum_{i=1}^{k-1}\left(\ln\frac{c\big(\Gamma_{n+1}/(\Gamma_{n+1}-\Gamma_{n-i+1})\big)}{c\big(\Gamma_{n+1}/(\Gamma_{n+1}-\Gamma_{n-k+1})\big)}\right. \\ &\qquad\left. + \int_{(1-\Gamma_{n-k+1}/\Gamma_{n+1})^{-1}}^{(1-\Gamma_{n-i+1}/\Gamma_{n+1})^{-1}} \frac{\delta(u)}{u}\,du\right) \\ &= \beta_n^{(3)}+\beta_n^{(4)} \, .\end{aligned}$$

If we assume that $k=k(n)\to\infty$ and $k/n\to 0$ then, by the SLLN, uniformly for $i\le k$,

$$\frac{\Gamma_{n-i+1}}{\Gamma_{n+1}} = \frac{n^{-1}\Gamma_{n-i+1}}{n^{-1}\Gamma_{n+1}} \stackrel{\text{a.s.}}{\rightarrow} 1 \, .$$

This immediately implies that $\beta_n^{(3)} \stackrel{\text{a.s.}}{\rightarrow} 0$. Set

$$C_n = \sup\left\{|\delta(u)| : u \ge (1-\Gamma_{n-k+1}/\Gamma_{n+1})^{-1}\right\}$$

and notice that, by the remark above, $C_n \stackrel{\text{a.s.}}{\to} 0$. Thus

$$\begin{aligned}\beta_n^{(4)} &\le C_n \frac{1}{k-1}\sum_{i=1}^{k-1}\int_{(1-\Gamma_{n-k+1}/\Gamma_{n+1})^{-1}}^{(1-\Gamma_{n-i+1}/\Gamma_{n+1})^{-1}} \frac{1}{u}\,du \\ &= C_n\,\alpha\,\beta_n^{(1)}\,.\end{aligned}$$

This shows that $\beta_n^{(2)} \stackrel{P}{\to} 0$ provided $k = k(n) \to \infty$ and $k/n \to 0$. This, together with $\beta_n^{(1)} \stackrel{P}{\to} \alpha^{-1}$, proves the consistency of Hill's estimator $\widehat{\alpha}_n^{-1}$ whatever the slowly varying function L_0 in the tail $\overline{F}(x) = x^{-\alpha}L_0(x)$. Under conditions on the growth of $k(n)$ (e.g. $k(n) = [n^\gamma]$ for some $\gamma \in (0,1)$) it can even be shown that $\widehat{\alpha}_n^{-1} \stackrel{\text{a.s.}}{\to} \alpha$. We refer to Mason [445], whose arguments we followed closely in the discussion above, and to Deheuvels, Häusler and Mason [170].

From the course of the proof it is clear that, in order to show a CLT for $\widehat{\alpha}_n^{-1}$, one has to prove $\sqrt{k}\beta_n^{(2)} \stackrel{P}{\to} 0$. This means one has to impose some condition on the decay to zero of the function $\delta(\cdot)$ in the representation (4.15). Alternatively, one needs some regular variation condition with remainder term which has to be specified. We do not intend to go into detail here, but we refer to the discussion in Section 6.4.2 on the Hill estimator and related topics where sufficient conditions for the asymptotic normality (also under dependence of the X_n) are given.

Finally, we want to illustrate that the Hill estimator can perform very poorly if the slowly varying function in the tail is far away from a constant. For the sake of argument, assume that

$$F^{\leftarrow}(y) = (1-y)^{-1/\alpha}(-\ln(1-y))\,, \quad y \in (0,1)\,. \tag{4.16}$$

Observe that

$$\left(\frac{\Gamma_{n+1}-\Gamma_{n-i+1}}{\Gamma_{n+1}}\right)_{i=1,\dots,k-1} \stackrel{d}{=} \left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)_{i=1,\dots,k-1}\,.$$

Then

$$\begin{aligned}\beta_n^{(2)} &= \frac{1}{k-1}\sum_{i=1}^{k-1}\ln\frac{\ln(1-\Gamma_{n-i+1}/\Gamma_{n+1})}{\ln(1-\Gamma_{n-k+1}/\Gamma_{n+1})} \\ &\stackrel{d}{=} \frac{1}{k-1}\sum_{i=1}^{k-1}\ln\frac{\ln(\Gamma_i/\Gamma_{n+1})}{\ln(\Gamma_k/\Gamma_{n+1})} \\ &= \frac{1}{k-1}\sum_{i=1}^{k-1}\ln\frac{1-\ln\Gamma_i/\ln\Gamma_{n+1}}{1-\ln\Gamma_k/\ln\Gamma_{n+1}}\,.\end{aligned}$$

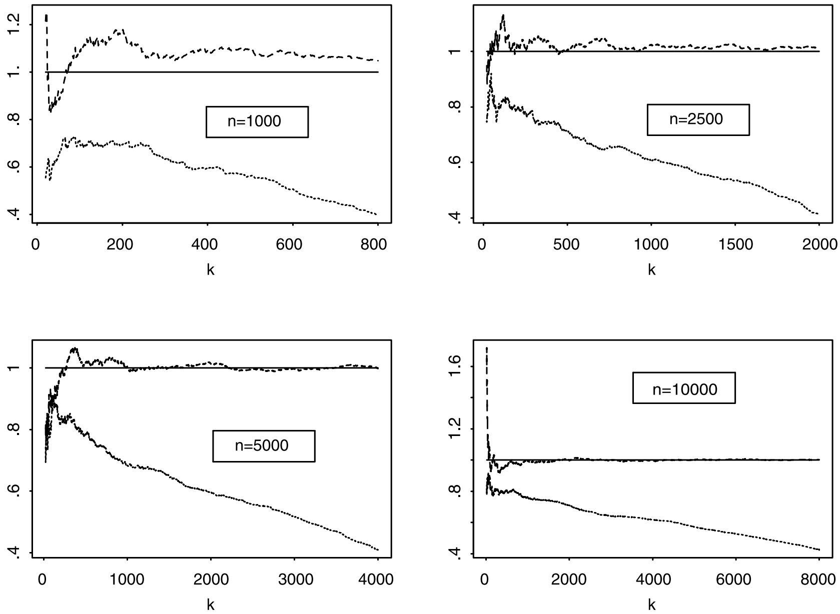


Figure 4.1.13 *A "Hill horror plot": the Hill estimator $\widehat{\alpha}_n^{-1}$ from n iid realisations with distribution tail $\overline{F}_1(x) = 1/x$ (top line) and $\overline{F}_2(x) = 1/(x \ln x)$ (bottom line). The solid line corresponds to $\alpha = 1$. The performance of the Hill estimate for F_2 is very poor. The value k is the number of upper order statistics used for the construction of the Hill estimator* (4.12).

The SLLN and a Taylor–expansion argument applied to the last relation show that, with probability 1, the rhs can be estimated as follows

$$\begin{aligned} &= -(1+o(1))\frac{1}{\ln \varGamma_{n+1}}\frac{1}{k-1}\sum_{i=1}^{k-1}\ln\frac{\varGamma_i}{\varGamma_k} \\ &\stackrel{d}{=} -(1+o(1))\frac{1}{\ln n}\frac{1}{k-1}\sum_{i=1}^{k-1}\ln U_i \\ &\stackrel{d}{=} (1+o(1))(\ln n)^{-1}(k-1)^{-1}\varGamma_{k-1} = O((\ln n)^{-1})\,. \end{aligned}$$

This means that $\beta_n^{(2)} \stackrel{P}{\rightarrow} 0$ at a logarithmic rate. Moreover, if we wanted to construct asymptotic confidence bands via a CLT for the Hill estimator, we would have to compensate for the (essentially $(1/\ln n)$–term) $\beta_n^{(2)}$ in the centring constants. Thus the centring constants in the CLT would depend on the (usually unknown) slowly varying function L. In other words, *there is no standard CLT for $\widehat{\alpha}_n^{-1}$ in the class of regularly varying tails.* These two facts concerning the quality of the Hill estimator should be a warning to everybody applying tail estimates. We also include a "Hill horror plot" (Figure 4.1.13) for the situation as in (4.16). For a further discussion of the Hill estimator we refer to Section 6.4.2. □

The asymptotic properties of the upper order statistic $X_{k,n}$ naturally enter when one studies tail and quantile estimators; see Chapter 6.

Proposition 4.1.14 (Almost sure convergence of order statistics)
Let F be a df with right (left) endpoint $x_F \le \infty$ ($\widetilde{x}_F \ge -\infty$) and $(k(n))$ a non–decreasing integer sequence such that

$$\lim_{n\to\infty} n^{-1}k(n) = c \in [0,1]\,.$$

(*a*) *Then* $X_{k(n),n} \stackrel{\text{a.s.}}{\to} x_F$ ($\widetilde{x}_F$) *according as* $c = 0$ ($c = 1$).

(*b*) *Assume that* $c \in (0,1)$ *is such that there is a unique solution* $x(c)$ *of the equation* $\overline{F}(x) = c$. *Then*

$$X_{k(n),n} \stackrel{\text{a.s.}}{\to} x(c).$$

Proof. We restrict ourselves to showing (b), the proof for (a) goes along the same lines. By (4.1) and the SLLN,

$$\begin{aligned} P(X_{k(n),n} \le x \quad \text{i.o.}) &= P\left(\frac{n}{k(n)}\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i > x\}} < 1 \quad \text{i.o.}\right) \\ &= P\left(\overline{F}(x)(1+o(1)) < c \quad \text{i.o.}\right). \end{aligned}$$

The latter probability is 0 or 1 according as $x < x(c)$ or $x > x(c)$. Hence $\liminf_{n\to\infty} X_{k(n),n} = x(c)$ a.s. In an analogous way one can show that the relation $\limsup_{n\to\infty} X_{k(n),n} = x(c)$ a.s. holds. This proves the proposition. □

Notes and Comments

A standard book on order statistics is David [156], while a more recent one is Arnold, Balakrishnan and Nagaraja [20]. Empirical distributions and processes are basic to all this material. Hence books such as Pollard [504], Shorack and Wellner [579], and van der Vaart and Wellner [628] provide the fundamentals for this section as well as others, and indeed go far beyond. Reiss [526] investigates in particular the link with statistical procedures based on extreme value theory in much greater detail. The latter reference also contains a wealth of interesting bibliographical notes. Two seminal papers on spacings were written by Pyke [518, 519]. They had a great impact on the field and are still worth reading.

4.2 The Limit Distribution of Upper Order Statistics

Let $X_1, \ldots, X_n$ be iid with df F. Recall from Proposition 3.1.1 that for a sequence (u_n) of thresholds and $0 \leq \tau \leq \infty$,

$$\lim_{n\to\infty} P\left(X_{1,n} \leq u_n\right) = e^{-\tau} \quad \Leftrightarrow \quad \lim_{n\to\infty} n\overline{F}\left(u_n\right) = \tau \,. \tag{4.17}$$

In this section we ask:

Can we extend relation (4.17) *to any upper order statistic* $X_{k,n}$ *for a fixed* $k \in \mathbb{N}$*?*

Or even

Can we obtain joint limit probabilities for a fixed number k *of upper order statistics* $X_{k,n}, \ldots, X_{1,n}$*?*

Consider for $n \in \mathbb{N}$ the number of exceedances of the threshold u_n by $X_1, \ldots, X_n$:

$$B_n = \sum_{i=1}^{n} I_{\{X_i > u_n\}} \,.$$

Then B_n is a binomial rv with parameters n and $\overline{F}(u_n)$. In Proposition 4.1.2 we used this quantity for finite n to calculate the df of the kth upper order statistic. Basic to the following result is the fact that exceedances $\{X_i > u_n\}$ tend to become rarer when we raise the threshold. On the other hand, we raise the sample size. We balance two effects so that $EB_n = n\overline{F}(u_n) \to \tau$ as $n \to \infty$, and hence immediately the classical theorem of Poisson applies: $B_n \stackrel{d}{\to} Poi(\tau)$. The thresholds u_n are chosen such that the expected number of exceedances converges. The following result shows that $n\overline{F}(u_n) \to \tau$ is also necessary for the Poisson approximation to hold.

Theorem 4.2.1 (Limit law for the number of exceedances)
Suppose (u_n) *is a sequence in* $\mathbb{R}$ *such that* $n\overline{F}(u_n) \to \tau$ *for some* $\tau \in [0, \infty]$ *as* $n \to \infty$*. Then*

$$\lim_{n\to\infty} P\left(B_n \leq k\right) = e^{-\tau} \sum_{r=0}^{k} \frac{\tau^r}{r!} \,, \qquad k \in \mathbb{N}_0 \,. \tag{4.18}$$

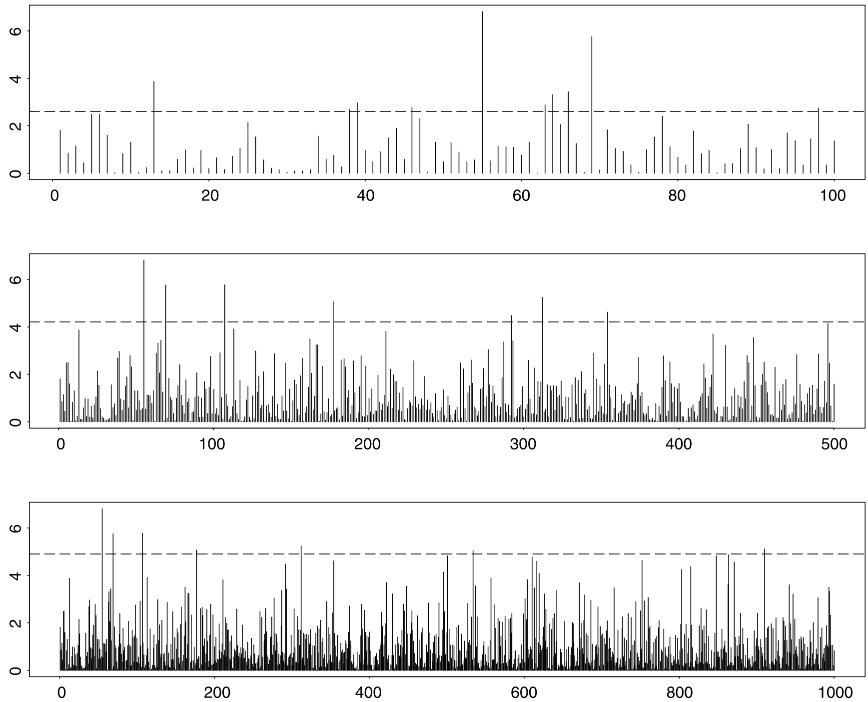


Figure 4.2.2 *Visualisation of the Poisson approximation for extremes of iid standard exponential rvs. The threshold increases with the sample size* $n = 100, 500, 1\,000$. *Notice that the first sample also appears at the beginning of the second and the second at the beginning of the third.*

For $\tau = 0$ *we interpret the rhs as* 1, *for* $\tau = \infty$ *as* 0.

If (4.18) *holds for some* $k \in \mathbb{N}_0$, *then* $n\overline{F}(u_n) \to \tau$ *as* $n \to \infty$, *and thus* (4.18) *holds for all* $k \in \mathbb{N}_0$.

Proof. For $\tau \in (0, \infty)$, sufficiency is simply the Poisson limit theorem as indicated above. For $\tau = 0$, we have

$$P(B_n \le k) \ge P(B_n = 0) = \left(1 - \overline{F}(u_n)\right)^n = \left(1 + o\left(\frac{1}{n}\right)\right)^n \to 1\,.$$

For $\tau = \infty$ we have for arbitrary $\theta > 0$ that $n\overline{F}(u_n) \ge \theta$ for large n. Since the binomial df is decreasing in θ, we obtain

$$P(B_n \le k) \le \sum_{r=0}^{k} \binom{n}{r} \left(\frac{\theta}{n}\right)^r \left(1 - \frac{\theta}{n}\right)^{n-r} .$$

Thus for k fixed,

$$\limsup_{n\to\infty} P(B_n \le k) \le e^{-\theta} \sum_{r=0}^{k} \frac{\theta^{\,r}}{r!} \to 0\,, \qquad \theta \to \infty\,.$$

Hence $P(B_n \le k) \to 0$ as $n \to \infty$.

For the converse assume that (4.18) holds for some $k \in \mathbb{N}_0$, but $n\overline{F}(u_n) \not\to \tau$. Then there exists some $\tau' \ne \tau$ in $[0,\infty]$ and a subsequence (n_k) such that $n_k\overline{F}(u_{n_k}) \to \tau'$ as $k \to \infty$, and thus B_{n_k} converges weakly to a Poisson rv with parameter τ', contradicting (4.18). □

The Poisson approximation (4.18) allows us to derive asymptotics for the kth order statistic. The definition of B_n and (4.1) imply

$$P(B_n < k) = P(X_{k,n} \le u_n)\,, \quad 1 \le k \le n\,, \tag{4.19}$$

which by (4.18) gives immediately the following result.

Theorem 4.2.3 (Limit probabilities for an upper order statistic)
Suppose (u_n) is a sequence in $\mathbb{R}$ such that $n\overline{F}(u_n) \to \tau \in [0,\infty]$ as $n \to \infty$. Then

$$\lim_{n\to\infty} P(X_{k,n} \le u_n) = e^{-\tau} \sum_{r=0}^{k-1} \frac{\tau^r}{r!}\,, \quad k \in \mathbb{N}. \tag{4.20}$$

For $\tau = 0$ we interpret the rhs as 1 and for $\tau = \infty$ as 0.

If (4.20) holds for some $k \in \mathbb{N}$, then $n\overline{F}(u_n) \to \tau$ as $n \to \infty$, and thus (4.20) holds for all $k \in \mathbb{N}$. □

For $u_n = c_n x + d_n$ and $\tau = \tau(x) = -\ln H(x)$ as in Proposition 3.3.2 we obtain the following corollary:

Corollary 4.2.4 (Limit distribution of an upper order statistic)
Suppose $F \in \mathrm{MDA}(H)$ with norming constants $c_n > 0$ and $d_n \in \mathbb{R}$. Define

$$H^{(k)}(x) = H(x) \sum_{r=0}^{k-1} \frac{(-\ln H(x))^r}{r!}\,, \quad x \in \mathbb{R}\,.$$

For x such that $H(x) = 0$ we interpret $H^{(k)}(x) = 0$. Then for each $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P\left(c_n^{-1}(X_{k,n} - d_n) \le x\right) = H^{(k)}(x)\,. \tag{4.21}$$

On the other hand, if for some $k \in \mathbb{N}$

$$\lim_{n\to\infty} P\left(c_n^{-1}(X_{k,n} - d_n) \le x\right) = G(x), \quad x \in \mathbb{R},$$

for a non–degenerate df G, then $G = H^{(k)}$ for some extreme value distribution H and (4.21) holds for all $k \in \mathbb{N}$. □

Example 4.2.5 (Upper order statistics of the Gumbel distribution)
By partial integration,

$$H^{(k)}(x) = \frac{1}{(k-1)!} \int_{-\ln H(x)}^{\infty} e^{-t}\, t^{k-1}\, dt = \Gamma_k\left(-\ln H(x)\right), \quad x \in \mathbb{R},$$

where Γ_k denotes the incomplete gamma function. In particular, if H is the Gumbel distribution Λ, then

$$\Lambda^{(k)}(x) = \frac{1}{(k-1)!} \int_{e^{-x}}^{\infty} e^{-t}\, t^{k-1}\, dt = P\left(\sum_{i=1}^{k} E_i > e^{-x}\right)$$

for $E_1, \ldots, E_k$ iid standard exponential rvs, where we used the well–known fact that $\sum_{i=1}^k E_i$ is $\Gamma(k,1)$–distributed. Hence, if $Y^{(k)}$ has df $\Lambda^{(k)}$, then $Y^{(k)} \stackrel{d}{=} -\ln \sum_{i=1}^k E_i$. □

The limit distribution of the kth upper order statistic was obtained by considering the number of exceedances of a level u_n by $X_1, \ldots, X_n$. Similar arguments can be adapted to prove convergence of the joint distribution of several upper order statistics.

To this end let for $k \in \mathbb{N}$ the levels $u_n^{(k)} \le \cdots \le u_n^{(1)}$ satisfy

$$\lim_{n\to\infty} n\overline{F}(u_n^{(i)}) = \tau_i, \quad i = 1, \ldots, k, \tag{4.22}$$

where $0 \le \tau_1 \le \tau_2 \le \cdots \le \tau_k \le \infty$, and define

$$B_n^{(j)} = \sum_{i=1}^{n} I_{\left\{X_i > u_n^{(j)}\right\}}, \quad j = 1, \ldots, k,$$

i.e. $B_n^{(j)}$ is the number of exceedances of $u_n^{(j)}$ by $X_1, \ldots, X_n$.

Theorem 4.2.6 (Multivariate limit law for the number of exceedances)
Suppose that the sequences $(u_n^{(j)})$ satisfy (4.22) for $j = 1, \ldots, k$. Then for $l_1, \ldots, l_k \in \mathbb{N}_0$,

$$\begin{aligned} &\lim_{n\to\infty} P\left(B_n^{(1)} = l_1, B_n^{(2)} = l_1 + l_2, \ldots, B_n^{(k)} = l_1 + \cdots + l_k\right) \\ &= \frac{\tau_1^{l_1}}{l_1!} \frac{(\tau_2 - \tau_1)^{l_2}}{l_2!} \cdots \frac{(\tau_k - \tau_{k-1})^{l_k}}{l_k!}\, e^{-\tau_k}. \end{aligned} \tag{4.23}$$

The rhs is interpreted as 0 if $\tau_k = \infty$.

Proof. We write $p_{n,j} = \overline{F}(u_n^{(j)})$. Using the defining properties of the multinomial distribution, we find that the lhs probability of (4.23) equals

$$\binom{n}{l_1} p_{n,1}^{l_1} \binom{n-l_1}{l_2} \left(p_{n,2} - p_{n,1}\right)^{l_2} \cdots$$
$$\cdots \binom{n - l_1 - \cdots - l_{k-1}}{l_k} \left(p_{n,k} - p_{n,k-1}\right)^{l_k} \left(1 - p_{n,k}\right)^{n - l_1 - \cdots - l_k} .$$

If $\tau_k < \infty$ then we obtain from (4.22) that

$$\binom{n}{l_1} p_{n,1}^{l_1} \sim \frac{\left(n\, p_{n,1}\right)^{l_1}}{l_1!} \to \frac{\tau_1^{l_1}}{l_1!} ,$$

$$\begin{aligned} \binom{n - l_1 - \cdots - l_{i-1}}{l_i} \left(p_{n,1} - p_{n,i-1}\right)^{l_i} &\sim \frac{\left(np_{n,i} - np_{n,i-1}\right)^{l_i}}{l_i!} \\ &\to \frac{\left(\tau_i - \tau_{i-1}\right)^{l_1}}{l_i!} , \quad \text{for } 2 \le i \le k , \end{aligned}$$

$$\left(1 - p_{n,k}\right)^{n - l_1 - \cdots - l_k} \sim \left(1 - \frac{np_{n,k}}{n}\right)^n \to e^{-\tau_k} ,$$

giving (4.23).

If $\tau_k = \infty$, the probability in (4.23) does not exceed $P(B_n^{(k)} = \sum_{i=1}^k l_i)$. By Theorem 4.2.1, the latter converges to 0. □

Clearly, as in (4.19),

$$\begin{aligned} & P\left(X_{1,n} \le u_n^{(1)}, \ldots, X_{k,n} \le u_n^{(k)}\right) \\ = \;& P\left(B_n^{(1)} = 0,\, B_n^{(2)} \le 1, \ldots, B_n^{(k)} \le k-1\right) , \end{aligned} \tag{4.24}$$

and thus the joint asymptotic distribution of the k upper order statistics can be obtained directly from Theorem 4.2.6. In particular, if $c_n^{-1}(X_{1,n} - d_n)$ converges weakly, then so does the vector

$$\left(c_n^{-1}\left(X_{1,n} - d_n\right), \ldots, c_n^{-1}\left(X_{k,n} - d_n\right)\right) .$$

Although for small k the joint limit distribution of the k upper order statistics can easily be derived from (4.24) and Theorem 4.2.6, the general case is rather complicated. If the df F is absolutely continuous with density f satisfying certain regularity conditions the following heuristic argument can be made precise (for details see Reiss [526], Theorem 5.3.4): suppose $F \in \mathrm{MDA}(H)$ with density f, then the df of the maximum $F^n(c_n x + d_n) = P(c_n^{-1}(X_{1,n} - d_n) \le x)$ has also a density such that for almost all $x \in \mathbb{R}$,

$$nc_n f\left(c_n x+d_n\right) F^{n-1}\left(c_n x+d_n\right) \sim nc_n f\left(c_n x+d_n\right) H(x) \rightarrow h(x)\,,$$

where h is the density of the extreme value distribution H. Furthermore, for $k \in \mathbb{N}$ the weak limit of the random vector $(c_n^{-1}(X_{j,n}-d_n))_{j=1,\ldots,k}$ has, by Theorem 4.1.3, the density

$$\begin{aligned}
&\lim_{n\rightarrow\infty} F^{n-k}\left(c_n x_k+d_n\right) \prod_{j=1}^{k}\Big(\left(n-j+1\right) c_n\, f\left(c_n x_j+d_n\right)\Big)\\
&= \quad H\left(x_k\right) \prod_{j=1}^{k} \frac{h\left(x_j\right)}{H\left(x_j\right)}\,, \qquad x_k<\cdots<x_1\,. \qquad (4.25)
\end{aligned}$$

Definition 4.2.7 (k–dimensional H–extremal variate)
For any extreme value distribution H with density h define for $x_k<\cdots<x_1$ in the support of H

$$h^{(k)}\left(x_1,\ldots,x_k\right)=H\left(x_k\right) \prod_{j=1}^{k} \frac{h\left(x_j\right)}{H\left(x_j\right)}\,.$$

*A vector $\left(Y^{(1)},\ldots,Y^{(k)}\right)$ of rvs with joint density $h^{(k)}$ is called a k–*dimensional H–extremal variate. □

The heuristic argument (4.25) can be made precise and formulated as follows:

Theorem 4.2.8 (Joint limit distribution of k upper order statistics)
Assume that $F\in \mathrm{MDA}(H)$ with norming constants $c_n>0$ and $d_n\in\mathbb{R}$. Then, for every fixed $k\in\mathbb{N}$,

$$\left(c_n^{-1}\left(X_{i,n}-d_n\right)\right)_{i=1,\ldots,k} \quad \stackrel{d}{\rightarrow} \quad (Y^{(i)})_{i=1,\ldots,k}\,, \qquad n\rightarrow\infty\,,$$

where $(Y^{(1)},\ldots,Y^{(k)})$ is a k–dimensional H–extremal variate. □

Example 4.2.9 (Density of a k–dimensional H–extremal variate)

$$\begin{aligned}
H=\Phi_\alpha: \quad & \varphi_\alpha(x_1,\cdots,x_k) \;=\; \alpha^k \exp\Big\{-x_k^{-\alpha}-(\alpha+1)\sum_{j=1}^{k}\ln x_j\Big\}\,,\\
& \qquad 0<x_k<\cdots<x_1\,,\\
H=\Psi_\alpha: \quad & \psi_\alpha(x_1,\cdots,x_k) \;=\; \alpha^k \exp\Big\{-(-x_k)^{\alpha}+(\alpha-1)\sum_{j=1}^{k}\ln(-x_j)\Big\},\\
& \qquad x_k<\cdots<x_1<0\,,\\
H=\Lambda: \quad & \lambda(x_1,\ldots,x_k) \;=\; \exp\Big\{-e^{-x_k}-\sum_{j=1}^{k}x_j\Big\}, \quad x_k<\cdots<x_1\,.
\end{aligned}$$

□

In Example 4.1.5 we investigated the spacings of an exponential sample. Now we may ask

What is the joint limit df of the spacings of a sample of extremal rvs?

Example 4.2.10 (Spacings of Gumbel variables)
The exponential distribution is in MDA(Λ) (see Example 3.2.7) and hence for iid standard exponential rvs $E_1, \ldots, E_n$ we obtain

$$(E_{i,n} - \ln n)_{i=1,\ldots,k+1} \quad \xrightarrow{d} \quad (Y^{(i)})_{i=1,\ldots,k+1}\,, \qquad n \to \infty\,,$$

where $(Y^{(1)}, \ldots, Y^{(k+1)})$ is the $(k+1)$–dimensional Λ–extremal variate with density

$$h^{(k+1)}(x_1, \ldots, x_{k+1}) = \exp\left\{-e^{-x_{k+1}} - \sum_{i=1}^{k+1} x_i\right\}, \quad x_{k+1} < \cdots < x_1\,. \tag{4.26}$$

The continuous mapping theorem (Theorem A2.6) implies for exponential spacings that

$$\begin{aligned}(E_{i,n} - E_{i+1,n})_{i=1,\ldots,k} \quad &= \quad ((E_{i,n} - \ln n) - (E_{i+1,n} - \ln n))_{i=1,\ldots,k} \\ &\xrightarrow{d} \quad \left(Y^{(i)} - Y^{(i+1)}\right)_{i=1,\ldots,k}\,, \quad n \to \infty\,.\end{aligned}$$

By Example 4.1.5 we obtain the representation

$$\left(Y^{(i)} - Y^{(i+1)}\right)_{i=1,\ldots,k} \stackrel{d}{=} \left(i^{-1}\, E_i\right)_{i=1,\ldots,k} \tag{4.27}$$

for iid standard exponential rvs $E_1, \ldots, E_k$. □

Corollary 4.2.11 (Joint limit distribution of upper spacings in MDA(Λ))
Suppose $F \in$ MDA(Λ) with norming constants $c_n > 0$, then

$$(a) \quad \left(c_n^{-1}\,(X_{i,n} - X_{i+1,n})\right)_{i=1,\ldots,k} \quad \xrightarrow{d} \quad \left(i^{-1}\,E_i\right)_{i=1,\ldots,k} \quad \textit{for } k \geq 1\,,$$

$$(b) \quad c_n^{-1}\left(\sum_{i=1}^{k} X_{i,n} - k\,X_{k+1,n}\right) \quad \xrightarrow{d} \quad \sum_{i=1}^{k} E_i \qquad \textit{for } k \geq 2\,,$$

where $E_1, \ldots, E_k$ are iid standard exponential rvs.

Proof. (a) This follows by the same argument as for the exponential rvs in Example 4.2.10.
(b) We apply the continuous mapping theorem (Theorem A2.6):

$$\begin{aligned}
c_n^{-1}\left(\sum_{i=1}^{k} X_{i,n} - k\,X_{k+1,n}\right) &= c_n^{-1}\sum_{i=1}^{k}\left(X_{i,n} - X_{k+1,n}\right) \\
= c_n^{-1}\sum_{i=1}^{k}\sum_{j=i}^{k}\left(X_{j,n} - X_{j+1,n}\right) &= \sum_{i=1}^{k} c_n^{-1} i\left(X_{i,n} - X_{i+1,n}\right) \\
\stackrel{d}{\to} \sum_{i=1}^{k} i\left(Y^{(i)} - Y^{(i+1)}\right) &\stackrel{d}{=} \sum_{i=1}^{k} E_i\,.
\end{aligned}$$

$\square$

Example 4.2.12 (Spacings of Fréchet variables)
The joint density of the spacings of the $(k+1)$–dimensional Fréchet variate $(Y^{(1)},\ldots,Y^{(k+1)})$ can also be calculated. We start with the joint density of $Y^{(1)} - Y^{(2)},\ldots,Y^{(k)} - Y^{(k+1)},\ Y^{(k+1)}$. Define the transformation

$$T\left(x_1,\ldots,x_{k+1}\right) = \left(x_1 - x_2, x_2 - x_3,\ldots,x_k - x_{k+1}, x_{k+1}\right)\,,$$

for $x_{k+1} < \cdots < x_1$. Then $\det(\partial T(\mathbf{x})/\partial \mathbf{x}) = 1$ and for $x_1, x_2,\ldots,x_{k+1} \in \mathbb{R}$ we obtain

$$T^{-1}\left(x_1,\ldots,x_{k+1}\right) = \left(\sum_{j=1}^{k+1} x_j, \sum_{j=2}^{k+1} x_j,\ldots,x_k + x_{k+1}, x_{k+1}\right)\,.$$

For the spacings of the $(k+1)$–dimensional Fréchet variate $(Y^{(1)},\ldots,Y^{(k+1)})$ this yields the density

$$\begin{aligned}
& g_{Y^{(1)}-Y^{(2)},\ldots,Y^{(k)}-Y^{(k+1)},\,Y^{(k+1)}}\left(x_1,\ldots,x_{k+1}\right) \\
&= \alpha^{k+1}\exp\left\{-x_{k+1}^{-\alpha}\right\} x_{k+1}^{-\alpha-1}\left(x_{k+1}+x_k\right)^{-\alpha-1}\cdots\left(x_{k+1}+\cdots+x_1\right)^{-\alpha-1}
\end{aligned}$$

for $x_1,\ldots,x_{k+1} > 0$.

From this density it is obvious that the spacings of the $(k+1)$–dimensional Fréchet variate are dependent. Hence such an elegant result as (4.27) cannot be expected for $F \in \mathrm{MDA}(\Phi_\alpha)$. $\square$

By analogous calculations as above we find the joint limit density of the spacings of the upper order statistics of a sample from a df $F \in \mathrm{MDA}(\Phi_\alpha)$.

Corollary 4.2.13 (Joint limit distribution of upper spacings in $\mathrm{MDA}(\Phi_\alpha)$)
Suppose $F \in \mathrm{MDA}(\Phi_\alpha)$ with norming constants $c_n > 0$. Let $(Y^{(1)},\ldots,Y^{(k+1)})$ be the $(k+1)$–dimensional Fréchet variate. Then

$$(a) \quad \left(c_n^{-1}\left(X_{i,n}-X_{i+1,n}\right)\right)_{i=1,\ldots,k} \quad \stackrel{d}{\rightarrow} \quad \left(Y^{(i)}-Y^{(i+1)}\right)_{i=1,\ldots,k}\,, \quad k\geq 1\,,$$

$$(b) \quad c_n^{-1}\left(\sum_{i=1}^{k} X_{i,n}-k\,X_{k+1,n}\right) \quad \stackrel{d}{\rightarrow} \quad \sum_{i=1}^{k} i\left(Y^{(i)}-Y^{(i+1)}\right)\,, \quad k\geq 2\,.$$

The limit variables in (a) *and* (b) *are defined by the spacings* $Y^{(1)}-Y^{(2)},\ldots,$ $Y^{(k)}-Y^{(k+1)}$ *which have joint density*

$$\begin{aligned} & g_{Y^{(1)}-Y^{(2)},\ldots,Y^{(k)}-Y^{(k+1)}}(x_1,\ldots,x_k) \\ &= \quad \alpha^{k+1}\int_0^\infty \exp\left\{-y^{-\alpha}\right\}\left(y\left(y+x_k\right)\cdots\left(y+x_k+\cdots+x_1\right)\right)^{-\alpha-1}dy \end{aligned}$$

for $x_1,\ldots,x_k>0$. □

Notes and Comments

The Poisson approximation which we applied in this section in order to prove weak limit laws for upper order statistics is a very powerful tool. Its importance in this field, particularly for the investigation of extremes of dependent sequences and stochastic processes, is uncontested. More generally, exceedances of a threshold can be modelled by a point process in the plane. This yields limit laws for maxima of stochastic sequences, allowing us to explain cluster effects in extremes of certain processes. The principle tool is weak convergence of point processes. In Chapter 5 an introduction to this important subject can be found. There also the extremal behaviour of special processes is treated.

There are many other applications of the Poisson approximation in various fields. Recent books are Aldous [7] and Barbour, Holst and Janson [43]; see also the review paper by Arratia, Goldstein and Gordon [22].

4.3 The Limit Distribution of Randomly Indexed Upper Order Statistics

In this section we compare the weak limit behaviour of a finite number of upper order statistics and of randomly indexed maxima for an iid sequence (X_n) of rvs with common df F. As usual, $(N(t))_{t\geq 0}$ is a process of integer–valued rvs which we also suppose to be independent of (X_n). We write

$$X_{n,n}\leq\cdots\leq X_{1,n} \quad \text{and} \quad X_{N(t),N(t)}\leq\cdots\leq X_{1,N(t)}$$

for the order statistics of the samples $X_1, \ldots, X_n$ and $X_1, \ldots, X_{N(t)}$, respectively, and we also use

$$M_n = X_{1,n} \quad \text{and} \quad M_{N(t)} = X_{1,N(t)}$$

for the corresponding sample maxima.

If F belongs to the maximum domain of attraction of the extreme value distribution H ($F \in \mathrm{MDA}(H)$), there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$c_n^{-1}\,(M_n - d_n) \xrightarrow{d} H\,. \tag{4.28}$$

It is a natural question to ask:

Does relation (4.28) *remain valid along the random index set* $(N(t))$*?*

From Lemma 2.5.6 we already know that (4.28) implies

$$c_{N(t)}^{-1}\left(M_{N(t)} - d_{N(t)}\right) \xrightarrow{d} H$$

provided $N(t) \xrightarrow{P} \infty$, but we want to keep the old norming sequences (c_n), (d_n) instead of the random processes $(c_{N(t)})$, $(d_{N(t)})$. This can be done under quite general conditions as we will soon see. However, the limit distribution will also then change.

We proceed as in Section 4.2. We introduce the variables

$$B_t^{(i)} = \sum_{j=1}^{N(t)} I_{\{X_j > u_t^{(i)}\}}\,, \quad i = 1, \ldots, k\,,$$

which count the number of exceedances of the (non–random) thresholds

$$u_t^{(k)} \le \cdots \le u_t^{(1)}\,, \quad t \ge 0\,, \tag{4.29}$$

by $X_1, \ldots, X_{N(t)}$. We also suppose that there exist numbers

$$0 \le \tau_1 \le \cdots \le \tau_k \le \infty$$

such that for $i = 1, \ldots, k$,

$$t\,p_{t,i} \;=\; t\,\overline{F}(u_t^{(i)}) \;\to\; \tau_i\,, \quad t \to \infty\,. \tag{4.30}$$

The following result is analogous to Theorem 4.2.6:

Theorem 4.3.1 (Multivariate limit law for the number of exceedances)
Suppose that $(u_t^{(i)})_{t\ge 0}$, $i=1,\ldots,k$, *satisfy* (4.29) *and* (4.30). *Assume there exists a non–negative rv* Z *such that*

$$\frac{N(t)}{t} \xrightarrow{P} Z\,, \quad t\to\infty\,. \tag{4.31}$$

Then, for all integers $l_i \ge 0$, $i=1,\ldots,k$,

$$\begin{aligned}&\lim_{t\to\infty} P\left(B_t^{(1)}=l_1, B_t^{(2)}=l_1+l_2,\ldots,B_t^{(k)}=l_1+\cdots+l_k\right)\\ &= E\left[\frac{(Z\tau_1)^{l_1}}{l_1!}\,\frac{(Z(\tau_2-\tau_1))^{l_2}}{l_2!}\cdots\frac{(Z(\tau_k-\tau_{k-1}))^{l_k}}{l_k!}\,e^{-Z\tau_k}\right].\end{aligned}$$

The rhs is interpreted as 0 *if* $\tau_k=\infty$.

Proof. We proceed in a similar way as for the proof of Theorem 4.2.6. For the sake of simplicity we restrict ourselves to the case $k=2$. We condition on $N(t)$, use the independence of $(N(t))$ and (X_n) and apply (4.30) and (4.31):

$$\begin{aligned}&P\left(B_t^{(1)}=l_1, B_t^{(2)}=l_1+l_2\,\middle|\, N(t)\right) \qquad (4.32)\\
&= \binom{N(t)}{l_1}p_{t,1}^{l_1}\binom{N(t)-l_1}{l_2}(p_{t,2}-p_{t,1})^{l_2}(1-p_{t,2})^{N(t)-l_1-l_2}\\
&= (1+o_P(1))\,\frac{(N(t)p_{t,1})^{l_1}}{l_1!}\,\frac{(N(t)(p_{t,2}-p_{t,1}))^{l_2}}{l_2!}(1-p_{t,2})^{N(t)}\\
&= (1+o_P(1))\,\frac{\left(\frac{N(t)}{t}(tp_{t,1})\right)^{l_1}}{l_1!}\,\frac{\left(\frac{N(t)}{t}(t(p_{t,2}-p_{t,1}))\right)^{l_2}}{l_2!}\times\\
&\qquad\times\exp\left\{\frac{N(t)}{t}\Big(t\ln(1-p_{t,2})\Big)\right\}\\
&\xrightarrow{P} \frac{(Z\tau_1)^{l_1}}{l_1!}\,\frac{(Z(\tau_2-\tau_1))^{l_2}}{l_2!}\,e^{-Z\tau_2}\,, \qquad t\to\infty\,. \qquad (4.33)\end{aligned}$$

Notice that the expressions in (4.32) are uniformly integrable and that (4.33) is integrable. Hence we may conclude (see for instance Karr [373], Theorem 5.17) that

$$\begin{aligned}P\left(B_t^{(1)}=l_1, B_t^{(2)}=l_1+l_2\right) &= E\left[P\left(B_t^{(1)}=l_1, B_t^{(2)}=l_1+l_2\,\middle|\, N(t)\right)\right]\\
&\to E\left[\frac{(Z\tau_1)^{l_1}}{l_1!}\,\frac{(Z(\tau_2-\tau_1))^{l_2}}{l_2!}\,e^{-Z\tau_2}\right]\end{aligned}$$

as $t \to \infty$, which concludes the proof. □

Now one could use the identity

$$P\left(X_{1,N(t)} \le u_t^{(1)}, \ldots, X_{k,N(t)} \le u_t^{(k)}\right)$$
$$= P\left(B_t^{(1)} = 0,\, B_t^{(2)} \le 1, \ldots, B_t^{(k)} \le k-1\right)$$

and Theorem 4.3.1 to derive the limit distribution of the vector of upper order statistics $(X_{1,N(t)}, \ldots, X_{k,N(t)})$. This, however, leads to quite complicated formulae, and so we restrict ourselves to some particular cases.

First we study the limit distribution of a single order statistic $X_{k,N(t)}$ for fixed $k \in \mathbb{N}$. For this reason we suppose that $F \in \mathrm{MDA}(H)$, i.e. (4.28) is satisfied for appropriate constants $c_n > 0$ and $d_n \in \mathbb{R}$. From Proposition 3.3.2 we know that (4.28) is equivalent to

$$\lim_{n\to\infty} n\overline{F}\,(c_n x + d_n) = -\ln H(x)\,, \quad x \in \mathbb{R}\,. \tag{4.34}$$

Under (4.34) it follows for every $k \in \mathbb{N}$ that the relation

$$\lim_{n\to\infty} P\left(c_n^{-1}\,(X_{k,n} - d_n) \le x\right) \;=\; \Gamma_k(-\ln H(x))\,, \qquad x \in \mathbb{R}\,,$$

holds, where Γ_k denotes the incomplete gamma function; see Corollary 4.2.4 and Example 4.2.5. A similar statement is true for randomly indexed upper order statistics.

Theorem 4.3.2 (Limit distribution of the kth upper order statistic in a randomly indexed sample)
Suppose that $N(t)/t \xrightarrow{P} Z$ holds for a non–negative rv Z with df F_Z and that (4.34) is satisfied. Then

$$\begin{aligned} &\lim_{n\to\infty} P\left(c_n^{-1}\left(X_{k,N(n)} - d_n\right) \le x\right) \\ &\quad= \int_0^\infty \Gamma_k(-z\ln H(x))\, dF_Z(z) \qquad (4.35) \\ &\quad= E\left[\Gamma_k\left(-\ln H^Z(x)\right)\right]\,, \qquad x \in \mathbb{R}\,. \end{aligned}$$

Proof. We use the same ideas as in the proof of Theorem 4.3.1. Write

$$B_n = \sum_{j=1}^{N(n)} I_{\{X_j > c_n x + d_n\}}\,.$$

Conditioning on $N(n)$, we find that

$$
\begin{aligned}
&P\left(c_n^{-1}\left(X_{k,N(n)}-d_n\right)\leq x \mid N(n)\right)\\
&= \quad P\left(B_n \leq k-1 \mid N(n)\right)\\
&= \quad \sum_{i=0}^{k-1}\binom{N(n)}{i}\left(F\left(c_n x+d_n\right)\right)^{N(n)-i}\left(\overline{F}\left(c_n x+d_n\right)\right)^i\\
&= \quad (1+o_P(1))\sum_{i=0}^{k-1}\frac{1}{i!}\left(\frac{N(n)}{n}\left(n\overline{F}\left(c_n x+d_n\right)\right)\right)^i\times\\
&\qquad\qquad\times\exp\left\{\frac{N(n)}{n}\left(n\ln\left(1-\overline{F}\left(c_n x+d_n\right)\right)\right)\right\}\\
&\xrightarrow{P} \quad \sum_{i=0}^{k-1}\frac{1}{i!}(-Z\ln H(x))^i e^{Z\ln H(x)}\\
&= \quad H^Z(x)\sum_{i=0}^{k-1}\frac{\left(-\ln H^Z(x)\right)^i}{i!}\,.
\end{aligned}
$$

Taking expectations in the limit relation above, we arrive at (4.35). □

Example 4.3.3 Let $\tilde{N}=(\tilde{N}(t))_{t\geq 0}$ be a homogeneous Poisson process with intensity 1 and let Z be a positive rv independent of $\tilde{N}$. Then

$$N(t)=\tilde{N}(Zt)\,,\qquad t\geq 0,$$

defines a so–called *mixed Poisson process*. The latter class of processes has been recognized as important in insurance; see Section 1.3.3 and Grandell [284]. Notice that, conditionally upon Z, N is a homogeneous Poisson process with intensity Z. Hence, by the SLLN for renewal counting processes (Theorem 2.5.10),

$$P\left(\frac{N(t)}{t}\to Z\,\middle|\, Z\right)=1\quad \text{a.s.}$$

Thus taking expectations on both sides,

$$P\left(\frac{N(t)}{t}\to Z\right)=1.$$

This shows that Theorems 4.3.1 and 4.3.2 are applicable to the order statistics of a sample indexed by a mixed Poisson process. □

For practical purposes, it often suffices to consider processes $(N(t))$ satisfying

$$\frac{N(t)}{t}\quad\xrightarrow{P}\quad\lambda \tag{4.36}$$

for some constant $\lambda > 0$. For example, the renewal counting processes, including the important homogeneous Poisson process, satisfy (4.36) under general conditions; see Section 2.5.2. Analogous arguments to those in Section 4.2 combined with the ones in the proof of Theorem 4.3.1 lead to the following:

Theorem 4.3.4 (Limit distribution of a vector of randomly indexed upper order statistics)
Assume that (4.36) *holds for a positive constant* λ *and that* $F \in \mathrm{MDA}(H)$ *for an extreme value distribution* H *such that* (4.28) *is satisfied. Then*

$$\left(c_n^{-1}\left(X_{i,N(n)} - d_n\right)\right)_{i=1,\ldots,k} \quad \stackrel{d}{\to} \quad (Y_\lambda^{(i)})_{i=1,\ldots,k}\,,$$

where $(Y_\lambda^{(1)},\ldots,Y_\lambda^{(k)})$ *denotes the* k*–dimensional extremal variate corresponding to the extreme value distribution* H^λ*. In particular,*

$$\lim_{n\to\infty} P\left(c_n^{-1}\left(X_{k,n} - d_n\right) \le x\right) = \Gamma_k\left(-\ln H^\lambda(x)\right)\,, \quad x \in \mathbb{R}\,. \qquad \square$$

Notes and Comments

The limit distribution of randomly indexed maxima and order statistics under general dependence assumptions between $(N(t))$ and (X_n) has been studied in Galambos [249] and in Barakat and El–Shandidy [42]. General randomly indexed sequences of rvs have been considered in Korolev [404]; see also the list of references therein.

Randomly indexed maxima and order statistics occur in a natural way when one is interested in the extreme value theory of the individual claims in an insurance portfolio up to time t. Randomly indexed order statistics are of particular interest for reinsurance where they occur explicitly as quantities in reinsurance treaties, as for instance when a reinsurer will cover the k largest claims of a company over a given period of time. This issue is discussed in more detail in Section 8.7.

4.4 Some Extreme Value Theory for Stationary Sequences

One of the natural generalisations of an iid sequence is a strictly stationary process: we say that the sequence of rvs (X_n) is *strictly stationary* if its finite–dimensional distributions are invariant under shifts of time, i.e.

$$(X_{t_1},\ldots,X_{t_m}) \stackrel{d}{=} (X_{t_1+h},\ldots,X_{t_m+h})$$

for any choice of indices $t_1 < \cdots < t_m$ and integers h; see also Appendix A2.1. It is common to define (X_n) with index set $\mathbb{Z}$. We can think of (X_n) as a time series of observations at discrete equidistant instants of time where the distribution of a block $(X_t, X_{t+1}, \ldots, X_{t+h})$ of length h is the same for all integers t.

For simplicity we use throughout the notion of a "stationary" sequence for a "strictly stationary" one. A strictly stationary sequence is naturally also *stationary in the wide sense* or *second order stationary* provided the second moment of $X = X_0$ is finite, i.e. $EX_n = EX$ for all n and $\operatorname{cov}(X_n, X_m) = \operatorname{cov}(X_0, X_{|n-m|})$ for all n and m.

It is impossible to build up a general extreme value theory for the class of all stationary sequences. Indeed, one has to specify the dependence structure of (X_n). For example, assume $X_n = X$ for all n. This relation defines a stationary sequence and

$$P\left(M_n \le x\right) = P(X \le x) = F(x)\,, \quad x \in \mathbb{R}\,.$$

Thus the distribution of the sample maxima can be *any* distribution F. This is not a reasonable basis for a general theory.

The other extreme of a stationary sequence occurs when the X_n are mutually independent, i.e. (X_n) is an iid sequence. In that case we studied the weak limit behaviour of the upper order statistics in Section 4.2. In particular, we know that there exist only three types of different limit laws: the *Fréchet distribution* Φ_α, the *Weibull distribution* Ψ_α and the *Gumbel distribution* Λ (Fisher–Tippett Theorem 3.2.7). The dfs of the type of $\Phi_\alpha, \Psi_\alpha, \Lambda$ are called extreme value distributions. In this section we give conditions on the stationary sequence (X_n) which ensure that its sample maxima (M_n) and the corresponding maxima $(\widetilde{M}_n)$ of an iid sequence $(\widetilde{X}_n)$ with common df $F(x) = P(\widetilde{X} \le x)$ exhibit a similar limit behaviour. We call $(\widetilde{X}_n)$ *an iid sequence associated with* (X_n) or simply *an associated iid sequence*. As before we write $F \in \mathrm{MDA}(H)$ for any of the extreme value distributions H if there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that $c_n^{-1}(\widetilde{M}_n - d_n) \xrightarrow{d} H$. For the derivation of the limit probability of $P(\widetilde{M}_n \le u_n)$ for a sequence of thresholds (u_n) we made heavy use of the following factorisation property:

$$\begin{aligned} P\left(\widetilde{M}_n \le u_n\right) &= P^n\left(\widetilde{X} \le u_n\right) \qquad (4.37) \\ &= \exp\left\{n \ln\left(1 - P\left(\widetilde{X} > u_n\right)\right)\right\} \\ &\approx \exp\left\{-n\overline{F}\left(u_n\right)\right\}\,. \end{aligned}$$

In particular, we concluded in Proposition 3.1.1 that, for any $\tau \in [0, \infty]$, $P(\widetilde{M}_n \le u_n) \to \exp\{-\tau\}$ if and only if $n\overline{F}(u_n) \to \tau \in [0, \infty]$. It is clear that we cannot directly apply (4.37) to maxima of a dependent stationary sequence. However, to overcome this problem we assume that there is *a specific type of asymptotic independence*:

Condition $D(u_n)$**:** *For any integers p, q and n*

$$1 \le i_1 < \cdots < i_p < j_1 < \cdots < j_q \le n$$

such that $j_1 - i_p \ge l$ *we have*

$$\left| P\left(\max_{i \in A_1 \cup A_2} X_i \le u_n \right) - P\left(\max_{i \in A_1} X_i \le u_n \right) P\left(\max_{i \in A_2} X_i \le u_n \right) \right| \le \alpha_{n,l} \,,$$

where $A_1 = \{i_1, \ldots, i_p\}$*,* $A_2 = \{j_1, \ldots, j_q\}$ *and* $\alpha_{n,l} \to 0$ *as* $n \to \infty$ *for some sequence* $l = l_n = o(n)$*.*

This condition as well as $D'(u_n)$ below and their modifications have been intensively applied to stationary sequences in the monograph by Leadbetter, Lindgren and Rootzén [418]. Condition $D(u_n)$ is a distributional mixing condition, weaker than most of the classical forms of dependence restrictions. A discussion of the role of $D(u_n)$ as a specific mixing condition can be found in Leadbetter et al. [418], Sections 3.1 and 3.2. Condition $D(u_n)$ implies, for example, that

$$P\left(M_n \le u_n\right) = P^k\left(M_{[n/k]} \le u_n\right) + o(1) \tag{4.38}$$

for constant or slowly increasing k. This relation already indicates that the limit behaviour of (M_n) and its associated sequence $(\widetilde{M}_n)$ must be closely related. The following result (Theorem 3.3.3 in Leadbetter et al. [418]) even shows that the classes of possible limit laws for the normalised and centred sequences (M_n) and $(\widetilde{M}_n)$ coincide.

Theorem 4.4.1 (Limit laws for maxima of a stationary sequence)
Suppose $c_n^{-1}(M_n - d_n) \xrightarrow{d} G$ *for some distribution* G *and appropriate constants* $c_n > 0$*,* $d_n \in \mathbb{R}$*. If the condition* $D(c_n x + d_n)$ *holds for all real* x*, then* G *is an extreme value distribution.*

Proof. Recall from Theorems 3.2.2, 3.2.3 and from Definition 3.2.6 that G is an extreme value distribution if and only if G is max–stable. By (4.38),

$$P\left(M_{nk} \le c_n x + d_n\right) \quad = \quad P^k\left(M_n \le c_n x + d_n\right) + o(1) \quad \to \quad G^k(x)$$

for every integer $k \ge 1$, and every continuity point x of G. On the other hand,

$$P\left(M_{nk} \le c_{nk} x + d_{nk}\right) \quad \to \quad G(x)\,.$$

Now we may proceed as in the proof of Theorem 3.2.2 to conclude that G is max–stable. □

Remark. 1) Theorem 4.4.1 does *not* mean that the relations $c_n^{-1}(M_n - d_n) \xrightarrow{d} G$ and $c_n^{-1}(\widetilde{M}_n - d_n) \xrightarrow{d} H$ hold with $G = H$. We will see later that G is often of the form H^θ for some $\theta \in [0,1]$ (see for instance Example 4.4.2 and Section 8.1); θ is then called *extremal index.* □

Thus max–stability of the limit distribution is *necessary* under the conditions $D(c_n x + d_n)$, $x \in \mathbb{R}$. Next we want to find *sufficient* conditions for convergence of the probabilities $P(M_n \le u_n)$ for a given threshold sequence (u_n) satisfying

$$n\overline{F}\,(u_n) \to \tau \tag{4.39}$$

for some $\tau \in [0,\infty)$. From Proposition 3.1.1 we know that (4.39) and $P(\widetilde{M}_n \le u_n) \to \exp\{-\tau\}$ are equivalent. But may we replace $(\widetilde{M}_n)$ by (M_n) under $D(u_n)$? The answer is, unfortunately, NO. All one can derive is

$$\liminf_{n\to\infty} P\,(M_n \le u_n) \ge e^{-\tau}\,;$$

see the proof of Proposition 4.4.3 below.

Example 4.4.2 (See also Figure 4.4.5.) Assume that (Y_n) is a sequence of iid rvs with df $\sqrt{F}$ for some df F. Define the sequence (X_n) by

$$X_n = \max\,(Y_n, Y_{n+1})\,, \quad n \in \mathbb{N}\,.$$

Then (X_n) is a stationary sequence and X_n has df F for all $n \ge 1$. From this construction it is clear that maxima of (X_n) appear as pairs at consecutive indices.

Now assume that for $\tau \in (0,\infty)$ the sequence u_n satisfies $u_n \uparrow x_F$ (x_F is the right endpoint of F) and (4.39). Then $F(u_n) \to 1$ and

$$nP\,(Y_1 > u_n) = n\left(1 - \sqrt{F(u_n)}\right) = \frac{n\overline{F}\,(u_n)}{1 + \sqrt{F\,(u_n)}} \to \frac{\tau}{2}\,.$$

Hence, by Proposition 3.1.1,

$$\begin{aligned} P\,(M_n \le u_n) &= P\,(\max\,(Y_1, \ldots, Y_n, Y_{n+1}) \le u_n) \\ &= P\,(\max\,(Y_1, \ldots, Y_n) \le u_n)\,\sqrt{F\,(u_n)} \\ &\to e^{-\tau/2}\,. \end{aligned}$$

Condition $D(u_n)$ is naturally satisfied: if A_1 and A_2 are chosen as in $D(u_n)$ and $l \ge 2$, then we can take $\alpha_{n,l} = 0$. □

This example supports the introduction of a second technical condition.

Condition $D'(u_n)$**:** *The relation*

$$\lim_{k\to\infty}\limsup_{n\to\infty}\; n\sum_{j=2}^{[n/k]} P\left(X_1 > u_n, X_j > u_n\right) = 0\,.$$

Remark. 2) $D'(u_n)$ is an "anti–clustering condition" on the stationary sequence (X_n). Indeed, notice that $D'(u_n)$ implies

$$E\sum_{1\le i<j\le [n/k]} I_{\{X_i>u_n, X_j>u_n\}} \le [n/k]\sum_{j=2}^{[n/k]} EI_{\{X_1>u_n, X_j>u_n\}} \to 0\,,$$

so that, in the mean, joint exceedances of u_n by pairs (X_i, X_j) become very unlikely for large n. □

Now we have introduced the conditions which are needed to formulate the following analogue of Proposition 3.1.1; see Theorem 3.4.1 in Leadbetter et al. [418]:

Proposition 4.4.3 (Limit probabilities for sample maxima)
Assume that the stationary sequence (X_n) *and the threshold sequence* (u_n) *satisfy* $D(u_n)$, $D'(u_n)$. *Suppose* $\tau \in [0,\infty)$. *Then condition* (4.39) *holds if and only if*

$$\lim_{n\to\infty} P\left(M_n \le u_n\right) = e^{-\tau}\,. \tag{4.40}$$

Proof. We restrict ourselves to the sufficiency part in order to illustrate the use of the conditions $D(u_n)$ and $D'(u_n)$. The necessity follows by similar arguments.
We have, for any $l \ge 1$,

$$\begin{aligned} &\sum_{i=1}^{l} P\left(X_i > u_n\right) - \sum_{1\le i<j\le l} P\left(X_i > u_n, X_j > u_n\right) \\ &\le \quad P\left(M_l > u_n\right) \quad \le \quad \sum_{i=1}^{l} P\left(X_i > u_n\right)\,. \end{aligned} \tag{4.41}$$

Exploiting the stationarity of (X_n) we see that

$$\sum_{i=1}^{l} P\left(X_i > u_n\right) \quad = \quad l\,\overline{F}\left(u_n\right)\,,$$

$$\sum_{1\le i<j\le l} P\left(X_i > u_n, X_j > u_n\right) \quad \le \quad l\sum_{j=2}^{l} P\left(X_1 > u_n, X_j > u_n\right)\,.$$

Combining this and (4.41) for $l = [n/k]$ ($[x]$ denotes the integer part of x) and for a fixed k, we derive upper and lower estimates for $P(M_{[n/k]} \le u_n)$:

$$\begin{aligned} 1 - [n/k]\,\overline{F}(u_n) &\le P\left(M_{[n/k]} \le u_n\right) \\ &\le 1 - [n/k]\,\overline{F}(u_n) + [n/k] \sum_{j=2}^{[n/k]} P\left(X_1 > u_n, X_j > u_n\right) . \end{aligned}$$

From (4.39) we immediately have

$$[n/k]\,\overline{F}(u_n) \to \tau/k\,, \quad n \to \infty\,,$$

and, by condition $D'(u_n)$,

$$\limsup_{n\to\infty} [n/k] \sum_{j=2}^{[n/k]} P\left(X_1 > u_n, X_j > u_n\right) = o(1/k)\,, \quad k \to \infty\,.$$

Thus we get the bounds

$$1 - \frac{\tau}{k} \le \liminf_{n\to\infty} P\left(M_{[n/k]} \le u_n\right) \le \limsup_{n\to\infty} P\left(M_{[n/k]} \le u_n\right) \le 1 - \frac{\tau}{k} + o(1/k)\,.$$

This and relation (4.38) imply that

$$\begin{aligned} \left(1 - \frac{\tau}{k}\right)^k &\le \liminf_{n\to\infty} P\left(M_n \le u_n\right) \\ &\le \limsup_{n\to\infty} P\left(M_n \le u_n\right) \le \left(1 - \frac{\tau}{k} + o(1/k)\right)^k . \end{aligned}$$

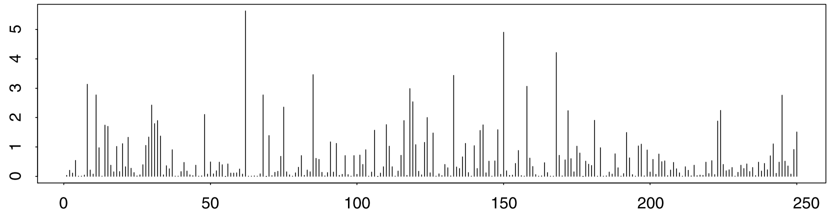
Letting $k \to \infty$ we see that

$$\lim_{n\to\infty} P\left(M_n \le u_n\right) = e^{-\tau}\,.$$

This concludes the proof. □

Example 4.4.4 (Continuation of Example 4.4.2)
We observed in Example 4.4.2 that condition (4.39) implies $P(M_n \le u_n) \to \exp\{-\tau/2\}$. We have already checked that $D(u_n)$ is satisfied. Thus $D'(u_n)$ must go wrong. This can be easily seen: since X_1 and X_j are independent for $j \ge 2$ we conclude that

$$\begin{aligned} & n \sum_{j=2}^{[n/k]} P\left(X_1 > u_n, X_j > u_n\right) \\ &= nP\left(X_1 > u_n, X_2 > u_n\right) + n([n/k] - 2)P^2\left(X_1 > u_n\right) \\ &= nP\left(\max(Y_1, Y_2) > u_n\,,\, \max(Y_2, Y_3) > u_n\right) + \tau^2/k + o(1) \\ &= n\left(P\left(Y_2 > u_n, Y_3 \le u_n\right) + P\left(Y_2 > u_n \text{ or } Y_1 > u_n\,,\, Y_3 > u_n\right)\right) \\ &\quad + \tau^2/k + o(1)\,, \quad n \to \infty\,. \end{aligned}$$



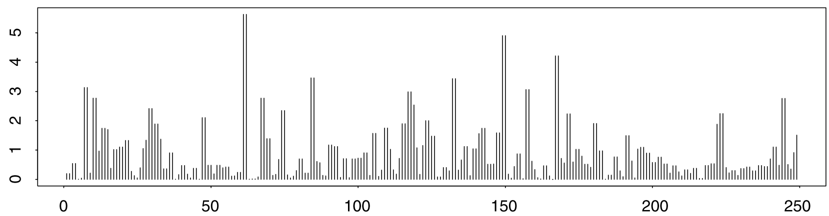


Figure 4.4.5 *A realisation of the sequences* (Y_n) *(top) and* (X_n) *(bottom) with* F *standard exponential as discussed in Examples 4.4.2 and 4.4.4. Extremes appear in clusters of size* 2.

We have

$$nP(Y_2 > u_n, Y_3 \le u_n) \sim nP(Y_1 > u_n) \to \tau/2\,.$$

Thus condition $D'(u_n)$ cannot be satisfied. The reason for this is that maxima in (X_n) appear in clusters of size 2. Notice that

$$E\left(\sum_{i=1}^{n} I_{\{X_i>u_n, X_{i+1}>u_n\}}\right) = nP(X_1 > u_n, X_2 > u_n) \to \tau/2 > 0\,,$$

so that in the long run the expected number of joint exceedances of u_n by the pairs (X_i, X_{i+1}) stabilises around a positive number. □

Proceeding precisely as in Section 3.3 we can now derive the limit distribution for the maxima M_n:

Theorem 4.4.6 (Limit distribution of maxima of a stationary sequence)
Let (X_n) *be a stationary sequence with common df* $F \in \mathrm{MDA}(H)$ *for some extreme value distribution* H*, i.e. there exist constants* $c_n > 0$, $d_n \in \mathbb{R}$ *such that*

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\ln H(x)\,, \quad x \in \mathbb{R}\,. \tag{4.42}$$

Assume that for $x \in \mathbb{R}$ the sequences $(u_n) = (c_n x + d_n)$ satisfy the conditions $D(u_n)$ and $D'(u_n)$. Then (4.42) *is equivalent to each of the following relations:*

$$c_n^{-1}\,(M_n - d_n) \quad \stackrel{d}{\to} \quad H\,, \tag{4.43}$$

$$c_n^{-1}(\widetilde{M}_n - d_n) \quad \stackrel{d}{\to} \quad H\,. \tag{4.44}$$

Proof. The equivalence of (4.42) and (4.44) is immediate from Proposition 3.3.2. The equivalence of (4.42) and (4.43) follows from Proposition 4.4.3. □

From the discussion above we are not surprised about the same limit behaviour of the maxima of a stationary sequence and its associated iid sequence; the conditions $D(c_n x + d_n)$ *and* $D'(c_n x + d_n)$ force the sequence (M_n) to behave very much like the maxima of an iid sequence. Notice that Theorem 4.4.6 also ensures that we can choose the sequences (c_n) and (d_n) in the same way as proposed in Section 3.3.

Thus the problem about the maxima of a stationary sequence has been reduced to a question about the extremes of iid rvs. However, now one has to verify the conditions $D(c_n x + d_n)$ and $D'(c_n x + d_n)$ which, in general, is tedious. Conditions $D(u_n)$ and $D'(u_n)$ have been discussed in detail in the monograph by Leadbetter et al. [418]. The case of a Gaussian stationary sequence is particularly nice: one can check $D(u_n)$ and $D'(u_n)$ via the asymptotic behaviour of the autocovariances

$$\gamma(h) = \operatorname{cov}(X_0, X_h)\,, \quad h \geq 0\,.$$

The basic idea is that the distributions of two Gaussian vectors are "close" to each other if their covariance matrices are "close". Leadbetter et al. [418] make this concept precise by a so–called *normal comparison lemma* (their Theorem 4.2.1), a particular consequence of which is the estimate

$$\begin{aligned} &\left|P\left(X_{i_1} \leq u_n, \ldots, X_{i_k} \leq u_n\right) - \Phi^k\left(u_n\right)\right| \\ \leq \quad & \text{const}\; n \sum_{h=1}^{n} |\gamma(h)| \exp\left\{\frac{-u_n^2}{1+|\gamma(h)|}\right\} \end{aligned}$$

for $1 \leq i_1 < \cdots < i_k \leq n$. Here (X_n) is stationary with marginal df the standard normal Φ, and it is assumed that $\sup_{h\geq 1} |\gamma(h)| < 1$. In particular,

$$\left|P\left(M_n \leq u_n\right) - \Phi^n\left(u_n\right)\right| \leq \text{const}\; n \sum_{h=1}^{n} |\gamma(h)| \exp\left\{\frac{-u_n^2}{1+|\gamma(h)|}\right\}\,. \tag{4.45}$$

Now it is not difficult to check conditions $D(u_n)$ and $D'(u_n)$. For details see Lemma 4.4.1 in Leadbetter et al. [418]:

Lemma 4.4.7 (Conditions for $D(u_n)$ and $D'(u_n)$ for a Gaussian stationary sequence)

Assume (X_n) is stationary Gaussian and let (u_n) be a sequence of real numbers.

(a) *Suppose the rhs in* (4.45) *tends to zero as $n \to \infty$ and* $\sup_{h \geq 1} |\gamma(h)| < 1$. *Then $D(u_n)$ holds.*
(b) *If in addition* $\limsup_{n\to\infty} n\overline{\Phi}(u_n) < \infty$ *then $D'(u_n)$ holds.*
(c) *If $\gamma(n)\ln n \to 0$ and* $\limsup_{n\to\infty} n\overline{\Phi}(u_n) < \infty$ *then both conditions $D(u_n)$ and $D'(u_n)$ are satisfied.* □

Now recall that the normal distribution Φ is in the maximum domain of attraction of the Gumbel law Λ; see Example 3.3.29. Then the following is a consequence of Lemma 4.4.7 and of Theorem 4.4.6. The constants c_n and d_n are chosen as in Example 3.3.29.

Theorem 4.4.8 (Limit distribution of the maxima of a Gaussian stationary sequence)
Let (X_n) be a stationary sequence with common standard normal df Φ. Suppose that

$$\lim_{n\to\infty} \gamma(n) \ln n = 0\,.$$

Then

$$\sqrt{2\ln n}\left(M_n - \sqrt{2\ln n} + \frac{\ln\ln n + \ln 4\pi}{2(2\ln n)^{1/2}}\right) \xrightarrow{d} \Lambda\,.$$

□

The assumption $\gamma(n)\ln n \to 0$ is called *Berman's condition* and is very weak. Thus Theorem 4.4.8 states that Gaussian stationary sequences have very much the same extremal behaviour as Gaussian iid sequences.

Example 4.4.9 (Gaussian linear processes)
An important class of stationary sequences is that of the linear processes (see Section 5.5 and Chapter 7), which have an infinite moving average representation

$$X_n = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j}\,, \quad n \in \mathbb{Z}\,, \tag{4.46}$$

where $(Z_n)_{n\in\mathbb{Z}}$ is an iid sequence and $\sum_j \psi_j^2 < \infty$. We also suppose that $EZ_1 = 0$ and $\sigma_Z^2 = \mathrm{var}(Z_1) < \infty$. If (Z_n) is Gaussian, so is (X_n). Conversely, most interesting Gaussian stationary processes have representation (4.46); see Brockwell and Davis [92], Theorem 5.7.1, in particular, the popular (causal) ARMA processes; see Example 7.1.1. In that case the coefficients ψ_j decrease to zero at an exponential rate. Hence the autocovariances of (X_n), i.e.

$$\gamma(h) = E(X_0 X_h) = \sigma_Z^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}\,, \quad h \geq 0\,,$$

decrease to zero exponentially as $h \to \infty$. Thus Theorem 4.4.8 is applicable to Gaussian ARMA processes.

Gaussian fractional ARIMA(p,d,q) processes with $p, q \geq 1, d \in (0, 0.5)$, enjoy a (causal) representation (4.46) with $\psi_j = j^{d-1} L(j)$ for a slowly varying function L; see Brockwell and Davis [92], Section 13.2. It is not difficult to see that the assumptions of Theorem 4.4.8 also hold in this case. Fractional ARIMA processes with $d \in (0, 0.5)$ are a standard example of long memory processes where the sequence $\gamma(h)$ is not supposed to be absolutely summable. This shows that the restriction $\gamma(n) \ln n \to 0$ is indeed very weak in the Gaussian case.

In Section 5.5 we will study the extreme value behaviour of linear processes with subexponential noise (Z_n) in MDA(Λ) or MDA(Φ_α). We will learn that the limit distributions of (M_n) are of the form H^θ for some $\theta \in (0,1]$ and an extreme value distribution H. This indicates the various forms of limit behaviour of maxima of linear processes, depending on their tail behaviour. □

Notes and Comments

Extreme value theory for stationary sequences has been treated in detail in Leadbetter et al. [418]. There one can also find some remarks on the history of the use of conditions $D(u_n)$ and $D'(u_n)$. A very recommendable review article is Leadbetter and Rootzén [419].

In summary, the conditions $D(u_n)$ and $D'(u_n)$ ensure that the extremes of the stationary sequence (X_n) have the same qualitative behaviour as the extremes of an associated iid sequence. The main problem is to verify conditions $D(u_n)$ and $D'(u_n)$. For Gaussian (X_n) this reduces to showing Berman's condition, namely that $\gamma(n) = \operatorname{cov}(X_0, X_n) = o(1/\ln n)$. It covers wide classes of Gaussian sequences, in particular ARMA and fractional ARIMA processes. We mention that Leadbetter et al. [418] also treated the cases $\gamma(n) \ln n \to c \in (0, \infty]$.

In Section 5.3.2 we will come back to stationary sequences satisfying the conditions $D(u_n)$ and $D'(u_n)$. There we will also study the behaviour of the upper order statistics.