# 5

# An Approach to Extremes via Point Processes

Point process techniques give insight into the structure of limit variables and limit processes which occur in the theory of summation (see Chapter 2), in extreme value theory (see Chapters 3 and 4) and in time series analysis (see Chapter 7).

One can think of a point process $N$ simply as a random distribution of points $X_i$ in space. For a given configuration $(X_i)$ and a set $A$, $N(A)$ counts the number of $X_i \in A$. It is convenient to imagine the distribution of $N$ as the probabilities

$$P\left(N(A_1) = k_1, \ldots, N(A_m) = k_m\right)$$

for all possible choices of nice sets $A_1, \ldots, A_m$ and all non–negative integers $k_1, \ldots, k_m$.

The most important point processes are those for which $N(A)$ is Poisson distributed. This leads to the notion of a Poisson random measure $N$ (see Definition 5.1.9) as a generalisation of the classical (homogeneous) Poisson process on $[0, \infty)$. Poisson random measures are basic for the understanding of links between extreme value theory and point processes; they occur in a natural way as weak limits of sample point processes $N_n$, say. This means, to over–simplify a little, that the relations

$$(N_n(A_1), \ldots, N_n(A_m)) \stackrel{d}{\rightarrow} (N(A_1), \ldots, N(A_m))$$

hold for any choice of sets $A_i$. Kallenberg's Theorem 5.2.2 gives surprisingly simple conditions for this convergence to hold.

P. Embrechts et al., *Modelling Extremal Events*, Stochastic Modelling and Applied Probability 33, DOI 10.1007/978-3-642-33483-2_6,

These are the fundamental notions which we need throughout. They are made precise in Sections 5.1 and 5.2. The interrelationship between extremes, point processes and weak convergence is perhaps best illustrated by the *point process of exceedances* of a given threshold by a sequence of rvs; see Example 5.1.3 and Section 5.3. Then the reader is urged to go through the beautiful results on exceedances, limits of upper order statistics, joint convergence of maxima and minima, records etc. in order to get a general impression about the method; see Sections 5.4 and 5.5. Point process methods yield a unified and relatively easy approach to extreme value theory. In contrast to the classical techniques as used in Chapters 3 and 4, they do allow for the treatment of extremes of sequences more general than iid in a straightforward way.

In this chapter we need some tools from functional analysis and from measure theory as well as certain arguments from weak convergence in metric spaces; see Appendix A2. In our presentation we try to reduce these technicalities to a minimum, but we cannot avoid them completely.

Our discussion below closely follows Resnick [530].

## 5.1 Basic Facts About Point Processes

### 5.1.1 Definition and Examples

In this section we are concerned with the question

*What is a point process, how can we describe its distribution, and what are simple examples?*

For the moment, consider a sequence $(X_n)$ of random vectors in the so–called *state space* $E$ and define for $A \subset E$

$$N(A) = \operatorname{card}\{i : X_i \in A\}\,,$$

i.e. $N(A)$ counts the number of $X_i$ falling into $A$. Naturally, $N(A) = N(A,\omega)$ is random for a given set $A$ and, under general conditions, $N(\cdot,\omega)$ defines a random counting measure with atoms $X_n$ on a suitable $\sigma$–algebra $\mathcal{E}$ of subsets of $E$. This is the intuitive meaning of the *point process* $N$.

For our purposes, the state space $E$, where the points live, is a subset of a finite–dimensional Euclidean space possibly including points with an infinite coordinate, and $E$ is equipped with the $\sigma$–algebra $\mathcal{E}$ of the Borel sets generated by the open sets. It is convenient to write a point process using *Dirac measure* $\varepsilon_x$ for $x \in E$:

$$\varepsilon_x(A) = \begin{cases} 1 & \text{if} \quad x \in A\,, \\ 0 & \text{if} \quad x \notin A\,, \end{cases} \qquad A \in \mathcal{E}\,.$$

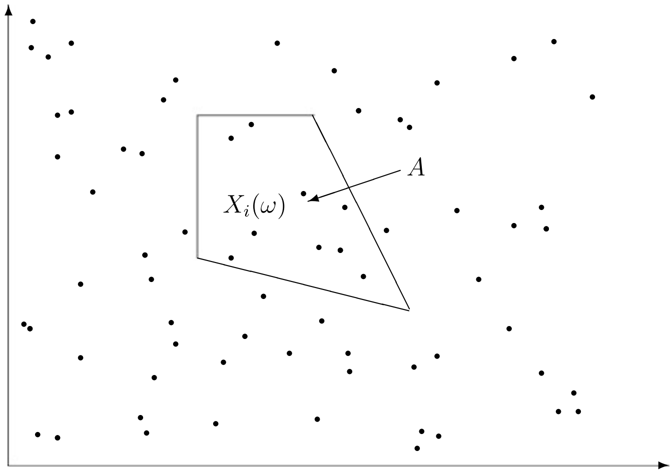


**Figure 5.1.1** *A configuration of random points $X_i$ in $\mathbb{R}_+ \times \mathbb{R}_+$. The number of points that fall into the set $A$ constitute the counting variable $N(A)$; in this case $N(A,\omega) = 9$.*

For a given sequence $(x_i)_{i\geq 1}$ in $E$,

$$m(A) = \sum_{i=1}^{\infty} \varepsilon_{x_i}(A) = \sum_{i:x_i \in A} 1 = \text{card}\,\{i : x_i \in A\}\ , \quad A \in \mathcal{E}\ ,$$

defines a *counting measure* on $\mathcal{E}$ which is called a *point measure* if $m(K) < \infty$ for all compact sets $K \in E$. Let $M_p(E)$ be the space of *all* point measures on $E$ equipped with an appropriate $\sigma$–algebra $\mathcal{M}_p(E)$.

**Definition 5.1.2** (Definition of a point process)
*A* point process *on $E$ is a measurable map*

$$N : [\Omega, \mathcal{F}, P] \to [M_p(E), \mathcal{M}_p(E)]\ . \qquad \square$$

**Remarks.** 1) The $\sigma$–algebra $\mathcal{M}_p(E)$ contains all sets of the form $\{m \in M_p(E) : m(A) \in B\}$ for $A \in \mathcal{E}$ and any Borel set $B \subset [0,\infty]$, i.e. it is the smallest $\sigma$–algebra making the maps $m \to m(A)$ measurable for all $A \in \mathcal{E}$.

2) A point process is a random element or a random function which assumes point measures as values. It is convenient to think of a point process as a collection $(N(A))_{A\in\mathcal{E}}$ of the extended rvs $N(A)$. (An extended rv can assume the value $\infty$ with positive probability.) Point processes are special *random measures*; see for instance Kallenberg [365].

3) The point processes we are interested in can often be written in the form

$$N = \sum_{i=1}^{\infty} \varepsilon_{X_i}$$

for a sequence $(X_n)$ of $d$–dimensional random vectors. Then, for each $\omega \in \Omega$,

$$N(A, \omega) = \sum_{i=1}^{\infty} \varepsilon_{X_i(\omega)}(A)\,, \quad A \in \mathcal{E}\,,$$

defines a point measure on $\mathcal{E}$.

4) Assume that $m = \sum_{i=1}^{\infty} \varepsilon_{x_i}$ is a point measure on $E$. Let $(y_i)$ be a subsequence of $(x_i)$ containing all mutually distinct values $(x_i)$ with no repeats. Define the *multiplicity of* $y_i$ as

$$n_i = \mathrm{card}\{j : j \geq 1\,,\, y_i = x_j\}\,.$$

Then we may write

$$m = \sum_{i=1}^{\infty} n_i \varepsilon_{y_i}\,.$$

If $n_i = 1$ for all $i$, then $m$ is called a *simple point measure*, and a *multiple* one, otherwise. Analogously, if the realisations of the point process $N$ are only simple point measures, then $N$ is a *simple point process*, and a *multiple* one, otherwise. Alternatively, a point process $N$ is simple if

$$P(N(\{x\}) \leq 1\,,\, x \in E) = 1\,. \qquad \square$$

**Example 5.1.3** (Point process of exceedances)
One of the point processes closely related to extreme value theory is the *point process of exceedances*: let $u$ be a real number and $(X_n)$ a sequence of rvs. Then the *point process of exceedances*

$$N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{n^{-1}i}(\cdot) I_{\{X_i > u\}}\,, \quad n = 1, 2, \ldots\,, \tag{5.1}$$

with state space $E = (0, 1]$ counts the number of exceedances of the threshold $u$ by the sequence $X_1, \ldots, X_n$. For example, take the whole interval $(0, 1]$. Then

$$\begin{aligned} N_n(0,1] &= \mathrm{card}\left\{i : 0 < n^{-1} i \leq 1 \quad \text{and} \quad X_i > u\right\} \\ &= \mathrm{card}\left\{i \leq n : X_i > u\right\}\,. \end{aligned}$$

Here and in the sequel we write for a measure $\mu$

$$\mu(a,b] = \mu((a,b])\,, \quad \mu[a,b] = \mu([a,b]) \quad \text{etc.}$$

We also see immediately the close link with extreme value theory. For example, let $X_{k,n}$ denote the $k$th largest order statistic of the sample $X_1, \ldots, X_n$. Then

$$\begin{aligned}
\{N_n(0,1] = 0\} &= \{\text{card}\,\{i \le n : X_i > u\} = 0\} \\
&= \{\text{None of the } X_i,\ i \le n, \text{ exceeds } u\} \\
&= \{\max(X_1, \ldots, X_n) \le u\} \\
\{N_n(0,1] < k\} &= \{\text{card}\,\{i \le n : X_i > u\} < k\} \\
&= \{\text{Fewer than } k \text{ among the } X_i,\ i \le n, \text{ exceed } u\} \\
&= \{\text{The order statistic } X_{k,n} \text{ does not exceed } u\} \\
&= \{X_{k,n} \le u\}\,.
\end{aligned}$$

We notice that the point process of exceedances can be written in the (perhaps more intuitively appealing) form

$$N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{n^{-1}i, X_i}(\cdot)\,, \quad n = 1, 2, \ldots\,, \tag{5.2}$$

with two–dimensional state space $E = (0,1] \cap (u, \infty)$. In our presentation we prefer version (5.1) on $E = (0,1]$, with the exception of Section 5.5.1. The advantage of this approach is that weak convergence of (5.1) can be dealt with by simpler means than for (5.2); compare for instance the difficulty of the proofs in Sections 5.3 and 5.5.1. In Section 5.3 our interest will focus on the point process of exceedances for a sequence of non–decreasing thresholds $u = u_n$ which we will choose in such a way that $(N_n)$ converges weakly, in the sense of Section 5.2, to a Poisson random measure; see Definition 5.1.9 below. □

**Example 5.1.4** (Renewal counting process)
Let $(Y_i)$ be a sequence of iid positive rvs, $T_n = Y_1 + \cdots + Y_n$, $n \ge 1$. Recall from Section 2.5.2 the *renewal counting process* generated by $(Y_i)$:

$$N(t) = \text{card}\{i : T_i \le t\}\,, \quad t \ge 0\,.$$

To this process we can relate the point process

$$N(A) = \sum_{i=1}^{\infty} \varepsilon_{T_i}(A)\,, \quad A \in \mathcal{E}\,,$$

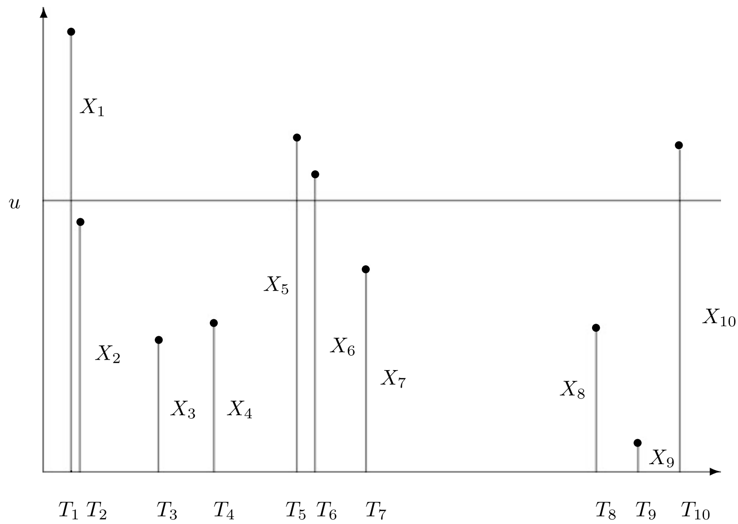


**Figure 5.1.5** *Visualisation of the point process of exceedances corresponding to the random sums from Example 5.1.6.*

with state space $E = [0, \infty)$. Notice that for $A = [0, t]$ we obtain

$$N(t) = N[0, t], \quad t \geq 0.$$

In this sense, every renewal counting process corresponds to a point process. The point process defined in this way is simple since $0 < T_1 < T_2 < \cdots$ with probability 1. Recall that a homogeneous Poisson process (see Example 2.5.2) is a particular renewal counting process with exponential rvs $Y_i$. Hence a Poisson process defines a "Poisson point process". □

**Example 5.1.6** (Random sums driven by a renewal counting process)
In Chapter 1 and Section 2.5.3 we considered random sums driven by a renewal counting process:

$$S(t) = \sum_{i=1}^{N(t)} X_i, \quad t \geq 0.$$

Here $(N(t))$ is a renewal counting process as defined in Example 5.1.4 and $(X_i)$ is an iid sequence independent of $(N(t))$. Recall from Chapter 1 that random sums are closely related to the renewal risk model in which we can interpret the rv $X_i$ as claim size arriving at time $T_i$. A point process related to $(S(t))$ is given by

$$\widetilde{N}(A) = \sum_{i=1}^{\infty} \varepsilon_{(T_i, X_i)}(A)\,, \quad A \in \mathcal{E}\,,$$

with state space $E = [0, \infty) \times \mathbb{R}$. For example, in the insurance context

$$\widetilde{N}\left((a, b] \times (u, \infty)\right) = \operatorname{card}\{i : a < T_i \le b,\ X_i > u\}$$

counts the number of claims arriving in the time interval $(a, b]$ and exceeding the threshold value $u$. Notice that $\widetilde{N}$ is very close in spirit to the point process of exceedances from Example 5.1.3. □

### 5.1.2 Distribution and Laplace Functional

The realisations of a point process $N$ are point measures. Therefore the *distribution* or the *probability law* of $N$ is defined on subsets of point measures:

$$P_N(A) = P(N \in A)\,, \quad A \in \mathcal{M}_p(E)\,.$$

This distribution is not easy to imagine. Fortunately, the distribution of $N$ is uniquely determined by the family of the distributions of the finite–dimensional random vectors

$$(N(A_1), \ldots, N(A_m)) \tag{5.3}$$

for any choice of $A_1, \ldots, A_m \in \mathcal{E}$ and $m \ge 1$; see Daley and Vere–Jones [153], Proposition 6.2.III. The collection of all these distributions is called the *finite–dimensional distributions of the point process*. We can imagine the finite–dimensional distributions much more easily than the distribution $P_N$ itself. Indeed, (5.3) is a random vector of integer–valued rvs which is completely given by the probabilities

$$P\left(N\left(A_1\right) = k_1, \ldots, N\left(A_m\right) = k_m\right)\,, \quad k_i \ge 0\,, \quad i = 1, \ldots, m\,.$$

From a course on probability theory we know that it is often convenient to describe the distribution of a rv or of a random vector by some analytical means. For example, one uses a whole class of transforms: chfs, Laplace–Stieltjes transforms, generating functions etc. A similar tool exists for point processes:

**Definition 5.1.7** (Laplace functional)
*The* Laplace functional *of the point process $N$ is given by*

$$\begin{aligned} \Psi_N(g) &= E \exp\left\{-\int_E g\, dN\right\} \qquad (5.4) \\ &= \int_{M_p(E)} \exp\left\{-\int_E g(x)\, dm(x)\right\} dP_N(m)\,. \end{aligned}$$

*It is defined for non–negative measurable functions $g$ on the state space $E$.* □

**Remarks.** 1) The Laplace functional $\Psi_N$ determines the distribution of a point process completely; see Example 5.1.8 below.

2) Laplace functionals are an important tool for discovering the properties of point processes; they are particularly useful for studying the weak convergence of point processes; see Section 5.2.

3) The integral $\int_E g dN$ in (5.4) is well defined as a Lebesgue–Stieltjes integral. Write $N = \sum_{i=1}^{\infty} \varepsilon_{X_i}$ for random vectors with values in $E$; then

$$\int_E g dN = \sum_{i=1}^{\infty} g(X_i).$$

In particular, $\int_A dN = \int_E I_A \, dN = N(A)$. □

**Example 5.1.8** (Laplace functional and Laplace transform)
To get an impression of the use of Laplace functionals we consider the special functions

$$g = z\, I_A\,, \quad z \ge 0\,, \quad A \in \mathcal{E}\,.$$

Then

$$\Psi_N(g) = E \exp\left\{ - \int_E g\, dN \right\} = E \exp\left\{ -z\, N(A) \right\}\,,$$

so that the notion of the ordinary Laplace transform of the rv $N(A)$ is embedded in the Laplace functional. Now suppose that $A_1, \ldots, A_m \in \mathcal{E}$. If we choose the functions

$$g = z_1\, I_{A_1} + \cdots + z_m\, I_{A_m}\,, \quad z_1 \ge 0\,, \ldots, z_m \ge 0\,,$$

then we obtain the joint Laplace transform of the finite–dimensional distributions, i.e. of the random vectors (5.3). From the remarks above we learnt that the finite–dimensional distributions determine the distribution of $N$. On the other hand, the finite–dimensional distributions are uniquely determined by their Laplace transforms, and hence the Laplace functional uniquely determines the distribution of $N$. □

### 5.1.3 Poisson Random Measures

Point processes are collections of counting variables. The simplest and perhaps most useful example of a counting variable is binomially distributed: $B_n = \sum_{i=1}^{n} I_{\{X_i \in A_n\}}$ for iid $X_i$ counts the number of "successes" $\{X_i \in A_n\}$ among $X_1, \ldots, X_n$, and $p_n = P(X_1 \in A_n)$ is the "success probability". Then Poisson's theorem tells us that $B_n \stackrel{d}{\to} Poi(\lambda)$ provided $p_n \sim \lambda/n$. This simple limit result also suggests the following definition of a *Poisson random measure* which occurs as natural limit of many point processes; see for instance Section 5.3.

Let $\mu$ be a *Radon measure* on $\mathcal{E}$, i.e. $\mu(A) < \infty$ for compact sets $A \subset E$.

**Definition 5.1.9** (Poisson random measure (PRM))
*A point process $N$ is called a* Poisson process *or a* Poisson random measure with mean measure $\mu$ (*we write* PRM($\mu$)) *if the following two conditions are satisfied:*

(*a*) *For $A \in \mathcal{E}$,*

$$P(N(A) = k) = \begin{cases} e^{-\mu(A)} \dfrac{(\mu(A))^k}{k!} & \text{if} \quad \mu(A) < \infty\,, \\ 0 & \text{if} \quad \mu(A) = \infty\,, \end{cases} \qquad k \geq 0\,.$$

(*b*) *For any $m \geq 1$, if $A_1, \ldots, A_m$ are mutually disjoint sets in $\mathcal{E}$ then $N(A_1), \ldots, N(A_m)$ are independent rvs.* □

**Remark.** The name *mean measure* is justified by the fact that $EN(A) = \mu(A)$. Since a Poisson distribution is determined by its mean value, it follows from the above definition that PRM($\mu$) is determined by its mean measure $\mu$. □

**Example 5.1.10** (Homogeneous PRM)
Recall the notion of a homogeneous Poisson process $(N(t))_{t\geq 0}$ with intensity $\lambda > 0$ from Example 2.5.2. It is a process with stationary, independent increments such that $N(t)$ is $Poi(\lambda t)$ distributed. Hence

$$P(N(t) = k) = e^{-\lambda t}\, \frac{(\lambda t)^k}{k!}\,, \quad k = 0, 1, \ldots\,.$$

Since $(N(t))_{t\geq 0}$ is a non–decreasing process the construction

$$N(s,t] = N(t) - N(s)\,, \quad 0 \leq s < t < \infty\,,$$

and the extension theorem for measures define a point process $N$ on the Borel sets of $E = [0,\infty)$. The stationary, independent increments of $(N(t))_{t\geq 0}$ imply that

$$\begin{aligned} & P\left(N\left(A_1\right) = k_1, \ldots, N\left(A_m\right) = k_m\right) \\ & = \quad e^{-\lambda|A_1|}\, \frac{(\lambda|A_1|)^{k_1}}{k_1!} \cdots e^{-\lambda|A_m|}\, \frac{(\lambda|A_m|)^{k_m}}{k_m!} \end{aligned}$$

for any mutually disjoint $A_i$ and integers $k_i \geq 0$. Here $|\cdot|$ denotes Lebesgue measure on $[0,\infty)$. This relation is immediate for disjoint intervals $A_i$, and in the general case one has to approximate the disjoint Borel sets $A_i$ by intervals.

Alternatively, we saw from Example 5.1.4 that a homogeneous Poisson process with intensity $\lambda$ can be defined as a simple point process $N = \sum_{i=1}^{\infty} \varepsilon_{T_i}$, where $T_i = Y_1 + \cdots + Y_i$ for iid exponential rvs $Y_i$ with mean value $1/\lambda$.

Notice that $N$ has mean measure

$$\mu(A) = \lambda|A| = \lambda \int_A dx\,, \quad A \in \mathcal{E}\,. \qquad \square$$

Now suppose that $N$ is PRM$(\lambda|\cdot|)$ with state space $E \subset \overline{\mathbb{R}}^d$ ($\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$), where $\lambda > 0$ and $|\cdot|$ denotes Lebesgue measure on $E$. As a generalisation of the homogeneous Poisson process on $[0,\infty)$ we call $N$ a *homogeneous PRM* or *homogeneous Poisson process with intensity* $\lambda$. Moreover, if the mean measure $\mu$ of a PRM is absolutely continuous with respect to Lebesgue measure, i.e. there exists a non–negative function $f$ such that

$$\mu(A) = \int_A f(x)\,dx\,, \quad A \in \mathcal{E}\,,$$

then $f$ is the *intensity* or the *rate of the PRM.*

Alternatively, we can define a PRM$(\mu)$ by its Laplace functional:

**Example 5.1.11** (Laplace functional of PRM$(\mu)$)

$$\Psi_N(g) = \exp\left\{-\int_E \left(1 - e^{-g(x)}\right) d\mu(x)\right\} \tag{5.5}$$

for any measurable $g \geq 0$. Formula (5.5) is a consequence of the more general Lemma 5.1.12 below. $\square$

**Lemma 5.1.12** *Let $N$ be* PRM$(\mu)$ *on $E \subset \overline{\mathbb{R}}^d$. Assume that the Lebesgue integral $\int_E (\exp\{f(x)\} - 1) d\mu(x)$ exists and is finite. Then $\int_E |f| dN < \infty$ a.s. and*

$$I_N(f) = E\exp\left\{\int_E f dN\right\} = \exp\left\{-\int_E \left(1 - e^{f(x)}\right) d\mu(x)\right\}.$$

**Proof.** For $A \in \mathcal{E}$ with $\mu(A) < \infty$ write $f = I_A$. Then

$$\begin{aligned} I_N(f) &= E\exp\left\{\int_E f\,dN\right\} = E\exp\{N(A)\} \\ &= \sum_{k=0}^{\infty} e^k \frac{(\mu(A))^k}{k!}\, e^{-\mu(A)} = e^{-\mu(A)(1-e)} \\ &= \exp\left\{-\int_E \left(1 - e^{f(x)}\right) d\mu(x)\right\}. \end{aligned}$$

For

$$f = \sum_{i=1}^{m} z_i \, I_{A_i} \,, \quad z_i \geq 0 \,, \quad i = 1, \ldots, m \,, \tag{5.6}$$

and disjoint $A_1, \ldots, A_m$ we can use the independence of $N(A_1), \ldots, N(A_m)$:

$$\begin{aligned} I_N(f) &= E \exp\left\{ \sum_{i=1}^{m} z_i N\left(A_i\right) \right\} \\ &= \prod_{i=1}^{m} \exp\left\{ - \int_E \left(1 - e^{z_i \, I_{A_i}}\right) d\mu(x) \right\} \\ &= \exp\left\{ - \int_E \left(1 - e^{f(x)}\right) d\mu(x) \right\} . \end{aligned}$$

General non–negative $f$ are the monotone limit of step functions $(f_n)$ as in (5.6). Thus, applying the monotone convergence theorem, we obtain

$$\begin{aligned} I_N(f) &= \lim_{n \to \infty} E \exp\left\{ \int_E f_n \, dN \right\} \\ &= \lim_{n \to \infty} \exp\left\{ - \int_E \left(1 - e^{f_n(x)}\right) \, d\mu(x) \right\} \\ &= \exp\left\{ - \int_E \left(1 - e^{f(x)}\right) \, d\mu(x) \right\} . \end{aligned}$$

Since the right–hand side is supposed to be finite, $E \exp\{\int_E f dN\} < \infty$, hence $\int_E f dN < \infty$ a.s.

For negative $f$ one can proceed similarly. For general $f$, write $f = f^+ - f^-$. Notice that $\int_E f^+ dN$ and $\int_E f^- dN$ are independent since $E_+ = \{x \in E : f(x) > 0\}$ and $E_- = \{x \in E : f(x) < 0\}$ are disjoint. Hence

$$\begin{aligned} I_N(f) &= I_N(f^+) I_N(-f^-) \\ &= \exp\left\{ - \int_{E_+} \left(1 - e^{f^+}\right) d\mu \right\} \exp\left\{ - \int_{E_-} \left(1 - e^{-f^-}\right) d\mu \right\} \\ &= \exp\left\{ - \int_E \left(1 - e^{f}\right) d\mu \right\} . \end{aligned}$$

This proves the lemma. □

PRM have an appealing property: they remain PRM under transformations of their points.

**Proposition 5.1.13** (Transformed PRM are PRM)
*Suppose $N$ is* PRM($\mu$) *with state space $E \subset \overline{\mathbb{R}}^d$. Assume that the points of $N$ are transformed by a measurable map $\widetilde{T} : E \to E'$, where $E' \subset \overline{\mathbb{R}}^m$ for some $m \geq 1$. Then the resulting transformed point process is* PRM($\mu(\widetilde{T}^{-1})$) *on $E'$, i.e. this PRM has mean measure $\mu(\widetilde{T}^{-1}(\cdot)) = \mu\{x \in E : \widetilde{T}(x) \in \cdot\}$.*

**Proof.** Assume that $N$ has representation $N = \sum_{i=1}^\infty \varepsilon_{X_i}$. Then the transformed point process can be written as

$$\widetilde{N} = \sum_{i=1}^\infty \varepsilon_{\widetilde{T}(X_i)}\,.$$

We calculate the Laplace functional of $\widetilde{N}$:

$$\begin{aligned}
\Psi_{\widetilde{N}}(g) &= E\exp\left\{-\int_{E'} g d\widetilde{N}\right\} \\
&= E\exp\left\{-\sum_{i=1}^\infty g(\widetilde{T}(X_i))\right\} \\
&= E\exp\left\{-\int_E g(\widetilde{T}) dN\right\} \\
&= \exp\left\{-\int_E \left(1 - e^{-g(\widetilde{T}(x))}\right) d\mu(x)\right\} \\
&= \exp\left\{-\int_{E'} \left(1 - e^{-g(y)}\right) d\mu(\widetilde{T}^{-1}(y))\right\}\,.
\end{aligned}$$

This is the Laplace functional of PRM($\mu(\widetilde{T}^{-1})$) on $E'$; see Example 5.1.11. □

**Example 5.1.14** Let $\Gamma_k$ be the points of a homogeneous Poisson process on $[0,\infty)$ with intensity $\lambda$ and $\widetilde{T}(x) = \exp\{x\}$. Then $\widetilde{N} = \sum_{i=1}^\infty \varepsilon_{\exp\{\Gamma_i\}}$ is PRM on $[1,\infty)$ with mean measure given by

$$\widetilde{\mu}(a,b] = \int_a^b d\mu(\widetilde{T}^{-1}(y)) = \lambda \int_{\ln a}^{\ln b} dx = \lambda \ln(b/a)\,, \quad 1 \leq a < b < \infty\,. \quad (5.7)$$

It is interesting to observe that the mean measure of the PRM $\widetilde{N}$ depends only on the fraction $b/a$, so that the mean measure is the same for all intervals $(ca, cb]$ for any $c > 0$.

Now assume that the PRM $\widetilde{N}$ is defined on the state space $\mathbb{R}_+$ where its mean measure is given by (5.7) for all $0 < a < b < \infty$. Since the distribution of a PRM is determined via its mean measure it follows that the PRM $\widetilde{N}(\cdot)$

and $\widetilde{N}(c\cdot)$ on $\mathbb{R}_+$ have the same distribution in $M_p(\mathbb{R}_+)$ for every positive constant $c$. □

**Example 5.1.15** (Compound Poisson process)
Let $(\Gamma_i)$ be the points of a homogeneous Poisson process $N$ on $[0,\infty)$ with intensity $\lambda > 0$ and $(\xi_i)$ be a sequence of iid non–negative integer–valued rvs, independent of $N$. Consider the multiple point process

$$\widetilde{N} = \sum_{i=1}^{\infty} \xi_i \varepsilon_{\Gamma_i}$$

and notice that

$$\widetilde{N}(0,t] = \sum_{i=1}^{\infty} \xi_i \varepsilon_{\Gamma_i}(0,t] = \sum_{i=1}^{N(t)} \xi_i\ , \quad t \geq 0\,.$$

This is nothing but a particular (i.e. integer–valued) compound Poisson process as used for instance in Chapter 1 for the Cramér–Lundberg model. Therefore we call the point process $\widetilde{N}$ a *compound Poisson process with intensity* $\lambda$ *and cluster sizes* $\xi_i$. The probabilities $\pi_k = P(\xi_1 = k)$, $k \geq 0$, are the *cluster probabilities* of $\widetilde{N}$.

The point process notion *compound Poisson process* as introduced above is perhaps not the most natural generalisation of the corresponding random sum process. One would like a *random measure* with property $\widetilde{N}(0,t] = \sum_{i=1}^{N(t)} \xi_i$ for iid $\xi_i$ with any distribution. Since $\widetilde{N}(0,t]$ could then assume any real value this calls for the introduction of a *signed random measure*. For details we refer to Kallenberg [365].

Compound Poisson processes frequently occur as limits of the point processes of exceedances of a strictly stationary sequence; see for instance Sections 5.5 and 8.4. □

## Notes and Comments

Point processes are special random measures; see Kallenberg [365]. Standard monographs on point processes and random measures are Cox and Isham [134], Daley and Vere–Jones [153], Kallenberg [365], Karr [372], Matthes, Kerstan and Mecke [447], Reiss [527]. Point processes are also treated in books on stochastic processes; see for instance Jacod and Shiryaev [352], Resnick [529, 531].

In our presentation we leave out certain details. This does not always leave the sufficient mathematical rigour. We are quite cavalier concerning measurability (for instance for point processes) and existence results (for instance for PRM), and we are not precise about compact sets in $E \subset \overline{\mathbb{R}}^d$. The disappointed reader is invited to read through Chapters 3 and 4 in Resnick [530] or to consult the first few chapters in Daley and Vere–Jones [153].

## 5.2 Weak Convergence of Point Processes

Weak convergence of point processes is a basic tool for dealing with the asymptotic theory of extreme values, linear time series and related fields. We give here a short introduction to the topic. First of all we have to clarify:

*What does weak convergence of point processes actually mean?*

This question cannot be answered at a completely elementary level. Consider point processes $N, N_1, N_2, \ldots$ on the same state space $E \subset \overline{\mathbb{R}}^d$. Then we know from Section 5.1.2 that the distribution of these point processes in $M_p(E)$, the space of all point measures on $E$, is determined by their finite–dimensional distributions. Thus a natural requirement for weak convergence of $(N_n)$ towards $N$ would be that, for any choice of "good" Borel sets $A_1, \ldots, A_m \in \mathcal{E}$ and for any integer $m \geq 1$,

$$P\left(N_n(A_1), \ldots, N_n(A_m)\right) \to P\left(N(A_1), \ldots, N(A_m)\right) . \tag{5.8}$$

On the other hand, every point process $N$ can be considered as a stochastic process, i.e. as a collection of rvs $N(A)$ indexed by the sets $A \in \mathcal{E}$. Thus $N$ is an infinite–dimensional object which must be treated in an appropriate way. A glance at Appendix A2 should convince us that we need something more than convergence of the finite–dimensional distributions, namely a condition which is called "tightness" meaning that the probability mass of the point processes $N_n$ should not disappear from "good" (compact) sets in $M_p(E)$. This may sound fine, but such a condition is not easily verified. For example, we would have to make clear in what sense we understand compactness. This has been done in Appendix A2.6 by introducing an appropriate (so–called *vague*) metric in $M_p(E)$.

Perhaps unexpectedly, point processes are user–friendly in the sense that tightness follows from the convergence of their finite–dimensional distributions; see for instance Daley and Vere–Jones [153], Theorem 9.1.VI. Hence we obtain quite an intuitive notion of weak convergence:

**Definition 5.2.1** (Weak convergence of point processes)
*Let $N, N_1, N_2, \ldots$ be point processes on the state space $E \subset \overline{\mathbb{R}}^d$ equipped with the $\sigma$–algebra $\mathcal{E}$ of the Borel sets. We say that $(N_n)$* converges weakly to $N$ in $M_p(E)$ *(we write $N_n \xrightarrow{d} N$) if* (5.8) *is satisfied for all possible choices of*

*sets* $A_i \in \mathcal{E}$ *satisfying* $P(N(\partial A_i) = 0) = 1$, $i = 1, \ldots, m$, $m \geq 1$. *(*$\partial A$ *denotes the boundary of* $A$.*)* □

Assume for the moment that the state space $E$ is an interval $(a, b] \subset \mathbb{R}$. Convergence of the finite–dimensional distributions can be checked by surprisingly simple means as the following result shows. Recall the notion of a *simple point process* from Remark 4 after Definition 5.1.2, i.e. it is a process whose points have multiplicity 0 or 1 with probability one.

**Theorem 5.2.2** (Kallenberg's theorem for weak convergence to a simple point process on an interval)
*Let* $(N_n)$ *and* $N$ *be point processes on* $E = (a, b] \subset \mathbb{R}$ *and let* $N$ *be simple. Suppose the following two conditions hold:*

$$EN_n(A) \to EN(A) \tag{5.9}$$

*for all intervals* $A = (c, d]$ *with* $a < c < d \leq b$ *and*

$$P(N_n(B) = 0) \to P(N(B) = 0) \tag{5.10}$$

*for all unions* $B = \cup_{i=1}^{k}(c_i, d_i]$ *of mutually disjoint intervals* $(c_i, d_i]$ *such that*

$$a < c_1 < d_1 < \cdots < c_k < d_k \leq b$$

*and for every* $k \geq 1$. *Then* $N_n \xrightarrow{d} N$ *in* $M_p(E)$. □

**Remarks.** 1) A result in the same spirit can also be formulated for point processes on intervals in $\mathbb{R}^d$.

2) In Section 5.3 we apply Kallenberg's theorem to point processes of exceedances (see also Example 5.1.3) which are closely related to extreme value theory. □

The Laplace functional (see Definition 5.1.7) is a useful theoretical tool for verifying the weak convergence of point processes. In much the same way as the weak convergence of a sequence of rvs is equivalent to the pointwise convergence of their chfs or Laplace–Stieltjes transforms, so the weak convergence of a sequence of point processes is equivalent to the convergence of their Laplace functionals over a suitable family of functions $g$. Specifically, recall that the real–valued function $g$ has compact support if there exists a compact set $K \subset E$ such that $g(x) = 0$ on $K^c$, the complement of $K$. Then we define

$$\begin{aligned} C_K^+(E) \quad = \quad & \{g : g \text{ is a continuous, non–negative function on } E \\ & \text{with compact support}\}\,. \end{aligned}$$

Now we can formulate the following fundamental theorem; see Daley and Vere–Jones [153], Proposition 9.1.VII:

**Theorem 5.2.3** (Criterion for weak convergence of point processes via convergence of Laplace functionals)
*The point processes* $(N_n)$ *converge weakly to the point process* $N$ *in* $M_p(E)$ *if and only if the corresponding Laplace functionals converge for every* $g \in C_K^+(E)$ *as* $n \to \infty$, *i.e.*

$$\Psi_{N_n}(g) = E \exp\left\{ - \int_E g \, dN_n \right\} \to \Psi_N(g) = E \exp\left\{ - \int_E g \, dN \right\} . \quad (5.11)$$

$\square$

**Remark.** 3) We mention that (5.11) for every $g \in C_K^+(E)$ is equivalent to $\int_E g dN_n \xrightarrow{d} \int_E g dN$ for every $g \in C_K^+(E)$. Indeed, if $g \in C_K^+(E)$ then $zg \in C_K^+(E)$, $z > 0$. Thus (5.11) implies the convergence of the Laplace transforms of the rvs $\int_E g dN_n$ and vice versa. But convergence of the Laplace transforms of non–negative rvs is equivalent to their convergence in distribution. $\square$

We consider another class of point processes which is important for applications. Assume

$$N_n = \sum_{i=1}^{\infty} \varepsilon_{(n^{-1}i, \xi_{n,i})} , \quad n = 1, 2, \ldots , \quad (5.12)$$

where the random vectors $\xi_{n,i}$ are iid for every $n$. It is convenient to interpret $n^{-1}i$ as a scaled (deterministic) time coordinate and $\xi_{n,i}$ as a scaled (random) space coordinate.

**Theorem 5.2.4** (Weak convergence to a PRM)
*Suppose* $(N_n)$ *is a sequence of point processes* (5.12) *with state space* $\mathbb{R}_+ \times E$ *and* $N$ *is* $\mathrm{PRM}(|\cdot| \times \mu)$, *where* $|\cdot|$ *is Lebesgue measure on* $\mathbb{R}_+$. *Then*

$$N_n \xrightarrow{d} N , \quad n \to \infty ,$$

*in* $M_p(\mathbb{R}_+ \times E)$ *if and only if the relation*

$$nP\left(\xi_{n,1} \in \cdot\right) \xrightarrow{v} \mu(\cdot) , \quad n \to \infty , \quad (5.13)$$

*holds on* $\mathcal{E}$.

**Remark.** 4) In (5.13), the relation $\mu_n \xrightarrow{v} \mu$ denotes *vague convergence* of the measures $\mu_n$ to the measure $\mu$ on $E$. For our purposes, $E$ is a subset of $\overline{\mathbb{R}}^d$. Typically, $E = (0, \infty]$ or $E = [-\infty, \infty] \backslash \{0\}$ or $E = (-\infty, \infty]$. In this case, $\mu_n \xrightarrow{v} \mu$ amounts to showing that $\mu_n(a, b] \to \mu(a, b]$ for all $a < b$ ($b = \infty$ is possible). In the case $E = (-\infty, \infty] \backslash \{0\}$ the origin must not be included in $(a, b]$. A brief introduction to vague convergence and weak convergence is given in Appendix A2.6. For a general treatment we refer to Daley and Vere–Jones [153], Chapter 9, or Resnick [530], Chapter 3. $\square$

**Sketch of the proof.** We restrict ourselves to the sufficiency part and give only the basic idea. For a full proof we refer to Resnick [530], Proposition 3.21. Let $g \in C_K^+(\mathbb{R}_+ \times E)$ and consider the Laplace functional

$$\begin{aligned}\Psi_{N_n}(g) &= E\exp\left\{-\int_E g\,dN_n\right\}\\ &= E\exp\left\{-\sum_{i=1}^{\infty} g\left(n^{-1}i,\,\xi_{n,i}\right)\right\}\\ &= \prod_{i=1}^{\infty}\left(1-\int_E\left(1-e^{-g\left(n^{-1}i,\mathbf{x}\right)}\right)dP\left(\xi_{n,1}\le\mathbf{x}\right)\right).\end{aligned}$$

Passing to logarithms, making use of a Taylor expansion for $\ln(1-x)$ and utilising the vague convergence in (5.13) one can show that

$$\begin{aligned}-\ln\Psi_{N_n}(g) &= -\sum_i \ln\left(1-\int_E\left(1-e^{-g\left(n^{-1}i,\mathbf{x}\right)}\right)dP\left(\xi_{n,1}\le\mathbf{x}\right)\right)\\ &= n^{-1}\sum_i\int_E\left(1-e^{-g\left(n^{-1}i,\mathbf{x}\right)}\right)d\left[nP\left(\xi_{n,1}\le\mathbf{x}\right)\right]+o(1)\\ &\to \int_{\mathbb{R}_+}\int_E\left(1-e^{-g(s,\mathbf{x})}\right)ds\;d\mu(\mathbf{x})\,,\quad n\to\infty\,.\end{aligned}$$

A glance at formula (5.5) convinces us that the last line in the above display is just $-\ln\Psi_N(g)$ where $N$ is PRM$(|\cdot|\times\mu)$. Now an application of Theorem 5.2.3 yields the result. □

## Notes and Comments

Weak convergence of point processes and random measures has been treated in all standard texts on the topic. We again refer here to Daley and Vere–Jones [153], Kallenberg [365], Matthes, Kerstan and Mecke [447], and Resnick [529, 530]. Resnick [530] gives an account of point process techniques particularly suited to extreme value theory. Leadbetter, Lindgren and Rootzén [418] use point process techniques for extremes of stationary sequences, and they provide the necessary background from point process theory in their Appendix.

As mentioned above, a rigorous treatment of weak convergence of point processes requires us to consider them as random elements in an appropriate metric space. A brief introduction to this topic is given in Appendix A2; the general theory can be found for instance in Billingsley [69] or Pollard

[504]. One way to metrize weak convergence of point processes is via vague convergence of measures; see Appendix A2.6. A rigorous treatment is given in Daley and Vere–Jones [153], Chapter 9, or Resnick [530], Chapter 3.

Weak convergence of point processes and vague convergence are closely related to regular variation in $\mathbb{R}_+^d$, see for instance de Haan and Resnick [299, 300] and Stam [605], also Bingham, Goldie and Teugels [72].

Theorems 5.2.2 and 5.2.4 are the basic tools in Sections 5.3–5.5. Theorem 5.2.2 is slightly more general in the sense that no vague convergence (or regular variation) assumption on the tails of the underlying dfs is required. Theorem 5.2.2 has been utilised in the monograph by Leadbetter et al. [418] on extremes of stationary sequences; see also Section 5.3.2. Theorem 5.2.4 will prove very effective in the case that the underlying sequence of random points has a special structure which can in some way be relaxed to an iid sequence, as is the case of linear processes (see Section 5.5) which are special stationary processes.

Resnick [529], pp. 134–135, gives a short resumé of advantages and disadvantages of point process techniques which we cite here in part:

**Some Advantages:**

($a$) The methods are by and large dimensionless. Proofs work just as well in $\mathbb{R}^d$ as in $\mathbb{R}$.
($b$) Computations are kept to a minimum and are often replaced by continuity or structural arguments. This makes proofs simpler and more instructive.
($c$) The methods lend themselves naturally to proving weak convergence in a function–space setting. Functional limit theorems are more powerful and informative than the one–dimensional variety. Furthermore, they are often (despite common prejudices) simpler.

**Some Disadvantages:**

($a$) The methods are not so effective for showing that regular variation is a necessary condition.
($b$) The methods sail smoothly only when all random variables are non–negative.
($c$) The methods rely heavily on continuity. Sometimes this can be seen as an advantage, as discussed above. But heavy reliance on continuity is also a limitation in that many questions which deal with quality of convergence (local limit theorems, rates of convergence, large deviations) are beyond the capabilities of continuity arguments.

($d$) The point process technique cannot handle problems involving normality or Brownian motion.
($e$) Those who prefer analytical methods may not find the approaches described here (in [529]) attractive.
($f$) Effective use of weak convergence techniques depends on a detailed knowledge of the properties of the limit processes. Thus it is necessary to know something about stochastic processes.

Since the second list appears longer than the first, I am compelled to make some remarks about the disadvantages. Most serious in my view are (a) and (d). As for (a), there are notable exceptions to the remark that the methods are not suited to proving necessity. Regarding (d), it is sad that the point process technique fails miserably in the presence of normality, but other weak convergence methods often succeed admirably in this case. Disadvantage (d) is a nuisance, but one can usually avoid the obstacles created by two signs by using pruning techniques or random indices. As for (f) a method cannot handle problems for which it is inherently unsuited. The problem raised in (e) is simply one of taste. As for disadvantage (f), these weak convergence techniques frequently suggest interesting problems in stochastic processes. So if (f) were rephrased suitably, it could be moved to the plus column in the ledger.

## 5.3 Point Processes of Exceedances

In Example 5.1.3 we introduced the point process of exceedances of a threshold $u_n$ by the rvs $X_1, \ldots, X_n$:

$$N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{n^{-1}i}(\cdot) I_{\{X_i > u_n\}}\,, \quad n = 1, 2, \ldots. \tag{5.14}$$

We also indicated the close link with extreme value theory: let $X_{n,n} \le \cdots \le X_{1,n}$ denote the order statistics of the sample $X_1, \ldots, X_n$ and $M_n = X_{1,n}$. Then

$$\begin{aligned} \{N_n(0,1] = 0\} &= \{M_n \le u_n\}\,, \\ \{N_n(0,1] < k\} &= \{X_{k,n} \le u_n\}\,. \end{aligned} \tag{5.15}$$

In this section we show the weak convergence of a sequence $(N_n)$ of such point processes to a homogeneous Poisson process $N$ on the state space $E = (0,1]$. The sequence $(X_n)$ is supposed to be iid or strictly stationary satisfying the assumptions $D$ and $D'$ from Section 4.4. As a byproduct and for illustrative purposes we give alternative proofs of the limit results for maxima and upper order statistics provided in Chapters 3 and 4.

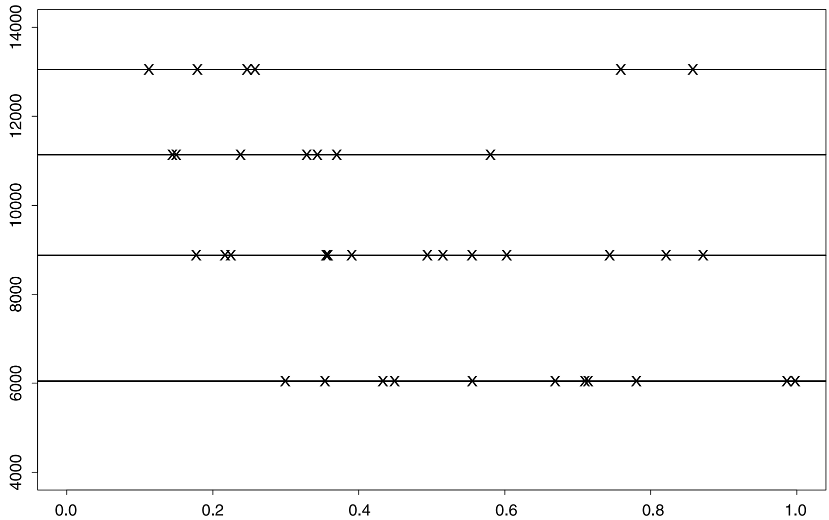


**Figure 5.3.1** *Visualisation of the point processes of exceedances of insurance claim data caused by water, $n = 1\,762$ observations. For the threshold $u_1 = 6\,046$ we chose $n/4$ data points, correspondingly $u_2 = 8\,880$ and $n/2$, $u_3 = 11\,131$ and $3n/4$, $u_4 = 13\,051$ and $n$.*

### 5.3.1 The IID Case

Assume that the $X_n$ are iid rvs and let $(u_n)$ be a sequence of real thresholds. Recall from Proposition 3.1.1 that, for any $\tau \in [0, \infty]$, the relation $P(M_n \leq u_n) \to \exp\{-\tau\}$ holds if and only if

$$n\overline{F}(u_n) = E \sum_{i=1}^{n} I_{\{X_i > u_n\}} \to \tau \,. \tag{5.16}$$

The latter condition ensures that there are on average roughly $\tau$ exceedances of the threshold $u_n$ by $X_1, \ldots, X_n$. The Poisson approximation for extremes is visualised in Figure 4.2.2; see also Figure 5.3.1. Condition (5.16) also implies weak convergence of the point processes $N_n$:

**Theorem 5.3.2** (Weak convergence of point processes of exceedances, iid case)

*Suppose that $(X_n)$ is a sequence of iid rvs with common df $F$. Let $(u_n)$ be threshold values such that* (5.16) *holds for some $\tau \in (0, \infty)$. Then the point processes of exceedances $N_n$, see* (5.14), *converge weakly in $M_p(E)$ to* *a homogeneous Poisson process $N$ on $E = (0,1]$ with intensity $\tau$, i.e. $N$ is* $\mathrm{PRM}(\tau|\cdot|)$*, where $|\cdot|$ denotes Lebesgue measure on $E$.*

**Proof.** We may and do assume that the limit process $N$ is embedded in a homogeneous Poisson process on $[0,\infty)$. In that case we argued that $N$ must be simple; see Example 5.1.10. Hence we can apply Kallenberg's Theorem 5.2.2. Notice that for $A = (a,b] \subset (0,1]$ the rv

$$\begin{aligned} N_n(A) &= \sum_{i=1}^{n} \varepsilon_{n^{-1}i}(A)\, I_{\{X_i > u_n\}} \\ &= \sum_{a < n^{-1} i \le b} I_{\{X_i > u_n\}} \\ &= \sum_{i=[na]+1}^{[nb]} I_{\{X_i > u_n\}} \end{aligned}$$

is binomial with parameters $([nb]-[na], \overline{F}(u_n))$. Here $[x]$ denotes the integer part of $x$. Thus, by assumption (5.16),

$$EN_n(A) = ([nb]-[na])\overline{F}\,(u_n) \sim (n(b-a))\,\left(n^{-1}\tau\right) = \tau(b-a) = EN(A)\,,$$

which proves (5.9).

Thus it remains to show (5.10). Since $N_n(A)$ is binomial and in view of (5.16) we have

$$\begin{aligned} P(N_n(A)=0) &= F^{[nb]-[na]}\,(u_n) \\ &= \exp\left\{([nb]-[na])\ln\left(1-\overline{F}\,(u_n)\right)\right\} \\ &\to \exp\left\{-\tau(b-a)\right\}\,. \end{aligned} \tag{5.17}$$

Recalling the definition of the set $B$ from (5.10) and taking the independence of the $X_i$ into account we conclude from (5.17) that

$$\begin{aligned} P(N_n(B)=0) &= P\left(N_n(c_i,d_i]=0\,, \quad i=1,\ldots,k\right) \\ &= P\left(\max_{[nc_i]<j\le[nd_i]} X_j \le u_n\,, \quad i=1,\ldots,k\right) \\ &= \prod_{i=1}^{k} P\left(\max_{[nc_i]<j\le[nd_i]} X_j \le u_n\right) \\ &= \prod_{i=1}^{k} P\left(N_n(c_i,d_i]=0\right) \end{aligned}$$

$$\to \prod_{i=1}^{k} \exp\{-\tau(d_i - c_i)\} .$$

On the other hand, by the Poisson property of $N$,

$$P(N(B) = 0) = \exp\{-\tau|B|\} = \exp\left\{-\tau \sum_{i=1}^{k} (d_i - c_i)\right\} .$$

This proves the theorem by virtue of Kallenberg's Theorem 5.2.2. $\square$

The following example shows the close link between extreme value theory and the point processes of exceedances.

**Example 5.3.3** (Continuation of Example 5.1.3)
An application of Theorem 5.3.2 together with (5.15) yields

$$P\left(X_{k,n} \le u_n\right) = P\left(N_n(0,1] < k\right) \to P(N(0,1] < k) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!} .$$

This was the content of Theorem 4.2.3. Similar arguments as for Corollary 4.2.4 also allow us to derive the limit distribution of the $k$th order statistic for dfs $F$ in the maximum domain of attraction of an extreme value distribution. $\square$

In Example 5.1.6 we considered iid sum processes indexed by a renewal counting process and a corresponding point process: let $(X_i)$ and $(Y_i)$ be two independent sequences of iid rvs, suppose $Y_1$ is positive with probability 1 and set $T_i = Y_1 + \cdots + Y_i$. Then $N'(t) = \operatorname{card}\{i : T_i \le t\}$ defines a renewal counting process and $S(t) = \sum_{i=1}^{N'(t)} X_i$ for $t > 0$ is the sum process. Here we consider the corresponding point process of exceedances

$$\widetilde{N}_n(\cdot) = \sum_{i=1}^{N'(n)} \varepsilon_{n^{-1}T_i}(\cdot) I_{\{X_i > u_n\}} \tag{5.18}$$

on the state space $E = (0,1]$. As before, $(u_n)$ is a real–valued threshold sequence. The strong law of large numbers implies $n^{-1}T_{[nx]} \stackrel{\text{a.s.}}{\to} xEY_1 = x\lambda^{-1}$ for $x \in (0,1]$, and so we may hope that a result similar to Theorem 5.3.2 holds in this situation. That is indeed the case:

**Theorem 5.3.4** (Weak convergence of point processes of exceedances, iid case and random index)
*Let $(\widetilde{N}_n)$ be the point processes of exceedances* (5.18) *of the threshold sequence $(u_n)$. Assume that $(u_n)$ satisfies* (5.16) *for some $\tau \in (0,\infty)$. Moreover, let $T_n = Y_1 + \cdots + Y_n$ be the points of a renewal counting process on $[0,\infty)$ with $EY_1 = \lambda^{-1} \in \mathbb{R}_+$. Then the relation $\widetilde{N}_n \stackrel{d}{\to} N$ holds in $M_p(E)$, where $N$ is a homogeneous Poisson process on $E = (0,1]$ with intensity $\tau\lambda$.*

**Proof.** For an application of Kallenberg's Theorem 5.2.2 it remains to show the following two relations:

$$E\widetilde{N}_n(a,b] \;\to\; EN(a,b] = \tau\lambda(b-a)\,, \quad 0 < a < b \le 1\,, \tag{5.19}$$

$$P(\widetilde{N}_n(B)=0) \;\to\; P(N(B)=0) \tag{5.20}$$

for all sets $B = \cup_{i=1}^m (c_i, d_i]$ with $0 < c_1 < d_1 < \cdots < c_k < d_k \le 1$, $k \ge 1$. As above, we write $(N'(t))$ for the renewal counting process generated by $(T_i)$. Then, conditioning on $N'$ and applying (5.16),

$$\begin{aligned}
E\widetilde{N}_n(a,b] &= E \sum_{i:a<n^{-1}T_i\le b} I_{\{X_i>u_n\}} \\
&= E \sum_{i=N'(na)+1}^{N'(nb)} I_{\{X_i>u_n\}} \\
&= E\left[E\left(\sum_{i=N'(na)+1}^{N'(nb)} I_{\{X_i>u_n\}} \,\middle|\, N'\right)\right] \\
&= E\left(\sum_{i=N'(na)+1}^{N'(nb)} E(I_{\{X_i>u_n\}})\right) \\
&= E\left(N'(nb)-N'(na)\right)\, P(X_1>u_n) \\
&= \left(n\overline{F}(u_n)\right)\left(n^{-1}E(N'(nb)-N'(na))\right) \\
&\to \tau\,\lambda\,(b-a)\,.
\end{aligned}$$

Here we also used that $n^{-1}E(N'(nb)-N'(na)) \sim \lambda(b-a)$; see Proposition 2.5.12. This proves (5.19). Next we turn to the proof of (5.20). For simplicity we restrict ourselves to the set $B=(c_1,d_1]\cup(c_2,d_2]$. Conditioning on $N'$ and using the independence of $(T_i)$ and $(X_i)$, we obtain

$$\begin{aligned}
&P\left(\widetilde{N}_n(B)=0\right) \\
&= P\left(\max_{i:c_1<n^{-1}T_i\le d_1} X_i \le u_n\,,\quad \max_{i:c_2<n^{-1}T_i\le d_2} X_i \le u_n\right) \\
&= P\left(\max_{N'(nc_1)<i\le N'(nd_1)} X_i \le u_n\,,\quad \max_{N'(nc_2)<i\le N'(nd_2)} X_i \le u_n\right)
\end{aligned}$$

$$
\begin{aligned}
&= E\left((F(u_n))^{(N'(nd_1)-N'(nc_1))+(N'(nd_2)-N'(nc_2))}\right)\\
&= E\exp\left\{\frac{(N'(nd_1)-N'(nc_1))+(N'(nd_2)-N'(nc_2))}{n} n\ln(1-\overline{F}(u_n))\right\}\\
&\to \exp\left\{-\lambda\tau\Big((d_1-c_1)+(d_2-c_2)\Big)\right\}\\
&= P(N(B)=0)\,.
\end{aligned}
$$

In the last step we also used the SLLN for renewal counting processes (Theorem 2.5.10) and Lebesgue dominated convergence. This proves (5.20) and, by Kallenberg's theorem, also the assertion. □

**Example 5.3.5** (Limit distribution for iid sequence with random index)
Let $(X_i)$ be iid and independent of the renewal counting process $(N'(t))$ on $[0,\infty)$ with $EY_1 = \lambda^{-1}$. Denote by $X_{N'(t),N'(t)} \le \cdots \le X_{1,N'(t)}$ the order statistics of the random sample $X_1,\ldots,X_{N'(t)}$. Then we may conclude from Theorem 5.3.4 that

$$
\begin{aligned}
P\left(X_{k,N'(n)} \le u_n\right) &= P\left(N_n(0,1] < k\right)\\
&\to e^{-\tau\lambda}\sum_{i=0}^{k-1}\frac{(\tau\lambda)^i}{i!}\,, \quad k=1,2,\ldots\,,
\end{aligned}
$$

provided $n\overline{F}(u_n) \to \tau \in (0,\infty)$. In particular, if $u_n = u_n(x) = c_n x + d_n$ and $n\overline{F}(u_n(x)) \to -\ln H(x)$, $x \in \mathbb{R}$, for some extreme value distribution $H$, then

$$
P\left(X_{k,N'(n)} \le c_n x + d_n\right) \to H^\lambda(x)\sum_{i=0}^{k-1}\frac{\left(-\ln H^\lambda(x)\right)^i}{i!}\,, \quad k=1,2,\ldots\,.
$$

This result was given in Theorem 4.3.2 in a more general set–up. □

### 5.3.2 The Stationary Case

In this section we approach the problem of finding the limit distribution of the maxima $M_n$ and of the upper order statistics of a sample from a strictly stationary sequence $(X_n)$ via the point process of exceedances as introduced in (5.14). We assume that the conditions $D(u_n)$ and $D'(u_n)$ from Section 4.4 hold for a threshold sequence $(u_n)$, and we cite them here for convenience:

**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* $l = l_n = o(n)$.

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

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

*holds as* $k \to \infty$.

**Remark.** For an interpretation of these conditions we refer to Section 4.4. We mention here that condition $D'(u_n)$ has an intuitive interpretation in the language of point processes: if $(u_n)$ is chosen to satisfy $n\overline{F}(u_n) \to \tau \in (0, \infty)$ then there are on average approximately $\tau$ exceedances of $u_n$ by $X_1, \ldots, X_n$, and hence $\tau/k$ among $X_1, \ldots, X_{[n/k]}$. Condition $D'(u_n)$ bounds the probability of more than one exceedance among $X_1, \ldots, X_{[n/k]}$. This will eventually ensure that there are no multiple points in the limiting Poisson process; i.e. this condition prevents clustering in the limit. In this context, Example 4.4.4 is quite instructive: condition $D'(u_n)$ is violated since maxima typically occur as pairs. □

Having in mind the results of Section 4.4 it is certainly not surprising that Theorem 5.3.2 remains valid for certain strictly stationary sequences:

**Theorem 5.3.6** (Weak convergence of point processes of exceedances, stationary case)
*Suppose* $(X_n)$ *is strictly stationary and* $(u_n)$ *is a sequence of threshold values such that* (5.16), $D(u_n)$ *and* $D'(u_n)$ *hold. Let* $(N_n)$ *be the processes* (5.14). *Then* $N_n \xrightarrow{d} N$ *in* $M_p(E)$, *where* $N$ *is a homogeneous PRM on* $E = (0, 1]$ *with intensity* $\tau$.

**Proof.** We proceed as in the proof of Theorem 5.3.2 or 5.3.6, applying Kallenberg's Theorem 5.2.2. The proof of (5.9) is the same as in the iid case. Thus it remains to show (5.10) making use of $D(u_n)$ and $D'(u_n)$. For simplicity we restrict ourselves to sets $B = (c_1, d_1] \cup (c_2, d_2]$ with $0 < c_1 < d_1 < c_2 < d_2 \le 1$. The general case can be dealt with analogously.

Take $(a, b] \subset (0, 1]$. Using the stationarity of $(X_n)$ and Proposition 4.4.3 we obtain

$$\begin{aligned} P\left(N_n(a,b]=0\right) &= P\left(\max_{i\le [nb]-[na]} X_i \le u_n\right) \\ &\to \exp\{-\tau(b-a)\} = P(N(a,b]=0)\,. \end{aligned} \tag{5.21}$$

From condition $D(u_n)$ we conclude that

$$\begin{aligned} & P\left(N_n(B)=0\right) \\ &= P\left(N_n(c_1,d_1]=0\,, N_n(c_2,d_2]=0\right) \\ &= P\left(\max_{c_1<n^{-1}i\le d_1} X_i \le u_n\,,\ \max_{c_2<n^{-1}i\le d_2} X_i \le u_n\right) \\ &= P\left(\max_{c_1<n^{-1}i\le d_1} X_i \le u_n\right) P\left(\max_{c_2<n^{-1}i\le d_2} X_i \le u_n\right) + o(1)\,. \end{aligned}$$

Indeed, the distance between the two index sets

$$A_1 = \{[nc_1]+1,\ldots,[nd_1]\} \quad \text{and} \quad A_2 = \{[nc_2]+1,\ldots,[nd_2]\}$$

exceeds $(c_2-d_1)n > l_n = o(n)$ which implies that $\alpha_{n,l_n} \to 0$. Hence, by (5.21),

$$P\left(N_n(B)=0\right) \to \exp\left\{-\tau\Big((d_1-c_1)+(d_2-c_2)\Big)\right\} = P(N(B)=0)\,,$$

which concludes the proof of (5.10) and, by Kallenberg's theorem, proves the assertion. □

The following is analogous to Example 5.3.3:

**Example 5.3.7** (Limit probabilities of upper order statistics)
As usual, let

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

denote the order statistics of the sample $X_1,\ldots,X_n$. Suppose that the assumptions of Theorem 5.3.6 hold. Then

$$P\left(X_{k,n}\le u_n\right) = P\left(N_n(0,1]<k\right) \to P(N(0,1]<k) = e^{-\tau}\sum_{i=0}^{k-1}\frac{\tau^i}{i!}\,.$$

This extends Proposition 4.4.3 to the upper order statistics of a strictly stationary sequence. □

Now it is immediate that we can derive the limit distribution of an upper order statistic $X_{k,n}$ by the usual folklore. Let $(\widetilde{X}_n)$ be an associated iid sequence such that $X \stackrel{d}{=} \widetilde{X}$, and denote its order statistics in the natural way by $\widetilde{X}_{k,n}$.

**Theorem 5.3.8** (Limit distribution of upper order statistics)
*Let $(X_n)$ be strictly stationary with common df $F \in \mathrm{MDA}(H)$ for an extreme value distribution $H$, i.e. there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that*

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

*Assume that the sequences $(u_n) = (c_n x + d_n)$, $x \in \mathbb{R}$, satisfy the conditions $D(u_n)$ and $D'(u_n)$. Then the relations*

$$P\left(c_n^{-1}\left(X_{k,n} - d_n\right) \le x\right) \quad \to \quad H(x) \sum_{i=0}^{k-1} \frac{(-\ln H(x))^i}{i!}\,, \quad x \in \mathbb{R}\,,$$

$$P(c_n^{-1}(\widetilde{X}_{k,n} - d_n) \le x) \quad \to \quad H(x) \sum_{i=0}^{k-1} \frac{(-\ln H(x))^i}{i!}\,, \quad x \in \mathbb{R}\,,$$

*hold for every $k \ge 1$.* □

Theorem 5.3.8 shows the similarity between the asymptotic behaviour of the extremes of the stationary sequence $(X_n)$ and of an associated iid sequence $(\widetilde{X}_n)$. This is again due to the conditions $D(u_n)$ and $D'(u_n)$.

In the following paragraphs we intend to generalise these results to a finite vector of order statistics. This means that we are interested in probabilities of the form

$$P\left(X_{1,n} \le u_n^{(1)}, \ldots, X_{k,n} \le u_n^{(k)}\right)$$

for $k$ sequences of real numbers

$$u_n^{(k)} \le \cdots \le u_n^{(1)}\,. \tag{5.22}$$

Since we are dealing with $k$ different sequences of thresholds $(u_n^{(i)})$, $i = 1, \ldots, k$, it seems appropriate to introduce a vector of $k$ point processes of exceedances, one for each threshold sequence. However, the exceedances of the levels $u_n^{(i)}$ are very much related to each other. For example, an exceedance of $u_n^{(r)}$ is automatically an exceedance of $u_n^{(r+1)}$, and so it is possible by a geometric argument to reduce the problem of $k$ exceedances to weak convergence of a point process on $(0,1] \times \mathbb{R}$. We refer to Leadbetter, Lindgren and Rootzén [418], Sections 5.5 and 5.6, for a complete description of their "thinning" procedure and omit details. We also omit the definition of the corresponding point processes and simply state the final result for the vector of exceedances. Before we can do this we have to introduce a $k$–dimensional analogue of condition $D(u_n)$ above. We suppose that the $k$ sequences (5.22) are given.

**Condition** $D_k(\mathbf{u_n})$**:** *For any fixed* $p, q$ *and for any integers*

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

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

$$\Big| P\left(X_{i_m} \le u_n^{(s_m)}\,,\, m=1,\ldots,p, \quad X_{j_r} \le u_n^{(s'_r)}, r=1,\ldots,q\right)$$

$$- \quad P\left(X_{i_m} \le u_n^{(s_m)}\,,\, m=1,\ldots,p\right) P\left(X_{j_r} \le u_n^{(s'_r)}, r=1,\ldots,q\right)\Big| \le \alpha_{n,l}\,,$$

*for any integers* $1 \le s_l, s'_r \le k$*, and* $\alpha_{n,l} \to 0$ *as* $n \to \infty$ *for some sequence* $l = l_n = o(n)$*.*

It will not be necessary to define an extended $D'(\mathbf{u_n})$ condition, since we shall simply need to assume that $D'(u_n^{(i)})$ holds separately for each $i = 1, \ldots, k$.

As in Section 4.2 we write

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

for the number of exceedances of $u_n^{(i)}$ by $X_1, \ldots, X_n$.

**Theorem 5.3.9** (Joint weak convergence of the number of exceedances, stationary case)
*Let* $(X_n)$ *be a strictly stationary sequence and suppose that the sequences* $(u_n^{(i)})$ *satisfy* (5.22) *and that* $n\overline{F}(u_n^{(i)}) \to \tau_i$ *for non-negative* $\tau_i$, $i = 1, \ldots, k$*. Assume* $D_k(\mathbf{u_n})$ *and* $D'(u_n^{(i)})$ *for* $i = 1, \ldots, k$*. Then, for* $\ell_1, \ldots, \ell_k \ge 0$*,*

$$P\left(B_n^{(1)} = \ell_1, B_n^{(2)} = \ell_1 + \ell_2, \ldots, B_n^{(k)} = \ell_1 + \cdots + \ell_k\right)$$

$$\to \quad \frac{\tau_1^{\ell_1}}{\ell_1!}\, \frac{(\tau_2 - \tau_1)^{\ell_2}}{\ell_2!} \cdots \frac{(\tau_k - \tau_{k-1})^{\ell_k}}{\ell_k!}\, e^{-\tau_k}\,, \qquad n \to \infty\,. \qquad \square$$

This theorem is completely analogous to the iid case; see Theorem 4.2.6. Moreover, as in the iid case, cf. Theorem 4.2.8, we obtain the joint limit law of the vector of upper order statistics:

**Corollary 5.3.10** (Joint limit law of upper order statistics, stationary case)
*Assume that* $F \in \mathrm{MDA}(H)$ *with normalising constants* $c_n > 0$ *and centring constants* $d_n \in \mathbb{R}$*. Moreover, suppose that* $D_k(\mathbf{u_n})$ *and* $D'(u_n)$ *are satisfied for all sequences* $u_n = c_n x + d_n$*,* $x \in \mathbb{R}$*. Then the limit relation*

$$(c_n^{-1}(X_{i,n} - d_n))_{i=1,\ldots,k} \xrightarrow{d} (Y^{(i)})_{i=1,\ldots,k}\,, \quad k \ge 1\,, \quad n \to \infty\,,$$

*holds, where* $(Y^{(1)}, \ldots, Y^{(k)})$ *is the* $k$*–dimensional extremal variate corresponding to the extreme value distribution* $H$*.* $\square$

Finally, we mention that all results for a vector of $k$ upper order statistics which were given in Section 4.2 for the iid case remain valid for the strictly stationary case as well, provided that $D$ and $D'$ are satisfied.

### Notes and Comments

The point process of exceeedances has been used extensively in the monograph by Leadbetter et al. [418] to build up an extreme value theory for iid and stationary sequences. There the theory presented above can be found in detail. In particular, they discuss the conditions $D(u_n)$ and $D'(u_n)$; see also Section 4.4. Further convergence results for the point process of exceedances are provided in Sections 5.5 and 8.4, where we consider linear and ARCH processes. In contrast to the present section the limiting point processes are not homogeneous Poisson but compound Poisson processes.

The point process techniques of this section could have been replaced by classical methods of extreme value theory. The latter were implicitly used for checking the assumptions of Kallenberg's theorem. Therefore the present section can be understood as an alternative approach to extreme value theory which is quite elegant in the case of stationary sequences. The real power of point process methods will become more transparent in Sections 5.4 and 5.5.

## 5.4 Applications of Point Process Methods to IID Sequences

In this section we apply point process techniques to the extremes of iid sequences (for some basic facts we refer to Chapters 3 and 4). We are mainly interested in records and record times. In Section 5.4.1 we give a short introduction to this topic. It is followed by some technical results (Section 5.4.2) which are used to embed the maxima of an iid sequence in an appropriate continuous–time process which in turn is a function of a PRM. This "coupling" construction is applied in Section 5.4.3 to derive limit results about the growth and the frequency of record times. In Section 5.4.4 we consider the weak convergence of maxima in a function space setting.

Throughout this section $X, X_1, X_2, \ldots$ is a sequence of iid rvs with common *continuous* df $F$. We also write

$$x_F^l = \inf\{x : F(x) > 0\} \quad \text{and} \quad x_F^r = \sup\{x : F(x) < 1\}$$

for the left and right endpoint of the distribution $F$. As usual, we denote the maximum of the first $n$ rvs by

$$M_1 = X_1\,, \quad M_n = \max_{i=1,\ldots,n}(X_1,\ldots,X_n)\,, \quad n \geq 2\,.$$

Later on we will sometimes find it convenient to use $\wedge$, $\vee$ for min, max, respectively. The rvs $\Gamma_i$ are always the points of a homogeneous Poisson process on $[0,\infty)$ with intensity 1. We can write them as

$$\Gamma_i = E_1 + \cdots + E_i\,, \quad i \geq 1\,,$$

for an iid sequence of standard exponential rvs $E_i$.

### 5.4.1 Records and Record Times

In daily life we hear quite often about records; they are indeed omnipresent in sports, science, economy, environment etc. We hear about records of pollution, records of governmental debts, records in sports events, record insurance claims or record gains/losses in finance. Some clever people collect information about all sorts of records and write books about them.

*What is a record in the context of extreme value theory?*

If we consider observations $X_n$ a record would be a temporary maximum (or minimum) in this sequence which will certainly change when time goes by. This is precisely the notion *record* which we intend to use in this chapter: a *record* $X_n$ occurs if $X_n > M_{n-1}$. Clearly, the new maximum $M_n$ coincides then with $X_n$. Notice that a record happens when there is a jump in the sequence $(M_n)$. The times $L_1 < L_2 < \cdots$ when these jumps occur are random. For obvious reasons, they are called the *record times* of $(X_n)$. In the insurance and financial context it is definitely an important issue to study both records and record times of sequences of rvs, dependent or independent. They give us some sort of prediction of the good or bad things which can happen in the future, in frequency and magnitude: big jumps in prices can lead to crashes of financial institutions; big claim sizes in an insurance portfolio can cause insolvency problems.

The following result describes the sequence of records $(X_{L_n})$ in terms of a PRM:

**Theorem 5.4.1** (Point process description of records)
*Let $F$ be a continuous df with left endpoint $x_F^l$ and right endpoint $x_F^r$. Then the records $(X_{L_n})$ of the iid sequence $(X_n)$ are the points of a* PRM$(\mu)$ *on $(x_F^l, x_F^r)$ with mean measure $\mu$ given by*

$$\mu(a,b] = R(b) - R(a)\,, \quad x_F^l < a \leq b < x_F^r\,, \quad \textit{where} \quad R(x) = -\ln \overline{F}(x)\,.$$

*In particular, if $F$ is standard exponential then $R(t) = t$ and $(X_{L_n}) \stackrel{d}{=} (\Gamma_n)$ are the points of a homogeneous Poisson process on $\mathbb{R}_+$ with intensity* 1.

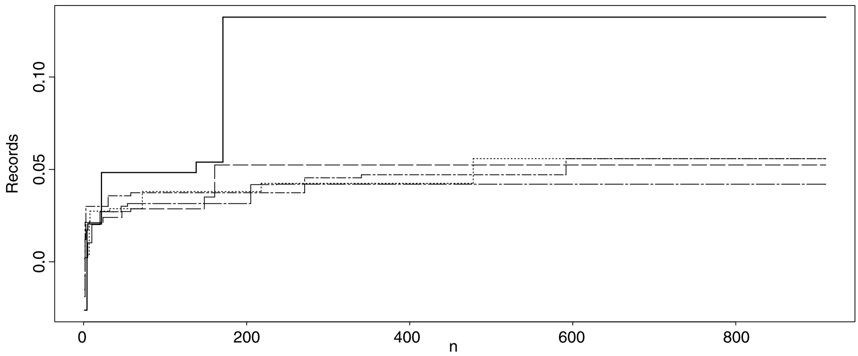


**Figure 5.4.2** *Records (solid top line) of* 910 *daily log–returns of the Japanese stock index* NIKKEI *(February 22, 1990 – October 8, 1993) compared with four sample paths of records from* 910 *iid rvs. The rvs in the latter sequence are Gaussian with mean zero and the same variance as the* NIKKEI *data.*

**Proof.** Since $F$ is continuous the function $R^{\leftarrow}$ is monotone increasing. Direct calculation yields

$$X_1 \stackrel{d}{=} R^{\leftarrow}(E_1)\,.$$

Indeed,

$$\begin{aligned} P\left(R^{\leftarrow}(E_1) \le x\right) &= P\left(E_1 \le R(x)\right) \\ &= 1 - e^{-R(x)} = F(x)\,. \end{aligned}$$

Hence the sequences $(M_n)$ and

$$\left(\bigvee_{i=1}^{n} R^{\leftarrow}(E_i)\right) = \left(R^{\leftarrow}\left(\bigvee_{i=1}^{n} E_i\right)\right)$$

have the same distribution. Moreover, denoting by $(\widetilde{L}_n)(\stackrel{d}{=} (L_n))$ the record times of the sequence $(R^{\leftarrow}(E_i))$, we have for the sequences of records that

$$(X_{L_n}) \stackrel{d}{=} \left(R^{\leftarrow}\left(E_{\widetilde{L}_n}\right)\right)\,.$$

If $F$ is standard exponential then the records $(R^{\leftarrow}(E_{\widetilde{L}_n})) = (E_{\widetilde{L}_n})$ are the points of a homogeneous Poisson process on $\mathbb{R}_+$ with intensity 1. This follows from the observation that $(E_{\widetilde{L}_n})$ is Markov with transition probabilities $\pi(x,(y,\infty)) = \exp\{-(y-x)\}$; see Resnick [530], Proposition 4.1. In view of Proposition 5.1.13, $(R^{\leftarrow}(E_{\widetilde{L}_n}))$ are then the points of a PRM with mean measure of $(a,b]$ given by

$$\left|\left(R^{\leftarrow}\right)^{-1}(a,b]\right| = |\{s : a < R^{\leftarrow}(s) \le b\}| = R(b) - R(a)\,,$$

where $|\cdot|$ denotes Lebesgue measure. This concludes the proof. □

### 5.4.2 Embedding Maxima in Extremal Processes

The sequence $(M_n)_{n\ge 1}$ defines a discrete–time stochastic process on the integers. We consider the finite–dimensional distributions of this process. We start with two dimensions: let $x_1 < x_2$ be real numbers and $t_1 < t_2$ be positive integers. Then

$$\begin{aligned} P\left(M_{t_1} \le x_1, M_{t_2} \le x_2\right) &= P\left(M_{t_1} \le x_1, \bigvee_{i=t_1+1}^{t_2} X_i \le x_2\right) \\ &= P\left(M_{t_1} \le x_1\right) P\left(M_{t_2-t_1} \le x_2\right) \\ &= F^{t_1}(x_1) F^{t_2-t_1}(x_2)\,. \end{aligned}$$

Moreover, if $x_1 > x_2$,

$$P\left(M_{t_1} \le x_1, M_{t_2} \le x_2\right) = F^{t_2}(x_2)\,.$$

Hence

$$P\left(M_{t_1} \le x_1, M_{t_2} \le x_2\right) = F^{t_1}\left(x_1 \wedge x_2\right)\ F^{t_2-t_1}\left(x_2\right)\,.$$

By induction we obtain

$$\begin{aligned} & P\left(M_{t_1} \le x_1, M_{t_2} \le x_2, \ldots, M_{t_m} \le x_m\right) \\ &= F^{t_1}\left(\bigwedge_{i=1}^m x_i\right) F^{t_2-t_1}\left(\bigwedge_{i=2}^m x_i\right) \cdots F^{t_m-t_{m-1}}\left(x_m\right) \end{aligned} \tag{5.23}$$

for all positive integers $t_1 < t_2 < \cdots < t_m$, every $m \ge 1$ and real numbers $x_i$. From this representation it is not difficult to see that $(M_n)$ is a Markov process; see Resnick [530], Section 4.1 or Breiman [90], Chapter 15.

We take (5.23) as the starting point for the definition of an *F–extremal process*: if we do not restrict ourselves to the non–negative integers, but if we allow for general real numbers $0 < t_1 < t_2 < \cdots < t_m$ then (5.23) defines a consistent family of distributions which, in view of Kolmogorov's consistency theorem, determines the distribution of a continuous–time process $Y$ on $\mathbb{R}_+$.

**Definition 5.4.3** ($F$–extremal process)
*The process* $Y = (Y(t))_{t>0}$ *with finite–dimensional distributions* (5.23) *is called an* extremal process generated by the df $F$ *or an* $F$–extremal process. □

Thus the discrete–time process of the sample maxima $(M_n)$ can be embedded in the continuous–time extremal process $Y$ in the sense that

$$(M_n)_{n\geq 1} \stackrel{d}{=} (Y(n))_{n\geq 1}\,.$$

The latter relation is checked by a glance at the finite–dimensional distributions of $(M_n)$ and $Y$ at integer instants of time. The continuous–time process $Y$ inherits the distributional properties of the sequence of maxima; it is a convenient tool for dealing with them.

An extremal process can be understood as a function of a PRM. Indeed, let

$$N = \sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k)} \tag{5.24}$$

be PRM$(|\cdot|\times\mu)$ with state space $E = \mathbb{R}_+ \times \mathbb{R}$, where $|\cdot|$ denotes Lebesgue measure, and $\mu$ is given by the relation $\mu(a,b] = \ln F(b) - \ln F(a)$ for $a<b$. It is convenient to interpret $(t_k, j_k)$ as coordinates of time (i.e. $t_k$) and space (i.e. $j_k$). Recall the definition of the Skorokhod space $\mathbb{D}$ of cadlag functions from Appendix A2.3 and define the mapping $\widetilde{T}_1 : M_p(E) \to \mathbb{D}\,(0,\infty)$ by

$$\widetilde{T}_1(N) = \widetilde{T}_1\left(\sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k)}\right) = \sup\{j_k : t_k \leq \cdot\}\,. \tag{5.25}$$

**Proposition 5.4.4** (Point process representation of $F$–extremal processes)
*The $F$–extremal process $Y = (Y(t))_{t>0}$ has representation*

$$Y(\cdot) \stackrel{d}{=} \sup\{j_k : t_k \leq \cdot\}$$

*with respect to the* PRM$(|\cdot|\times\mu)$ *defined in* (5.24).

**Sketch of the proof.** In view of the constructive definition of $Y$ given above it suffices to show that the finite–dimensional distributions of $Y$ and $\widetilde{T}_1(N)$ coincide. Fix $t>0$. Notice that

$$\{\sup\{j_k : t_k \leq t\} \leq x\} \quad = \quad \{N((0,t]\times(x,\infty)) = 0\}\,.$$

Thus we have by definition of a PRM that

$$\begin{aligned} P(\sup\{j_k : t_k \leq t\} \leq x) &= P(N((0,t]\times(x,\infty)) = 0) \\ &= \exp\{-EN((0,t]\times(x,\infty))\} \\ &= \exp\{-t\,\mu(x,\infty)\} \\ &= F^t(x) \\ &= P(Y(t) \leq x)\,. \end{aligned}$$

Similar arguments yield the finite–dimensional distributions in the general case. (The reader is urged to calculate them at least for two dimensions.) They can be shown to coincide with (5.23) which determine the whole distribution of $Y$. This concludes the proof. □

In the following we need another representation of an $F$–extremal process. It is a consequence of the following auxiliary result:

**Lemma 5.4.5** *Assume $F$ is continuous. Let $N$ be the* PRM$(|\cdot|\times\mu)$ *on* $(0,t_0]\times(x_F^l,x_F^r)$, $t_0>0$, *as defined in* (5.24). *Then $N$ has representation*

$$N'=\sum_{i=1}^{\infty}\varepsilon_{(U_i,Q^{\leftarrow}(\Gamma_i/t_0))}\,,$$

*where $(U_i)$ are iid uniform on $(0,t_0)$, independent of the points $(\Gamma_i)$ of a homogeneous Poisson process on $[0,\infty)$ with intensity* 1, *and*

$$Q^{\leftarrow}(y)=\inf\{s:Q(s)\le y\}\,,\quad Q(x)=-\ln F(x)\,.$$

**Proof.** In view of Remark 1 after Definition 5.1.7 it suffices to show that the Laplace functionals of $N$ and $N'$ coincide. Since $N$ is PRM we know from Example 5.1.11 that

$$\Psi_N(g)=\exp\left\{-\int_{(0,t_0]}\int_{(x_F^l,x_F^r)}\left(1-e^{-g(t,x)}\right)d(\ln F(x))\,dt\right\}\,. \tag{5.26}$$

Now, since $F$ is continuous, $Q^{\leftarrow}$ is monotone decreasing. Conditioning on $(\Gamma_i)$ and writing

$$g_1(x)=\frac{1}{t_0}\int_0^{t_0}\exp\{-g(t,x)\}\,dt=E\exp\{-g(U_1,x)\}\,,$$

we obtain

$$\begin{aligned}
\Psi_{N'}(g) &= E\exp\left\{-\int_{(0,t_0]\times(x_F^l,x_F^r)}g\,dN'\right\}\\
&= E\exp\left\{-\sum_{i=1}^{\infty}g\left(U_i,Q^{\leftarrow}(\Gamma_i/t_0)\right)\right\}\\
&= E\prod_{i=1}^{\infty}g_1\left(Q^{\leftarrow}(\Gamma_i/t_0)\right)\\
&= E\exp\left\{\sum_{i=1}^{\infty}\ln g_1\left(Q^{\leftarrow}(\Gamma_i/t_0)\right)\right\}\,.
\end{aligned}$$

An application of Lemma 5.1.12 yields

$$\begin{aligned} \Psi_{N'}(g) &= \exp\left\{-\int_{\mathbb{R}_+} \left(1 - g_1\left(Q^{\leftarrow}(z/t_0)\right)\right) dz\right\} \\ &= \exp\left\{-\int_{(0,t_0]}\int_{\mathbb{R}_+} \left(1 - e^{-g(t,Q^{\leftarrow}(z))}\right) dz\, dt\right\}. \end{aligned}$$

Substituting $x$ for $Q^{\leftarrow}(z)$ we arrive at the right–hand side of (5.26) which concludes the proof. □

An immediate consequence of Lemma 5.4.5 and Proposition 5.4.4 is the following

**Corollary 5.4.6** *Let $F$ be a continuous df, $Y$ an $F$–extremal process. Then $Y$ has representation*

$$Y(t) = \sup\left\{Q^{\leftarrow}(\Gamma_i/t_0) : U_i \le t\right\}, \quad t \in (0, t_0],$$

*where $(U_i)$ and $(\Gamma_i)$ are defined in Lemma 5.4.5.* □

The *jump times* $\tau_n$ of an $F$–extremal process are of particular interest since we may hope that jumps of $(M_n)$ (the *records*) and of $Y$ occur almost at the same time. This intuition will be made precise by a coupling argument in Section 5.4.3.

**Theorem 5.4.7** (Point process of the jump times of an extremal process) *If $F$ is continuous then*

$$N_\infty = \sum_{n=1}^{\infty} \varepsilon_{\tau_n} \tag{5.27}$$

*is* $\mathrm{PRM}(\mu)$ *on* $\mathbb{R}_+$ *with intensity* $f(t) = 1/t$, *i.e.*

$$\mu(a,b] = \int_a^b f(t)\, dt = \ln b - \ln a \quad \text{for } a < b\,.$$

**Proof.** It suffices to show that $N_\infty$ is $\mathrm{PRM}(\mu)$ on $(0, t_0]$ for every fixed $t_0 > 0$. In view of Corollary 5.4.6 we may assume that the $F$–extremal process $Y$ has representation

$$Y(t) = \sup\left\{Q^{\leftarrow}(\Gamma_i/t_0) : U_i \le t\right\}, \quad t \in (0, t_0]\,.$$

Since $F$ is continuous, $Q^{\leftarrow}$ is monotone decreasing, hence the jump times of the processes $Y$ and $N_1(t) = \inf\{n \ge 1 : U_n \le t\}$ are identical. We may write

$$\begin{aligned}N_1(t) &= \inf\{n \geq 1 : U_n^{-1} \geq t^{-1}\} \\ &= \inf\left\{n \geq 1 : \bigvee_{i=1}^{n} U_i^{-1} \geq t^{-1}\right\}.\end{aligned}$$

Hence the jump times of $N_1$ in $(0, t_0]$ must be the records of $\max_{i=1,\ldots,n} U_i^{-1}$ in $[t_0^{-1}, \infty)$. By Theorem 5.4.1, the records of $\max_{i=1,\ldots,n} U_i^{-1}$ are the points of a PRM on $[t_0, \infty)$ with mean measure of $(a, b]$ given by

$$\begin{aligned}-\ln P\left(U_1^{-1} > b\right) - \left(-\ln P\left(U_1^{-1} > a\right)\right) &= -\ln(b^{-1}/t_0) + \ln(a^{-1}/t_0) \\ &= \ln(b/a)\,.\end{aligned}$$

This concludes the proof. □

### 5.4.3 The Frequency of Records and the Growth of Record Times

In this section we use a special "coupling" construction of the jump times $L_n$ of $(M_n)$ and $\tau_n$ of the $F$–extremal process $Y$ to derive information about the record times of the iid sequence $(X_n)$. This will allow us to compare $(L_n)$ and $(\tau_n)$ not only in distribution but also path by path.

By definition of $Y$ (see Definition 5.4.3) $(M_n) \stackrel{d}{=} (Y(n))$. This relation allows us to assume that $(L_n)$ and $(\tau_n)$ are defined on the same probability space in such a way that a jump of $(M_n)$ (i.e. a *record*) at $L_n$ (the *record time*) is also a jump of $Y$ but the converse is not necessarily true. Indeed, $Y$ is a continuous–time process, and so it may also have jumps in the open intervals $(L_n - 1, L_n)$. Recall the definition of the point process $N_\infty$ of the jump times of $Y$ from (5.27) and define the *point process of the record times* of $(X_n)$ by

$$N = \sum_{i=1}^{\infty} \varepsilon_{L_i}\,. \tag{5.28}$$

Then, given the above coupling construction of $(L_n)$ and $(\tau_n)$,

$$\begin{aligned}\{N(n-1, n] = 1\} &= \{(X_i) \text{ has a record at time } n.\} \\ &= \{N_\infty(n-1, n] \geq 1\}\,.\end{aligned} \tag{5.29}$$

The following question arises naturally:

*How often does it actually happen that $N_\infty(n-1, n] > N(n-1, n]$?*

The following result ensures that the sequences $(N_\infty(n-1, n])$ and $(N(n-1, n])$ are identical starting from a certain random index.

**Proposition 5.4.8** (Coupling of $N_\infty$ and $N$)
*Assume the df $F$ is continuous and that $(L_n)$ and $(\tau_n)$ are constructed as above. Then there exists an integer–valued rv $N_0$ such that for almost every $\omega \in \Omega$,*

$$N((n,n+1],\omega) = N_\infty((n,n+1],\omega)\,, \quad n \geq N_0(\omega)\,. \tag{5.30}$$

**Proof.** It suffices to show (see (5.29)) that the event $\{N_\infty(n,n+1] > 1\}$ occurs only finitely often with probability 1. By the Borel–Cantelli lemma, see Section 3.5, this is the case if

$$\sum_{n=1}^{\infty} P\left(N_\infty(n,n+1] > 1\right) < \infty\,. \tag{5.31}$$

Since $N_\infty$ is PRM($\mu$) with $\mu(a,b] = \ln(b/a)$ (see Theorem 5.4.7), direct calculation shows that

$$\begin{aligned}
& P\left(N_\infty(n,n+1] > 1\right) \\
&= 1 - P\left(N_\infty(n,n+1] = 0\right) - P\left(N_\infty(n,n+1] = 1\right) \\
&= 1 - e^{-\ln(1+n^{-1})} - e^{-\ln(1+n^{-1})}\ln(1+n^{-1}) \\
&= 1 - (1+n^{-1})^{-1}\left(1 + \ln(1+n^{-1})\right) \\
&\leq n^{-2}\,, \quad n \geq 1\,,
\end{aligned}$$

and (5.31) follows, which concludes the proof. □

**Remark.** The coupling relation (5.30) can be reformulated as follows: for almost every $\omega \in \Omega$ there exists an integer $j(\omega)$ such that

$$N_\infty((1,n],\omega) = j(\omega) + N((1,n],\omega)\,, \quad n \geq N_0(\omega)\,. \tag{5.32}$$

□

We use the coupling argument to answer the following question:

*How often do records happen in a given period of time?*

A first answer is supported by the following Poisson approximation to the point process $N$ of the record times (see (5.28)):

**Theorem 5.4.9** (Weak convergence of the point process of record times)
*The limit relation*

$$N_n(\cdot) = N(n\cdot) = \sum_{i=1}^{\infty} \varepsilon_{n^{-1}L_i}(\cdot) \xrightarrow{d} N_\infty(\cdot) = \sum_{i=1}^{\infty} \varepsilon_{\tau_i}(\cdot)$$

*holds in $M_p(\mathbb{R}_+)$.*

**Proof.** According to Theorem 5.2.3 and Remark 3 afterwards it suffices to show that

$$I_n = \int_{\mathbb{R}_+} g(x)dN_n(x) \xrightarrow{d} \int_{\mathbb{R}_+} g(x)dN_\infty(x)\,, \quad g \in C_K^+(\mathbb{R}_+)\,. \tag{5.33}$$

Since $g$ has compact support, there exists an interval $[a,b] \subset \mathbb{R}_+$ such that $g(x) = 0$ for $x \notin [a,b]$. Hence

$$\begin{aligned} I_n &= \sum_{i=1}^\infty g(n^{-1}L_i) \\ &= \sum_{i:a\le n^{-1}i\le b} g(n^{-1}i) I_{\{N(i-1,i]=1\}}\,. \end{aligned}$$

Recalling the special construction (5.30) we obtain, for $na > N_0(\omega)$,

$$\begin{aligned} I_n &= \sum_{i:a\le n^{-1}i\le b} g(n^{-1}i)\ I_{\{N_\infty(i-1,i]=1\}} \\ &= \sum_{i:a\le n^{-1}i\le b} g(n^{-1}i)\ N_\infty(i-1,i] \\ &= \int_{\mathbb{R}_+} g_n(x)dN_\infty(x) = J_n\,, \end{aligned}$$

where $g_n(x) = \sum_{i=1}^\infty g(n^{-1}i) I_{(i-1,i]}(x)$. Thus we have shown that $I_n - J_n \xrightarrow{\text{a.s.}} 0$. By a Slutsky argument (see Appendix A2.5) it remains, for (5.33), to show that

$$J_n \xrightarrow{d} \int_{\mathbb{R}_+} g(x)dN_\infty(x)\,. \tag{5.34}$$

Recall from Example 5.1.14 that $N_\infty(\cdot)$ and $N_\infty(n\cdot)$ have the same distribution. Then

$$\begin{aligned} J_n &\stackrel{d}{=} \int_{\mathbb{R}_+} g_n(x)dN_\infty(nx) \\ &= \sum_{i=1}^\infty g(n^{-1}i)\ N_\infty\left(n^{-1}(i-1), n^{-1}i\right] \\ &\xrightarrow{\text{a.s.}} \int_{\mathbb{R}_+} g(x)dN_\infty(x)\,. \end{aligned}$$

In the last step we have used the defining properties of a Lebesgue integral, which here exists since $g$ is continuous, has compact support and is bounded. This proves (5.34) and thus the theorem. □

It is an immediate consequence of this theorem that $N_n(a,b]$ is approximately $Poi(\ln(b/a))$ distributed:

$$\begin{aligned} N_n(a,b] &= \text{card}\left\{i : a < n^{-1}L_i \leq b\right\} \\ &\xrightarrow{d} N_\infty(a,b] \stackrel{d}{=} Poi(\ln(b/a)) . \end{aligned}$$

Alternatively, the frequency of records in a given interval can be described by limit theorems for $N_\infty(1,t]$. *In the following we assume that* $\tau_1 > 1$. Otherwise we may consider only that part of the sequence $(\tau_n)$ for which $\tau_i > 1$. Since $N_\infty$ is PRM($\mu$) on $\mathbb{R}_+$ with $\mu(a,b] = \ln(b/a)$ we may work with the representation (see Example 5.1.14)

$$N_\infty = \sum_{i=1}^{\infty} \varepsilon_{\exp\{\Gamma_i\}} , \tag{5.35}$$

where, as usual, $(\Gamma_i)$ are the points of a homogeneous Poisson process on $[0,\infty)$ with intensity 1. Thus

$$N_\infty(1,t] = \text{card}\left\{i : 1 < e^{\Gamma_i} \leq t\right\} = \text{card}\left\{i : 0 < \Gamma_i \leq \ln t\right\} .$$

It is immediate that we can now apply the whole limit machinery for renewal counting processes from Section 2.5.2. For the time–changed renewal counting process $(N_\infty(1,t])$ we obtain the following: let $\Phi$ denote the standard normal distribution. Then

$$\left.\begin{array}{ll} \text{SLLN} & \lim\limits_{t\to\infty} (\ln t)^{-1} N_\infty(1,t] = 1 \text{ a.s.}, \\ \text{LIL} & \limsup\limits_{t\to\infty} \, (2\ln t \, \ln\ln\ln t)^{-1/2} \, (N_\infty(1,t] - \ln t) \\ & = -\liminf\limits_{t\to\infty} \, (2\ln t \, \ln\ln\ln t)^{-1/2} \, (N_\infty(1,t] - \ln t) = 1 \text{ a.s.}, \\ \text{CLT} & (\ln t)^{-1/2} \, (N_\infty(1,t] - \ln t) \xrightarrow{d} \Phi . \end{array}\right\} \tag{5.36}$$

The coupling construction (5.32) immediately implies that

$$c_n^{-1} \left(N_\infty(1,n] - N(1,n]\right) \stackrel{\text{a.s.}}{\rightarrow} 0$$

provided $c_n \to \infty$. This ensures that we may replace $N_\infty$ by $N$ and $t$ by $n$, everywhere in (5.36):

**Theorem 5.4.10** (Limit results for the frequency of records)
*Suppose $F$ has a continuous distribution, let $(X_n)$ be an iid sequence with record times $(L_n)$ and let $N$ be the corresponding point process* (5.28). *Then the following relations hold:*

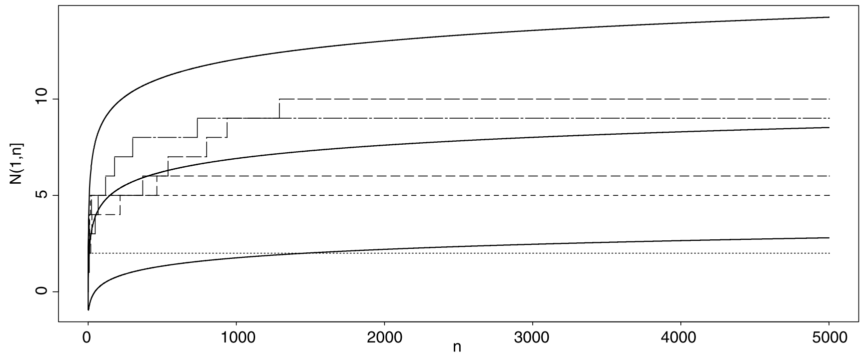


**Figure 5.4.11** *The number of records $N(1,n]$, $n \le 5\,000$, from iid standard normal rvs. Five sample paths are given. The solid lines indicate the graphs of $\ln n$ (middle) and the 95% asymptotic confidence bands (top and bottom) based on Theorem 5.4.10.*

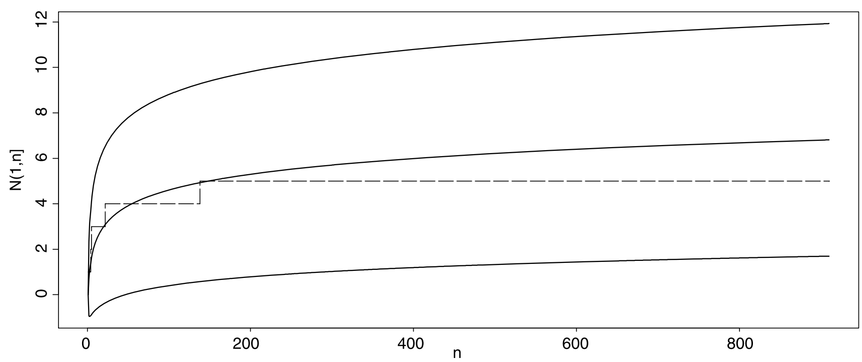


**Figure 5.4.12** *The number of records $N(1,n]$, $n \le 910$, from 910 daily log–returns of the Japanese stock index* NIKKEI *(February 22, 1990 – October 8, 1993). The solid lines are the graphs of $\ln n$ (middle) and the 95% asymptotic confidence bands for the iid case; see Theorem 5.4.10.*

$$
\begin{aligned}
&\text{SLLN} \quad \lim_{t\to\infty} (\ln t)^{-1} N(1,t] = 1 \quad \text{a.s.}\,,\\
&\text{LIL} \quad \limsup_{t\to\infty}\, (2\ln t\;\ln\ln\ln t)^{-1/2}\,(N(1,t] - \ln t)\\
&\qquad\quad = -\liminf_{t\to\infty}\, (2\ln t\;\ln\ln\ln t)^{-1/2}\,(N(1,t] - \ln t) = 1 \text{ a.s.}\,,\\
&\text{CLT} \quad (\ln t)^{-1/2}\,(N(1,t] - \ln t) \xrightarrow{d} \Phi\,,
\end{aligned}
$$

*where $\Phi$ denotes the standard normal distribution.* □

Finally, we attack the following problem:

*When do the records of $(X_n)$ occur?*

The coupling construction (5.32) again gives an answer: for $n \geq N_0(\omega)$ and almost every $\omega$,

$$\left|L_n(\omega) - \tau_{n+j(\omega)}(\omega)\right| \leq 1\,.$$

Hence, by (5.35),

$$\begin{aligned}\ln L_n &= \ln\left(e^{\Gamma_{n+j}}\left(1 + O\left(e^{-\Gamma_{n+j}}\right)\right)\right)\\ &= \Gamma_{n+j} + o(1) \quad \text{a.s.}\end{aligned}$$

since $\Gamma_{n+j} = O(n)$ a.s. by the SLLN. We learnt in Example 3.5.6 that

$$\lim_{n\to\infty} (\ln n)^{-1} \max(E_1, \ldots, E_n) = 1 \quad \text{a.s.}$$

Hence

$$\ln L_n = \Gamma_n + (\Gamma_{n+j} - \Gamma_n) + o(1) = \Gamma_n + O(\ln n) \quad \text{a.s.}$$

This and the classical limit theory for sums of iid rvs (see Sections 2.1 and 2.2) yield the following:

**Theorem 5.4.13** (Limit results for the growth of record times)
*Assume $F$ is continuous. Then the following relations hold for the record times $L_n$ of an iid sequence $(X_n)$ :*

$$\begin{array}{ll}\text{SLLN} & \lim\limits_{n\to\infty} n^{-1}\ln L_n = 1 \quad \text{a.s.}\,,\\ \text{LIL} & \limsup\limits_{n\to\infty}\,(2n\,\ln\ln n)^{-1/2}(\ln L_n - n)\\ & = -\liminf\limits_{n\to\infty}\,(2n\,\ln\ln n)^{-1/2}(\ln L_n - n) = 1 \text{ a.s.}\,,\\ \text{CLT} & n^{-1/2}(\ln L_n - n) \xrightarrow{d} \Phi\,,\end{array}$$

*where $\Phi$ denotes the standard normal distribution.* □

In summary, the number of records in the interval $(1,t]$ is roughly of the order $\ln t$. Thus records become more and more unlikely for large $t$. Alternatively, the record times $L_n$ grow roughly exponentially like $\exp\{\Gamma_n\}$ (or $\exp\{n\}$) and thus the period between two successive records becomes bigger and bigger.

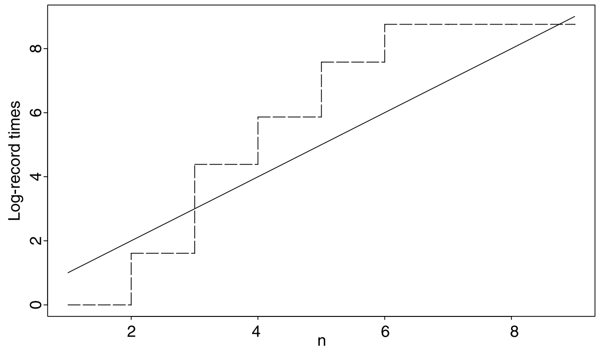


**Figure 5.4.14** *The logarithmic record times of* 1 864 *daily log–returns of the* S&P *index. According to Theorem 5.4.13, the logarithmic record times should grow roughly linearly provided that they come from iid data.*

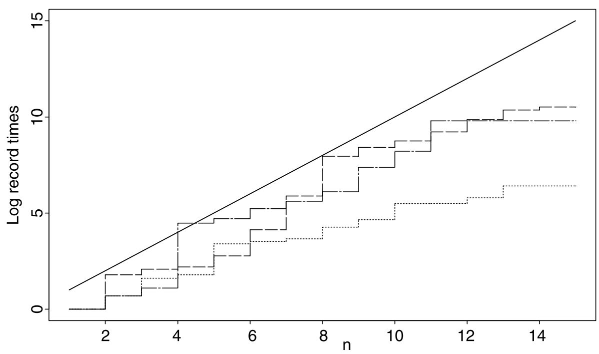


**Figure 5.4.15** *The logarithmic record times of* 100 000 *iid standard normal rvs. Three sample paths are given. The straight line indicates the ideal asymptotic behaviour of these record times; see Theorem 5.4.13.*

## 5.4.4 Invariance Principle for Maxima

In Section 5.4.2 we embedded the sequence of the sample maxima $(M_n)$ in a continuous–time $F$–extremal process $Y$. This was advantageous because we could make use of the hidden Poisson structure of $Y$ to derive limit results about records and record times. In the sequel we are interested in the question:

*How can we link the weak convergence of sample maxima with the weak convergence of point processes?*

Intuitively, we try to translate the problem about the extremes of the sequence $(X_n)$ for some particular df $F$ into a question about the extremes of an iid sequence with common extreme value distribution $H$. Since there are only three standard extreme value distributions $H$, but infinitely many dfs $F$ in the maximum domain of attraction of $H$ ($F \in \mathrm{MDA}(H)$) this is quite a promising approach.

To make this idea precise suppose that $F$ belongs to the maximum domain of attraction of $H$, i.e. there exist constants $d_n$ and $c_n > 0$ such that

$$c_n^{-1}\left(M_n - d_n\right) \xrightarrow{d} H\,, \quad n \to \infty\,, \tag{5.37}$$

where $H$ is one of the standard extreme value distributions (Weibull, Fréchet, Gumbel) as introduced in Definition 3.2.6. Set

$$Y_n(t) = \begin{cases} c_n^{-1}\left(M_{[nt]} - d_n\right) & \text{if} \quad t \geq n^{-1}\,, \\ c_n^{-1}\left(X_1 - d_n\right) & \text{if} \quad 0 < t < n^{-1}\,, \end{cases}$$

where $[x]$ denotes the integer part of $x$. Recall the notion of weak convergence in the Skorokhod space $\mathbb{D}\,(0,\infty)$ from Appendix A2.

The processes $(Y_n)$ obey a result which parallels very much the Donsker invariance principle for sums of iid random variables; see Theorem 2.4.4.

**Theorem 5.4.16** (Invariance principle for maxima)
*Let $H$ be one of the extreme value distributions and $Y = (Y(t))_{t>0}$ the corresponding $H$–extremal process. Then the relation*

$$Y_n \xrightarrow{d} Y\,, \quad n \to \infty\,,$$

*holds in $\mathbb{D}\,(0,\infty)$ if and only if (5.37) is satisfied.*

**Sketch of the proof.** For a detailed proof see Resnick [530], Proposition 4.20. Take $t = 1$. Then $Y_n \xrightarrow{d} Y$ obviously implies (5.37), i.e.

$$Y_n(1) = c_n^{-1}\left(M_n - d_n\right) \xrightarrow{d} Y(1)\,,$$

where $Y(1)$ has distribution $H$.
Now suppose that (5.37) holds. This is known to be equivalent to

$$n\overline{F}\left(c_n x + d_n\right) \to -\ln H(x) \tag{5.38}$$

on the support $S$ of $H$; see Proposition 3.1.1. We define

$$\xi_{n,j} = \begin{cases} c_n^{-1}\left(X_j - d_n\right) & \text{if } c_n^{-1}\left(X_j - d_n\right) \in S\,, \\ \inf S & \text{otherwise}\,, \end{cases}$$

and

$$\mu(a,b] = \ln H(b) - \ln H(a)$$

for $(a,b] \subset S$. Topologising the state space $E$ in the right way, (5.38) just means (see Proposition A2.12) that

$$nP\left(\xi_{n,1} \in \cdot\right) \xrightarrow{v} \mu(\cdot) \tag{5.39}$$

on the Borel sets of $S$. Now define

$$N_n = \sum_{k=1}^{\infty} \varepsilon_{(n^{-1}k,\xi_{n,k})}\,, \qquad N = \sum_{k=1}^{\infty} \varepsilon_{(t_k,j_k)}\,,$$

where $N$ is PRM$(|\cdot| \times \mu)$ on $\mathbb{R}_+ \times S$ and $|\cdot|$ denotes Lebesgue measure. Then Theorem 5.2.4 and (5.39) imply that $N_n \xrightarrow{d} N$. Recall the definition of the mapping $\widetilde{T}_1$ from (5.25). If we restrict ourselves to path spaces in which both $N_n$ and $N$ live then $\widetilde{T}_1$ can be shown to be a.s. continuous. An application of the continuous mapping theorem (see Theorem A2.6) yields that

$$\widetilde{T}_1\left(N_n\right) = \bigvee_{n^{-1}k\le\cdot} \xi_{n,k} \quad \xrightarrow{d} \quad \widetilde{T}_1(N) = \bigvee_{t_k\le\cdot} j_k$$

in $\mathbb{D}\left(0,\infty\right)$. Note that in view of Proposition 5.4.4

$$Y(\cdot) \stackrel{d}{=} \bigvee_{t_k\le\cdot} j_k\,.$$

Moreover, one can show that in $\mathbb{D}\left(0,\infty\right)$ the relation

$$Y_n(\cdot) - \bigvee_{n^{-1}k\le\cdot} \xi_{n,k} \xrightarrow{P} 0$$

is valid. This proves that $Y_n \xrightarrow{d} Y$. □

**Remark.** In the course of the proof above we left out all messy details. We also swept certain problems under the carpet which are related to the fact that the $\xi_{n,k}$ can be concentrated on the whole real line. This requires for instance a special treatment for $F \in \mathrm{MDA}(\Phi_\alpha)$ (equivalently, $\overline{F} \in \mathcal{R}_{-\alpha}$) since a regular variation assumption on the right tail does naturally not influence the left tail of the distribution. Read Resnick [530], Section 4.4.2! □

This invariance principle encourages one to work with the $H$–extremal process $Y$ instead of the process $Y_n$ of sample maxima for $F$ in the maximum domain of attraction of $H$. Thus, in an asymptotic sense, we are allowed to work with the distribution of $Y$ instead of the one for $Y_n$. We stop here the discussion and refer to Section 5.5 where the weak convergence of extremal processes and of the underlying point processes is used to derive limit results about the upper extremes of dependent sequences.

### Notes and Comments

We have seen in this section that point process techniques are very elegant tools for dealing with extremal properties of sequences of iid rvs. They allow us to derive deep results about the structure of extremal processes, of their jump times, about records, record times, exceedances etc. The basic idea is always to find the right point process, to show weak convergence to a PRM and possibly to apply the continuous mapping theorem in a suitable way.

The elegance of the method is one side of the coin. We have seen from the above outline of proofs that we have to be familiar with many tools from functional analysis, measure theory and stochastic processes. In particular, the proof of the a.s. continuity of the $\widetilde{T}$–mappings is never trivial and requires a deep understanding of stochastic processes. The a.s. continuity of the $\widetilde{T}$–mappings was treated for instance in Mori and Oodaira [468], Resnick [529], Serfozo [577].

Excellent references for extreme value theory in the context of point processes are Falk, Hüsler and Reiss [225], Leadbetter, Lindgren and Rootzén [418], Reiss [527] and Resnick [529, 530]. We followed closely the last source in our presentation.

In Section 6.2.5 we consider records as an exploratory statistical tool. There we also give some further references to literature on records.

## 5.5 Some Extreme Value Theory for Linear Processes

In Sections 4.4 and 5.3.2 we found conditions which ensured that the extremal behaviour of the strictly stationary sequence $(X_n)$ is the same as that of an associated iid sequence $(\widetilde{X}_n)$, i.e. an iid sequence with the same common df $F$ as $X = X_0$. Intuitively, those conditions $D(u_n)$ and $D'(u_n)$ guaranteed that high level exceedances by the sequence $(X_n)$ were separated in time; i.e. clustering of extremes was avoided. This will change dramatically for the special class of strictly stationary sequences which we consider in this section. We suppose that $(X_n)$ has representation as a *linear process*, i.e.

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

where the *noise* sequence or the *innovations* $(Z_n)$ are iid and the $\psi_j$ are real numbers to be specified later. For simplicity we set $Z = Z_0$. Here we study linear processes from the point of view of extreme value theory. In Chapter 7 they are reconsidered from the point of view of time series analysis. Linear processes are basic in classical time series analysis. In particular, every ARMA process is linear, see Example 7.1.1, and most interesting Gaussian stationary sequences have a linear process representation.

Again we are interested in exceedances of a given deterministic sequence of thresholds $(u_n)$ by the process $(X_n)$, and in the joint distribution of a finite number of upper order statistics of a sample $X_1, \ldots, X_n$. We compare sequences of sample maxima for the noise $(Z_n)$, the stationary sequence $(X_n)$ and an associated iid sequence $(\widetilde{X}_n)$. As usual, $(M_n)$ denotes the sequence of the sample maxima of $(X_n)$.

### 5.5.1 Noise in the Maximum Domain of Attraction of the Fréchet Distribution $\Phi_\alpha$

We assume that $Z$ satisfies the following condition:

$$\overline{F}_Z(x) = P(Z > x) = \frac{L(x)}{x^\alpha}, \quad x > 0, \tag{5.40}$$

for some $\alpha > 0$ and a slowly varying function $L$, i.e. $L(x)x^{-\alpha}$ is regularly varying with index $-\alpha$; see Appendix A3. By Theorem 3.3.7 this is equivalent to $Z \in \mathrm{MDA}(\Phi_\alpha)$ where

$$\Phi_\alpha(x) = e^{-x^{-\alpha}}, \quad x > 0,$$

denotes the standard Fréchet distribution which is one of the extreme value distributions; see Definition 3.2.6. Moreover, we assume that the tails are balanced in the sense that

$$\lim_{x\to\infty} \frac{P(Z > x)}{P(|Z| > x)} = p, \quad \lim_{x\to\infty} \frac{P(Z \le -x)}{P(|Z| > x)} = q, \tag{5.41}$$

for some $0 < p \le 1$ and such that $p + q = 1$. Thus we can combine (5.40) and (5.41):

$$\overline{F}_Z(x) = \frac{L(x)}{x^\alpha}, \quad x > 0, \quad F_Z(-x) \sim \frac{q}{p}\frac{L(x)}{x^\alpha}, \quad x \to \infty. \tag{5.42}$$

We also suppose that

$$\sum_{j=-\infty}^{\infty} |\psi_j|^\delta < \infty \quad \text{for some } 0 < \delta < \min(\alpha, 1). \tag{5.43}$$

This condition implies the absolute a.s. convergence of the linear process representation of $X_n$ for every $n$; see also the discussion in Section 7.2. Note that the conditions here are very much like in Sections 7.2–7.5, but there we restrict ourselves to symmetric $\alpha$–stable ($s\alpha s$) $Z_n$ for some $\alpha < 2$. In that case,

$$\overline{F}_Z(x) \sim \frac{c}{x^\alpha}\,, \quad x \to \infty\,, \qquad \lim_{x\to\infty} \frac{P(Z > x)}{P(|Z| > x)} = \frac{1}{2}\,,$$

hence (5.40) and (5.41) are naturally satisfied. We also mention that, if $\alpha < 2$, then the conditions (5.40) and (5.41) imply that $Z$ has a distribution in the domain of attraction of an $\alpha$–stable law; see Section 2.2.

We plan to reduce the study of the extremes of $(X_n)$ to the study of the extremes of the iid sequence $(Z_n)$. We choose the normalisation

$$c_n = (1/\overline{F}_Z)^{\leftarrow}(n)\,, \tag{5.44}$$

where $f^{\leftarrow}$ denotes the generalised inverse of the function $f$. By (5.40) this implies that $\overline{F}_Z(c_n) \sim n^{-1}$. Then we also know that

$$c_n = n^{1/\alpha} L_1(n)$$

for a slowly varying function $L_1$. Moreover, from Theorem 3.3.7 we are confident of the limit behaviour

$$c_n^{-1} \max(Z_1, \ldots, Z_n) \xrightarrow{d} \Phi_\alpha\,.$$

So we may hope that $c_n$ is also the right normalisation for the maxima $M_n$ of the linear process $(X_n)$.

We first embed $(c_n^{-1} X_k)_{k\geq 1}$ in a point process and show its weak convergence to a function of a PRM. This is analogous to the proof of Theorem 5.4.16. Then we can proceed as in the iid case to derive information about the extremes of the sequence $(X_n)$.

**Theorem 5.5.1** (Weak convergence of the point processes of the embedded linear process)
*Let* $\sum_{k=1}^\infty \varepsilon_{(t_k, j_k)}$ *be* $\mathrm{PRM}(|\cdot| \times \mu)$ *on* $\mathbb{R}_+ \times E$, *where* $E = [-\infty, \infty]\backslash\{0\}$, $|\cdot|$ *is Lebesgue measure and the measure* $\mu$ *on the Borel sets of* $E$ *has density*

$$\alpha x^{-\alpha-1} I_{(0,\infty]}(x) + qp^{-1}\alpha(-x)^{-\alpha-1} I_{[-\infty,0)}(x)\,, \quad x \in \mathbb{R}\,. \tag{5.45}$$

*Suppose the conditions* (5.42) *and* (5.43) *are satisfied. Then*

$$\sum_{k=1}^\infty \varepsilon_{\left(n^{-1}k,\, c_n^{-1} X_k\right)} \xrightarrow{d} \sum_{k=1}^\infty \sum_{i=-\infty}^\infty \varepsilon_{(t_k, \psi_i j_k)}\,, \quad n \to \infty\,,$$

*in* $M_p(\mathbb{R}_+ \times E)$.

**Sketch of the proof.** For a complete proof we refer to Davis and Resnick [160]; see also Resnick [530], Section 4.5.

We notice that condition (5.42) is equivalent to

$$nP\left(c_n^{-1} Z \in \cdot\right) \xrightarrow{v} \mu(\cdot)$$

on the Borel sets of $E$, where the measure $\mu$ on $E$ is determined by (5.45). This holds by virtue of Proposition A2.12 and since, as $n \to \infty$,

$$nP\left(c_n^{-1}Z > x\right) \to x^{-\alpha} \quad \text{and} \quad nP\left(c_n^{-1}Z \le -x\right) \to qp^{-1}x^{-\alpha}\,, \quad x > 0\,.$$

It is then a consequence of Theorem 5.2.4 (see also the proof of Theorem 5.4.16) that

$$\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k,c_n^{-1}Z_k\right)} \xrightarrow{d} \sum_{k=1}^{\infty} \varepsilon_{(t_k,j_k)}\,, \quad n \to \infty\,, \tag{5.46}$$

in $M_p\left(\mathbb{R}_+ \times E\right)$, where the limit is PRM$(|\cdot| \times \mu)$.

The process $X_n = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j}$ is a (possibly infinite) moving average of the iid noise $(Z_n)$. A naive argument suggests that we should first consider finite moving averages

$$X_n^{(m)} = \sum_{j=-m}^{m} \psi_j Z_{n-j}\,, \quad n \in \mathbb{Z}\,,$$

for a fixed integer $m$, then apply a Slutsky argument (see Appendix A2.5) and let $m \to \infty$.

For simplicity we restrict ourselves to the case $m = 1$ and we further assume that $(X_n^{(1)})$ is a moving average process of order 1 (MA(1)):

$$X_n^{(1)} = Z_n + \psi_1 Z_{n-1}\,, \quad n \in \mathbb{Z}\,.$$

We embed $(X_n^{(1)})$ in a point process which will be shown to converge weakly. We notice that $X_n^{(1)}$ is just a functional of the 2–dimensional vector

$$\mathbf{Z}_n = (Z_{n-1}, Z_n) = Z_{n-1}\mathbf{e}_1 + Z_n\mathbf{e}_2\,,$$

and so it is natural to consider the point process

$$\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k,c_n^{-1}\mathbf{Z}_k\right)}\,. \tag{5.47}$$

By some technical arguments it can be shown that (5.47) has the same weak limit behaviour as

$$\sum_{k=1}^{\infty} \left(\varepsilon_{\left(n^{-1}k,c_n^{-1}Z_k\mathbf{e}_1\right)} + \varepsilon_{\left(n^{-1}k,c_n^{-1}Z_k\mathbf{e}_2\right)}\right).$$

Then an application of (5.46) and the continuous mapping theorem (see Theorem A2.6) yield that the point processes (5.47) converge weakly to

$$\sum_{k=1}^{\infty} \left(\varepsilon_{(t_k, j_k \mathbf{e}_1)} + \varepsilon_{(t_k, j_k \mathbf{e}_2)}\right) .$$

Since we want to deal with the MA(1) process $(X_n^{(1)})$ we have to stick the coordinates of $\mathbf{Z}_n$ together, and this is again guaranteed by an a.s. continuous mapping $\widetilde{T}_2$, say, acting on the point processes:

$$\begin{aligned}
&\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1}(Z_k + \psi_1 Z_{k-1})\right)} \\
&\quad = \quad \widetilde{T}_2 \left( \sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1}\mathbf{Z}_k\right)} \right) \\
&\quad \approx \quad \widetilde{T}_2 \left( \sum_{k=1}^{\infty} \left( \varepsilon_{\left(n^{-1}k, c_n^{-1} Z_k \mathbf{e}_1\right)} + \varepsilon_{\left(n^{-1}k, c_n^{-1} Z_k \mathbf{e}_2\right)} \right) \right) \\
&\quad \stackrel{d}{\to} \quad \widetilde{T}_2 \left( \sum_{k=1}^{\infty} \left( \varepsilon_{(t_k, j_k \mathbf{e}_1)} + \varepsilon_{(t_k, j_k \mathbf{e}_2)} \right) \right) \\
&\quad = \quad \sum_{k=1}^{\infty} \left( \varepsilon_{(t_k, j_k)} + \varepsilon_{(t_k, \psi_1 j_k)} \right) .
\end{aligned}$$

Similar arguments prove that

$$\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1} X_k^{(m)}\right)} \stackrel{d}{\to} \sum_{k=1}^{\infty} \sum_{i=-m}^{m} \varepsilon_{(t_k, \psi_i j_k)}$$

for every $m \geq 1$, and a Slutsky argument as $m \to \infty$ concludes the proof. $\square$

It is now our goal to consider some applications of this theorem. We suppose throughout that the assumptions of Theorem 5.5.1 are satisfied.

### Extremal Processes and Limit Distributions of Maxima

Analogously to iid sample maxima we consider the continuous–time process

$$Y_n(t) = \begin{cases} c_n^{-1} M_{[nt]} & \text{if} \quad t \geq n^{-1} , \\ c_n^{-1} X_1 & \text{if} \quad 0 < t < n^{-1} , \end{cases}$$

which is constructed from the sample maxima

$$M_n = \max(X_1, \ldots, X_n) , \quad n \geq 1 .$$

Note that $M_n$ is now the maximum of $n$ dependent rvs. Define

$$\psi_+ = \max_j \, (\psi_j \vee 0)\,, \quad \psi_- = \max_j \, ((-\psi_j) \vee 0)\,. \tag{5.48}$$

Recall the definition of the mapping $\widetilde{T}_1$ from (5.25): for a point process $\sum_{k=1}^{\infty} \varepsilon_{(r_k, s_k)}$ set

$$\widetilde{T}_1 \left( \sum_{k=1}^{\infty} \varepsilon_{(r_k,s_k)} \right) = \sup \{ s_k : r_k \le \cdot \}\,.$$

It is an a.s. continuous mapping from $M_p(\mathbb{R}_+ \times E)$ to $\mathbb{D}\,(0,\infty)$. This relation, Theorem 5.5.1 and the continuous mapping theorem yield that

$$\begin{aligned}
\widetilde{T}_1 \left( \sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1} X_k\right)} \right) & \stackrel{d}{=} & Y_n(\cdot) \quad \stackrel{d}{\to} \\
\widetilde{T}_1 \left( \sum_{k=1}^{\infty} \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, \psi_i j_k)} \right) & = & \bigvee_{t_k \le \cdot} \left( \bigvee_{i=-\infty}^{\infty} \psi_i j_k \right) \\
& = & \bigvee_{t_k \le \cdot} \Big( (\psi_+ j_k) \vee (-\psi_- j_k) \Big) = Y(\cdot)\,.
\end{aligned}$$

The process $Y$ defined thus is indeed an extremal process (see Definition 5.4.3 and Proposition 5.4.4) since $Y = \widetilde{T}_1(\widetilde{N})$ where

$$\widetilde{N} = \sum_{k=1}^{\infty} \varepsilon_{(t_k, \psi_+ j_k)} + \sum_{k=1}^{\infty} \varepsilon_{(t_k, -\psi_- j_k)}\,,$$

i.e. $\widetilde{N}$ is a PRM with mean measure of $(0,t] \times (x,\infty)$ equal to

$$t\left(\psi_+^{\alpha} + \psi_-^{\alpha} q p^{-1}\right) x^{-\alpha} \quad \text{for} \quad t > 0\,, x > 0\,.$$

By the definition of a PRM, for $t > 0$, $x > 0$,

$$\begin{aligned}
P\left(Y(t) \le x\right) & = & P\left( \widetilde{N}\Big((0,t] \times (x,\infty)\Big) = 0 \right) \\
& = & \exp\left\{ -E\widetilde{N}\Big((0,t] \times (x,\infty)\Big) \right\} \\
& = & \exp\left\{ -t\left(\psi_+^{\alpha} + \psi_-^{\alpha} q p^{-1}\right) x^{-\alpha} \right\}\,.
\end{aligned}$$

Summarising the facts above we obtain an invariance principle for sample maxima which in the iid case is analogous to Theorem 5.4.16:

**Theorem 5.5.2** (Invariance principle for the maxima of a linear process with noise in MDA($\Phi_\alpha$))
*Assume either $\psi_+ p > 0$ or $\psi_- q > 0$, that the conditions* (5.42) *and* (5.43) *hold and let* $(c_n)$ *be defined by* (5.44)*. Then*

$$Y_n \xrightarrow{d} Y\,, \quad n \to \infty\,,$$

*where $Y$ is the extremal process generated by the extreme value distribution*

$$\Phi_\alpha^{\psi_+^\alpha + \psi_-^\alpha qp^{-1}}(x) = \exp\left\{-\left(\psi_+^\alpha + \psi_-^\alpha qp^{-1}\right)x^{-\alpha}\right\}\,, \quad x > 0\,. \qquad \square$$

**Remarks.** 1) For $(X_n)$ iid, $\psi_i = 0$ for $i \neq 0$ and $\psi_0 = 1$. Then Theorem 5.5.2 degenerates into the case of a $\Phi_\alpha$–extremal process $Y$.

2) The above method can be extended to get joint convergence of the processes generated by a finite number of upper extremes in the sample $X_1, \ldots, X_n$. $\square$

**Corollary 5.5.3** (Limit laws for the maxima of a linear process with noise in MDA($\Phi_\alpha$))
*Assume that $Z \in \mathrm{MDA}(\Phi_\alpha)$ for some $\alpha > 0$ and choose $(c_n)$ according to* (5.44)*. Then*

$$c_n^{-1} \max\left(Z_1, \ldots, Z_n\right) \xrightarrow{d} \Phi_\alpha\,, \tag{5.49}$$

*and, under the conditions of Theorem* 5.5.2*,*

$$c_n^{-1} M_n \xrightarrow{d} \Phi_\alpha^{\psi_+^\alpha + \psi_-^\alpha qp^{-1}}\,. \tag{5.50}$$

*Moreover, let $(\widetilde{X}_n)$ be an iid sequence associated with $(X_n)$. Then*

$$c_n^{-1} \widetilde{M}_n \xrightarrow{d} \Phi_\alpha^{\|\psi\|_\alpha^\alpha}\,, \tag{5.51}$$

*where*

$$\|\psi\|_\alpha^\alpha = \sum_{j=-\infty}^{\infty} |\psi_j|^\alpha \left(I_{\{\psi_j > 0\}} + qp^{-1} I_{\{\psi_j < 0\}}\right).$$

**Proof.** (5.49) and (5.50) follow from Theorem 5.5.2 and the fact that

$$Y_n(1) \stackrel{d}{=} c_n^{-1} M_n \xrightarrow{d} Y(1).$$

(5.51) is a consequence of (5.49), taking into consideration (see Lemma A3.26) that

$$P\left(\sum_{j=-\infty}^{\infty} \psi_j Z_j > x\right) \sim \|\psi\|_\alpha^\alpha P(|Z| > x)\,. \qquad \square$$

The latter relation suggests that classical estimators for the tail index $\alpha$ might also work for the tail of $X_t$. This is unfortunately not the case; see for instance Figure 5.5.4.

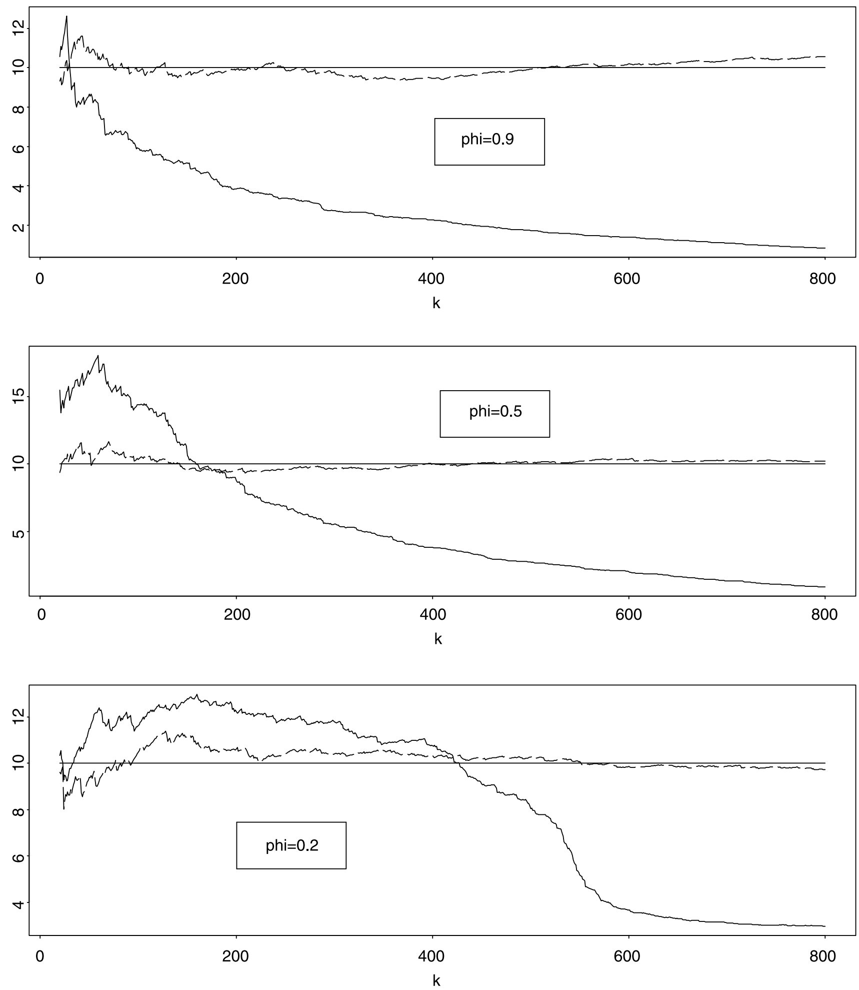


**Figure 5.5.4** *A comparative study of the Hill–plots for* $1\,000$ *iid simulated data from an* $\mathrm{AR}(1)$ *process* $X_t = \phi X_{t-1} + Z_t$, $\phi \in \{0.9, 0.5, 0.2\}$. *The noise sequence* $(Z_t)$ *comes from a symmetric distribution with exact Pareto tail* $P(Z > x) = 0.5x^{-10}$, $x \geq 1$. *According to Lemma A3.26,* $P(X > x) \sim c\,x^{-10}$. *The solid line corresponds to the Hill estimator of the* $X_t$ *as a function of the* $k$ *upper order statistics. The dotted line corresponds to the Hill estimator of the residuals* $\widehat{Z}_t = X_t - \widehat{\phi} X_{t-1}$, *where* $\widehat{\phi}$ *is the Yule–Walker estimator of* $\phi$. *Obviously, the Hill estimator of the residuals yields much more accurate values. These figures indicate that the Hill estimator for correlated data has to be used with extreme care. Even for* $\phi = 0.2$ *the Hill estimator of the* $X_t$ *cannot be considered as a satisfactory tool for estimating the index of regular variation. The corresponding theory for the Hill estimator of linear processes can be found in Resnick and Stărică* [535, 537].

**Exceedances**

Theorem 5.5.1 also allows us to derive results about the observations $X_k/c_n$ exceeding a given threshold $x$, or equivalently about the linear process $(X_k)$ exceeding the threshold $u_n = c_n x$. Without loss of generality we will assume that $|\psi_j| \le 1$ for all $j$.

Applying Theorem 5.5.1 and the continuous mapping theorem we find that the point process of points with ordinates bigger than $x > 0$ converges as $n \to \infty$. Thus let

$$E_x^+ = (x,\infty)\,, \quad E_x^- = (-\infty,-x)\,, \quad E_x = E_x^+ \cup E_x^-\,, \quad x>0\,;$$

then

$$\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k,c_n^{-1}X_k\right)}\left(\cdot \cap \mathbb{R}_+ \times E_x^+\right) \quad \stackrel{d}{\to} \quad \sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty} \varepsilon_{(t_k,\psi_i j_k)}\left(\cdot \cap \mathbb{R}_+ \times E_x^+\right) \tag{5.52}$$

in $M_p(\mathbb{R}_+ \times E_x^+)$. We can interpret this as weak convergence of the point processes of exceedances of $xc_n$ by $(X_k)$. We need the following auxiliary result.

**Lemma 5.5.5** *The following relation holds in $M_p(\mathbb{R}_+ \times E_x)$ :*

$$N_1 = \sum_{k=1}^{\infty} \varepsilon_{(t_k,j_k)} \stackrel{d}{=} N_2 = \sum_{k=1}^{\infty} \varepsilon_{(\Gamma_k,J_k)}\,,$$

*where $(\Gamma_k)$ is the sequence of points of a homogeneous Poisson process on $\mathbb{R}_+$ with intensity $\lambda = p^{-1}x^{-\alpha}$, independent of the iid sequence $(J_k)$ with common density*

$$\begin{aligned} g(y) &= \left(\alpha y^{-\alpha-1} I_{(x,\infty)}(y) + qp^{-1}\alpha(-y)^{-\alpha-1} I_{(-\infty,-x)}(y)\right) p x^{\alpha} \\ &= f(y)\lambda^{-1}\,, \quad y \in \mathbb{R}\,. \end{aligned}$$

**Proof.** It suffices to show that the Laplace functionals of the point processes $N_1$ and $N_2$ coincide; see Example 5.1.8. Since $N_1$ is PRM$(|\cdot| \times \mu)$ on $\mathbb{R}_+ \times E_x$ we have by Example 5.1.11 that

$$\Psi_{N_1}(h) = \exp\left\{ -\int_{\mathbb{R}_+}\int_{\mathbb{R}} \left(1 - e^{-h(t,z)}\right) f(z)dzdt \right\}.$$

On the other hand, conditioning on $(\Gamma_k)$ and writing

$$h_1(t) = \int_{\mathbb{R}} \exp\{-h(t,z)\}\, g(z)dz = E\exp\{-h(t,J_1)\}\,, \quad t>0\,,$$

we obtain

$$\begin{aligned}\Psi_{N_2}(h) &= E\exp\left\{-\int_{\mathbb{R}_+\times\mathbb{R}} h\,dN_2\right\}\\ &= E\exp\left\{-\sum_{k=1}^{\infty} h(\Gamma_k, J_k)\right\}\\ &= E\prod_{k=1}^{\infty} h_1(\Gamma_k)\\ &= E\exp\left\{\sum_{k=1}^{\infty}\ln h_1(\Gamma_k)\right\}.\end{aligned}$$

The rvs $\Gamma_k$ are the points of a homogeneous Poisson process with intensity $\lambda = p^{-1}x^{-\alpha}$. This and Lemma 5.1.12 yield

$$\begin{aligned}\Psi_{N_2}(h) &= E\exp\left\{-\lambda\int_{\mathbb{R}_+}(1-h_1(t))\,dt\right\}\\ &= \exp\left\{-\int_{\mathbb{R}_+}\int_{\mathbb{R}}\left(1-e^{-h(t,z)}\right)f(z)dzdt\right\}.\end{aligned}$$

This proves the lemma. □

Therefore the limit process in (5.52) has representation

$$\sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty}\varepsilon_{(t_k,\psi_i j_k)} \stackrel{d}{=} \sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty}\varepsilon_{(\Gamma_k,\psi_i J_k)}$$

in $M_p\left(\mathbb{R}_+\times(x,\infty)\right)$. Finally, we define the iid rvs $\xi_k = \operatorname{card}\{i : \psi_i J_k > x\}$. Now we can represent the limit process in (5.52) as a point process on $\mathbb{R}_+$:

$$N_\infty = \sum_{k=1}^{\infty}\xi_k\varepsilon_{\Gamma_k}$$

for independent $(\Gamma_k)$ and $(\xi_k)$. For any Borel set $A$ in $\mathbb{R}_+$ this means that

$$N_\infty(A) = \sum_{k=1}^{\infty}\xi_k\varepsilon_{\Gamma_k}(A) = \sum_{k:\Gamma_k\in A}\xi_k\,,$$

i.e. $N_\infty$ is a multiple point process with iid *multiplicities* or *cluster sizes* $\xi_k$. In particular, it is a compound Poisson process as defined in Example 5.1.15. This is completely different from the point processes of exceedances of iid or weakly dependent rvs (cf. Section 5.3) where the limit is homogeneous Poisson, hence simple. Thus the special dependence structure of linear processes yields clusters in the point process of exceedances.

**Example 5.5.6** (AR(1) process)
We consider the AR(1) process $X_t = \varphi X_{t-1} + Z_t\,, t \in \mathbb{Z}$, for some $\varphi \in (0,1)$. It has a linear process representation

$$X_t = \sum_{j=0}^{\infty} \varphi^j Z_{t-j}\,, \quad t \in \mathbb{Z}.$$

The iid cluster sizes $\xi_k$ have the following distribution:

$$\pi_0 = P\,(\xi_1 = 0) = P(J_1 \le x, \varphi J_1 \le x, \ldots) = P\,(J_1 \le x) = q\,, \quad x > 0\,,$$

and for $\ell \ge 1$,

$$\begin{aligned} \pi_\ell &= P(\xi_1 = \ell) \\ &= P(J_1 > x, \ldots, \varphi^{\ell-1} J_1 > x, \varphi^\ell J_1 \le x) \\ &= P(\varphi^{\ell-1} J_1 > x, \varphi^\ell J_1 \le x) \\ &= p\varphi^{\alpha(\ell-1)}(1-\varphi^\alpha). \end{aligned}$$

□

**Example 5.5.7** (MA(1) process)
We consider the MA(1) process $X_t = Z_t + \theta Z_{t-1}$, $t \in \mathbb{Z}$. Assume first $\theta > 0$. Direct calculation yields

$$\begin{aligned} P\,(\xi_1 = 0) &= q \\ P\,(\xi_1 = 1) &= (1 - \theta^\alpha \wedge 1)\,p \\ P\,(\xi_1 = 2) &= (\theta^\alpha \wedge 1)\,p\,. \end{aligned}$$

Thus the cluster sizes $\xi_k$ may assume the values 0, 1 and 2 with positive probability for $\theta \in (0,1)$, whereas for $\theta > 1$ only the values 0 and 2 may occur.
Now assume $\theta < 0$, then

$$\begin{aligned} P(\xi_1 = 0) &= (1 - (|\theta|^\alpha \wedge 1))\; q\,, \\ P(\xi_1 = 1) &= p + (|\theta|^\alpha \wedge 1)\; q\,. \end{aligned}$$

Thus the cluster sizes $\xi_k$ may assume only the values 0 and 1 for $\theta \le -1$, whereas for $\theta \in (-1,0)$, $\xi_k = 1$ a.s.

This means (in an asymptotic sense) that exceedances may only occur in clusters of 2 values if $\theta > 1$, whereas the cluster size may be 1 or 2 for $\theta \in (0,1)$. For $\theta < 0$ the point process of exceedances does not cluster. □

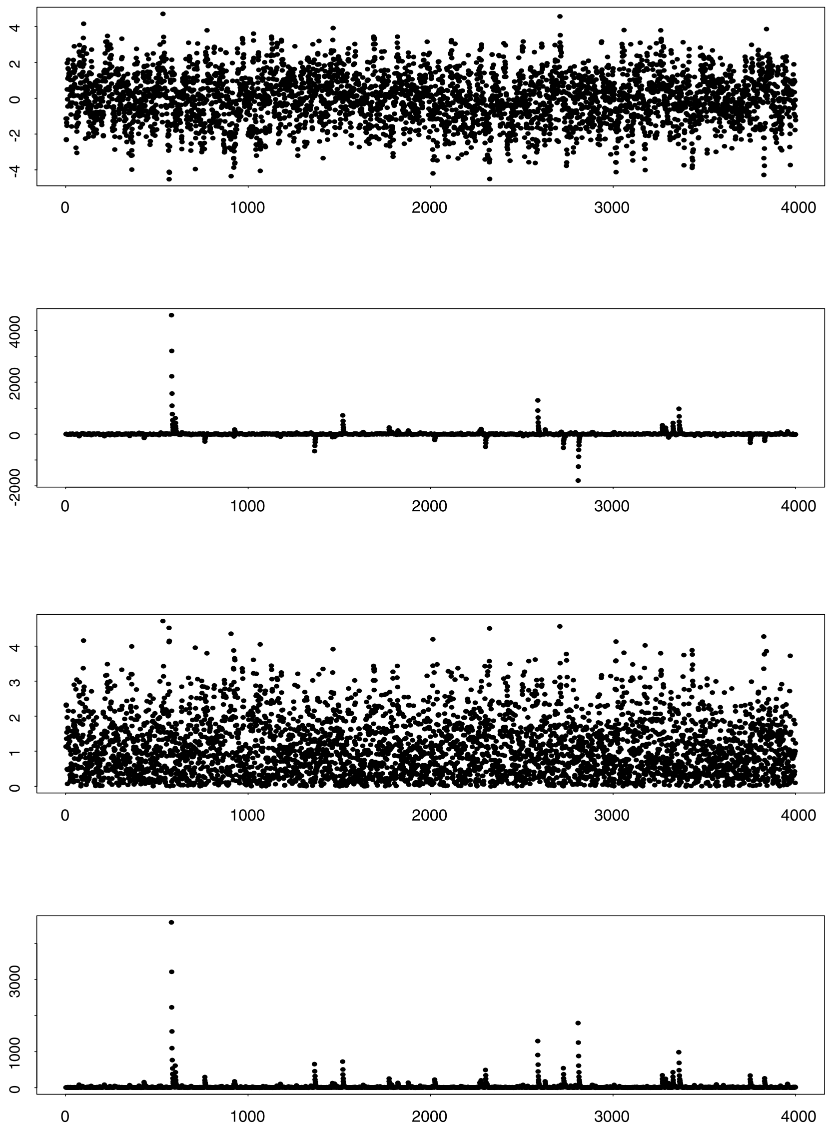


**Figure 5.5.8** *Realisations of the* AR(1) *process* $X_t = 0.7X_{t-1}+Z_t$ (*top two*) *and of the corresponding absolute values* (*bottom two*). *In each pair of figures, the upper one corresponds to iid standard normal noise* $(Z_t)$, *the lower one to iid standard Cauchy noise. In the Cauchy case extremes tend to occur in clusters; see Example 5.5.6. In the Gaussian case clustering effects of extremal values are not present.*

**Maxima and Minima**

We consider the joint limit behaviour of the maxima $(M_n)$ and of the minima

$$W_1 = X_1\,, \quad W_n = \min\left(X_1, \ldots, X_n\right)\,, \quad n \geq 2\,.$$

Choose $x > 0$ and $y < 0$ and write

$$A = (0,1] \times [(-\infty, y) \cup (x, \infty)]\ .$$

Then, by Theorem 5.5.1,

$$\begin{aligned} & P\left(c_n^{-1} M_n \leq x\,, c_n^{-1} W_n > y\right) \\ &= P\left(\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1} X_k\right)}(A) = 0\right) \\ &\to P\left(\sum_{k=1}^{\infty} \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, \psi_i j_k)}(A) = 0\right)\,. \end{aligned} \tag{5.53}$$

We consider the event in (5.53) in detail. Notice that

$$\begin{aligned} & \left\{\mathrm{card}\left\{(k,i) : 0 < t_k \leq 1 \text{ and } (\psi_i j_k < y \text{ or } \psi_i j_k > x)\right\} = 0\right\} \\ = & \left\{\mathrm{card}\left\{k :\ 0 < t_k \leq 1 \text{ and}\right.\right. \\ & \left.\left.(j_k < -x/\psi_- \text{ or } j_k > x/\psi_+ \text{ or } j_k < y/\psi_+ \text{ or } j_k > -y/\psi_-)\right\} = 0\right\}. \end{aligned}$$

Write

$$B = (0,1] \times \Big[\Big(-\infty, (-x/\psi_-) \vee (y/\psi_+)\Big) \cup \Big((x/\psi_+) \wedge (-y/\psi_-)\,, \infty\Big)\Big]\,.$$

Then the right–hand side in (5.53) translates into

$$\begin{aligned} & P\left(\sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k)}(B) = 0\right) \\ = & \exp\Big\{-\mu\Big(\big(-\infty, (-x/\psi_-) \vee (y/\psi_+)\big) \cup \big((x/\psi_+) \wedge (-y/\psi_-)\,, \infty\big)\Big)\Big\} \\ = & \exp\left\{-\left(\left[\psi_+^{\alpha} x^{-\alpha} \vee \psi_-^{\alpha}(-y)^{-\alpha}\right] + q p^{-1}\left[\psi_-^{\alpha} x^{-\alpha} \vee \psi_+^{\alpha}(-y)^{-\alpha}\right]\right)\right\}\,, \end{aligned} \tag{5.54}$$

where $\psi_+, \psi_-$ were defined in (5.48). Now introduce the two–dimensional df

$$G(x_1,x_2)=\begin{cases}\exp\left\{-\psi_+^\alpha x_1^{-\alpha}\right\}\wedge\exp\left\{-\psi_-^\alpha x_2^{-\alpha}\right\} & \text{for } x_1>0,\, x_2>0\,,\\ 0 & \text{otherwise}\,.\end{cases} \tag{5.55}$$

Thus the right–hand side of (5.54) can be written in the form

$$G(x,-y)G^{q/p}(-y,x)\,,$$

and using the relation

$$P\left(c_n^{-1}M_n\le x\,,c_n^{-1}W_n\le y\right)=P\left(c_n^{-1}M_n\le x\right)-P\left(c_n^{-1}M_n\le x\,,c_n^{-1}W_n>y\right)$$

we obtain the following:

**Theorem 5.5.9** (Joint limit distribution of sample maxima and minima of linear process)
*Assume that the conditions of Theorem 5.5.2 hold and let $(c_n)$ be defined by* (5.44)*. Then, for all real $x,y$,*

$$P\left(c_n^{-1}M_n\le x\,,c_n^{-1}W_n\le y\right)\to G(x,\infty)G^{q/p}(\infty,x)-G(x,-y)G^{q/p}(-y,x)\,,$$

*where $G(x,y)$ is defined by* (5.55)*.* □

**Summary**

Assume that $Z\in\mathrm{MDA}(\Phi_\alpha)$, i.e.

$$P(Z>x)=\frac{L(x)}{x^\alpha}\,,\quad x>0\,,$$

for some $\alpha>0$, and that

$$P(Z\le -x)\sim\frac{q}{p}\frac{L(x)}{x^\alpha}\,,\quad x\to\infty\,,$$

for non–negative $p,q$ such that $p+q=1$ and $p>0$. Choose the constants $c_n$ by

$$c_n=(1/\overline{F}_Z)^{\leftarrow}(n)\,.$$

Then

$$c_n^{-1}\max\left(Z_1,\ldots,Z_n\right)\xrightarrow{d}\Phi_\alpha$$

for the Fréchet distribution $\Phi_\alpha(x)=e^{-x^{-\alpha}}$, $x>0$. Moreover, under the conditions of Theorem 5.5.2,

$$c_n^{-1}M_n\xrightarrow{d}\Phi_\alpha^{\psi_+^\alpha+\psi_-^\alpha qp^{-1}}\,,\quad x>0\,,$$

where $\psi_+,\psi_-$ are defined in (5.48). The point process of the exceedances of the threshold $c_nx$ by the linear process $(X_k)$ converges weakly to a compound Poisson point process with iid cluster sizes which depend on the coefficients $\psi_j$.

### 5.5.2 Subexponential Noise in the Maximum Domain of Attraction of the Gumbel Distribution $\Lambda$

In this section we again consider the linear process $X_n = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j}$ driven by iid noise $(Z_n)$ with common df $F_Z$. In contrast to Section 5.5.1 we assume that $F_Z$ belongs to the maximum domain of attraction of the Gumbel distribution

$$\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$$

We know from Section 3.3.3 and Example 3.3.35 that MDA($\Lambda$) contains a wide range of distributions with quite different tail behaviour. Indeed, $F_Z$ may be subexponential (for instance the lognormal distribution), exponential or superexponential (for instance the normal distribution). We found in Example 4.4.9 that fairly general Gaussian linear processes $(X_n)$ exhibit the same asymptotic extremal behaviour as their associated iid sequence $(\widetilde{X}_n)$. This changes dramatically for linear processes with subexponential noise as we have already learnt in Section 5.5.1 for regularly varying $\overline{F}_Z$. A similar statement holds when $F_Z \in \text{MDA}(\Lambda) \cap \mathcal{S}$, where $\mathcal{S}$ denotes the class of distributions $F_Z$ with subexponential positive part $Z^+$; for the definition and properties of $\mathcal{S}$ see Section 1.3.2 and Appendix A3.2.

Before we state the main results for $F_Z \in \text{MDA}(\Lambda) \cap \mathcal{S}$ we introduce some conditions on the coefficients $\psi_j$ and on the distribution $F_Z$. Throughout we suppose that the tail balance condition

$$\lim_{x\to\infty} \frac{P(Z > x)}{P(|Z| > x)} = p, \quad \lim_{x\to\infty} \frac{P(Z \le -x)}{P(|Z| > x)} = q \tag{5.56}$$

holds with $0 < p \le 1$, $p + q = 1$. We also assume

$$\sum_{j=-\infty}^{\infty} |\psi_j|^{\delta} < \infty \quad \text{for some } \delta \in (0,1). \tag{5.57}$$

We have that $E|Z| < \infty$, which follows from the tail balance condition (5.56) and from the fact that $E(Z^+)^{\beta} < \infty$, $\beta > 0$, for $F_Z \in \text{MDA}(\Lambda)$; see Corollary 3.3.32. This and (5.57) guarantee the absolute a.s. convergence of the series $X_n$ for every $n$. Without loss of generality we assume that

$$\max_j |\psi_j| = 1, \tag{5.58}$$

since otherwise we may consider the re–scaled process $X_n / \max_j |\psi_j|$. Then one or more of the $\psi_j$ have absolute value one. The quantities

$$k^+ = \text{card}\{j : \psi_j = 1\}, \quad k^- = \text{card}\{j : \psi_j = -1\} \tag{5.59}$$

are crucial for the extremal behaviour of the sequence $(X_n)$. The above conditions lead to the following result which is analogous to Theorem 5.5.1. Theorem 5.5.10 below is proved in Davis and Resnick [163], Theorem 3.3, in a more general situation.

**Theorem 5.5.10** (Weak convergence of the point processes of the embedded linear process)
*Suppose* $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$. *Then there exist constants* $c_n > 0$ *and* $d_n \in \mathbb{R}$ *such that*

$$n\overline{F}_Z\left(c_n x + d_n\right) \to -\ln \Lambda(x)\,, \quad x \in \mathbb{R}\,. \tag{5.60}$$

*Furthermore, assume that conditions* (5.56)–(5.58) *hold. Then*

$$\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k,\, c_n^{-1}(X_k - d_n)\right)} \quad \xrightarrow{d} \quad k^+ N_1 + k^- N_2$$

*in* $M_p\left(\mathbb{R}_+ \times E\right)$ *with* $E = (-\infty, \infty]$. *Here*

$$N_i = \sum_{k=1}^{\infty} \varepsilon_{(t_{ki}, j_{ki})}\,, \quad i = 1, 2\,,$$

*are two independent* $\mathrm{PRM}(|\cdot| \times \mu_i)$ *on* $\mathbb{R}_+ \times E$, $\mu_1$ *has density* $f_1(x) = e^{-x}$ *and* $\mu_2$ *has density* $f_2(x) = (q/p)e^{-x}$, *both with respect to Lebesgue measure.* □

**Remarks.** 1) If $k^+ > 1$ or $k^- > 1$, the limit point process $k^+N_1 + k^-N_2$ is multiple with constant multiplicities $k^+$, $k^-$. The two independent processes $k^+N_1$ and $k^-N_2$ are due to the contributions of those innovations $Z_n$ for which $\psi_n = 1$ or $\psi_n = -1$.

2) A comparison of Theorems 5.5.10 and 5.5.1 shows that the limit point processes for $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ and $F_Z \in \mathrm{MDA}(\Phi_\alpha)$ are completely different although in both cases $F_Z$ is subexponential. For $F_Z \in \mathrm{MDA}(\Phi_\alpha)$ the limit depends on *all* coefficients $\psi_j$ whereas for $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ only the numbers $k^+$ and $k^-$ defined by (5.59) are of interest. The differences are due to the completely different tail behaviour; $F_Z \in \mathrm{MDA}(\Phi_\alpha)$ implies regular variation of $\overline{F}_Z$, $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ rapid variation of $\overline{F}_Z$; see Corollary 3.3.32. This has immediate consequences for $P(X > x)$; see Appendix A3.3. □

In the sequel we again apply some standard arguments to derive information from Theorem 5.5.10 about the extremal behaviour of the linear process $(X_n)$.

**Extremal Processes and Limit Distribution of Maxima**

Analogously to iid sample maxima we define the process

$$Y_n(t) = \begin{cases} c_n^{-1}\left(M_{[nt]} - d_n\right) & \text{if} \quad t \geq n^{-1}\,, \\ c_n^{-1}\left(X_1 - d_n\right) & \text{if} \quad 0 < t < n^{-1}\,. \end{cases}$$

Let

$$Y(t) = \bigvee_{t_{k1} \leq t} j_{k1} \vee \bigvee_{t_{k2} \leq t} j_{k2} = Y^+(t) \vee Y^-(t)\,, \quad t > 0\,.$$

We use the convention that $\max \emptyset = -\infty$. Then an application of the a.s. continuous mapping $\widetilde{T}_1$ from (5.25), the continuous mapping theorem and Theorem 5.5.10 yield that

$$\begin{aligned} Y_n &= \widetilde{T}_1\left(\sum_{k=1}^{\infty} \varepsilon_{\left(n^{-1}k, c_n^{-1}(X_k - d_n)\right)}\right) \\ &\stackrel{d}{\to} \widetilde{T}_1\left(k^+ \sum_{k=1}^{\infty} \varepsilon_{(t_{k1}, j_{k1})} + k^- \sum_{k=1}^{\infty} \varepsilon_{(t_{k2}, j_{k2})}\right) \\ &= Y \end{aligned}$$

in $\mathbb{D}\,(0,\infty)$. The cadlag processes $Y^+$ and $Y^-$ are independent extremal processes and $Y$ (being the maximum of them) is again an extremal process. Remember that $\sum_{k=1}^{\infty} \varepsilon_{(t_{k1}, j_{k1})}$ is PRM with the mean measure of $(0,t] \times (x,\infty)$ equal to $te^{-x}$ and likewise $\sum_{k=1}^{\infty} \varepsilon_{(t_{k2}, j_{k2})}$ is PRM with the mean measure of $(0,t] \times (x,\infty)$ equal to $t(q/p)e^{-x}$. Thus $Y^+$ is $\Lambda$–extremal and $Y^-$ is $\Lambda^{q/p}$–extremal. Hence $Y = Y^+ \vee Y^-$ is $\Lambda^{1+q/p}$–extremal. Then, for $t > 0$, $x \in \mathbb{R}$,

$$P(Y(t) \leq x) = \exp\left\{-t(1+q/p)e^{-x}\right\} = \exp\left\{-tp^{-1}e^{-x}\right\}\,.$$

**Theorem 5.5.11** (Invariance principle for the maxima of a linear process with noise in $\mathrm{MDA}(\Lambda) \cap \mathcal{S}$)
*Assume that $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ and that conditions (5.56)–(5.58) hold. Choose the constants $c_n$, $d_n$ according to (5.60). Then*

$$Y_n \stackrel{d}{\to} Y\,, \qquad n \to \infty\,,$$

*where $Y$ is the extremal process generated by the extreme value distribution*

$$\Lambda^{p^{-1}}(x) = \exp\left\{-p^{-1}e^{-x}\right\}\,, \quad x \in \mathbb{R}\,. \qquad \square$$

An immediate consequence is the following.

**Corollary 5.5.12** (Limit laws for the maxima of a linear process with noise in $\mathrm{MDA}(\Lambda) \cap \mathcal{S}$)
*Under the conditions of Theorem 5.5.11 the following limit relations hold:*

$$c_n^{-1} \left(\max(Z_1, \ldots, Z_n) - d_n\right) \xrightarrow{d} \Lambda \,, \tag{5.61}$$

$$c_n^{-1} \left(M_n - d_n\right) \xrightarrow{d} \Lambda^{p^{-1}} \,. \tag{5.62}$$

$$c_n^{-1} \left(\widetilde{M}_n - d_n\right) \xrightarrow{d} \Lambda^{k^+ + k^- q p^{-1}} \,. \tag{5.63}$$

**Proof.** (5.61) and (5.62) follow from Theorem 5.5.11, while (5.63) is a consequence of (5.61) taking into consideration (see Lemma A3.27) that

$$P\left(\sum_{j=-\infty}^{\infty} \psi_j Z_j > x\right) \sim (k^+ p + k^- q) P\left(|Z| > x\right) \,. \qquad \square$$

**Exceedances**

For $x \in \mathbb{R}$ the point process of exceedances of $c_n x + d_n$ by the linear process $(X_k)$ is given by

$$N_n(\cdot) = \sum_{k=1}^{\infty} \varepsilon_{n^{-1}k}(\cdot) I_{\left\{c_n^{-1}(X_k - d_n) > x\right\}} \,.$$

As a consequence of Theorem 5.5.10 and of the continuous mapping theorem we conclude that

$$N_n \xrightarrow{d} k^+ \sum_{k=1}^{\infty} \varepsilon_{t_{k1}} I_{\{j_{k1} > x\}} + k^- \sum_{k=1}^{\infty} \varepsilon_{t_{k2}} I_{\{j_{k2} > x\}} = k^+ N^+ + k^- N^- \tag{5.64}$$

in $M_p(\mathbb{R}_+)$. With a glance at the finite–dimensional distributions or at the Laplace functionals it is not difficult to check that $N^+$ and $N^-$ are homogeneous Poisson processes on $\mathbb{R}_+$ with intensity $e^{-x}$ and $(q/p)e^{-x}$, respectively. If $(\Gamma_k^+)$ and $(\Gamma_k^-)$ denote the sequences of the points of $N^+$ and $N^-$ then we obtain the following result from (5.64):

**Theorem 5.5.13** *Suppose that the assumptions of Theorem* 5.5.11 *hold. Then the point processes of exceedances of* $c_n x + d_n$ *by the linear process* $(X_k)$ *converge weakly in* $M_p(\mathbb{R}_+)$ *as* $n \to \infty$*:*

$$\sum_{k=1}^{\infty} \varepsilon_{n^{-1}k} I_{\left\{c_n^{-1}(X_k - d_n) > x\right\}} \xrightarrow{d} \sum_{k=1}^{\infty} \left(k^+ \varepsilon_{\Gamma_k^+} + k^- \varepsilon_{\Gamma_k^-}\right) .$$

*Here* $(\Gamma_{k^+})$ *and* $(\Gamma_{k^-})$ *are the sequences of the points of two independent homogeneous Poisson processes on* $\mathbb{R}_+$ *with corresponding intensities* $e^{-x}$ *and* $(q/p)e^{-x}$*.* $\square$

We notice that the limit process of the point processes of exceedances is the sum of two independent compound Poisson processes where the cluster sizes are just constants $k^+, k^-$. This is in contrast to the iid or weakly dependent stationary case where the limit point process is a (simple) homogeneous Poisson process (see Section 5.3), but it is also different from the situation when $F_Z \in \mathrm{MDA}(\Phi_\alpha)$. In the latter case the limit point process is compound Poisson with random cluster sizes (see Section 5.5.1).

### Maxima and Minima

As in Section 5.5.1 point process methods can be used to derive the joint limit distribution of maxima and minima of linear processes with $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$. The approach is similar to the one in Section 5.5.1. We omit details and simply state a particular result. Let $W_n = \bigwedge_{i=1}^n X_i$ and suppose that $k^- = 0$, i.e. there is no index $j$ with $\psi_j = -1$. Then

$$P\left(c_n^{-1}(M_n - d_n) \le x, \quad c_n^{-1}(W_n + d_n) \ge y\right) \to \Lambda(x)\Lambda^{q/p}(-y)$$

for $x, y > 0$. In general, the limit distribution depends on the fact whether $k^+ = 0$ or $k^- = 0$. For more details see Davis and Resnick [163].

### Summary

Assume that $F_Z \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ with constants $c_n$ and $d_n$ chosen according to Theorem 3.3.26, i.e.

$$c_n^{-1}(\max(Z_1, \ldots, Z_n) - d_n) \xrightarrow{d} \Lambda\,,$$

where $\Lambda$ denotes the Gumbel distribution $\Lambda(x) = e^{-e^{-x}}$, $x \in \mathbb{R}$. Then, under the conditions of Theorem 5.5.11,

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Lambda^{p^{-1}}\,.$$

Furthermore, the point processes of exceedances of the threshold $c_n x + d_n$ by the linear process $(X_k)$ converge weakly to a multiple point process with constant multiplicities.

### Notes and Comments

Asymptotic extreme value theory for linear processes with regularly varying tails is given in Resnick [530], Chapter 4.5. The latter is based on Davis and Resnick [160, 161, 162] who also treat more general aspects of time series analysis, see Chapter 7, and on Rootzén [549] and Leadbetter, Lindgren and Rootzén [418] who consider exceedances of linear processes.

Extremes of linear processes with exponential and subexponential noise variables were treated in Davis and Resnick [163]. Further interesting work in this context is due to Leadbetter and Rootzén [419] and Rootzén [549, 550].

Note that both the present section and Sections 4.4 and 5.3.2 deal with strictly stationary sequences. However, the assumptions and results are of different nature. The central conditions in the present section are regular variation of the tails $\overline{F}_Z$ or subexponentiality of the df $F_Z$. This allows one to embed the linear process in a point process and to derive elegant results which yield much information about the extremal behaviour of a linear process. The assumptions on the tails are much weaker in Sections 4.4 and 5.3.2. In particular, the df does not have to belong to any maximum domain of attraction. Thus more general classes of dfs can be covered. On the other hand, conditions of type $D(u_n)$ or $D'(u_n)$ ensure that we do not go too far away from the iid case. Linear processes seem to allow for "more dependence" in the sequence $(X_n)$ although the kind of dependence is quite specific. We can also compare the different point processes of exceedances. In the case of linear processes we obtain multiple PRM in the limit. This is in contrast to Section 5.3.2, where the limit is a homogeneous Poisson process.