Extreme values of random processes

Lecture Notes

Anton Bovier
Institut für Angewandte Mathematik
Wegelerstrasse 6
53115 Bonn, Germany

Contents

Preface

These lecture notes are compiled for the course “Extremes of stochastic sequences and processes”. that I have been teaching repeatedly at the Technical University Berlin for advanced undergraduate students. This is a one semester course with two hours of lectures per week. I used to follow largely the classical monograph on the subject by Leadbetter, Lindgren, and Róotzen [7], but on the one hand, not all material of that book can be covered, and on the other hand, as time went by I tended to include some extra stuff, I felt that it would be helpfull to have typed notes, both for me, and for the students. As I have been working on some problems and applications of extreme value statistics myself recently, my own experience also will add some personal flavour to the exposition.

The current version is updated for a cours in the Master Programme at Bonn University. THIS is NOT the FINAL VERSION.

Be aware that this is not meant to replace a textbook, and that at times this will be rather sketchy at best.

1

Extreme value distributions of iid sequences

> Sedimentary evidence reveals that the maximal flood levels are getting higher and higher with time.....
>
> *An un-named physicist.*

Records and extremes are not only fascinating us in all areas of live, they are also of tremendous importance. We are constantly interested in knowing how big, how, small, how rainy, how hot, etc. things may possibly be. This it not just vain curiosity, it is and has been vital for our survival. In many cases, these questions, relate to very variable, and highly unpredictable phenomena. A classical example are levels of high waters, be it flood levels of rivers, or high tides of the oceans. Probably everyone has looked at the markings of high waters of a river when crossing some bridge. There are levels marked with dates, often very astonishing for the beholder, who sees these many meters above from where the water is currently standing: looking at the river at that moment one would never suspect this to be likely, or even possible, yet the marks indicate that in the past the river has risen to such levels, flooding its surroundings. It is clear that for settlers along the river, these historical facts are vital in getting an idea of what they might expect in the future, in order to prepare for all eventualities.

Of course, historical data tell us about (a relatively remote) past; what we would want to know is something about the future: given the past observations of water levels, what can we say about what to expect in the future?

A look at the data will reveal no obvious “rules”; annual flood levels appear quite “random”, and do not usually seem to suggest a strict pattern. We will have little choice but to model them as a *stochastic process*, and hence, our predictions on the future will be in nature *statistical*: we will make assertions on the probability of certain events. But note that the events we will be concerned with are rather particular: they will be *rare* events, and relate to the worst things that may happen, in other words, to *extremes*. As a statistician, we will be asked to answer questions like this: What is the probability that for the next 500 years the level of this river will not exceed a certain mark? To answer such questions, an entire branch of statistics, called extreme value statistics, was developed, and this is the subject of this course.

1.1 Basic issues

As usual in statistics, one starts with a set of observations, or "data", that correspond to partial observations of some sequence of events. Let us assume that these events are related to the values of some random variables, X_i, $i \in \mathbb{Z}$, taking values in the real numbers. Problem number one would be to devise from the data (which could be the observation of N of these random variables) a *statistical model* of this process, i.e., a probability distribution of the infinite random sequence $\{X_i\}_{i\in\mathbb{Z}}$. Usually, this will be done partly empirically, partly by prejudice; in particular, the dependence structure of the variables will often be assumed a priori, rather than derived strictly from the data. At the moment, this basic statistical problem will not be our concern (but we will come back to this later). Rather, we will assume this problem to be solved, and now ask for consequences on the properties of extremes of this sequence. Assuming that $\{X_i\}_{i\in\mathbb{Z}}$ is a stochastic process (with discrete time) whose joint law we denote be $\mathbb{P}$, our first question will be about the distribution of its *maximum*: Given $n \in \mathbb{N}$, define the maximum up to time n,

$$M_n \equiv \max_{i=1}^{n} X_i. \tag{1.1}$$

We then ask for the distribution of this new random variable, i.e. we ask what is $\mathbb{P}(M_n \leq x)$? As often, we will be interested in this question particularly when n is large, i.e. we are interested in the asymptotics as $n \uparrow \infty$.

The problem should remind us of a problem from any first course in probability: what is the distribution of $S_n \equiv \sum_{i=1}^{n} X_i$? In both problems, the question has to be changed slightly to receive an answer. Namely, certainly S_n and possibly M_n may tend to infinity, and their distribution may have no reasonable limit. In the case of S_n, we learned that the correct procedure is (most often), to subtract the mean and to divide by $\sqrt{n}$, i.e. to consider the random variable

$$Z_n \equiv \frac{S_n - \mathbb{E}S_n}{\sqrt{n}} \tag{1.2}$$

The most celebrated result of probability theory, the *central limit theorem*, says then that (if, say, X_i are iid and have finite second moments) Z_n converges to a Gaussian random variable with mean zero and variance that of X_1. This result has two messages: there is a natural rescaling (here dividing by the square root of n), and then there is a *universal limiting distribution*, the Gaussian distribution, that emerges (largely) independently of what the law of the variable X_i is. Recall that this is of fundamental importance for statistics, as it suggests a class of distributions, depending on only two parameters (mean and variance) that will be a natural candidate to fit any random variables that are expected to be sums of many independent random variables!

The natural first question about M_n are thus: first, can we rescale M_n in some way such that the rescaled variable converges to a random variable, and second, is there a universal class of distributions that arises as the distribution of the limits? If that is the case, it will again be a great value for statistics! To answer these questions will be our first target.

A second major issue will be to go beyond just the maximum value. Coming back to the marks of flood levels under the bridge, we do not just see one, but a whole bunch of marks. can we say something about their joint distribution? In other words, what is the law of the maximum, the second largest, third largest, etc.? Is there, possibly again a universal law of how this process of extremal marks looks like? This will be the second target, and we will see that there is again an answer to the affirmative.

1.2 Extremal distributions

We will consider a family of real valued, independent identically distributed random variables X_i, $i \in \mathbb{N}$, with common distribution function

$$F(x) \equiv \mathbb{P}[X_i \leq x] \tag{1.3}$$

Recall that by convention, $F(x)$ is a non-decreasing, right-continuous function $F : \mathbb{R} \to [0,1]$. Note that the distribution function of M_n,

$$\mathbb{P}[M_n \leq x] = \mathbb{P}[\forall_{i=1}^n X_i \leq x] = \prod_{i=1}^n \mathbb{P}[X_i \leq x] = (F(x))^n \tag{1.4}$$

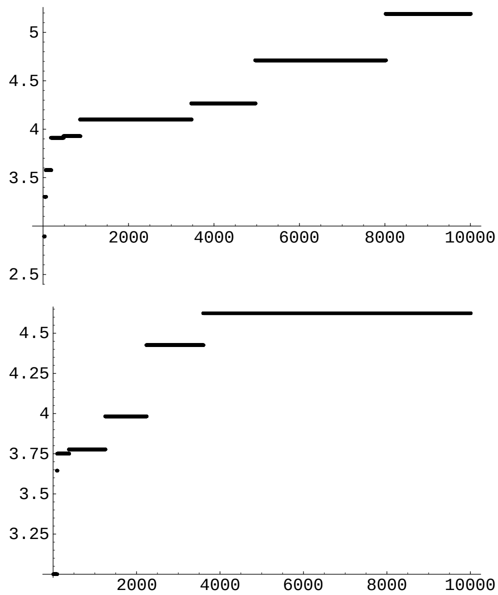

Fig. 1.1. Two sample plots of M_n against n for the Gaussian distribution.

As n tends to infinity, this will converge to a trivial limit

$$\lim_{n\uparrow\infty}(F(x))^n = \begin{cases} 0, & \text{if } F(x) < 1 \\ 1, & \text{if } F(x) = 1 \end{cases} \tag{1.5}$$

which simply says that any value that the variables X_i can exceed with positive probability will eventually exceeded after sufficiently many independent trials.

To illustrate a little how extremes behave, Figures 1.2 and 1.2 show the plots of samples of M_n as functions of n for the Gaussian and the exponential distribution, respectively.

As we have already indicated above, to get something more interesting, we must rescale. It is natural to try something similar to what is done in the central limit theorem: first subtract an n-dependent constant, then rescale by an n-dependent factor. Thus the first question is whether one can find two sequences, b_n, and a_n, and a non-trivial distribution function, $G(x)$, such that

$$\lim_{n\uparrow\infty}\mathbb{P}\left[a_n(M_n - b_n)\right] = G(x), \tag{1.6}$$

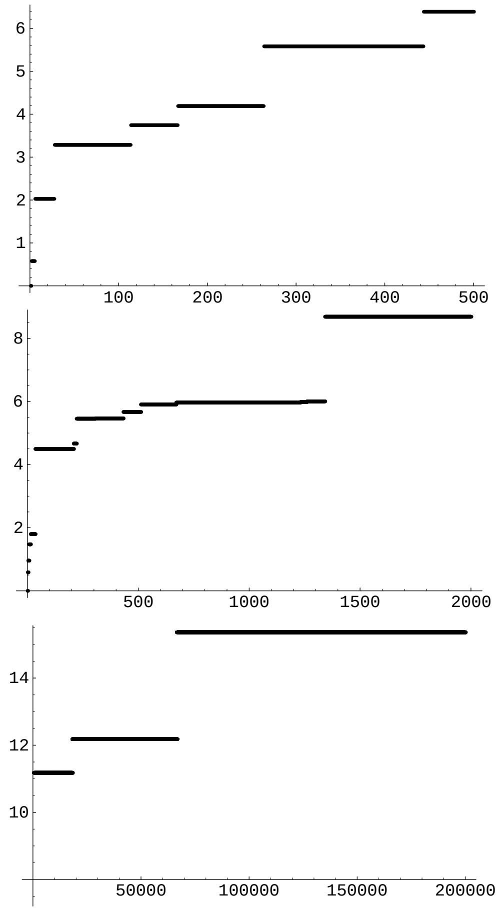

Fig. 1.2. Three plots of M_n against n for the exponential distribution over different ranges.

Example. The Gaussian distribution. In probability theory, it is always natural to start playing with the example of a Gaussian distribution. So we now assume that our X_i are Gaussian, i.e. that $F(x) = \Phi(x)$, where

$$\phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \tag{1.7}$$

We want to compute

$$\mathbb{P}\left[a_n(M_n - b_n) \leq x\right] = \mathbb{P}\left[M_n \leq a_n^{-1}x + b_n\right] = \left(\Phi\left(a_n^{-1}x + b_n\right)\right)^n \tag{1.8}$$

Setting $x_n \equiv a_n^{-1}x + b_n$, this can be written as

$$\left(1 - (1 - \Phi(x_n)\right)^n \tag{1.9}$$

For this to converge, we must choose x_n such that

$$(1 - \Phi(x_n)) = n^{-1}g(x) + o(1/n) \tag{1.10}$$

in which case

$$\lim_{n\uparrow\infty} \left(1 - (1 - \Phi(x_n)\right)^n = e^{-g(x)} \tag{1.11}$$

Thus our task is to find x_n such that

$$\frac{1}{\sqrt{2\pi}} \int_{x_n}^{\infty} e^{-y^2/2} dy = n^{-1}g(x) \tag{1.12}$$

At this point it will be very convenient to use an approximation for the function $1 - \Phi(u)$ when u is large, namely

$$\frac{1}{u\sqrt{2\pi}} e^{-u^2/2} \left(1 - 2u^{-2}\right) \leq 1 - \Phi(u) \leq \frac{1}{u\sqrt{2\pi}} e^{-u^2/2} \tag{1.13}$$

Using this, our problem simplifies to solving

$$\frac{1}{x_n\sqrt{2\pi}} e^{-x_n^2/2} = n^{-1}g(x), \tag{1.14}$$

that is

$$n^{-1}g(x) = \frac{e^{-\frac{1}{2}(a_n^{-1}x+b_n)^2}}{\sqrt{2\pi}(a_n^{-1}x + b_n)} = \frac{e^{-b_n^2/2 - a_n^{-2}x^2/2 - a_n^{-1}b_n x}}{\sqrt{2\pi}(a_n^{-1}x + b_n)} \tag{1.15}$$

Setting $x = 0$, we find

$$\frac{e^{-b_n^2/2}}{\sqrt{2\pi}b_n} = n^{-1}g(0) \tag{1.16}$$

Let us make the ansatz $b_n = \sqrt{2\ln n} + c_n$. Then we get for c_n

$$e^{-\sqrt{2\ln n}c_n - c_n^2/2} = \sqrt{2\pi}(\sqrt{2\ln n} + c_n) \tag{1.17}$$

It is convenient to choose $g(0) = 1$. Then, the leading terms for c_n are given by

$$c_n = -\frac{\ln\ln n + \ln(4\pi)}{2\sqrt{2\ln n}} \tag{1.18}$$

The higher order corrections to c_n can be ignored, as they do not affect the validity of (1.10). Finally, inspecting (1.13), we see that we can choose $a_n = \sqrt{2 \ln n}$. Putting all things together we arrive at the following assertion.

Lemma 1.2.1 *Let X_i, $i \in \mathbb{N}$ be iid normal random variables. Let*

$$b_n \equiv \sqrt{2 \ln n} - \frac{\ln \ln n + \ln(4\pi)}{2\sqrt{2 \ln n}} \tag{1.19}$$

and

$$a_n = \sqrt{2 \ln n} \tag{1.20}$$

Then, for any $x \in \mathbb{R}$,

$$\lim_{n \uparrow \infty} \mathbb{P}\left[a_n(M_n - b_n) \leq x\right] = e^{-e^{-x}} \tag{1.21}$$

Remark 1.2.1 It will be sometimes convenient to express (1.21) in a slightly different, equivalent form. With the same constants, a_n, b_n, define the function

$$u_n(x) \equiv b_n + x/a_n \tag{1.22}$$

Then

$$\lim_{n \uparrow \infty} \mathbb{P}\left[M_n \leq u_n(x)\right] = e^{-e^{-x}} \tag{1.23}$$

This is our first result on the convergence of extremes, and the function $e^{-e^{-x}}$, that is called the *Gumbel distribution* is the first *extremal distribution* that we encounter.

Let us take some basic messages home from these calculations:

- Extremes grow with n, but rather slowly; for Gaussians they grow like the square root of the logarithm only!
- The distribution of the extremes concentrates in absolute terms around the typical value, at a scale $1/\sqrt{\ln n}$; note that this feature holds for Gaussians and is not universal. In any case, to say that for Gaussians, $M_n \sim \sqrt{2 \ln n}$ is a quite precise statement when n (or rather $\ln n$) is large!

The next question to ask is how "typical" the result for the Gaussian distribution is. From the computation we see readily that we made no use of the Gaussian hypothesis to get the general form $\exp(-g(x))$ for any possible limit distribution. The fact that $g(x) = \exp(-x)$, however, depended on the particular form of Φ. We will see next that, remarkably, only two other types of functions can occur.

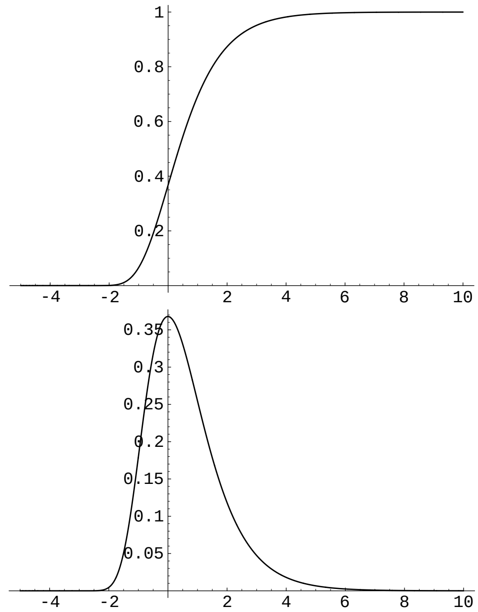

Fig. 1.3. The distribution function of the Gumbel distribution its derivative.

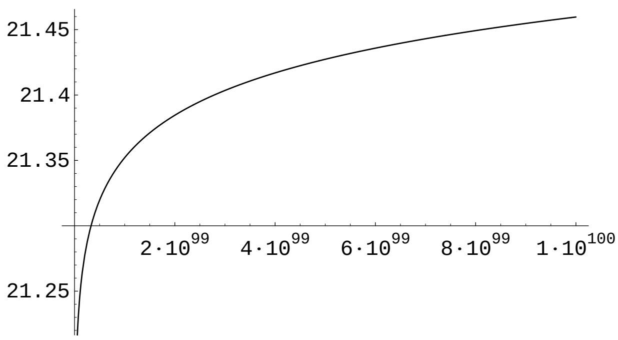

Fig. 1.4. The function $\sqrt{2 \ln n}$.

Some technical preparation. Our goal will be to be as general as possible with regard to the allowed distributions F. Of course we must anticipate that in some cases, no limiting distributions can be constructed (e.g. think of the case of a distribution with support on the two points 0 and 1!). Nonetheless, we are not willing to limit ourselves to random variables with continuous distribution functions, and this will introduce a little bit of complication, that, however, can be seen as a useful exercise.

Before we continue, let us explain where we are heading. In the Gaussian case we have seen already that we could make certain choices at various places. In general, we can certainly multiply the constants a_n by a finite number and add a finite number to the choice of b_n. This will clearly result in a different form of the extremal distribution, which, however, we think as morally equivalent. Thus, when classifying extremal distributions, we will think of two distributions, G, F, as equivalent if

$$F(ax + b) = G(x) \tag{1.24}$$

The distributions we are looking for arise as limits of the form

$$F^n(a_n x + b_n) \to G(x)$$

We will want to use that such limits have particular properties, namely that for some choices of α_n, β_n,

$$G^n(\alpha_n x + \beta_n) = G(x) \tag{1.25}$$

This property will be called *max-stability*. Our program will then be reduced to classify all max-stable distributions modulo the equivalence (1.24) and to determine their domains of attraction. Note the similarity of the characterisation of the Gaussian distribution as a stable distribution under addition of random variables.

Let us first comment on the notion of convergence of probability distribution functions. The common notion we will use is that of *weak convergence*:

Definition 1.2.1 A sequence, F_n, of probability distribution functions is said converge weakly to a probability distribution function F,

$$F_n \stackrel{w}{\to} F$$

iff and only if

$$F_n(x) \to F(x)$$

for all points x where F is continuous.

The next thing we want to do is to define the notion of the (left-continuous) inverse of a non-deceasing, right-continuous function (that may have jumps and flat pieces).

Definition 1.2.2 Let $\psi : \mathbb{R} \to \mathbb{R}$ be a monotone increasing, right-continuous function. Then the inverse function ψ^{-1} is defined as

$$\psi^{-1}(y) \equiv \inf\{x : \psi(x) \geq y\} \tag{1.26}$$

We will need the following properties of ψ^{-1}.

Lemma 1.2.2 *Let ψ be as in the definition, and $a > c$ and b real constants. Let $H(x) \equiv \psi(ax + b) - c$. Then*

(i) ψ^{-1} *is left-continuous.*
(ii) $\psi(\psi^{-1}(x)) \geq x$.
(iii) If ψ^{-1} *is continuous at* $\psi(x) \in \mathbb{R}$*, then* $\psi^{-1}(\psi(x)) = x$.
(iv) $H^{-1}(y) = a^{-1}\left(\psi^{-1}(y + c) - b\right)$
(v) If G is a non-degenerate distribution function, then there exist $y_1 < y_2$, such that $G^{-1}(y_1) < G^{-1}(y_2)$.

Proof (i) First note that ψ^{-1} is increasing. Let $y_n \uparrow y$. Assume that $\lim_n \psi^{-1}(y_n) < \psi^{-1}(y)$. This means that for all y_n, $\inf\{x : \psi(x) \geq y_n\} < \inf\{x : \psi(x) \geq y\}$. This means that there is a number, $x_0 < \psi^{-1}(y)$, such that, for all n, $\psi(x_0) \leq y_n$ but $\psi(x_0)) > y$. But this means that $\lim_n y_n \geq y$, which is in contradiction to the hypothesis. Thus ψ^{-1} is left continuous.

(ii) is immediate from the definition.

(iii) $\psi^{-1}(\psi(x)) = \inf\{x' : \psi(x') \geq \psi(x)\}$, thus obviously $\psi^{-1}(\psi(x)) \leq x$. On the other hand, for any $\epsilon > 0$, $\psi^{-1}(\psi(x) + \epsilon) = \inf\{x' : \psi(x') \geq \psi(x) + \epsilon\}$. But $\psi(x')$ can only be strictly greater than $\psi(x)$ if $x' > x$, so for any $y' > \psi(x)$, $\psi^{-1}(y') \geq x$. Thus, if ψ^{-1} is continuous at $\psi(x)$, this implies that $\psi^{-1}(\psi(x)) = x$.

(iv) The verification of the formula for the inverse of H is elementary and left as an exercise.

(v) If G is not degenerate, then there exist $x_1 < x_2$ such that $0 < G(x_1) \equiv y_1 < G(x_2) \equiv y_2 \leq 1$. But then $G^{-1}(y_1) \leq x_1$, and $G^{-1}(y_2) = \inf\{x : G(x) \geq G(x_2)\}$. If the latter equals x_1, then for all $x \geq x_1$, $G(x) \geq G(x_2)$, and since G is right-continuous, $G(x_1) = G(x_2)$, which is a contradiction. $\square$

For our purposes, the following corollary will be important.

Corollary 1.2.3 *If G is a non-degenerate distribution function, and there are constants $a > 0, \alpha > 0$, and $b, \beta \in \mathbb{R}$, such that, for all $x \in \mathbb{R}$,*

$$G(ax + b) = G(\alpha x + \beta) \tag{1.27}$$

then $a = \alpha$ and $b = \beta$.

Proof Let us call set $H(x) \equiv G(ax + b)$. Then, by (i) of the preceding lemma,

$$H^{-1}(y) = a^{-1}(G^{-1}(y) - b)$$

but by (1.27) also

$$H^{-1}(y) = \alpha^{-1}(G^{-1}(y) - \beta)$$

On the other hand, by (v) of the same lemma, there are at least two values of y such that $G^{-1}(y)$ are different, i.e. there are $x_1 < x_2$ such that

$$a^{-1}(x_i - b) = \alpha^{-1}(x_i - \beta)$$

which obviously implies the assertion of the corollary. □

Remark 1.2.2 Note that the assumption that G is non-degenerate is necessary. If, e.g., $G(x)$ has a single jump from 0 to 1 at a point a, then it holds that $G(5x - 4a) = G(x)$!

The next theorem is known as Khintchine's theorem:

Theorem 1.2.4 *Let F_n, $n \in \mathbb{N}$, be distribution functions, and let G be a non-degenerate distribution function. Let $a_n > 0$, and $b_n \in \mathbb{R}$ be sequences such that*

$$F_n(a_n x + b_n) \xrightarrow{w} G(x) \tag{1.28}$$

Then it holds that there are constants $\alpha_n > 0$, and $\beta_n \in \mathbb{R}$, and a non-degenerate distribution function G_*, such that

$$F_n(\alpha_n x + \beta_n) \xrightarrow{w} G_*(x) \tag{1.29}$$

if and only if

$$a_n^{-1}\alpha_n \to a, \quad (\beta_n - b_n)/a_n \to b \tag{1.30}$$

and

$$G_*(x) = G(ax + b) \tag{1.31}$$

Remark 1.2.3 This theorem makes the comment made above precise, saying that different choices of the scaling sequences a_n, b_n can lead only to distributions that are related by a transformation (1.31).

Proof By changing F_n, we can assume for simplicity that $a_n = 1$, $b_n = 0$. Let us first show that if $\alpha_n \to a, \beta_n \to b$, then $F_n(\alpha_n x + \beta_n) \to G_*(x)$.

Let $ax + b$ be a point of continuity of G.

$$F_n(\alpha_n x + \beta_n) = F_n(\alpha_n x + \beta_n) - F_n(ax + b) + F_n(ax + b) \qquad (1.32)$$

By assumption, the last term converges to $G(ax + b)$. Without loss of generality we may assume that $\alpha_n x + \beta_n$ is monotone increasing. We want to show that

$$F_n(\alpha_n x + \beta_n) - F_n(ax + b) \uparrow 0 \qquad (1.33)$$

Otherwise, there would be a constant, $\delta > 0$, such that along a subsequence n_k, $\lim_k F_{n_k}(\alpha_{n_k} x + \beta_{n_k}) - F_{n_k}(ax + b) < -\delta$. But since $\alpha_{n_k} x + \beta_{n_k} \uparrow ax + b$, this implies that for any $y < ax + b$, $\lim_k F_{n_k}(y) - F_{n_k}(ax + b) < -\delta$. Now, if G is continuous at y, this implies that $G(y) - G(ax + b) < -\delta$. But this implies that either F is discontinuous at $ax + b$, or there exists a neighborhood of $ax + b$ such that $G(x)$ has no point of continuity within this neighborhood. But this is impossible since a probability distribution function can only have countably many points of discontinuity. Thus (1.33) must hold, and hence

$$F_n(\alpha_n x + \beta_n) \xrightarrow{w} G(ax + b) \qquad (1.34)$$

which proves (1.29) and (1.31).

Next we want to prove the converse, i.e. we want to show that (1.29) implies (1.30). Note first that (1.29) implies that the sequence $\alpha_n x + \beta_n$ is bounded, since otherwise there would be subsequences converging to plus or minus infinity, along those $F_n(\alpha_n x + \beta_n)$ would converge to 0 or 1, contradicting the assumption. This implies that the sequence has converging subsequences $\alpha_{n_k}, \beta_{n_k}$, along which

$$\lim_k F_{n_k}(\alpha_{n_k} x + \beta_{n_k}) \to G_*(x) \qquad (1.35)$$

Then the preceding results shows that $a_{n_k} \to a'$, $b_{n_k} \to b'$, and $G'_*(x) = G(a'x + b')$. Now, if the sequence does not converge, there must be another convergent subsequence $a_{n'_k} \to a''$, $b_{n'_k} \to b''$. But then

$$G_*(x) = \lim_k F_{n'_k}(\alpha_{n'_k} x + \beta_{n'_k}) \to G(a''x + b') \qquad (1.36)$$

Thus $G(a'x + b') = G(a''x + b'')$. and so, since G is non-degenerate, Corollary 1.2.3 implies that $a' = a''$ and $b' = b''$, contradicting the assumption that the sequences do not converge. This proves the theorem

$\square$

max-stable distributions. We are now prepared to continue our search for extremal distributions. Let us formally define the notion of max-stable distributions.

Definition 1.2.3 A non-degenerate probability distribution function, G, is called *max-stable*, if for all $n \in \mathbb{N}$, there exists $a_n > 0, b_n \in \mathbb{R}$, such that, for all $x \in \mathbb{R}$,

$$G^n(a_n^{-1}x + b_n) = G(x) \tag{1.37}$$

The next proposition gives some important equivalent formulations of max-stability and justifies the term.

Proposition 1.2.5*(i) A probability distribution, G, is max-stable, if and only if there exists probability distributions F_n and constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, for all $k \in \mathbb{N}$,*

$$F_n(a_{nk}^{-1}x + b_{nk}) \xrightarrow{w} G^{1/k}(x) \tag{1.38}$$

(ii) G is max-stable, if and only if there exists a probability distribution function, F, and constants $a_n > 0$, $b_n \in \mathbb{R}$, such that

$$F^n(a_n^{-1}x + b_n) \xrightarrow{w} G(x) \tag{1.39}$$

Proof We first prove (i). If (1.38) holds, then by Khintchine's theorem, there exist constants, α_k, β_k, such that

$$G^{1/k}(x) = G(\alpha_k x + \beta_k)$$

for all $k \in \mathbb{N}$, and thus G is max-stable. Conversely, if G is max-stable, set $F_n = G^n$, and let a_n, b_n the constants that provide for (1.37). Then

$$F_n(a_{nk}^{-1}x + b_{nk}) = \left[G^{nk}(a_{nk}^{-1}x + b_{nk})\right]^{1/k} = G^{1/k}$$

which proves the existence of the sequence F_n and the respective constants.

Now let us prove (ii). Assume first that G is max-stable. Then choose $F = G$. Then the fact that $\lim_n F^n(a_n^{-1}x + b_n) = G(x)$ follows if the constants from the definition of max-stability are used trivially.

Next assume that (1.39) holds. Then, for any $k \in N$,

$$F^{nk}(a_{nk}^{-1}x + b_{nk}) \xrightarrow{w} G(x)$$

and so

$$F^n(a_{nk}^{-1}x + b_{nk}) \xrightarrow{w} G^{1/k}(x)$$

so G is max-stable by (i)! □

There is a slight extension to this result.

Corollary 1.2.6 *If G is max-stable, then there exist functions $a(s) > 0, b(s) \in \mathbb{R}$, $s \in \mathbb{R}_+$, such that*

$$G^s(a(s)x + b(s)) = G(x) \tag{1.40}$$

Proof This follows essentially by interpolation. We have that

$$G^{[ns]}(a_{[ns]}x + b_{[ns]}) = G(x)$$

But

$$\begin{aligned} G^n(a_{[ns]}x + b_{[ns]} &= G^{[ns]/s}(a_{[ns]}x + b_{[ns]})G^{n-[ns]/s}(a_{[ns]}x + b_{[ns]}) \\ &= G^{1/s}(x)G^{n-[ns]/s}(a_{[ns]}x + b_{[ns]}) \end{aligned}$$

As $n \uparrow \infty$, the last factor tends to one (as the exponent remains bounded), and so

$$G^n(a_{[ns]}x + b_{[ns]}) \xrightarrow{w} G^{1/s}(x)$$

and

$$G^n(a_n x + b_n) \xrightarrow{w} G(x)$$

Thus by Khintchine's theorem,

$$a_{[ns]}/a_n \to a(s), \quad (b_n - b_{[ns]})/a_n \to b(s)$$

and

$$G^{1/s}(x) = G(a(s)x + b(s))$$

□

The extremal types theorem.

Definition 1.2.4 Two distribution functions, G, H, are called "of the same type", if and only if there exists $a > 0, b \in \mathbb{R}$ such that

$$G(x) = H(ax + b) \tag{1.41}$$

We have seen that the only distributions that can occur as extremal distributions are max-stable distributions. We will now classify these distributions.

Theorem 1.2.7 *Any max-stable distribution is of the same type as one of the following three distributions:*

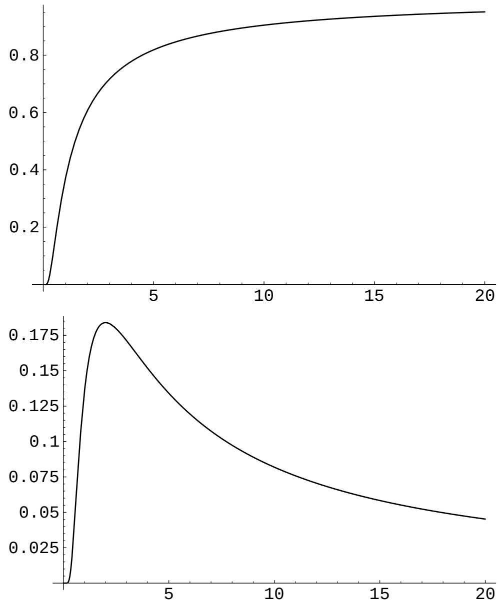

Fig. 1.5. The distribution function and density of the Fréchet with $\alpha = 1$.

(I) The Gumbel distribution,

$$G(x) = e^{-e^{-x}} \tag{1.42}$$

(II) The Fréchet distribution with parameter $\alpha > 0$*,*

$$G(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-x^{-\alpha}}, & \text{if } x > 0 \end{cases} \tag{1.43}$$

(III) The Weibull distribution with parameter $\alpha > 0$*,*

$$G(x) = \begin{cases} e^{-(-x)^{\alpha}}, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases} \tag{1.44}$$

Proof Let us check that the three types are indeed max-stable. For the Gumbel distribution this is already obvious as it appears as extremal distribution in the Gaussian case. In the case of the Fréchet distribution, note that

$$G^n(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-nx^{-\alpha}} = e^{-(n^{-1/\alpha}x)^{-\alpha}}, & \text{if } x > 0 \end{cases} = G(n^{-1/\alpha}x)$$

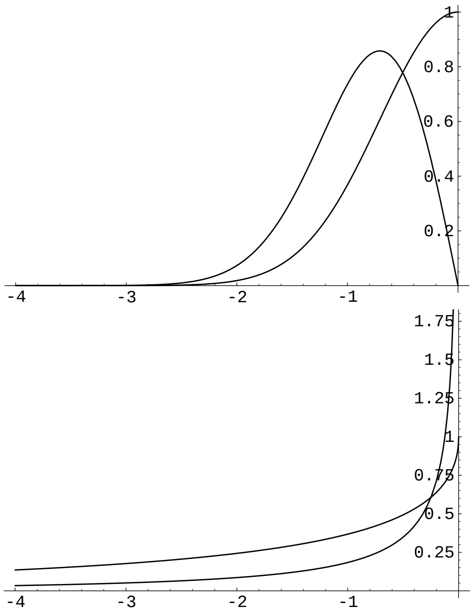

Fig. 1.6. The distribution function and density of the Weibull distribution with $\alpha = 2$ and $\alpha = 0.5$.

which proves max-stability. The Weibull case follows in exactly the same way.

To prove that the three types are the only possible cases, we use Corollary 1.2.6. Taking the logarithm, it implies that, if G is max-stable, then there must be $a(s), b(s)$, such that

$$-s \ln (G(a(s)x + b(s))) = -\ln G(x)$$

One more logarithm leads us to

$$\begin{aligned} &-\ln [-s \ln (G(a(s)x + b(s)))] \\ &= -\ln [-\ln (G(a(s)x + b(s)))] - \ln s \stackrel{!}{=} -\ln [-\ln G(x)] \equiv \psi(x) \end{aligned} \quad (1.45)$$

or equivalently

$$\psi(a(s)x + b(s)) - \ln s = \psi(x)$$

Now ψ is an increasing function such that $\inf_x \psi(x) = -\infty$, $\sup_x \psi(x) = +\infty$. We can define the inverse $\psi^{-1}(y) \equiv U(y)$. Using (iv) Lemma 1.2.2, we get that

$$\frac{U(y + \ln s) - b(s)}{a(s)} = U(y)$$

and subtracting the same equation for $y = 0$,

$$\frac{U(y + \ln s) - U(\ln s)}{a(s)} = U(y) - U(0)$$

Setting $\ln s = z$, this gives

$$U(y + z) - U(z) = [U(y) - U(0)]\, a(e^z) \tag{1.46}$$

To continue, we distinguish the case $a(s) \equiv 1$ and $a(s) \neq 1$ for some s.
Case 1. If $a(s) \equiv 1$, then

$$U(y + z) - U(z) = U(y) - U(0) \tag{1.47}$$

whose only solutions are

$$U(y) = \rho y + b \tag{1.48}$$

with $\rho > 0$, $b \in \mathbb{R}$. To see this, let $x_1 < x_2$ be any two points and let $\bar{x}$ be the middle point of $[x_1, x_2]$. Then (1.47) implies that

$$U(x_2) - U(\bar{x}) = U(x_2 - \bar{x}) - U(0) = U(\bar{x}) - U(x_1), \tag{1.49}$$

and thus $U(\bar{x}) = (U(x_2) - U(x_1))/2$. Iterating this proceedure implies readily that on all points of the form $x_k^{(n)} x_1 + k2^{-n}(X_2 - x_1)$ we have that $U(x_k) = U(x_1) + k2^{-n}(U(x_2) - U(x_1))$; that is, on a dense set of points (1.48) holds. But since U is also monotonous, it is completely determined by its values on a dense set, so U is a linear function.

But then $\psi(x) = \rho^{-1}x - b$, and

$$G(x) = \exp\left(-\exp\left(-\rho^{-1}x - b\right)\right)$$

which is of the same type as the Gumbel distribution.
Case 2. Set $\widetilde{U}(y) \equiv U(y) - U(0)$, Then subtract from (1.46) the same equation with y and z exchanged. This gives

$$-\widetilde{U}(z) + \widetilde{U}(y) = a(e^z)\widetilde{U}(y) - a(e^y)\widetilde{U}(z)$$

or

$$\widetilde{U}(z)\,(1 - a(e^y)) = \widetilde{U}(y)\,(1 - a(e^z))$$

Now chose z such that $a(e^z) \neq 1$. Then

$$\widetilde{U}(y) = \widetilde{U}(z)\frac{1 - a(e^y)}{1 - a(e^z)} \equiv c(z)(1 - a(e^y))$$

Now we insert this result again into (1.46). We get

$$\begin{aligned}\widetilde{U}(y+z) &= c(z)\left(1-a(e^{y+z})\right) && (1.50)\\ &= \widetilde{U}(z)+\widetilde{U}(y)a(e^z) && (1.51)\\ &= c(z)\left(1-a(e^z)\right)+c(z)\left(1-a(e^y)\right)a(e^z) && (1.52)\end{aligned}$$

which yields an equation for a, namely,

$$a(e^{y+z}) = a(e^y)a(e^z)$$

The only functions satisfying this equation are the powers, $a(x) = x^\rho$. Therefore,

$$U(y) = U(0) + c(1 - e^{\rho y})$$

Setting $U(0) = \nu$, going back to G this gives

$$G(x) = \exp\left(-\left(1 - \frac{x-\nu}{c}\right)^{-1/\rho}\right) \tag{1.53}$$

for those x where the right-hand side is < 1.

To conclude the proof, it suffices to discuss the two cases $-1/\rho \equiv \alpha > 0$ and $-1/\rho \equiv -\alpha < 0$, which yield the Fréchet, resp. Weibull types. □

Let us state as an immediate corollary the so-called *extremal types theorem.*

Theorem 1.2.8 *Let X_i, $i \in \mathbb{N}$ be a sequence of i.i.d. random variables. If there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$, and a non-degenerate probability distribution function, G, such that*

$$\mathbb{P}\left[a_n(M_n - b_n) \leq x\right] \stackrel{w}{\to} G(x) \tag{1.54}$$

then $G(x)$ is of the same type as one of the three extremal-type distributions.

Note that it is not true, of course, that for arbitrary distributions of the variables X_i it is possible to obtain a nontrivial limit as in (1.54).

Domains of attraction of the extremal type distributions. Of course it will be nice to have simple, verifiable criteria to decide for a given distribution F to which distribution the maximum of iid variables with this distribution corresponds. We will say that, if X_i are distributed according to F, if (1.54) holds with an extremal distribution, G, that F *belongs to the domain of attraction of G.*

The following theorem gives necessary and sufficient conditions. We set $x_F \equiv \sup\{x : F(x) < 1\}$.

Theorem 1.2.9 *The following conditions are necessary and sufficient for a distribution function, F, to belong to the domain of attraction of the three extremal types:*

Fréchet: $x_F = +\infty$,

$$\lim_{t\uparrow\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \quad \forall_{x>0}, \alpha > 0 \tag{1.55}$$

Weibull: $x_F =< +\infty$,

$$\lim_{h\downarrow\infty} \frac{1 - F(x_F - xh)}{1 - F(x_F - h)} = x^{\alpha}, \quad \forall_{x>0}, \alpha > 0 \tag{1.56}$$

Gumbel: $\exists g(t) > 0$,

$$\lim_{t\uparrow x_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}, \quad \forall_x \tag{1.57}$$

Proof We will only prove the sufficiency of the criteria. As we have seen in the computations for the Gaussian distribution, the statements

$$n(1 - F(a_n^{-1}x + b_n)) \to g(x) \tag{1.58}$$

and

$$F^n(a_n^{-1}x + b_n) \to e^{-g(x)} \tag{1.59}$$

are equivalent. Thus we only have to check when (1.58) holds with which $g(x)$.

Let us assume that there is a sequence, γ_n, such that

$$n(1 - F(\gamma_n)) \to 1.$$

Since necessarily $F(\gamma_n) \to 1$, $\gamma_n \to x_F$, and we may choose $\gamma_n < x_F$, for all n. We now turn to the three cases.

Fréchet: We know that (for $x > 0$),

$$\frac{1 - F(\gamma_n x)}{1 - F(\gamma_n)} \to x^{-\alpha}$$

while $n(1 - F(\gamma_n)) \to 1$. Thus,

$$n(1 - F(\gamma_n x)) \to x^{-\alpha}$$

and so, for $x > 0$,

$$F^n(\gamma_n x) \to e^{-x^{-\alpha}}.$$

Since $\lim_{x\downarrow 0} e^{-x^{-\alpha}} = 0$, it must be true that, for $x \leq 0$,

$$F^n(\gamma_n x) \to 0$$

which concludes the argument.

Weibull: Let now $h_n = x_F - \gamma_n$. By the same argument as above, we get, for $x > 0$

$$n(1 - F(x_F - h_n x)) \to x^\alpha$$

and so

$$F^n(x_F - x(x_F - \gamma_n)) \to e^{-x^\alpha}$$

or equivalently, for $x < 0$,

$$F^n((x_F - \gamma_n)x + x_F) \to e^{-(-x)^\alpha}$$

Since, for $x \uparrow 0$, the right-hand side tends to 1, it follows that, for $x \geq 0$,

$$F^n(x(x_F - \gamma_n) - x_F) \to 1$$

Gumbel: In exactly the same way we conclude that

$$n(1 - F(\gamma_n + xg(\gamma_n))) \to e^{-x}$$

from which the conclusion is now obvious, with $a_n = 1/g(\gamma_n)$, $b_n = \gamma_n$.

We are left with proving the existence of γ_n with the desired property. If F had no jumps, we could choose γ_n simply such that $F(\gamma_n) = 1 - 1/n$ and we would be done. The problem becomes more subtle since we want to allow for more general distribution functions. The best approximation seems to be

$$\gamma_n \equiv F^{-1}(1 - 1/n) = \inf\{x : F(x) \geq 1 - 1/n)\}$$

Then we get immediately that

$$\limsup n(1 - F(\gamma_n)) \leq 1.$$

But for $x < \gamma_n$, $F(x) \leq 1 - 1/n$, and so $n(1 - F(\gamma_n^-)) \geq 1$. Thus we may just show that

$$\liminf_n \frac{1 - F(\gamma_n)}{1 - F(\gamma_n^-)} \geq 1.$$

This, however, follows in all cases from the hypotheses on the functions F, e.g.

$$\frac{1 - F(x\gamma_n)}{1 - F(\gamma_n)} \to x^{-\alpha}$$

which tends to 1 as $x \uparrow 1$. This concludes the proof of the sufficiency in the theorem. $\square$

Remark 1.2.4 The proof of the necessity of the conditions of the theorem can be found in the book by Resnik [10].

Examples. Let us see how the theorem works in some examples. *normal distribution.* In the normal case, the criterion for the Gumbel distribution is

$$\frac{1-F(t+xg(t))}{1-F(t)} \sim \frac{e^{-(t+xg(t))^2/2}t}{e^{-t^2/2}(t+xg(t))} = \frac{e^{-x^2g^2(t)/2-xtg(t)}}{1+xg(t)/t}$$

which converges with the choice $g(t) = 1/t$, to $\exp(-x)$. Also, the choice of γ_n, $\gamma_n = F^{-1}(1-1/n)$ gives $\exp(-\gamma_n^2/2)/(\sqrt{2\pi}\gamma_n) = n^{-1}$, which is the same criterion as found before.

Exponential distribution. We should again expect the Gumbel distribution. In fact, since $F(x) = 1 - e^{-x}$,

$$\frac{e^{-(t+xg(t))}}{e^{-t}} = e^{-x}$$

if $g(t) = 1$. γ_n will here be simply $\gamma_n = \ln n$, so that $a_n = 1, b_n = \ln n$, and

$$\mathbb{P}\left[M_n - \ln n \leq x\right] \stackrel{w}{\to} e^{-e^{-x}}$$

Pareto distribution. Here

$$F(x) = \begin{cases} 1 - Kx^{-\alpha}, & \text{if } x \geq K^{1/\alpha} \\ 0, & \text{else} \end{cases}$$

Here

$$\frac{1-F(tx)}{1-F(t)} = \frac{x^{-\alpha}t^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}$$

for positive x, so it falls in the domain of attraction of the Fréchet distribution. Moreover, we get

$$\gamma_n = (nK)^{1/\alpha}$$

so that here

$$\mathbb{P}\left[(nK)^{-1/\alpha}M_n \leq x\right] \stackrel{w}{\to} e^{-x^{-\alpha}}$$

Thus, here, $M_n \sim (nK)^{1/\alpha}$, i.e. the maxima grow much faster than in the Gaussian or exponential situation!

Uniform distribution. We consider $F(x) = 1 - x$ on $[0,1]$. Here $x_F = 1$, and

$$\frac{1-F(1-xh)}{1-F(1-h)} = x$$

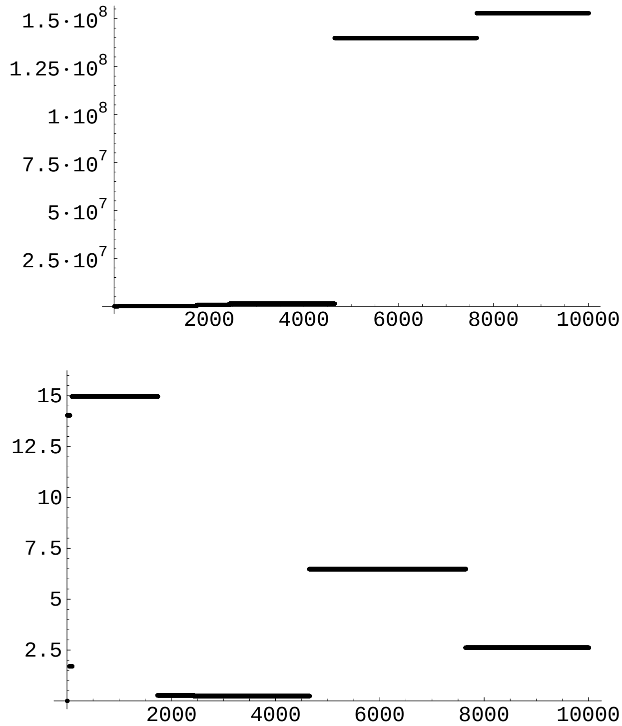

Fig. 1.7. Records of the Pareto distribution with $K = 1$, $\alpha = 0.5$. Second picture shows M_n/a_n.

so we are in the case of the Weibull distribution with $\alpha = 1$. We find $\gamma_n = 1 - 1/n$, $a_n = 1/n$, $b_n = 1$, and so

$$\mathbb{P}\left[(n(M_n - 1) \leq x\right] \xrightarrow{w} e^x, \quad x \leq 0$$

Bernoulli distribution. Consider

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1/2, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } x \geq 1 \end{cases}$$

Clearly $x_F = 1$, but

$$\frac{1 - F(1 - hx)}{1 - F(1 - h)} = 1$$

so it is impossible that this converges to x^α, with $\alpha \neq 0$. Thus, as

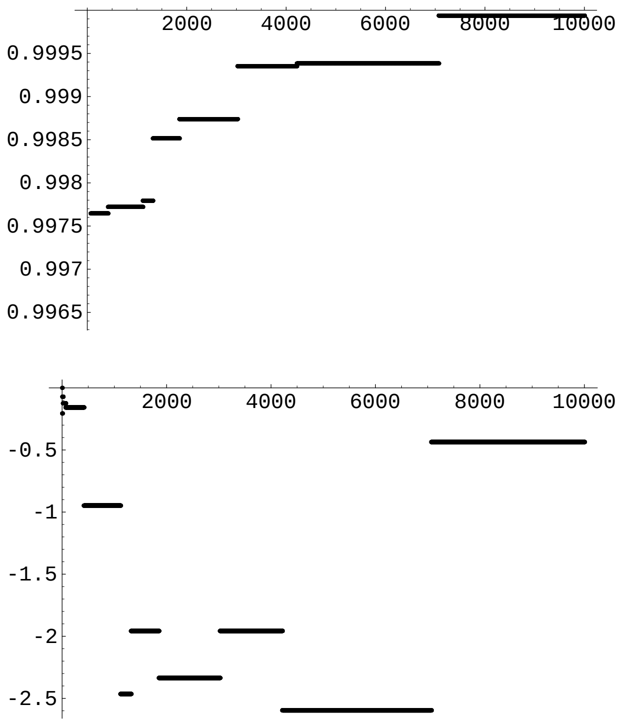

Fig. 1.8. Records of the uniform distribution. Second picture shows $n(M_n - 1)$.

expected, the Bernoulli distribution does not permit any convergence of its maximum to a non-trivial distribution.

In the proof of the previous theorem we have seen that the existence of sequences γ_n such that $n(1-F(\gamma_n)) \to 1$ was crucial for the convergence to an extremal distribution. We will now extend this discussion and ask for criteria when there will be sequences for which $n(1 - F(\gamma_n))$ tends to an arbitrary limit. Naturally, this must be related to the behaviour of F near the point x_F.

Theorem 1.2.10 *Let F be a distribution function. Then there exists a sequence, γ_n, such that*

$$n(1 - F(\gamma_n)) \to \tau, \quad 0 < \tau < \infty, \tag{1.60}$$

if and only if

$$\lim_{x \uparrow x_F} \frac{1-F(x)}{1-F(x^-)} = 1 \tag{1.61}$$

Remark 1.2.5 To see what is at issue, note that

$$\frac{1-F(x)}{1-F(x^-)} = 1 + \frac{p(x)}{1-F(x^-)},$$

where $p(x)$ is the probability of the "atom" at x, i.e. the size of the jump of F at x. Thus, (1.61) says that the size of jumps of F should diminish faster, as x approaches the upper boundary of the support of F, than the total mass beyond x.

Proof Assume that (1.60) holds, but

$$\frac{p(x)}{1-F(x^-)} \not\to 0.$$

Then there exists $\epsilon > 0$ and a sequence, $x_j \uparrow x_F$, such that

$$p(x_j) \geq 2\epsilon(1 - F(x_j^-)).$$

Now chose n_j such that

$$1 - \frac{\tau}{n_j} \leq \frac{F(x_j^-) + F(x_j)}{2} \leq 1 - \frac{\tau}{n_j + 1}. \tag{1.62}$$

The gist of the argument (given in detail below) is as follows: Since the upper and lower limit in (1.62) differ by only $O(1/n_j^2)$, the term in the middle must equal, up to that error, $F(\gamma_{n_j})$; but $F(x_j)$ and $F(x_j^-)$ differ (by hypothesis) by ϵ/n_j, and since F takes no value between these two, it is impossible that $\frac{F(x_j^-)+F(x_j)}{2} = F(\gamma_{n_j})$ to the precision required. Thus (1.61) must hold.

Let us formalize this argument. Now it must be true that either

(i) $\gamma_{n_j} < x_j$ i.o., or
(ii) $\gamma_{n_j} \geq x_j$ i.o.

In case (i), it holds that for these j,

$$n_j(1 - F(\gamma_{n_j})) > n_j(1 - F(x_j^-)). \tag{1.63}$$

Now replace in the right-hand side

$$F(x_j^-) = \frac{F(x_j^-) + F(x_j) - p(x_j)}{2}$$

and write

$$1 = \tau/n_j + 1 - \tau/n_j$$

to get

$$\begin{aligned} n_j(1 - F(x_j^-)) &= \tau + n_j \left(1 - \frac{\tau}{n_j} - \frac{F(x_j^-) + F(x_j) - p(x_j)}{2}\right) \\ &\geq \tau + \frac{n_j p(x_j)}{2} - n_j \left(\frac{\tau}{n_j} - \frac{\tau}{n_j + 1}\right) \\ &\geq \tau + \epsilon n_j (1 - F(x_j^-)) - \frac{\tau}{n_j + 1}. \end{aligned}$$

Thus

$$n_j(1 - F(x_j^-)) \geq \tau \frac{1 - 1/(n_j + 1)}{1 - \epsilon}.$$

For large enough j, the right-hand side will be strictly larger than τ, so that

$$\limsup_j n_j(1 - F(x_j^-)) > \tau,$$

and in view of (1.63), a fortiori

$$\limsup_j n_j(1 - F(\gamma_j^-)) > \tau,$$

in contradiction with the assumption.

In case (ii), we repeat the same argument mutando mutandis, to conclude that

$$\liminf_j n_j(1 - F(\gamma_j^-)) < \tau,$$

To prove the converse assertion, choose

$$\gamma_n \equiv F^{-1}(1 - \tau/n).$$

Using (1.61), one deduces (1.60) exactly as in the special case $\tau = 1$ in the proof of Theorem 1.2.9. □

Example. Let us show that in the case of the Poisson distribution condition (1.61) is not satisfied. The Poisson distribution with parameter λ has only jumps at the positive integers, k, of values $p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$. Thus

$$\frac{p(n)}{1 - F(n^-)} = \frac{\lambda^n/n!}{\sum_{k=n}^{\infty} \lambda^k/k!} = \frac{1}{1 + \sum_{k=n+1}^{\infty} \lambda^{n-k} \frac{n!}{k!}},$$

But

$$\sum_{k=n+1}^{\infty} \lambda^{n-k} \frac{n!}{k!} = \sum_{k=+1}^{\infty} \lambda^k \frac{n!}{(n+k)!} \leq \sum_{k=+1}^{\infty} \lambda^k n^{-k} = \frac{\lambda/n}{1-\lambda/n} \downarrow 0,$$

so that $\frac{p(n)}{1-F(n^-)} \to 1$. Thus, for the Poisson distribution, we cannot construct a non-trivial extremal distribution.

1.3 Level-crossings and the distribution of the k-th maxima.

In the previous section we have answered the question of the distribution of the maximum of n iid random variables. It is natural to ask for more, i.e. for the joint distribution of the maximum, the second largest, third largest, etc.

From what we have seem, the levels u_n for which $\mathbb{P}[X_n > u_n] \sim \tau/n$ will play a crucial rôle. A natural variable to study is $\mathcal{M}_n^k$, the value of the k-th largest of the first n variables X_i.

It will be useful to introduce here the notion of *order statistics*.

Definition 1.3.1 Let $X_1, \ldots, X_n$ be real numbers. Then we denote by $\mathcal{M}_n^1, \ldots, \mathcal{M}_n^n$ its order statistics, i.e. for some permutation, π, of n numbers, $\mathcal{M}_n^k = X_{\pi(k)}$, and

$$\mathcal{M}_n^n \leq \mathcal{M}_n^{n-1} \leq \cdots \leq \mathcal{M}_n^2 \leq \mathcal{M}_n^1 \equiv M_n \tag{1.64}$$

We will also introduce the notation

$$S_n(u) \equiv \#\{i \leq n : X_i > u\} \tag{1.65}$$

for the number of exceedences of the level u. Obviously we have the relation

$$\mathbb{P}\left[\mathcal{M}_n^k \leq u\right] = \mathbb{P}\left[S_n(u) < k\right] \tag{1.66}$$

The following result states that the number of exceedances of an extremal level u_n is Poisson distributed.

Theorem 1.3.1 *Let X_i be iid random variables with common distribution F. If u_n is such that*

$$n(1-F(u_n)) \to \tau, \quad 0 < \tau < \infty,$$

then

$$\mathbb{P}\left[\mathcal{M}_n^k \leq u_n\right] = \mathbb{P}\left[S_n(u_n) < k\right] \to e^{-\tau} \sum_{s=0}^{k-1} \frac{\tau^s}{s!} \tag{1.67}$$

Proof The proof of this lemma is quite simple. We just need to consider all possible ways to realise the event $\{S_n(u_n) = s\}$. Namely

$$\begin{aligned}
\mathbb{P}\left[S_n(u_n) = s\right] &= \sum_{\{i_1,\dots,i_s\}\subset\{1,\dots,n\}} \prod_{\ell=1}^{s} \mathbb{P}\left[X_{i_\ell} > u_n\right] \prod_{j\notin\{1_i,\dots,i_s\}} \mathbb{P}\left[X_j \leq u_n\right] \\
&= \binom{n}{s}(1-F(u_n))^s F(u_n)^{n-s} \\
&= \frac{1}{s!}\frac{n!}{n^s(n-s)!}\left[n(1-F(u_n))\right]^s \left[F^n(u_n)\right]^{1-s/n}.
\end{aligned}$$

But, for any s fixed, $n(1-F(u_n)) \to \tau$, $F^n(u_n) \to e^{-\tau}$, $s/n \to 0$, and $\frac{n!}{n^s(n-s)!} \to 1$. Thus

$$\mathbb{P}\left[S_n(u_n) = s\right] \to \frac{\tau^s}{s!}e^{-\tau}.$$

Summing over all $s < k$ gives the assertion of the theorem. □

Using very much the same sort of reasoning, one can generalise the question answered above to that of the numbers of exceedances of several extremal levels.

Theorem 1.3.2 *Let $u_n^1 > n_n^2 \cdots > u_n^r$ such that*

$$n(1-F(u_n^\ell)) \to \tau_\ell,$$

with

$$0 < \tau_1 < \tau_2 < \ldots, < \tau_r < \infty.$$

Then, under the assumptions of the preceding theorem, with $S_n^i \equiv S_n(u_n^i)$,

$$\begin{aligned}
&\mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \dots, S_n^r - S_n^{r-1} = k_r\right] \to \\
&\frac{\tau_1^{k_1}}{k_1!}\frac{(\tau_2-\tau_1)^{k_2}}{k_2!}\dots\frac{(\tau_r-\tau_{r-1})^{k_r}}{k_r!}e^{-\tau_r}
\end{aligned} \tag{1.68}$$

Proof Again, we just have to count the number of arrangements that will place the desired number of variables in the respective intervals.

Then

$$\begin{aligned}
&\mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \dots, S_n^r - S_n^{r-1} = k_r\right] \\
&= \binom{n}{k_1, \dots, k_r} \mathbb{P}\big[X_1, \dots, X_{k_1} > u_n^1 \geq X_{k_1+1}, \dots, X_{k_1+k_2} > u_n^2, \dots \\
&\dots, u_n^{r-1} \geq X_{k_1+\dots+k_{r-1}+1}, \dots, X_{k_1+\dots+k_r} > u_n^r \geq X_{k_1+\dots+k_r+1}, \dots, X_n\big] \\
&= \binom{n}{k_1, \dots, k_r} (1 - F(u_n^1))^{k_1} \left[F(u_n^1) - F(u_n^2)\right]^{k_2} \dots \left[F(u_n^{r-1}) - F(u_n^r)\right]^{k_r} \\
&\qquad\qquad \times F^{n-k_1-\dots-k_r}(u_n^r)
\end{aligned}$$

Now we write

$$\left[F(u_n^{\ell-1}) - F(u_n^\ell)\right] = \frac{1}{n}\left[n(1 - F(u_n^\ell)) - n(1 - F(u_n^{\ell-1}))\right]$$

and use that $\left[n(1 - F(u_n^\ell)) - n(1 - F(u_n^{\ell-1}))\right] \to \tau_\ell - \tau_{\ell-1}$. Proceeding otherwise as in the proof of Theorem 1.3.1, we arrive at (1.68) □

2

Extremes of stationary sequences.

2.1 Mixing conditions and the extremal type theorem.

One of the classic settings that generalise the case of iid sequences of random variables are *stationary* sequences.

We recall the definition:

Definition 2.1.1 An infinite sequence of random variables X_i, $i \in \mathbb{Z}$ is called stationary, if, for any finite collection of indices, $i_1, \dots, i_m$, and and positive integer k, the collections of random variables

$$\{X_{i_1}, \dots, X_{i_m}\}$$

and

$$\{X_{i_1+k}, \dots, X_{i_m+k}\}$$

have the same distribution.

It is clear that there cannot be any general results on the sole condition of stationarity. E.g., the constant sequence $X_i = X$, for all $i \in \mathbb{Z}$ is stationary, and here clearly the distribution of the maximum is the distribution of X. Generally, one will want to ask what the effect of correlation on the extremes is, and the first natural question is, of course, whether for sufficiently weak dependence, the effect may simply be nil. This is, in fact, the question most works on extremes address, and we will devote some energy to this. From a practical point of view, this question is also very important. Namely, it is in practice quite difficult to determine for a given random process its precise dependence structure, simply because there are so many parameters that would need to be estimated. Under simplifying assumptions, e.g. assume a Gaussian multivariate distribution, one may limit the number of parameters, but still it is a rather difficult task. Thus it will be very helpful not to have to do this, and rather get some control on the dependences that will ensure that, as far as extremes are concerned, we need not worry about the details.

In the case of stationary sequences, one introduces traditionally some *mixing conditions*, called Condition D and the weaker Condition $D(u_n)$.

Definition 2.1.2 A stationary sequence, X_i, of random variables satisfies Condition D, if there exists a sequence, $g(\ell) \downarrow 0$, such that, for all $p, q \in \mathbb{N}$, $i_1 < i_2 < \cdots < i_p$, and $j_1 < j_2 < \cdots < j_q$, such that $j_1 - i_q > \ell$, for all $u \in \mathbb{R}$,

$$\Big| \mathbb{P}\left[X_{i_1} \leq u, \ldots, X_{i_p} \leq u, X_{j_1} \leq u, \ldots X_{j_q} \leq u\right] \tag{2.1}$$
$$- \mathbb{P}\left[X_{i_1} \leq u, \ldots, X_{i_p} \leq u\right] \mathbb{P}\left[X_{j_1} \leq u, \ldots X_{j_q} \leq u\right] \Big| \leq g(\ell)$$

A weaker and often useful condition is adapted to a given extreme level.

Definition 2.1.3 A stationary sequence, X_i, of random variables satisfies Condition $D(u_n)$, for a sequence u_n, $n \in \mathbb{N}$, if there exists a sequences, $\alpha_{n,\ell}$, satisfying for some $\ell_n = o(n)$ $\alpha_{n,\ell_n} \downarrow 0$, such that, for all $p, q \in \mathbb{N}$, $i_1 < i_2 < \cdots < i_p$, and $j_1 < j_2 < \cdots < j_q$, such that $j_1 - i_q > \ell$, for all $u \in \mathbb{R}$,

$$\Big| \mathbb{P}\left[X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \ldots X_{j_q} \leq u_n\right] \tag{2.2}$$
$$- \mathbb{P}\left[X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n\right] \mathbb{P}\left[X_{j_1} \leq u_n, \ldots X_{j_q} \leq u_n\right] \Big| \leq \alpha_{n,\ell}$$

Note that in both mixing conditions, the decay rate of the correlation does only depend on the distance, ℓ, between the two blocks of variables, and not on the number of variables involved. This will be important, since the general strategy of our proofs will be to remove a "small" fraction of the variable from consideration such that the remaining ones form sufficiently separated blocks, that, due to the mixing conditions, behave as if they were independent. The following proposition provides the basis for this strategy.

Proposition 2.1.1 *Assume that a sequence of random variables X_i satisfies $D(u_n)$. Let $E_1, \ldots, E_r$ a finite collection of disjoint subsets of $\{1, \ldots, n\}$. Set*

$$M(E) \equiv \max_{i \in E} X_i$$

If, for all $1 \leq i, j \leq r$, $\mathrm{dist}(\mathrm{E_i}, \mathrm{E_j}) \geq \mathrm{k}$, then

$$\left|\mathbb{P}\left[\cap_{i=1}^{r} M(E_i) \leq u_n\right] - \prod_{i=1}^{r} \mathbb{P}\left[M(E_i) \leq u_n\right]\right| \leq (r-1)\alpha_{n,k} \tag{2.3}$$

Proof The proof is simply by induction over r. By assumption, (2.3) holds for $r = 2$. We will show that, if it holds for $r-1$, then it holds for r. Namely,

$$\mathbb{P}\left[\cap_{i=1}^{r} M(E_i) \leq u_n\right] = \mathbb{P}\left[\cap_{i=1}^{r-1} M(E_i) \leq u_n \cap M(E_r) \leq u_n\right]$$

But by assumption,

$$\left|\mathbb{P}\left[\cap_{i=1}^{r-1} M(E_i) \leq u_n \cap M(E_r) \leq u_n\right] - \mathbb{P}\left[\cap_{i=1}^{r-1} M(E_i) \leq u_n\right] \mathbb{P}\left[M(E_r) \leq u_n\right]\right| \leq \alpha_{n,k}$$

and by induction hypothesis

$$\left|\mathbb{P}|\mathbb{P}\left[\cap_{i=1}^{r-1} M(E_i) \leq u_n\right] - \prod_{i=1}^{r-1} \mathbb{P}\left[M(E_i) \leq u_n\right]\right| \leq (r-2)\alpha_{n,k}$$

Putting both estimates together using the triangle inequality yields (2.3). □

A first consequence of this observation is the so-called extremal type theorem that asserts that the our extremal types keep their importance for weakly dependent stationary sequences.

Theorem 2.1.2 *Let X_i be a stationary sequence of random variables and assume that there are sequences $a_n > 0, b_n \in \mathbb{R}$ be such that*

$$\mathbb{P}\left[a_n(M_n - b_n) \leq x\right] \stackrel{w}{\to} G(x),$$

where $G(x)$ is a non-degenerate distribution function. Then, if X_i satisfies condition $D(a_n x + b_n)$ for all $x \in \mathbb{R}$, then G is of the same type as one of the three extremal distributions.

Proof The strategy of the proof is to show that G must be max-stable. To do this, we show that, for all $k \in \mathbb{N}$,

$$\mathbb{P}\left[a_{nk}(M_n - b_{nk})) \leq x\right] \stackrel{w}{\to} G^{1/k}(x). \tag{2.4}$$

Now (2.4) means that we have to show that

$$\mathbb{P}\left[M_{kn} \leq x/a_{nk} + b_{nk}\right] - \left(\mathbb{P}\left[M_n \leq x/a_{nk} + b_{nk}\right]\right)^k \to 0$$

This calls for Proposition 2.1.1. Naively, we would group the segment $(1, \ldots, kn)$ into k blocks of size n. The problem is that there would be no distance between them. The solution is to remove from each of the blocks the last piece of size m, so that we have k blocks

$$I_\ell \equiv \{n\ell+1, \dots, n\ell+(n-m)\}$$

Let us denote the remaining pieces by

$$I'_\ell \equiv \{n\ell+m+1, \dots, n\ell+(n-1\}$$

Then (we abbreviate $x/a_{nk}+b_{nk} \equiv u_{nk}$),

$$\begin{aligned}
&\mathbb{P}\left[M_{kn} \leq u_{nk}\right] = \mathbb{P}\left[\{\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\} \bigcap \{\cap_{i=0}^{k-1} M(I'_i) \leq u_{nk}\}\right] \quad (2.5)\\
&= \mathbb{P}\left[\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\right]\\
&+ \mathbb{P}\left[\{\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\} \bigcap \{\cap_{i=0}^{k-1} M(I'_i) \leq u_{nk}\}\right] - \mathbb{P}\left[\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\right]
\end{aligned}$$

The last term can be written as

$$\begin{aligned}
&\left|\mathbb{P}\left[\{\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\} \bigcap \{\cap_{i=0}^{k-1} M(I'_i) \leq u_{nk}\}\right] - \mathbb{P}\left[\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\right]\right|\\
&= \mathbb{P}\left[\{\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\} \bigcap \{\cup_{i=0}^{k-1} M(I'_i) > u_{nk}\}\right]\\
&\leq k\mathbb{P}\left[M(I_1) \leq u_{nk} < M(I'_1)\right] \qquad (2.6)
\end{aligned}$$

This term should be small, because is requires the maximum of the small interval I'_1 of exceed the level u_{nk}, while on the much larger interval I_1 this level is not exceeded. This would be obvious if we knew that $(1-F(u_{nk})) \sim 1/n$, but of course we have not made such an assumption. The problem is, however, easily solved by using again condition $D(u_n)$. In fact, it suffices to show that the interval I_i contains a number r of well separated subintervals of the same size as I_1, where r can be taken as large as desired, as n goes to infinity. In fact, for any any integer $r < (n-2m)/2$, we can find r intervals $E_1, \dots, E_r$ in I_1, such that $|E_i| = m$, and $\mathrm{dist(E_i,E_j) \geq m}$, and $\mathrm{dist(E_i,I'_1) \geq m}$. Then, using Proposition 2.1.1,

$$\begin{aligned}
&\mathbb{P}\left[M(I_1) \leq u_{nk} < M(I'_1)\right] \leq \mathbb{P}\left[\{\cap_{j=1}^{r} M(E_j) \leq u_{nk}\} \bigcap \{M(I'_1) > u_{nk}\}\right]\\
&=\leq \mathbb{P}\left[\cap_{j=1}^{r} M(E_j) \leq u_{nk}\right] - \mathbb{P}\left[\{\cap_{j=1}^{r} M(E_j) \leq u_{nk}\} \bigcap \{M(I'_1) \leq u_{nk}\}\right]\\
&\leq \mathbb{P}\left[M(E_1) \leq u_{nk}\right]^r - \mathbb{P}\left[M(E_1) \leq u_{nk}\right]^{r+1} + r\alpha_{nk,m}\\
&\leq 1/r + r\alpha_{nk,m} \qquad (2.7)
\end{aligned}$$

In the last line we used the elementary fact that, for any $0 \leq p \leq 1$,

$$0 \leq p^r(1-p) \leq 1/r$$

To deal with the first term in (2.5), we use again Proposition 2.1.1

$$\left|\mathbb{P}\left[\cap_{i=0}^{k-1} M(I_i) \leq u_{nk}\right] - \mathbb{P}\left[M(I_1) \leq u_{nk}\right]^k\right| \leq k\alpha_{kn,m}$$

Since by the same argument as in (2.6),

$$|\mathbb{P}\left[M(I_i \cup I_1') \leq u_{nk}\right] - \mathbb{P}\left[M(I_1') \leq u_{nk}\right]| \leq 1/r + r\alpha_{nk,m},$$

we arrive easily at

$$\left|\mathbb{P}\left[M_{kn} \leq u_{nk}\right] - \mathbb{P}\left[M_n \leq u_{nk}\right]^k\right| \leq 2k\left((r+1)\alpha_{kn,m} + 1/r\right)$$

It suffices to choose $r \ll m \ll n$ such that $r \uparrow \infty$, and $r\alpha_{kn,m} \downarrow 0$, which is possible by assumption. $\square$

2.2 Equivalence to iid sequences. Condition D'

The extremal types theorem is a strong statement about universality of extremal distributions whenever some nontrivial rescaling exists that leads to convergence of the distribution of the maximum. But when is this the case, and more particularly, when do we have the same behaviour as in the iid case, i.e. when does $n(1 - F(u_n)) \to \tau$ imply $\mathbb{P}\left[M - n \leq u_n\right] \to e^{-\tau}$? It will turn out that $D(u_n)$ is not a sufficient condition.

An sufficient additional condition will turn out to be the following.

Definition 2.2.1 A stationary sequence of random variables X_i is said to satisfy, for a sequence, $u_n \in \mathbb{R}$, condition $D'(u_n)$, if

$$\lim_{k\uparrow\infty} \limsup_{n\uparrow\infty} n \sum_{j=1}^{[n/k]} \mathbb{P}\left[X_1 > u_n, X_j > u_n\right] = 0 \tag{2.8}$$

Proposition 2.2.1 *Let X_i be a stationary sequence of random variables, and assume that u_n is a sequence such that X_i satisfy $D(u_n)$ and $D'(u_n)$. Then, for $0 \leq \tau < \infty$,*

$$\lim_{n\uparrow\infty} \mathbb{P}\left[M_n \leq u_n\right] = e^{-\tau} \tag{2.9}$$

if and only if

$$\lim_{n\uparrow\infty} n\mathbb{P}\left[X_1 > u_n\right] = \tau. \tag{2.10}$$

Proof Let $n' \equiv [n/k]$. We show first that (2.10) implies (2.9). We have seen in the proof of the preceding theorem that, under condition $D(u_n)$,

$$\mathbb{P}\left[M_n \leq u_n\right] \sim \left(\mathbb{P}\left[M_{n'} \leq u_n\right]\right)^k. \tag{2.11}$$

Thus, (2.9) will follow if we can show that

$$\mathbb{P}\left[M_{n'} \leq u_n\right] \sim (1-\tau/k).$$

Now clearly

$$\mathbb{P}\left[M_{n'} \leq u_n\right] = 1 - \mathbb{P}\left[M_{n'} > u_n\right]$$

and

$$\mathbb{P}\left[M_{n'} > u_n\right] \leq \sum_{i=1}^{n'} \mathbb{P}\left[X_i > u_n\right] = \frac{n'}{n} n\mathbb{P}\left[X_1 > u_n\right] \to \tau/k$$

On the other hand, we also have the converse bound[1]

$$\mathbb{P}\left[M_{n'} > u_n\right] \geq \sum_{i=1}^{n'} \mathbb{P}\left[X_i > u_n\right] - \sum_{i<j}^{n'} \mathbb{P}\left[X_i > u_n, X_j > u_n\right].$$

All we need is to show that the extra term vanishes faster than the first one. But this is ensured by $D'(u_n)$:

$$\sum_{i<j}^{n'} \mathbb{P}\left[X_i > u_n, X_j > u_n\right] \leq n' \sum_{j=2}^{n'} \mathbb{P}\left[X_1 > u_n, X_j > u_n\right] \leq \frac{1}{k} o(1),$$

where $o(1)$ tends to zero as $k \uparrow \infty$. Thus (2.9) follows.

To prove the converse direction, note that (2.9) together with (2.11) implies that

$$1 - \mathbb{P}\left[M_{n'} \leq u_n\right] \sim 1 - e^{-\tau/k}$$

But we have just seen that under $D'(u_n)$,

$$1 - \mathbb{P}\left[M_{n'} \leq u_n\right] \sim n'(1 - F(u_n))$$

and so

$$n'(1 - F(u_n) \sim k^{-1} n(1 - F(u_n) \sim 1 - e^{-\tau/k},$$

so that, letting $k \uparrow \infty$, $n(1 - F(u_n) \to \tau$ follows. □

2.3 Two approximation results

In this section we collect some results that are rather technical but that will be convenient later.

[1] By the inclusion-exclusion principle, see Section 3.1.

Lemma 2.3.1 *Let X_i be a stationary sequence of random variables with marginal distribution function F, and let u_n, v_n sequences of real numbers. Assume that*

$$\lim_{n\uparrow\infty} n\left(F(u_n) - F(v_n)\right) = 0 \tag{2.12}$$

Then the following hold:

(i) If I_n are interval of length $\nu_n = O(n)$, then

$$\mathbb{P}\left[M(I_n) \leq u_n\right] - \mathbb{P}\left[M(I_n) \leq v_n\right] \to 0 \tag{2.13}$$

(ii) Conditions $D(u_n)$ and $D(v_n)$ are equivalent.

Proof Let us define

$$F_{k_1,\dots,k_m}(u) \equiv \mathbb{P}\left[\cap_{i=1}^m X_{k_i} \leq u\right]. \tag{2.14}$$

Assume, without loss of generality, that $v_n \leq u_n$. Then

$$|F_{k_1,\dots,k_m}(u_n) - F_{k_1,\dots,k_m}(v_n)| \leq \mathbb{P}\left[\cup_{i=1}^m v_n < X_{k_i} < u_n\right] \leq m\,|F(u_n) - F(v_n)|.$$

Thus, if for some $K < \infty$, $m \leq Kn$,

$$|F_{k_1,\dots,k_m}(u_n) - F_{k_1,\dots,k_m}(v_n)| \to 0.$$

Choosing $m = \nu_n$, this implies immediately (i). To prove (ii), assume $D(u_n)$.

Set

$$\mathbf{i} = i_1, \dots, i_p, \quad \mathbf{j} = j_1, \dots, j_q$$

with $i_1 < i_2 < \dots i_p < j_1 < \dots < j_q$ with $j_i - i_p > \ell$. Then

$$|F_{\mathbf{ij}}(u_n) - F_{\mathbf{i}}(u_n)F_{\mathbf{j}}(u_n)| \leq \alpha_{n,\ell}$$

But

$$\begin{aligned}
&|F_{\mathbf{ij}}(v_n) - F_{\mathbf{i}}(v_n)F_{\mathbf{j}}(v_n)| \\
&\leq |F_{\mathbf{ij}}(v_n) - F_{\mathbf{ij}}(u_n)| + F_{\mathbf{i}}(u_n)\,|F_{\mathbf{j}}(v_n) - F_{\mathbf{j}}(u_n)| + F_{\mathbf{j}}(v_n)\,|F_{\mathbf{i}}(u_n) - F_{\mathbf{i}}(v_n)| \\
&+ |F_{\mathbf{ij}}(u_n) - F_{\mathbf{i}}(u_n)F_{\mathbf{j}}(u_n)|.
\end{aligned}$$

But all terms tend to zero as n and ℓ tend to infinity, so $D(v_n)$ holds $\square$

Lemma 2.3.2 *Let u_n be a sequence such that $n(1 - F(u_n)) \to \tau$. Let $v_n \equiv u_{[n/\theta]}$, for some $\theta > 0$. Then the following hold:*

(i) $n(1 - F(v_n)) \to \theta\tau$,

(ii) *if* $\theta \leq 1$*, then* $D(u_n) \Rightarrow D(v_n)$*,*
(iii) *if* $\theta \leq 1$*, then* $D'(u_n) \Rightarrow D'(v_n)$*, and,*
(iv) *if for* w_n*,* $n(1 - F(w_n)) \to \tau' \leq \tau$*, then* $D(u_n) \Rightarrow D(w_n)$*.*

Proof The proof is fairly straightforward. (i):

$$n(1 - F(v_n)) = n(1 - F(u_{[n/\theta]})) = \frac{n}{[n/\theta]}[n/\theta](1 - F(u_{[n/\theta]})) \to \theta\tau$$

(ii): with $\mathbf{i}, \mathbf{j}$ as in the preceding proof,

$$|F_{\mathbf{ij}}(v_n) - F_{\mathbf{i}}(v_n)F_{\mathbf{j}}(v_n)| = |F_{\mathbf{ij}}([n/\theta]) - F_{\mathbf{i}}([n\theta])F_{\mathbf{j}}([n/\theta])| \leq \alpha_{[n/\theta],\ell}$$

which implies $D(v_n)$.
(iii): If $\theta \leq 1$,

$$\begin{aligned} & n \sum_{i=1}^{[n/k]} \mathbb{P}\left[X_1 > v_n, X_i > v_n\right] \\ & \leq \frac{n}{[n/\theta]}[n/\theta] \sum_{i=1}^{[[n/\theta]/k]} \mathbb{P}\left[X_1 > u_{[n/\theta]}, X_i > u_{[n/\theta]}\right] \downarrow 0. \end{aligned}$$

(iv): Let $\tau' = \theta\tau$. By (iii), $D(v_n)$ holds, and $n(1 - F(v_n)) \to \theta\tau = \tau'$. This by (ii) of Lemma 2.3.1 implies $D(w_n)$. $\square$

The following assertion is now immediate.

Theorem 2.3.3 *Let* u_n, v_n *be such that* $n(1 - F(u_n)) \to \tau$ *and* $n(1 - F(v_n)) \to \theta\tau$*. Assume* $D(v_n)$ *and* $D'(v_n)$*. Then, for intervals* I_n *with* $|I_n| = [\theta n]$*,*

$$\lim_{n \uparrow \infty} \mathbb{P}\left[M(I_n) \leq u_n\right] = e^{-\theta\tau}. \tag{2.15}$$

We leave the proof as an exercise.

2.4 The extremal index

We have seen that under conditions $D(u_n), D'(u_n)$, extremes of stationary dependent sequences behave just as if the sequences were independent. Of course it will be interesting to see what can be said if these conditions do not hold. The following important theorem tells us what $D(u_n)$ alone can imply.

Theorem 2.4.1 *Assume that, for all* $\tau > 0$*, there is* $u_n(\tau)$*, such that* $n(1 - F(u_n(\tau)) \to \tau$*, and that* $D(u_n(\tau))$ *holds for all* $\tau > 0$*. Then there exists* $0 \leq \theta \leq \theta' \leq 1$*, such that*

$$\limsup_{n \uparrow \infty} \mathbb{P}\left[M_n \leq u_n(\tau)\right] = e^{-\theta\tau} \tag{2.16}$$

$$\liminf_{n \uparrow \infty} \mathbb{P}\left[M_n \leq u_n(\tau)\right] = e^{-\theta'\tau} \tag{2.17}$$

Moreover, if, for some τ*,* $\mathbb{P}\left[M_n \leq u_n(\tau)\right]$ *converges, then* $\theta' = \theta$*.*

Proof We had seen that under $D(u_n)$,

$$\mathbb{P}\left[M_n \leq u_n\right] - \left(\mathbb{P}\left[M_{[n/k]} \leq u_n\right]\right)^k \to 0,$$

and so, if

$$\limsup_{n \uparrow \infty} \mathbb{P}\left[M_n \leq u_n(\tau)\right] = \psi(\tau),$$

then

$$\limsup_{n \uparrow \infty} \mathbb{P}\left[M_{[n/k]} \leq u_n(\tau)\right] = \psi^{1/k}(\tau).$$

It also holds that

$$\limsup_{n \uparrow \infty} \mathbb{P}\left[M_{[n/k]} \leq u_{[n/k]}(\tau/k)\right] = \psi(\tau/k).$$

Thus, if we can show that

$$\limsup_{n \uparrow \infty} \mathbb{P}\left[M_{[n/k]} \leq u_{[n/k]}(\tau/k)\right] = \limsup_{n \uparrow \infty} \mathbb{P}\left[M_{[n/k]} \leq u_n(\tau)\right], \tag{2.18}$$

then $\psi^k(\tau/k) = \psi(\tau)$ for all τ and all k, which has as its only solutions $\psi(\tau) = e^{-\theta\tau}$. To show (2.18), assume without loss of generality $u_{[n/k]}(\tau/k) \geq u_n(\tau)$. Then

$$\begin{aligned}
&\left|\mathbb{P}\left[M_{[n/k]} \leq u_{[n/k]}(\tau/k)\right] - \mathbb{P}\left[M_{[n/k]} \leq u_n(\tau)\right]\right| \\
&\leq [n/k]\left|F(u_{[n/k]}(\tau/k)) - F(u(\tau))\right| \\
&= \frac{[n/k]}{n}\left|\frac{n}{[n/k]}[n/k](1 - F(u_{[n/k]}(\tau/k))) - n(1 - F(u(\tau)))\right| \\
&= \frac{[n/k]}{n}\left|k(\tau/k) - \tau + o(1)\right| \downarrow 0
\end{aligned}$$

Thus we have proven the assertion for the limsup. The assertion for the liminf is completely analogous, with possibly a different value, θ'.

Clearly, if for some τ, the limsup and the liminf agree, then $\theta = \theta'$. $\square$

Definition 2.4.1 If a sequence of random variables, X_i, has the property that there exist $u_n(\tau)$ such that $n(1-F(u_n(\tau))) \to \tau$ and $\mathbb{P}[M_n \leq u_n(\tau)] \to e^{-\theta\tau}$, $0 \leq \theta \leq 1$, one says that the sequence X_i has *extremal index* θ.

The extremal index can be seen as a measure of the effect of dependence on the maximum.

One can give a slightly different version of the preceding theorem in which the idea that we are *comparing* a stationary sequence with an iid sequence becomes even more evident. If X_i is a stationary random sequence with marginal distribution function F, denote by $\widehat{X}_i$ a sequence of iid random variables that have F as their common distribution function. Let M_n and $\widehat{M}_n$ denote the respective maxima.

Theorem 2.4.2 *Let X_i be a stationary sequence that has extremal index $\theta \leq 1$. Let v_n be a real sequence and $0 \leq \rho \leq 1$. Then,*

(i) for $\theta > 0$, if

$$\mathbb{P}\left[\widehat{M}_n \leq v_n\right] \to \rho, \quad \text{then } \mathbb{P}[M_n \leq v_n] \to \rho^\theta \tag{2.19}$$

(ii) for $\theta = 0$,

$$\text{(a) if } \liminf_{n\uparrow\infty} \mathbb{P}\left[\widehat{M}_n \leq v_n\right] > 0, \quad \text{then } \mathbb{P}[M_n \leq v_n] \to 1,$$

$$\text{(b) if } \limsup_{n\uparrow\infty} \mathbb{P}[M_n \leq v_n] < 1, \quad \text{then } \mathbb{P}\left[\widehat{M}_n \leq v_n\right] \to 0.$$

Proof (i): Choose $\tau > 0$ such that $e^{-\tau} < \rho$. Then

$$\mathbb{P}\left[\widehat{M}_n \leq u_n(\tau)\right] \to e^{-\tau} \quad \text{and } \mathbb{P}\left[\widehat{M}_n \leq v_n\right] \to \rho > e^{-\tau}.$$

Therefore, for n large enough, $v_n \geq u_n(\tau)$, and so

$$\liminf_{n\uparrow\infty} \mathbb{P}[M_n \leq v_n] \geq \lim_{n\uparrow\infty} \mathbb{P}[M_n \leq u_n(\tau)] \to e^{-\theta\tau}.$$

As this holds whenever $e^{-\tau} > \rho$, it follows that

$$\liminf_{n\uparrow\infty} \mathbb{P}[M_n \leq v_n] \geq \rho^\theta.$$

In much the same way we show also that

$$\limsup_{n\uparrow\infty} \mathbb{P}[M_n \leq v_n] \leq \rho^\theta,$$

which concludes the argument for (i).

(ii): Since $\theta = 0$, $\mathbb{P}[M_n \le u_n(\tau)] \to 1$ for all $\tau > 0$. If $\liminf_{n\uparrow\infty} \mathbb{P}\left[\widehat{M}_n \le v_n\right] = \rho > 0$, and $e^{-\tau} < \rho$, then $v_n > u_n(\tau)$ for all large n, and thus

$$\liminf_{n\uparrow\infty} \mathbb{P}[M_n \le v_n] \ge \liminf_{n\uparrow\infty} \mathbb{P}[M_n \le u_n(\tau)] = 1,$$

which implies (a). If, on the other hand, $\limsup_{n\uparrow\infty} \mathbb{P}[M_n \le v_n] < 1$, while for all $\tau < \infty$, $\mathbb{P}[M_n \le u_n(\tau)] \to 1$, then, for all $\tau > 0$, and for almost all n, $v_n < u_n(\tau)$, so that

$$\limsup_{n\uparrow\infty} \mathbb{P}\left[\widehat{M}_n \le v_n\right] \le \lim_{n\uparrow\infty} \mathbb{P}\left[\widehat{M}_n \le u_n(\tau)\right] = e^{-\tau},$$

from which (b) follows by letting $\tau \uparrow \infty$. □

Let us make some observations that follow easily from the preceding theorems. First, if a stationary sequence has extremal index $\theta > 0$, then $\widehat{M}_n$ has a non-degenerate limiting distribution if and only if M_n does, and these are of the same type. It is possibly to use the same scaling constants in both cases.

On the contrary, if a sequence of random variables has extremal index $\theta = 0$, then it is impossible that M_n and $\widehat{M}_n$ have non-degenerate limiting distributions with the same scaling constants.

An autoregressive sequence. A nice example of an sequence with extremal index less than one is given by the stationary first-order autoregressive sequence, ξ_n, defined by

$$\xi_n = r^{-1}\xi_{n-1} + r^{-1}\epsilon_n, \tag{2.20}$$

where $r \ge 2$ is an integer, and the ϵ_n are iid random variables that are uniformly distributed on the set $\{0, 1, 2, \ldots, r-1\}$. ϵ_n is independent of ξ_{n-1}.

Note that if we assume that ξ_0 is uniformly distributed on $[0, 1]$, then the same holds true for all ξ_n, $n \ge 0$. Thus, with $u_n(\tau) = 1 - \tau/n$, $n\mathbb{P}[\xi_n > u_n(\tau)] = \tau$.

The following result was proven by Chernick [3].

Theorem 2.4.3 *For the sequence ξ_n defined above, for any $x \in \mathbb{R}_+$,*

$$\mathbb{P}[M_n \le 1 - x/n] \to \exp\left(-\frac{r-1}{r}x\right) \tag{2.21}$$

The proof of this theorem relies on the following key technical lemma.

Lemma 2.4.4 *In the setting above, if m is such that $1 > r^m x/n$, then*

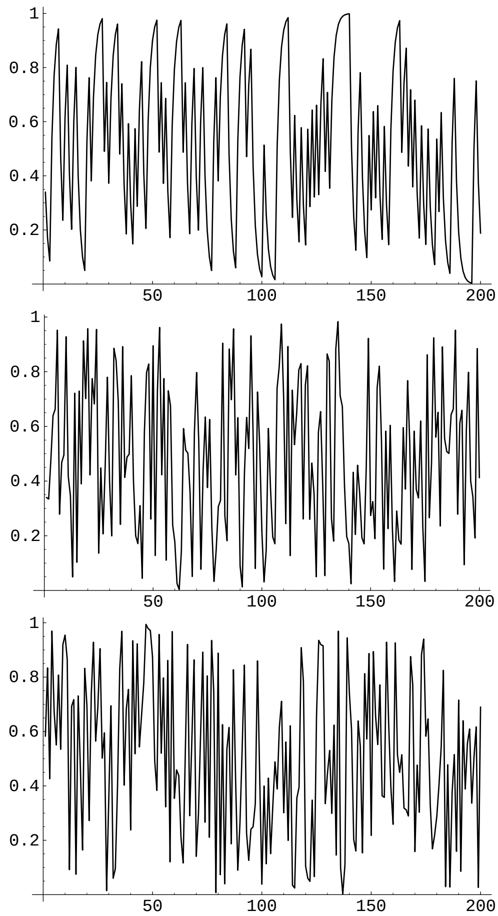

Fig. 2.1. The ARP process with $r = 2$, $r = 7$; for comparison an iid uniform sequence. Note the pronounced double peaks in the case $r = 2$.

$$\mathbb{P}\left[M_m \leq 1 - x/n\right] = 1 - \frac{(m+1)r - m}{rn}x \tag{2.22}$$

Proof The basic idea of the proof is of course to use the recursive definition of the variables ξ_n to derive a recursion for the distribution of their maxima. Apparently,

$$\begin{aligned}\mathbb{P}\left[M_m \le 1-x/n\right] &= \mathbb{P}\left[M_{m-1} \le 1-x/n, \xi_m \le 1-x/n\right] \qquad (2.23)\\ &= \mathbb{P}\left[M_{m-1} \le 1-x/n, r^{-1}\xi_{m-1} + r^{-1}\epsilon_m \le 1-x/n\right]\\ &= \mathbb{P}\left[M_{m-1}x \le 1-x/n, \xi_{m-1} \le r-\epsilon_m - xr/n\right]\\ &= \sum_{\epsilon=0}^{r-1} r^{-1}\mathbb{P}\left[M_{m-1} \le 1-x/n, \xi_{m-1} \le r-\epsilon - xr/n\right]\end{aligned}$$

Now let $rx/n < 1$. Then, for all $\epsilon \le r-2$, it is true that $r-\epsilon-rx/n \ge 2-rx/n > 1$, and so, since $\xi_{m-1} \le M_{m-1} \le 1-x/n$, for all these ϵ, the condition $\xi_{m-1} \le r-\epsilon-xr/n$ is trivially satisfied. Thus, for these x,

$$\begin{aligned}\mathbb{P}\left[M_m \le 1-x/n\right] &= \frac{r-1}{r}\mathbb{P}\left[M_{m-1} \le 1-x/n\right] \qquad (2.24)\\ &\quad + r^{-1}\mathbb{P}\left[M_{m-1} \le 1-x/n, r^{-1}\xi_{m-1} \le 1-xr/n\right]\end{aligned}$$

We see that even with this restriction we do not get closed formula involving only the M_m. But, in the same way as before, we see that, for $i \ge 1$, if $r^{i+1}x/n < 1$, then

$$\begin{aligned}\mathbb{P}\left[M_m \le 1-x/n, \xi_m < 1-r^i x/n\right] &= \frac{r-1}{r}\mathbb{P}\left[M_{m-1} \le 1-x/n\right]\\ + r^{-1}\mathbb{P}\left[M_{m-1} \le 1-x/n, r^{-1}\xi_{m-1} \le 1-xr^{i+1}/n\right] & \qquad (2.25)\end{aligned}$$

That, is, if we set

$$\mathbb{P}\left[M_m \le 1-x/n, \xi_m < 1-r^i x/n\right] \equiv A_{m,i},$$

we have the recursive set of equations

$$A_{m,i} = \frac{r-1}{r}A_{m-1,0} + \frac{1}{r}A_{m-1,i+1}. \qquad (2.26)$$

If we iterate this relation k times, we clearly get an expression for $A_{m,0}$ of the form

$$A_{m,0} = \sum_{\ell=0}^{k} C_{k,\ell}A_{m-k,\ell} \qquad (2.27)$$

with constants $C_{k,\ell}$ that we will now determine. To do this, use (2.26) to re-express the right-hand side as

$$\sum_{\ell=0}^{k} C_{k,\ell}\left[\frac{r-1}{r}A_{m-k-1,0}+\frac{1}{r}A_{m-k-1,\ell+1}\right] \tag{2.28}$$

$$=\frac{r-1}{r}\sum_{\ell=0}^{k} C_{k,\ell}A_{m-k-1,0}+\frac{1}{r}\sum_{\ell=1}^{k+1} C_{k,\ell-1}A_{m-k-1,\ell} \tag{2.29}$$

$$=\sum_{\ell=0}^{k+1} C_{k+1,\ell}A_{m-k-1,\ell}, \tag{2.30}$$

where

$$C_{k+1,0}=\frac{r-1}{r}\sum_{\ell=0}^{k} C_{k,\ell} \tag{2.31}$$

$$C_{k+1,\ell}=r^{-1}C_{k,\ell-1}, \quad \text{for } \ell \geq 1 \tag{2.32}$$

Solving this recursion turns out very easy. Namely, if we set $x=0$, then of course all $A_{k,\ell}=1$, and therefore, for all k, $\sum_{\ell=0}^{k} C_{k,\ell}=1$, so that

$$C_{k,0}=\frac{r-1}{r}, \quad \text{for all } k\geq 1.$$

Also, obviously $C_{0,0}=1$. Iterating the second equation, we see that

$$C_{k,\ell}=r^{-\ell}C_{k-\ell,0}=\begin{cases} r^{-\ell-1}(r-1), & \text{if } k>\ell \\ r^{-\ell}, & \text{if } k=\ell \end{cases}$$

We can now insert this into Eq.(2.27), to get

$$\begin{aligned}
\mathbb{P}\left[M_m \leq 1-x/n\right] &= \sum_{\ell=0}^{m} C_{m,\ell}\mathbb{P}\left[M_0\leq 1-x/n, \xi_0<1-r^{\ell}x/n\right]\\
&= \sum_{\ell=0}^{m} C_{m,\ell}\mathbb{P}\left[\xi_0<1-r^{\ell}x/n\right]\\
&= \sum_{\ell=0}^{m} C_{m,\ell}[1-r^{\ell}x/n]\\
&= \sum_{\ell=0}^{m-1} (r-1)r^{-\ell-1}[1-r^{\ell}x/n]+r^{-m}[1-r^{m}x/n]\\
&= (r-1)\frac{1-r^{-m}}{r-1}-m\frac{r-1}{r}\frac{x}{n}+r^{-m}-\frac{x}{n}\\
&= 1-\frac{r(m+1)-m}{rn}x
\end{aligned} \tag{2.33}$$

which proves the lemma. □

We can now readily prove the theorem.

Proof We would like to use that, for m satisfying the hypothesis of the lemma,

$$\mathbb{P}\left[M_n \leq 1 - x/n\right] \sim \mathbb{P}\left[M_m \leq 1 - x/n\right]^{n/m}. \tag{2.34}$$

The latter, by the assertion of the lemma, converges to $\exp\left(-\frac{r-1}{r}x\right)$. To prove (2.34), we show that $D(1-x/n)$ holds. In fact, what we will see is that the correlations of the variables ξ_i decay exponentially fast. By construction, if $j > i$, the variable ξ_j consists of a large piece that is independent of ξ_i, plus $r^{-j-i}\xi_i$.

With the notation introduced earlier, consider

$$\begin{aligned} &F_{\mathbf{ij}}(u_n) - F_{\mathbf{i}}(u_n)F_{\mathbf{j}}(u_n) = \\ &F_{\mathbf{i}}(u_n)\Big(\mathbb{P}\left[\xi_{j_1} \leq u_n, \dots, \xi_{j_q} \leq u_n \middle| \xi_{i_1} \leq u_n, \dots, \xi_{i_p} \leq u_n\right] \\ &\qquad - \mathbb{P}\left[\xi_{j_1} \leq u_n, \dots, \xi_{j_q} \leq u_n\right]\Big) \end{aligned} \tag{2.35}$$

Now

$$\xi_{j_k} = r^{-1}\xi_{j_k-1} + r^{-1}\epsilon_{j_k} = \dots = r^{-\ell}\xi_{j_k-\ell} + W_{j_k}^{(\ell)}$$

where $W_{j_k}^{(\ell)}$ is independent of all ξ_i with $i \leq j_k - \ell$. Thus

$$\begin{aligned} &\mathbb{P}\left[\xi_{j_1} \leq u_n, \dots, \xi_{j_q} \leq u_n \middle| \xi_{i_1} \leq u_n, \dots, \xi_{i_p} \leq u_n\right] \\ &= \mathbb{P}\Big[W_{j_1}^{(j_1-i_p)} + r^{-(j_1-i_p)}\xi_{i_p} \leq u_n, \dots, \\ &\qquad W_{j_q}^{(j_q-i_p)} + r^{-(j_q-i_p)}\xi_{i_p} \leq u_n \Big| \xi_{i_1} \leq u_n, \dots, \xi_{i_p} \leq u_n\Big] \\ &\leq \mathbb{P}\left[W_{j_1}^{(j_1-i_p)} \leq u_n, \dots, W_{j_q}^{(j_q-i_p)} \leq u_n\right] \end{aligned}$$

and

$$\begin{aligned} &\mathbb{P}\left[\xi_{j_1} \leq u_n, \dots, \xi_{j_q} \leq u_n \middle| \xi_{i_1} \leq u_n, \dots, \xi_{i_p} \leq u_n\right] \\ &\geq \mathbb{P}\left[W_{j_1}^{(j_1-i_p)} + r^{-(j_1-i_p)} \leq u_n, \dots, W_{j_q}^{(j_q-i_p)} + r^{-(j_q-i_p)} \leq u_n\right] \end{aligned}$$

But similarly,

$$\begin{aligned} &\mathbb{P}\left[\xi_{j_1} \leq u_n, \dots, \xi_{j_q} \leq u_n\right] \\ &\leq \mathbb{P}\left[W_{j_1}^{(j_1-i_p)} \leq u_n, \dots, W_{j_q}^{(j_q-i_p)} \leq u_n\right] \end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left[\xi_{j_1} \le u_n, \dots, \xi_{j_q} \le u_n\right] \\
&\ge \mathbb{P}\left[W_{j_1}^{(j_1-i_p)} + r^{-(j_1-i_p)} \le u_n, \dots, W_{j_q}^{(j_q-i_p)} + r^{-(j_q-i_p)} \le u_n\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Big|\mathbb{P}\left[\xi_{j_1} \le u_n, \dots, \xi_{j_q} \le u_n \middle| \xi_{i_1} \le u_n, \dots, \xi_{i_p} \le u_n\right] \\
&\qquad\qquad - \mathbb{P}\left[\xi_{j_1} \le u_n, \dots, \xi_{j_q} \le u_n\right]\Big| \\
&\le \Big|\mathbb{P}\left[W_{j_1}^{(j_1-i_p)} \le u_n, \dots, W_{j_q}^{(j_q-i_p)} \le u_n\right] \\
&- \mathbb{P}\left[W_{j_1}^{(j_1-i_p)} + r^{-(j_1-i_p)} \le u_n, \dots, W_{j_q}^{(j_q-i_p)} + r^{-(j_q-i_p)} \le u_n\right]\Big| \\
&\le \sum_{k=1}^{q} \mathbb{P}\left[u_n - r^{-(j_k-i_p)} \le W_{j_k}^{(j_k-i_p)} \le u_n\right]
\end{aligned}$$

But

$$\begin{aligned}
&\mathbb{P}\left[u_n - r^{-(j_k-i_p)} \le W_{j_k}^{(j_k-i_p)} \le u_n\right] \\
&\le \mathbb{P}\left[u_n - r^{-(j_k-i_p)} \le \xi_{j_k} \le u_n + r^{-(j_k-i_p)}\right] \le 2r^{-(j_k-i_p)}
\end{aligned}$$

which implies that

$$\begin{aligned}
&\Big|\mathbb{P}\left[\xi_{j_1} \le u_n, \dots, \xi_{j_q} \le u_n \middle| \xi_{i_1} \le u_n, \dots, \xi_{i_p} \le u_n\right] \\
&\qquad\qquad - \mathbb{P}\left[\xi_{j_1} \le u_n, \dots, \xi_{j_q} \le u_n\right]\Big| \le \sum_{k=1}^{q} r^{-(j_k-i_p)} \le \frac{r^{-\ell}}{r-1}
\end{aligned}$$

which implies $D(1-x/n)$. □

Remark 2.4.1 We remark that is it easy to see directly that condition $D'(1-x/n)$ does not hold. In fact,

$$\begin{aligned}
n \sum_{i=2}^{[n/k]} P\left[\xi_1 > u_n, \xi_j > u_n\right] &\ge n \sum_{i=2}^{[n/k]} P\left[\xi_1 > u_n, \forall_{2 \le j \le i} \epsilon_j = r-1\right] \\
&= \sum_{i=2}^{[n/k]} r^{-i+1} > 0 \qquad (2.36)
\end{aligned}$$

We see that the appearance of a non-trivial extremal index is related to strong correlations between the random variables with neighboring indices, a fact that condition D' is precisely excluding.

3

Non-stationary sequences

3.1 The inclusion-exclusion principle

One of the key relations in the analysis of iid sequences was the observation that

$$n(1 - F(u_n)) \to \tau \quad \Leftrightarrow \quad \mathbb{P}[M_n \leq u_n] \to e^{-\tau}. \tag{3.1}$$

This relation was also instrumental for the Poisson distribution of the number of crossings of extreme levels. The key to this relation was the fact that in the iid case,

$$\mathbb{P}[M_n \leq u_n] = F^n(u_n) = \left(1 - \frac{n(1 - F(u_n))}{n}\right)^n$$

which of course converges to $e^{-\tau}$. The first equality fails of course in the dependent case. However, this is equation is also far from necessary.

The following simple lemma gives a much weaker, and, as we will see, useful, criterium for convergence to the exponential function.

Lemma 3.1.1 *Assume that a sequence A_n satisfies, for any $s \in \mathbb{N}$, the bounds*

$$A_n \leq \sum_{\ell=0}^{2s} \frac{(-1)^\ell}{\ell!} a_\ell(n) \tag{3.2}$$

$$A_n \geq \sum_{\ell=0}^{2s+1} \frac{(-1)^\ell}{\ell!} a_\ell(n) \tag{3.3}$$

and, for any $\ell \in \mathbb{N}$,

$$\lim_{n\uparrow\infty} a_\ell(n) = a^\ell \tag{3.4}$$

Then

$$\lim_{n\uparrow\infty} A_n = e^{-a} \tag{3.5}$$

Proof Obviously the hypothesis of the lemma imply that, for all $s \in \mathbb{N}$,

$$\limsup_{n\uparrow\infty} A_n \leq \sum_{\ell=0}^{2s} \frac{(-\tau)^\ell}{\ell!} \tag{3.6}$$

$$\liminf_{n\uparrow\infty} A_n \geq \sum_{\ell=0}^{2s+1} \frac{(-\tau)^\ell}{\ell!} \tag{3.7}$$

But the upper and lower bounds are the partial series of the exponential function $e^{-\tau}$, which are absolutely convergent, and this implies convergence of A_n to this values. $\square$

The reason that one may expect $\mathbb{P}[M_n \leq u_n]$ to satisfy bounds of this form lies in the *inclusion-exclusion principle*

Theorem 3.1.2 *Let $\mathcal{B}_i$, $i \in \mathbb{N}$ be a sequence of events, and let $\mathbb{1}_\mathcal{B}$ denote the indicator function of $\mathcal{B}$. Then, for all $s \in \mathbb{N}$,*

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} \leq \sum_{\ell=0}^{2s} (-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{1}_{\cap_{r=1}^\ell \mathcal{B}_{j_r}^c} \tag{3.8}$$

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} \geq \sum_{\ell=0}^{2s+1} (-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{1}_{\cap_{r=1}^\ell \mathcal{B}_{j_r}^c} \tag{3.9}$$

Note that terms with $\ell > n$ are treated as zero.

Remark 3.1.1 Note that the sum over subsets $\{i_1, \dots, i_\ell\}$ is over all ordered subsets, i.e., $1 \leq i_1 < i_2 < \cdots < i_\ell \leq n$.

Proof We write first

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} = 1 - \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c}$$

We will prove the theorem by induction over n. The key observation is that

$$\begin{aligned} \mathbb{1}_{\cup_{i=1}^{n+1} \mathcal{B}_i^c} &= \mathbb{1}_{\mathcal{B}_{n+1}^c} + \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} \mathbb{1}_{\mathcal{B}_{n+1}} \\ &= \mathbb{1}_{\mathcal{B}_{n+1}^c} + \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} - \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} \mathbb{1}_{\mathcal{B}_{n+1}^c} \end{aligned} \tag{3.10}$$

To prove an upper bound of some $2s+1$, we now insert an upper bound of that order in the second term, and a lower bound of order $2s$ in the third term. It is a simple matter of inspection that this reproduces exactly the desired bounds for $n+1$. □

The inclusion-exclusion principle has an obvious corollary.

Corollary 3.1.3 *Let X_i be any sequence of random variables. Then*

$$\mathbb{P}\left[M_n \leq u\right] \leq \sum_{\ell=0}^{2s}(-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r} > u\right] \tag{3.11}$$

$$\mathbb{P}\left[M_n \leq u\right] \geq \sum_{\ell=0}^{2s+1}(-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r} > u\right] \tag{3.12}$$

Proof The proof is straightforward. □

Combining Lemma 3.1.1 and Corollary 3.1.3, we obtain a quite general criteria for triangular arrays of random variables [2].

Theorem 3.1.4 *Let X_i^n, $n \in \mathbb{N}$, $i \in \{1,\dots,n\}$ be a triangular array of random variables. Assume that, for any ℓ,*

$$\lim_{n\uparrow\infty} \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r}^n > u_n\right] = \frac{\tau^\ell}{\ell!} \tag{3.13}$$

Then,

$$\lim_{n\uparrow\infty} \mathbb{P}\left[M_n \leq u_n\right] = e^{-\tau} \tag{3.14}$$

Proof The proof of the theorem is again straightforward from the preceding results. □

Remark 3.1.2 In the iid case, (3.13) does of course hold, since here

$$\sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_r > u_n\right] = \binom{n}{\ell} n^{-\ell} \left(n(1-F(u_n))\right)^\ell$$

A special case where Theorem 3.1.4 gives an easily verifiable criterion is the case of *exchangeable random variables*.

Corollary 3.1.5 *Assume that X_i^n is a triangular array of random variables such that, for any n, the joint distribution of $X_1^n,\dots,X_n^n$ is invariant under permutation of the indices $i,\dots,n$. If, for any $\ell \in \mathbb{N}$,*

$$\lim_{n\uparrow\infty} n^\ell \mathbb{P}\left[\forall_{r=1}^{\ell} X_r^n > u_n\right] = \tau^\ell \tag{3.15}$$

Then,

$$\lim_{n\uparrow\infty} \mathbb{P}\left[M_n \leq u_n\right] = e^{-\tau} \tag{3.16}$$

Proof Again straightforward. □

Theorem 3.1.4 and its corollary have an obvious extension to the distribution the number of exceedances of extremal levels.

Theorem 3.1.6 *Let $u_n^1 > n_n^2 \cdots > u_n^r$, and let X_i^n, $n \in \mathbb{N}$, $i \in \{1,\dots,n\}$ be a triangular array of random variables. Assume that, for any $\ell \in \mathbb{N}$, and any $1 \leq s \leq r$,*

$$\lim_{n\uparrow\infty} \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r}^n > u_n^s\right] = \frac{\tau_s^\ell}{\ell!} \tag{3.17}$$

with

$$0 < \tau_1 < \tau_2 \dots, < \tau_r < \infty.$$

Then,

$$\begin{aligned}&\lim_{n\uparrow\infty} \mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \dots, S_n^r - S_n^{r-1} = k_r\right] \\ &= \frac{\tau_1^{k_1}}{k_1!} \frac{(\tau_2-\tau_1)^{k_2}}{k_2!} \dots \frac{(\tau_r-\tau_{r-1})^{k_r}}{k_r!} e^{-\tau_r}\end{aligned} \tag{3.18}$$

Proof □

In the following section we will give an application for the criteria developed in this chapter.

3.2 An application to number partitioning

The number partitioning problem is a classical optimization problem: Given N numbers $X_1, X_2, \dots, X_N$, find a way of distributing them into two groups, such that their sums in each group are as similar as possible. One can easily imagine that this problem occurs all the time in real life, albeit with additional complication: Imagine you want to stuff two moving boxes with books of different weights. You clearly have an interest in having both boxes have more or less the same weight, just so that none of them is too heavy. In computing, you want to distribute a certain number of jobs on, say, two processors, in such a way that all of your processes are executed in the fastest way, etc..

As pointed out by Mertens [8, 9], some aspects of the problem give rise to an interesting problem in extreme value theory, and in particular provide an application for our Theorem 3.1.4.

Let us identify any partition of the set $\{1, \ldots, N\}$ into two disjoint subsets, Λ_1, Λ_2, with a map, $\sigma : \{1, \ldots, N\} \to \{-1, +1\}$ via $\Lambda_1 \equiv \{i : \sigma_i = +1\}$ and $\Lambda_2 = \equiv \{i : \sigma_i = -1\}$. Then, the quantity to be minimised is

$$\left| \sum_{i \in \Lambda_1} X_i - \sum_{i \in \Lambda_2} X_i \right| = \left| \sum_{i=1}^{N} n_i \sigma_i \right| \equiv X_\sigma^{(N)}. \tag{3.19}$$

Note that our problem has an obvious symmetry: $X_\sigma^{(N)} = -X_{-\sigma}^{(N)}$. It will be reasonable to factor out this symmetry and consider σ to be an element of the set $\Sigma_N \equiv \{\sigma \in \{-1, 1\}^N : \sigma_1 = +1\}$.

We will consider, for simplicity, only the case where the n_i are replaced by independent, centered Gaussian random variables, X_i. More general cases can be treated with more analytic effort.

Thus define $Y_N(\sigma)$

$$Y_N(\sigma) \equiv N^{-1/2} \sum_{i=1}^{N} \sigma_i X_i \tag{3.20}$$

and let

$$X_\sigma^{(N)} = -|Y_N(\sigma)| \tag{3.21}$$

The first result will concern the distribution of the largestest values of $H_N(\sigma)$.

Theorem 3.2.7 *Assume that the random variables X_i are indendent, standard normal Gaussian random variables, i.e. $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$. Then,*

$$\mathbb{P}\left[\max_{s \in \Sigma_N} X_\sigma^{(N)} \leq C_N x \right] \to e^{-x} \tag{3.22}$$

We will now prove Theorem 3.2.7. In view of Theorem 3.1.4 we wil be done if we can prove the following:

Proposition 3.2.8 *Let $K_N = 2^N (2\pi)^{-1/2}$. We write $\sum_{\sigma^1, \ldots, \sigma^l \in \Sigma_N} (\cdot)$ for the sum over all possible ordered sequences of different elements of Σ_N. Then, for any $l \in \mathbb{N}$ and any constants $c_j > 0$, $j = 1, \ldots, \ell$, we have:*

$$\sum_{\sigma^1, \ldots, \sigma^\ell \in \Sigma_N} \mathbb{P}\left[K_N |Y_N(\sigma^j)| < c_j \ , \forall_{j=1}^{\ell} \right] \to \prod_{j=1, \ldots, \ell} c_j \tag{3.23}$$

Heuristics. Let us first outline the main steps of the proof. The random variables $Y_N(\sigma)$ are Gaussian random variables with mean zero covariance matrix $B_N(\sigma^1,\dots,\sigma^l)$, whose elements are

$$b_{m,n} = \mathrm{cov}\left(Y_N(\sigma^m), Y_N(\sigma^n)\right) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^m\sigma_i^n \tag{3.24}$$

In particular, $b_{m,m} = 1$. Moreover, for the vast majority of choices, $\sigma^1,\dots,\sigma^\ell$, $b_{i,j} = o(1)$, for all $i \neq j$; in fact, this fails only for an exponentially small fraction of configurations. Thus in the typicall choices, they should behace like independent random variables. The probability defined in (3.23) is then the probability that these these Gaussians belong to the exponentially small intervals $[-c_j 2^{-N}\sqrt{2\pi}, c_j 2^{-N}\sqrt{2\pi}]$, and is thus of the order

$$\prod_{j=1,\dots,\ell} c_j 2^{-N} \tag{3.25}$$

This estimate would yield the assertion of the proposition, if all remaining terms could be ignored.

Let us turn to the remaining tiny part of $\Sigma_N^{\otimes l}$ where $\sigma^l,\dots,\sigma^l$ are such that $b_{i,j} \not\to 0$ for some $i \neq j$ as $N \to \infty$. A priori, we would be inclined to believe that there should be no problem, since the number of terms in the sum is by an exponential factor smaller than the total number of terms. In fact, we only need to worry if the corresponding probability is also going to be exponentially larger than for the bulk of terms. As it turns out, the latter situation can only arise when the covariance matrix is degenerate.

Namely, if the covariance matrix, $B_N(\sigma^1,\dots,\sigma^l)$, is non-degenerate, the probability $\mathbb{P}[\cdot]$ is of the order

$$\left(\det B_N(\sigma^1,\dots,\sigma^l)\right)^{-1/2} \prod_{j=1,\dots,\ell} 2(2\pi)^{-1/2} c_j K_N^{-1} \tag{3.26}$$

But, from the definition of $b_{i,j}$, $(\det B_N(\sigma^1,\dots,\sigma^\ell))^{-1/2}$ may grow at most polynomially. Thus, the probability $\mathbb{P}[\cdot]$ is $K_N^{-\ell}$ up to a polynomial factor, while the number of sets $\sigma^1,\dots,\sigma^\ell$ in this part is exponentially smaller than K_N^ℓ. Hence, the contribution of all such $\sigma^1,\dots,\sigma^\ell$ in (3.23) is exponentially small.

The case when $\sigma^1,\dots,\sigma^\ell$ give rise to a degenerate $B(\sigma^1,\dots,\sigma^\ell)$ is more delicate. Degeneracy of the covariance implies that there are linear relations between the random variables $\{Y(\sigma^i)\}_{i=1,\dots,\ell}$, and hence the probabilities $\mathbb{P}[\cdot]$ can be exponentially bigger than $K_N^{-\ell}$. A detailed analysis shows, however, that the total contribution from such terms is still negligible.

Proof of Proposition 3.2.8. Let us denote by $C(\vec{\sigma})$ the $\ell \times N$ matrix with elements σ_i^r. Note that $B(\vec{\sigma}) = N^{-1}C^t(\vec{\sigma})C(\vec{\sigma})$.

We will split the sum of (3.23) into two terms

$$\sum_{\sigma^1,\dots,\sigma^l \in \Sigma_N} \mathbb{P}[\cdot] = \sum_{\substack{\sigma^1,\dots,\sigma^\ell \in \Sigma_N \\ \operatorname{rank} C(\vec{\sigma})=\ell}} \mathbb{P}[\cdot] + \sum_{\substack{\sigma^1,\dots,\sigma^l \in \Sigma_N \\ \operatorname{rank} C(\vec{\sigma})<\ell}} \mathbb{P}[\cdot] \tag{3.27}$$

and show that the first term converges to the right-hand side of (3.23) while the second term converges to zero.

Lemma 3.2.9 *Assume that the matrix $C(\vec{\sigma})$ contains all 2^ℓ possible different rows. Assume that a configuration $\tilde{\sigma}$ is such that it is a linear combination of the columns of the matrix $C(\vec{\sigma})$. Then, there exists $1 \leq j \leq \ell$ such that either $\tilde{\sigma} = \sigma^{(j)}$, or $\tilde{\sigma} = -\sigma^{(j)}$.*

Proof We are looking for solutions of the set of linear equations

$$\tilde{\sigma}_i = \sum_{r=1}^{\ell} z_r \sigma_i^\ell \tag{3.28}$$

We may assumme without loss of generality that $\sigma_i^1 \equiv +1$. By that assumption that all possible assignments of signs to the rows of $C(\vec{\sigma})$ occur, there exists i such that $\sigma_i^r = -\sigma_1^r$, for all $r \geq 2$. Hence we have, in particular,

$$\begin{aligned} \tilde{\sigma}_1 &= z_1 + \sum_{r=2}^{\ell} z_r \sigma_1^\ell \\ \tilde{\sigma}_i &= z_1 - \sum_{r=2}^{\ell} z_r \sigma_1^\ell. \end{aligned} \tag{3.29}$$

Adding these two equations, we find that $z_1 \in \{-1, 0, 1\}$. If $z_i = 0$, then we are done in the case $\ell = 2$, and can continue inductively otherwise. If $z_1 \neq 0$, then $\tilde{\sigma}_i = \tilde{\sigma}_1 = z_1$, and we obtain that

$$\sum_{r=2}^{\ell} z_r \sigma_1^\ell = 0.$$

In fact, $\sum_{r=2}^{\ell} z_r \sigma_j^\ell = 0$ for all j such that $\tilde{\sigma}_j = z_1$. Now assume that there is k such that $\tilde{\sigma}_k = -z_1$. Then there must exist k' such that $\sigma^r_{k'} = -\sigma^r_k$, for all $r \geq 2$. Hence we get

$$\sum_{r=2}^{\ell} z_r \sigma^{\ell}_k = -2z_1$$

and

$$-\sum_{r=2}^{\ell} z_r \sigma^{\ell}_k = \tilde{\sigma}_{k'} - z_1.$$

But this leads to $2z_1 = \tilde{\sigma}_{k'} - z_1$, which is impossible. Thus we are left with the case when for all j, $\tilde{\sigma}_j = z_1$, and so

$$\sum_{r=2}^{\ell} z_r \sigma^{\ell}_j = 0.$$

for all j. Now consider i such $\sigma^2_i = \sigma^2_j$ and $\sigma^r_i = \sigma^r_j$ for $r \geq 3$. Then by the same reasoning as before we find $z_2 = 0$, and so $\sum_{r=3}^{\ell} z_r \sigma^{\ell}_j = 0$, for all j.
In conclusion, if $z_1 \neq 0$, then $z_1 = \tilde{\sigma}_i$, for all i, and $z_r = 0$, for all $r \geq 2$. If $z_1 = 0$, then we continue the argument until we find a k such that $z_k = \tilde{\sigma}_i$, for all i, and all other z_r are again zero. This proves the lemma. $\square$

Lemma 3.2.9 implies the following: Assume that there are $r < \ell$ linearly independent vectors, $\sigma^{i_1}, \dots, \sigma^{i_r}$, among the ℓ vectors $\sigma^1, \dots, \sigma^{\ell}$. The number of such vectors is at most $(2^r - 1)^N$. In fact, if the matrix $C(\sigma^{i_1}, \dots, \sigma^{i_r})$ contains all 2^r different rows, then by Lemma 3.2.9 the remaining configurations, σ^j with $j \in \{1, \dots, l\} \backslash \{i_1, \dots, i_r\}$, would be equal to one of $\sigma^{i_1}, \dots, \sigma^{i_r}$, as elements of Σ_N, which is impossible, since we sum over *different* elements of Σ_N. Thus there can be at most $O((2^r - 1)^N)$ ways to construct these r columns. Furthermore, there is only an N-independent number of possibilities to complete the set of vectors by $\ell - r$ linear configurations of these columns to get $C(\sigma^1, \dots, \sigma^l)$.

The next lemma gives an a priori estimate on the probability corresponding to each of these terms.

Lemma 3.2.10 *There exists a constant, $C > 0$, independent of N, such that, for any distinct $\sigma^1, \dots, \sigma^{\ell} \in \Sigma_N$, any $r = \operatorname{rank} C(\sigma^1, \dots \sigma^{\ell}) \leq \ell$, and all $N > 1$,*

$$\mathbb{P}\left[\forall_{j=1}^{\ell} |Y(\sigma^j)| < \frac{c_j}{K_N}\right] \leq C K_N^{-r} N^{r/2} \tag{3.30}$$

Proof Let us remove from the matrix $C(\sigma^1,\dots,\sigma^\ell)$ linearly dependent columns and leave only r linearly independent columns. They correspond to a certain subset of r configurations, σ^j, $j \in A_r \equiv \{j_1,\dots,j_r\} \subset \{1,\dots,\ell\}$. We denote by $\bar{C}^r(\vec{\sigma})$ the $N \times r$ matrix composed by them, and by $B^r(\vec{\sigma})$ the corresponding covariance matrix. Then the probability in the right-hand side of (3.30) is not greater than the probability of the same events for $j \in A_r$ only.

$$\mathbb{P}\left[\forall_{j=1}^{\ell} \frac{|Y(\sigma^j)|}{\sqrt{\operatorname{var} X}} < \frac{c_j}{K_N}\right] \leq \mathbb{P}\left[\forall_{j\in A_r} \frac{|Y(\sigma^j)|}{\sqrt{\operatorname{var} X}} < \frac{c_j}{K_N}\right] \tag{3.31}$$

$$\leq \frac{1}{(2\pi)^{r/2}\sqrt{\det(B^r(\vec{\sigma}))}} \int_{-c_{j_1}/K_N}^{c_{j_1}/K_N} \cdots$$

$$\cdots \int_{-c_{j_r}/K_N}^{c_{j_r}/K_N} \prod_{j\in A_r} dx_j \exp\left(- \sum_{s,s'=1}^{r} x_{j_s} \left[B^r(\vec{\sigma})\right]^{-1}_{s,s'} x_{j_{s'}} \right)$$

$$\leq \frac{1}{(2\pi)^{r/2}\sqrt{\det(B^r(\vec{\sigma}))}} (K_N)^{-r} 2^r \prod_{s=1}^{r} c_{j_s}$$

Finally, note that the elements of the matrix $B^r(\vec{\sigma})$ are of the form N^{-1} times an integer. Thus $\det B^r(\vec{\sigma})$ is N^{-r} times the determinant of a matrix with only integer entries. But the determinant of an integer matrix is an integer, and since we have ensured that the rank of $B^r(\vec{\sigma})$ is r, this integer is different from zero. Thus $\det(B^r(\vec{\sigma})) \geq N^{-r}$. Inserting this bound, we get the conclusion of the lemma. $\square$

Lemma 3.2.10 implies that each term in the second sum in (3.27) is smaller than $CK_N^{-r}N^{r/2} \sim 2^{-Nr}$. It follows that the sum over these terms is of order $O\big([(2^r-1)2^{-r}]^N\big) \to 0$ as $N \to \infty$.

We now turn to the first sum in (3.27), where the covariance matrix is non-degenerate. Let us fix $\alpha \in (0,1/2)$ and introduce a subset, $\mathcal{R}^\alpha_{l,N} \subset \Sigma_N^{\otimes l}$, through

$$\mathcal{R}^\alpha_{N,\ell} = \left\{\sigma^1,\dots,\sigma^\ell \in \Sigma_N : \forall_{1\leq m<r\leq \ell}, \left|\sum_{i=1}^{N} \sigma_i^m \sigma_i^r\right| < N^{\alpha+1/2}\right\} \tag{3.32}$$

It is easy to estimate

$$|\Sigma_N^l \backslash \mathcal{R}^\alpha_{l,N}| \leq \ell^2 2^{Nl} \exp(-N^{2\alpha}) \tag{3.33}$$

By definition, for any $(\sigma^1,\dots,\sigma^\ell) \in \mathcal{R}^\alpha_{N,\ell}$, the elements, $b_{k,m}$, of the covariance matrix, (see (3.24)) satisfy, for all $k \neq m$,

$$|b_{k,m}| = \left| N^{-1} \sum_{i=1}^{N} \sigma_i^k \sigma_i^m \right| \leq N^{\alpha - 1/2} \tag{3.34}$$

Therefore, for any $\sigma^1, \ldots, \sigma^l \in \mathcal{R}_{l,n}^{\alpha}$, $\det B_N(\vec{\sigma}) = 1 + o(1)$ and, in particular, the rank of $C(\vec{\sigma})$ equals ℓ.

By Lemma 3.2.10 and the estimate (3.33),

$$\sum_{\substack{\sigma^1,\ldots,\sigma^\ell \notin \mathcal{R}_{\ell,N}^{\alpha} \\ \mathrm{rank} C(\sigma^1,\ldots,\sigma^\ell) = \ell}} \mathbb{P}[\cdot] \leq 2^{N\ell} e^{-N^{2\alpha}} C N^{3\ell/2} K_N^{-\ell} \to 0 \tag{3.35}$$

To complete the study of the first term of (3.27), let us show that

$$\sum_{\sigma^1,\ldots,\sigma^\ell \in \mathcal{R}_{\ell,N}^{\alpha}} \mathbb{P}[\cdot] \to \prod_{j=1,\ldots,\ell} c_j \tag{3.36}$$

This is of course again obvious fromm the representation (3.31) where, in the case $r = \ell$, the inequality signs can now be replaced by equalities, and the fact that the determinant of the covariance matric is now $1+o(1)$.

$\square$

4

Normal sequences

A particular class of random variables are of course Gaussian random variables. In this case, explicit computations are far more feasible.

In the stationary case, a normalized Gaussian sequence, X_i, is characterised by

$$\begin{aligned} \mathbb{E}X_i &= 0 \\ \mathbb{E}X_i^2 &= 1 \\ \mathbb{E}X_iX_j &= r_{i-j} \end{aligned} \tag{4.1}$$

where $r_k = r_{|k|}$. The sequence must of course be such that the infinite dimensional matrix with entries $c_{ij} = r_{i-j}$ is positive definite.

Our main goal here is to show that under the so-called Berman condition,

$$r_n \ln n \downarrow 0,$$

the extremes of a stationary normal sequences behave like those of the corresponding iid normal sequence. A very nice tool for the analysis of Gaussian processes is the so-called *normal comparison lemma*.

4.1 Normal comparison

In the context of Gaussian random variables, a recurrent idea is to compare one Gaussian process to another, simpler one. The simplest one to compare with are, of course, iid variables, but the concept goes much farther.

Let us consider a family of Gaussian random variables, $\xi_1, \dots, \xi_n$, normalised to have mean zero and variance one (we refer to such Gaussian random variables as *centered normal random variables*), and let Λ^1 denote their covariance matrix. Let similarly $\eta_1, \dots, \eta_n$ be centered normal random variables with covariance matrix Λ^0.

Generally speaking, one is interested in comparing functions of these two processes that typically will be of the form

$$\mathbb{E}F(X_1, \dots, X_n)$$

where $F : \mathbb{R}^n \to \mathbb{R}$. For us the most common case would be

$$F(X_1, \dots, X_n) = \mathbb{1}_{X_1 \leq x_1, \dots, X_n \leq x_n}$$

An extraordinary efficient tool to compare such processes turns out to be *interpolation*. Given ξ and η, we define $X_1^h, \dots, X_n^h$, for $h \in [0,1]$, such X^h is normal and has covariance

$$\Lambda^h = h\Lambda^1 + (1-h)\Lambda^0,$$

i.e. $X^1 = \xi$ and $X^0 = \eta$. Clearly we can realize

$$X^h = \sqrt{h}\xi + \sqrt{1-h}\eta.$$

This leads to the following general Gaussian comparison lemma.

Lemma 4.1.1 *Let η, ξ, X^h be as above. Let $F : \mathbb{R}^n \to \mathbb{R}$ be differentiable and of moderate growth. Set $f(h) \equiv \mathbb{E}F(X_1^h, \dots, X_n^h)$. Then*

$$f(1)-f(0) = \frac{1}{2}\int_0^1 dh \sum_{i \neq j} (\Lambda_{ij}^1 - \Lambda_{ij}^0) \mathbb{E}\left(\frac{\partial^2 F}{\partial x_i \partial x_j}(X_1^h, \dots, X_n^h)\right) \tag{4.2}$$

Proof Trivially,

$$f(1) - f(0) = \int_0^1 dh \frac{d}{dh} f(h)$$

and

$$\frac{d}{dh} f(h) = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left(\frac{\partial F}{\partial x_i}\left(h^{-1/2}\xi_i - (1-h)^{-1/2}\eta_i\right)\right) \tag{4.3}$$

where of course $\frac{\partial F}{\partial x_i}$ is evaluated at X^h. To continue we use a remarkable formula for Gaussian processes, known as the *Gaussian integration by parts formula.*

Lemma 4.1.2 *Let X_i, $i \in \{1, \dots, n\}$ be a multivariate Gaussian process, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function of at most polynomial growth. Then*

$$\mathbb{E}g(X)X_i = \sum_{j=1}^n \mathbb{E}(X_i X_j) \mathbb{E}\frac{\partial}{\partial x_j} g(X) \tag{4.4}$$

We give the proof of this formula later. Applying it in (4.3) yields

$$\begin{aligned}\frac{d}{dh} f(h) &= \frac{1}{2}\sum_{i\neq j} \mathbb{E}\frac{\partial^2 F}{\partial x_j \partial x_i} \mathbb{E}\left(\xi_i\xi_j - \eta_i\eta_j\right) \qquad (4.5)\\ &= \frac{1}{2}\sum_{i\neq j}\left(\Lambda^1_{ji} - \Lambda^0_{ji}\right)\mathbb{E}\frac{\partial^2 F}{\partial x_j \partial x_i}\left(X^h_1, \ldots, X^h_n\right),\end{aligned}$$

which is the desired formula. □

The general comparison lemma can be put to various good uses. The first is a monotonicity result that is sometimes known as Kahane's theorem [5].

Theorem 4.1.3 *Let X and Y be two independent n-dimensional Gaussian vectors. Let D_1 and D_2 be subsets of $\{1, \ldots, n\} \times \{1, \ldots, n\}$. Assume that*

$$\begin{aligned}\mathbb{E}\xi_i\xi_j &\geq \mathbb{E}\eta_i\eta_j, \quad \text{if} \quad (i,j) \in D_1\\ \mathbb{E}\xi_i\xi_j &\leq \mathbb{E}\eta_i\eta_j, \quad \text{if} \quad (i,j) \in D_2 \qquad (4.6)\\ \mathbb{E}\xi_i\xi_j &= \mathbb{E}\eta_i\eta_j, \quad \text{if} \quad (i,j) \notin D_1 \cup D_2\end{aligned}$$

Let F be a function on $\mathbb{R}^n$, such that its second derivatives satisfy

$$\begin{aligned}\frac{\partial^2}{\partial x_i \partial x_j} F(x) &\geq 0, \quad \text{if} \quad (i,j) \in D_1\\ \frac{\partial^2}{\partial x_i \partial x_j} F(x) &\leq 0, \quad \text{if} \quad (i,j) \in D_2 \qquad (4.7)\end{aligned}$$

Then

$$\mathbb{E}f(\xi) \leq \mathbb{E}f(\eta) \qquad (4.8)$$

Proof The proof of the theorem can be trivially read off the preceding lemma by inserting the hypotheses into the right-hand side of (4.2). □

We will need two extensions of these results for functions that are not differentiable. The first is known as *Slepian's lemma* [11].

Lemma 4.1.4 *Let X and Y be two independent n-dimensional Gaussian vectors. Assume that*

$$\begin{aligned}\mathbb{E}\xi_i\xi_j &\geq \mathbb{E}\eta_i\eta_j, \quad \text{for all}\, i \neq j\\ \mathbb{E}\xi_i\xi_i &= \mathbb{E}\eta_i\eta_i, \quad \text{for all}\, i \qquad (4.9)\end{aligned}$$

Then

$$\mathbb{E}\max_{i=1}^{n}(\xi_i) \leq \mathbb{E}\max_{i=1}^{n}(\eta_i) \tag{4.10}$$

Proof Let

$$F_\beta(x_1,\dots,x_n) \equiv \beta^{-1}\ln\sum_{i=1}^{n} e^{\beta x_i}.$$

A simple computation shows that, for $i \neq j$,

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = -\beta \frac{e^{\beta(x_i+x_j)}}{\left(\sum_{k=1}^{n} e^{\beta x_k}\right)^2} < 0,$$

and so the theorem implies that, for all $\beta > 0$,

$$\mathbb{E}F_\beta(\xi) \leq \mathbb{E}F_\beta(\eta),$$

On the other hand,

$$\lim_{\beta\uparrow\infty} F_\beta(x_1,\dots,x_n) = \max_{i=1}^{n} x_i.$$

Thus

$$\lim_{\beta\uparrow\infty} \mathbb{E}F_\beta(\xi_1,\dots,\xi_n) \leq \lim_{\beta\uparrow\infty} \mathbb{E}F_\beta(\eta_1,\dots,\eta_n),$$

and hence (4.10) holds. □

As a second application, we want to study $\mathbb{P}[X_1 \leq u_1,\dots,X_n \leq u_n]$. This corresponds to choosing $F(X_1,\dots,X_n) = \mathbb{1}_{X_1\leq u_1,\dots,X_n\leq u_n}$.

Lemma 4.1.5 *Let ξ,η be as above. Set $\rho_{ij} \equiv \max(\Lambda^0_{ij},\Lambda^1_{ij})$, and denote by $x_+ \equiv \max(x,0)$. Then*

$$\begin{aligned} &\mathbb{P}[\xi_1 \leq u_1,\dots,\xi_n \leq u_n] - \mathbb{P}[\eta_1 \leq u_1,\dots,\eta_n \leq u_n] \\ &\leq \frac{1}{2\pi}\sum_{1\leq i<j\leq n} \frac{(\Lambda^1_{ij}-\Lambda^0_{ij})_+}{\sqrt{1-\rho_{ij}^2}} \exp\left(-\frac{u_i^2+u_j^2}{2(1+\rho_{ij})}\right). \end{aligned} \tag{4.11}$$

Proof Although the indicator function is not differentiable, we will proceed as if it was, setting

$$\frac{d}{dx}\mathbb{1}_{x\leq u} = \delta(x-u),$$

where δ denotes the Dirac delta function, i.e. $\int f(x)\delta(x-u)dx \equiv f(u)$. This can be justified e.g. by using smooth approximants of the indicator function and passing to the limit at the end (e.g. replace $\mathbb{1}_{x\leq u}$ by

$(2\pi\sigma^2)^{1/2}\int_{-\infty}^{x}\exp\left(-\frac{(z-u)^2}{2\sigma^2}\right)$, do all the computations, and pass to the limit $\sigma\downarrow 0$ at the end. With this convention, we have that, for $i\neq j$,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2 F}{\partial x_i\partial x_j}(X_1^h,\dots,X_n^h)\right) &= \mathbb{E}\prod_{k\neq i\vee j}\mathbb{1}_{X_k^h\leq u_k}\delta(X_i^h-u_i)\delta(X_j^h-u_j) \\ &\leq \mathbb{E}\delta(X_i^h-u_i)\delta(X_j^h-u_j)=\phi_h(u_i,u_j),\end{aligned} \tag{4.12}$$

where ϕ_h denotes the density of the bivariate normal distribution with covariance Λ_{ij}^h, i.e.

$$\phi_h(u_i,u_j)=\frac{1}{2\pi\sqrt{1-(\Lambda_{ij}^h)^2}}\exp\left(-\frac{u_i^2+u_j^2-2\Lambda_{ij}^h u_i u_j}{2(1-(\Lambda_{ij}^h)^2)}\right)$$

Now

$$\begin{aligned}\frac{u_i^2+u_j^2-2\Lambda_{ij}^h u_i u_j}{2(1-(\Lambda_{ij}^h)^2)} &= \frac{(u_i^2+u_j^2)(1-\Lambda_{ij}^h)+\Lambda_{ij}^h(u_i-u_j)^2}{2(1-(\Lambda_{ij}^h)^2)} \\ &\geq \frac{(u_i^2+u_j^2)}{2(1+|\Lambda_{ij}^h|)}\geq\frac{(u_i^2+u_j^2)}{2(1+\rho_{ij})},\end{aligned} \tag{4.13}$$

where $\rho_{ij}=\max(\Lambda_{ij}^0,\Lambda_{ij}^1)$. (To prove the first inequality, note that this is trivial if $\Lambda_{ij}^h\geq 0$. If $\Lambda_{ij}^h<0$, the result follows after some simple algebra). Inserting this into (4.12) gives

$$\mathbb{E}\left(\frac{\partial^2 F}{\partial x_i\partial x_j}(X_1^h,\dots,X_n^h)\right)\leq\frac{1}{2\pi\sqrt{1-\rho_{ij}^2}}\exp\left(-\frac{(u_i^2+u_j^2)}{2(1+\rho_{ij})}\right), \tag{4.14}$$

from which (4.11) follows immediately. □

Remark 4.1.1 It is often convenient to replace the assertion of Lemma 4.1.5 by

$$\begin{aligned}&|\mathbb{P}\left[\xi_1\leq u_1,\dots,\xi_n\leq u_n\right]-\mathbb{P}\left[\eta_1\leq u_1,\dots,\eta_n\leq u_n\right]| \\ &\leq\frac{1}{2\pi}\sum_{1\leq i<j\leq n}\frac{|\Lambda_{ij}^1-\Lambda_{ij}^0|}{\sqrt{1-\rho_{ij}^2}}\exp\left(-\frac{u_i^2+u_j^2}{2(1+\rho_{ij})}\right).\end{aligned} \tag{4.15}$$

A simple, but useful corollary is the specialisation of this lemma to the case when η_i are independent random variables.

Corollary 4.1.6 *Let ξ_i be centered normal variables with covariance matrix Λ, and let η_i be iid centered normal variables. Then*

$$\begin{aligned}&\mathbb{P}\left[\xi_1 \leq u_1, \ldots, \xi_n \leq u_n\right] - \mathbb{P}\left[\eta_1 \leq u_1, \ldots, \eta_n \leq u_n\right] \\ &\leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq n} \frac{(\Lambda_{ij})_+}{\sqrt{1-\Lambda_{ij}^2}} \exp\left(-\frac{u_i^2+u_j^2}{2(1+|\Lambda_{ij}|)}\right).\end{aligned} \tag{4.16}$$

In particular, if $|\Lambda_{ij}| < \delta \leq 1$,

$$\begin{aligned}&\left|\mathbb{P}\left[\xi_1 \leq u, \ldots, \xi_n \leq u\right] - [\Phi(u)]^n\right| \\ &\leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq n} \frac{|\Lambda_{ij}|}{\sqrt{1-\delta^2}} \exp\left(-\frac{u_i^2}{1+|\Lambda_{ij}|}\right).\end{aligned} \tag{4.17}$$

Proof The proof of the corollary is straightforward and left to the reader. □

Another simple corollary is a version of Slepian's lemma:

Corollary 4.1.7 *Let ξ, η be as above. Assume that $\Lambda_{ij}^0 \leq \Lambda_{ij}^1$, for all $i \neq j$. Then*

$$\mathbb{P}\left[\max_{i=1}^{n} \xi_i \leq u\right] - \mathbb{P}\left[\max_{i=1}^{n} \eta_i \leq u\right] \leq 0 \tag{4.18}$$

Proof The proof of the corollary is again obvious, since under our assumption, $(\Lambda_{ij}^0 - \Lambda_{ij}^1)_+ = 0$. □

4.2 Applications to extremes

The comparison results of Section 4.1 can readily be used to give criterion under which the extremes of correlated Gaussian sequences are distributed as in the independent case.

Lemma 4.2.1 *Let ξ_i, $i \in \mathbb{Z}$ be a stationary normal sequence with covariance r_n. Assume that $\sup_{n \geq 1} r_n \leq \delta < 1$. Let u_n be such that*

$$\lim_{n \uparrow \infty} n \sum_{i=1}^{n} |r_i| e^{-\frac{u_n^2}{1+|r_i|}} = 0 \tag{4.19}$$

Then

$$n\left(1-\Phi(u_n)\right) \to \tau \Leftrightarrow \mathbb{P}\left[M_n \leq u_n\right] \to e^{-\tau} \tag{4.20}$$

Proof Using Corollary 4.1.6, we see that (4.19) implies that

$$\left|\mathbb{P}\left[M_n \leq u_n\right] - \Phi(u_n)^n\right| \downarrow 0.$$

Since

$$n\,(1-\Phi(u_n)) \to \tau \Leftrightarrow \Phi(u_n)^n \to e^{-\tau},$$

the assertion of the lemma follows. $\square$

Since the condition $n\,(1-\Phi(u_n)) \to \tau$ determines u_n (if $0<\tau<\infty$), on can easily derive a criteria for (4.19) to hold.

Lemma 4.2.2 *Assume that $r_n \ln n \downarrow 0$, and that u_n is such that $n\,(1-\Phi(u_n)) \to \tau$, $0<\tau<\infty$. Then (4.19) holds.*

Proof We know that, if $n(1-\Phi(u_n)) \sim \tau$,

$$\exp\left(-\frac{1}{2}u_n^2\right) \sim K u_n n^{-1}.$$

and $u_n \sim \sqrt{2\ln n}$. Thus

$$n|r_i|e^{-\frac{u_n^2}{1+|r_i|}} = n|r_i|e^{-u_n^2}e^{\frac{u_n^2|r_i|}{1+|r_i|}}$$

Let $\alpha>0$, and $i \geq n^\alpha$. Then

$$n|r_i|e^{-u_n^2} \sim 2n^{-1}|r_i|\ln n$$

and

$$\frac{u_n^2|r_i|}{1+|r_i|} \leq 2|r_i|\ln n$$

But then

$$|r_i|\ln n = |r_i|\ln i\frac{\ln n}{\ln i} \leq |r_i|\ln i\frac{\ln n}{\ln n^\alpha} = \alpha^{-1}|r_i|\ln i$$

which tends to zero, since $r_n \ln n \to 0$, if $i \uparrow \infty$. Thus

$$\sum_{i\geq n^\alpha} n|r_i|e^{-\frac{u_n^2}{1+|r_i|}} \leq 2\alpha^{-1}\sup_{i\geq n^\alpha}|r_i|\ln i\exp\left(2\alpha^{-1}|r_i|\ln i\right) \downarrow 0,$$

as $n \uparrow \infty$. On the other hand, since there exists $\delta>0$, such that $1-r_i \geq \delta$,

$$\sum_{i\leq n^\alpha} n|r_i|e^{-\frac{u_n^2}{1+|r_i|}} \leq n^{1+\alpha}n^{-2/(2-\delta)}(2\ln n)^2$$

which tends to zero as well, provided we chose α such that $1+\alpha < 2/(2-\delta)$, i.e. $\alpha < \delta/(2+\delta)$. This proves the lemma. $\square$

The following theorem sumarizes the results on the stationary Gaussian case.

Theorem 4.2.3 *Let ξ_i be a stationary centered normal series with covariance r_i, such that $r_n \ln n \to 0$. Then*

(i) For $0 \leq \tau \leq \infty$,

$$n\,(1 - \Phi(u_n)) \to \tau \Leftrightarrow \Phi(u_n)^n \to e^{-\tau},$$

(ii) with a_n and b_n chosen as in the iid normal case,

$$\mathbb{P}\left[a_n(M_n - b_n) \leq x\right] \to e^{-e^{-x}}$$

Exercise. Give an alternative proof of Theorem 4.2.3 by verifying conditions $D(u_n)$ and $D'(u_n)$.

4.3 Appendix: Gaussian integration by parts

Before turning to applications, let us give a short proof of the Gaussian integration by parts formula, Lemma 4.1.2.

Proof (of Lemma 4.1.2). We first consider the scalar case, i.e.

Lemma 4.3.1 *Let X be a centered Gaussian random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable function of at most polynomial growth. Then*

$$\mathbb{E}g(X)X = \mathbb{E}(X^2)\mathbb{E}g'(X) \tag{4.21}$$

Proof Let $\sigma = \mathbb{E}X^2$. Then

$$\begin{aligned}
\mathbb{E}Xg(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(x) x e^{-\frac{x^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(x) \frac{d}{dx}\left(-\sigma^2 e^{-\frac{x^2}{2\sigma^2}}\right) dx \\
&= \sigma^2 \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{d}{dx} g(x) e^{-\frac{x^2}{2\sigma^2}}\, dx = \mathbb{E}X^2 \mathbb{E}g'(x)
\end{aligned} \tag{4.22}$$

where we used elementary integration by parts and the assumption that $g(x)e^{-\frac{x^2}{2\sigma^2}} \downarrow 0$, if $x \to \pm\infty$. □

To prove the multivariate case, we use a trick from [12]: We let X_i, $i, = 1, \ldots, n$ be a centered Gaussian vector, and let X be a centered Gaussian random variable. Set $X_i' \equiv X_i - X\frac{\mathbb{E}XX_i}{\mathbb{E}X^2}$. Then

$$\mathbb{E}X_i'X = \mathbb{E}X_iX - \mathbb{E}X_iX = 0$$

and so X is independent of the vector X_i'. Now compute

$$\mathbb{E}XF(X_1,\ldots,X_n) = \mathbb{E}XF\left(X_1' + X\frac{\mathbb{E}X_iX}{\mathbb{E}X^2},\ldots,X_n' + X\frac{\mathbb{E}X_nX}{\mathbb{E}X^2}\right)$$

Using in this expression Lemma 4.3.1 for the random variable X alone, we obtain

$$\begin{aligned}
\mathbb{E}XF(X_1,\ldots,X_n) &= \mathbb{E}X^2\mathbb{E}F'\left(X_1' + X\frac{\mathbb{E}X_iX}{\mathbb{E}X^2},\ldots,X_n' + X\frac{\mathbb{E}X_nX}{\mathbb{E}X^2}\right)\\
&= \mathbb{E}X^2\sum_{i=1}^{n}\frac{\mathbb{E}X_iX}{\mathbb{E}X^2}\mathbb{E}\frac{\partial F}{\partial x_i}\left(X_1' + X\frac{\mathbb{E}X_iX}{\mathbb{E}X^2},\ldots,X_n' + X\frac{\mathbb{E}X_nX}{\mathbb{E}X^2}\right)\\
&= \sum_{i=1}^{n}\mathbb{E}(X_iX)\mathbb{E}\frac{\partial F}{\partial x_i}(X_1,\ldots,X_n)
\end{aligned}$$

which proves Lemma 4.1.2. $\square$

5

Extremal processes

In this section we develop and complete the description of the collection of "extremal values" of a stochastic process that was started in Chapter 1 with the consideration of the joint distribution of k-largest values of a process. Here we will develop this theory in the language of *point processes*. We begin with some background on this subject and the particular class of processes that will turn out to be fundamental, the Poisson point processes. For more details, see [10, 6, 4].

5.1 Point processes

Point processes are designed to describe the probabilistic structure of point sets in some metric space, for our purposes $\mathbb{R}^d$. For reasons that may not be obvious immediately, a convenient way to represent a collection of points x_i in $\mathbb{R}^d$ is by associating to them a *point measure*.

Let us first consider a single point x. We consider the usual Borel-sigma algebra, $\mathcal{B} \equiv \mathcal{B}(\mathbb{R}^d)$, of $\mathbb{R}^d$, that is generated form the open sets in the open sets in the Euclidean topology of $\mathbb{R}^d$. Given $x \in \mathbb{R}^d$, we define the Dirac measure, δ_x, such that, for any Borel set $A \in \mathcal{B}$,

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases} \tag{5.1}$$

A point measure is now a measure, μ, on $\mathbb{R}^d$, such that there exists a countable collection of points , $\{x_i \in \mathbb{R}^d, i \in \mathbb{N}\}$, such that

$$\mu = \sum_{i=1}^{\infty} \delta_{x_i} \tag{5.2}$$

and, if K is compact, then $\mu(K) < \infty$.

Note that the points x_i need not be all distinct. The set $S_\mu \equiv \{x \in$

$\mathbb{R}^d : \mu(x) \neq 0\}$ is called the support of μ. A point measure such that for all $x \in \mathbb{R}^d$, $\mu(x) \leq 1$ is called simple.

We denote by $M_p(\mathbb{R}^d)$ the set of all point measures on $\mathbb{R}^d$. We equip this set with th sigma-algebra $\mathcal{M}_p(\mathbb{R}^d)$, the smallest sigma algebra that contains all subsets of $M_p(\mathbb{R}^d)$ of the form $\{\mu \in M_m(\mathbb{R}^d) : \mu(F) \in B\}$, where $F \in \mathcal{B}(\mathbb{R}^d)$ and $B \in \mathcal{B}([0,\infty))$. $\mathcal{M}_p(\mathbb{R}^d)$ is also characterized by saying that it is the smallest sigma-algebra that makes the evaluation maps, $\mu \to \mu(F)$, measurable for all Borel sets $F \in \mathcal{B}(\mathbb{R}^d)$.

A *point process*, N, is a random variable taking values in $M_p(\mathbb{R}^d)$, i.e. a measurable map, $N : (\Omega, \mathcal{F}; \mathbb{P}) \to M_p(\mathbb{R}^d)$, from a probability space to the space of point measures.

This looks very fancy, but in reality things are quite down-to-earth:

Proposition 5.1.1 *N is a point process, if and only if the map $N(\cdot, F)$: $\omega \to N(\omega, F)$, is measurable from $(\Omega, \mathcal{F}) \to \mathcal{B}([0,\infty))$, for any Borel set F, i.e. if $N(F)$ is a real random variable.*

Proof Let us first prove necessity, which should be obvious. In fact, since $\omega \to N(\omega, \cdot)$ is measurable into $(M_p(\mathbb{R}^d), \mathcal{M}_p(\mathbb{R}^p))$, and $\mu \to \mu(F)$ is measurable from this space into $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, the composition of these maps is also measurable.

Next we prove sufficiency. Define the set

$$\mathcal{G} \equiv \{A \in \mathcal{M}_p(\mathbb{R}^d) : N^{-1}A \in \mathcal{F}\}$$

This set is a sigma-algebra and N is measurable from $(\Omega, \mathcal{F}) \to (M_p(\mathbb{R}^d), \mathcal{G})$ by definition. But $\mathcal{G}$ contains all sets of the form $\{\mu \in M_p(\mathbb{R}^d) : \mu(F) \in B\}$, since

$$N^{-1}\{\mu \in M_p(\mathbb{R}^d) : \mu(F) \in B\} = \{\omega \in \Omega : N(\omega, F) \in B\} \in \mathcal{F},$$

since $N(\cdot, F)$ is measurable. Thus $\mathcal{G} \supset \mathcal{M}_p(\mathbb{R}^d)$, and N is measurable a fortiori as a map from the smaller sigma-algebra. $\square$

We will have need to find criteria for convergence of point processes. For this we recall some notions of measure theory. If $\mathcal{B}$ is a Borel-sigma algebra, of a metric space E, then $\mathcal{T} \subset \mathcal{B}$ is called a Π-system, if $\mathcal{T}$ is closed under finite intersections; $\mathcal{G} \subset \mathcal{B}$ is called a λ-system, or a sigma-additive class, if

(i) $E \in \mathcal{G}$,
(ii) If $A, B \in \mathcal{G}$, and $A \supset B$, then $A \setminus B \in \mathcal{G}$,
(iii) If $A_n \in \mathcal{G}$ and $A_n \subset A_{n+1}$, then $\lim_{n\uparrow\infty} A_n \in \mathcal{G}$.

The following useful observation is called Dynkin's theorem.

Theorem 5.1.2 *If $\mathcal{T}$ is a Π-system and $\mathcal{G}$ is a λ-system, then $\mathcal{G} \supset \mathcal{T}$ implies that $\mathcal{G}$ contains the smallest sigma-algebra containing $\mathcal{T}$.*

The most useful application of Dynkin's theorem is the observation that, if two probability measures are equal on a Π-system that generates the sigma-algebra, then they are equal on the sigma-algebra. (since the set on which the two measures coincide forms a λ-system containing $\mathcal{T}$).

As a consequence we can further restrict the criteria to be verified for N to be a Point process. In particular, we can restrict the class of F's for which $N(\cdot, F)$ need to be measurable to bounded rectangles.

Proposition 5.1.3 *Suppose that $\mathcal{T}$ are relatively compact sets in $\mathcal{B}$ satisfying*

(i) *$\mathcal{T}$ is a Π-system,*
(ii) *The smallest sigma-algebra containing $\mathcal{T}$ is $\mathcal{B}$,*
(iii) *Either, there exists $E_n \in \mathcal{T}$, such that $E_n \uparrow E$, or there exists a partition, $\{E_n\}$, of E with $\cup_n E_n = E$, with $E_n \subset \mathcal{T}$.*

Then N is a point process on $(\Omega, \mathcal{F})$ in $(E, \mathcal{B})$, if and only if the map $N(\cdot, I) : \omega \to N(\omega, I)$ is measurable for any $I \in \mathcal{T}$.

Exercise. Check that the set of all finite collections bounded (semi-open) rectangles forms indeed a Π-system for $E = \mathbb{R}^d$ that satisfies the hyopthesis of the proposition.

Corollary 5.1.4 *Let $\mathcal{T}$ satisfy the hypothesis of Proposition 5.1.3 and set*

$$\mathcal{G} \equiv \{\{\mu : \mu(I_j) = n_j, 1 \leq j \leq k\}, k \in \mathbb{N}, I_j \in \mathcal{T}, n_j \geq 0\}. \tag{5.3}$$

Then the smallest sigma-algebra containing $\mathcal{G}$ is $\mathcal{M}_p(\mathbb{R}^d)$ and $\mathcal{G}$ is a Π-system.

Next we show that the law, P_N, of a point process is determined by the law of the collections of random variables $N(F_n)$, $F_n \in \mathcal{B}(\mathbb{R}^d)$.

Proposition 5.1.5 *Let N be a point process in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$ and suppose that $\mathcal{T}$ is as in Proposition 5.1.3. Define the mass functions*

$$P_{I_1,\dots,I_k}(n_1, \dots, n_k) \equiv \mathbb{P}\left[N(I_j) = n_j, \forall 1 \leq j \leq k\right] \tag{5.4}$$

for $I_j \in \mathcal{T}$, $n_j \geq 0$. Then P_N is uniquely determined by the collection

$$\{P_{I_1,\dots,I_k}, k \in \mathbb{N}, I_j \in \mathcal{T}\}$$

We need some further notions. First, if N_1, N_2 are point processes, we say that they are independent, if and only if, for any collection $F_j \in \mathcal{B}$, $G_j \in \mathcal{B}$, the vectors

$$(N_1(F_j), 1 \leq j \leq k) \quad \text{and} \quad (N_2(G_j), 1 \leq j \leq \ell)$$

are independent random vectors.

The *intensity measure*, λ, of a point process N is defined as

$$\lambda(F) \equiv \mathbb{E}N(F) = \int_{M_p(\mathbb{R}^d)} \mu(F) P_N(d\mu) \tag{5.5}$$

for $F \in \mathcal{B}$.

For measurable functions $f : \mathbb{R}^d \to \mathbb{R}_+$, we define

$$N(\omega, f) \equiv \int_{\mathbb{R}^d} f(x) N(\omega, dx)$$

Then $N(\cdot, f)$ is a random variable. We have that

$$\mathbb{E}N(f) = \lambda(f) = \int_{\mathbb{R}^d} f(x) \lambda(dx).$$

5.2 Laplace functionals

If Q is a probability measure on $(M_p, \mathcal{M}_p)$, the *Laplace transform* of Q is a map, ψ from non-negative Borel functions on $\mathbb{R}^d$ to $\mathbb{R}_+$, defined as

$$\psi(f) \equiv \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(dx) \right) Q(d\mu). \tag{5.6}$$

If N is a point process, the *Laplace functional* of N is

$$\begin{aligned} \psi_N(f) \equiv \mathbb{E}e^{-N(f)} &= \int e^{-N(\omega,f)} \mathbb{P}(d\omega) \\ &= \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(dx) \right) P_N(d\mu) \end{aligned} \tag{5.7}$$

Proposition 5.2.1 *The Laplace functional, ψ_N, of a point process, N, determines N uniquely.*

Proof For $k \geq 1$, and $F_1, \dots, F_k \in \mathcal{B}$, $c_1, \dots, c_k \geq 0$, let $f = \sum_{i=1}^k c_i \mathbb{1}_{F_i}(x)$. Then

$$N(\omega, f) = \sum_{i=1}^k c_i N(\omega, F_i)$$

and

$$\psi_N(f) = \mathbb{E}\exp\left(-\sum_{i=1}^{k} c_i N(F_i)\right).$$

This is the Laplace transform of the vector $(N(F_i), 1 \leq i \leq k)$, that determines uniquely its law. Hence the proposition follows from Proposition 5.1.5 □

5.3 Poisson point processes.

The most important class of point processes for our purposes will be Poisson point processes.

Definition 5.3.1 Let λ be a σ-finite, positive measure on $\mathbb{R}^d$. Then a point process, N, is called a Poisson point process with intensity measure λ ($PPP(\lambda)$), if

(i) For any $F \in \mathcal{B}(\mathbb{R}^d)$, and $k \in \mathbb{N}$,

$$\mathbb{P}[N(F) = k] = \begin{cases} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!}, & \text{if} \lambda(F) < \infty \\ 0, & \text{if} \lambda(F) = \infty, \end{cases} \tag{5.8}$$

(ii) If $F, G \in \mathcal{B}$ are disjoint sets, then $N(F)$ and $N(G)$ are independent random variables.

In the next theorem we will assert the existence of a Poisson point process with any desired intensity measure. In the proof we will give an explicit construction of such a process.

Proposition 5.3.1*(i) $PPP(\lambda)$ exists , and its law is uniquely determined by the requirements of the definition.*

(ii) The Laplace functional of $PPP(\lambda)$ is given, for $f \geq 0$, by

$$\Psi_N(f) = \exp\left(-\int_{\mathbb{R}^d} (1 - e^{-f(x)})\lambda(dx)\right) \tag{5.9}$$

Proof Since we know that the Laplace functional determines a point process, in order to prove that the conditions of the definition uniquely determine the $PPP(\lambda)$, we show that they determine the form (5.9) of the Laplace functional. Thus suppose that N is a $PPP(\lambda)$. Let $f = c\mathbb{1}_F$.

Then

$$\begin{aligned}\Psi_N(f) &= \mathbb{E}\exp\left(-N(f)\right) = \mathbb{E}\exp\left(-cN(F)\right) \\ &= \sum_{k=0}^{\infty} e^{-ck} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!} = e^{(e^{-c}-1)\lambda(F)} \\ &= \exp\left(-\int (1-e^{-f(x)})\lambda(dx)\right),\end{aligned} \tag{5.10}$$

which is the desired form. Next, if F_i are disjoint, and $f = \sum_{i=1}^k c_i \mathbb{1}_{F-i}$, it is straightforward to

$$\Psi_N(f) = \mathbb{E}\exp\left(-\sum_{i=1}^k c_i N(F_i)\right) = \prod_{i=^1}^k \mathbb{E}\exp\left(-c_i N(F_i)\right)$$

due to the independence assumption (ii); a simple calculations shows that this yields again the desired form. Finally, for general f, we can choose a sequence, f_n, of the form considered, such that $f_n \uparrow f$. By monotone convergence then $N(f_n) \uparrow N(f)$. On the other hand, since $e^{-N(g)} \leq 1$, we get from dominated convergence that

$$\Psi_N(f_n) = \mathbb{E}e^{-N(f_n)} \to \mathbb{E}e^{-N(f)} = \Psi_N(f).$$

But, since $1-e^{-f_n(x)} \uparrow 1-e^{-f(x)}$, and monotone convergence gives once more

$$\Psi_N(f_n) = \exp\left(\int (1-e^{-f_n(x)})\lambda(dx)\right) \uparrow \exp\left(\int (1-e^{-f(x)})\lambda(dx)\right)$$

On the other hand, given the form of the Laplace functional, it is trivial to verify that the conditions of the definition hold, by choosing suitable functions f.

Finally we turn to the *construction* of $PPP(\lambda)$. Let us first consider the case $\lambda(\mathbb{R}^d) < \infty$. Then construct

(i) A Poisson random variable, τ, of parameter $\lambda(\mathbb{R}^d)$.
(ii) A family, X_i, $i \in \mathbb{N}$, of independent, $\mathbb{R}^d$-valued random variables with common distribution λ. This family is independent of τ.

Then set

$$N^* \equiv \sum_{i=1}^{\tau} \delta_{X_i} \tag{5.11}$$

It is not very hard to verify that N^* is a $PPP(\lambda)$.

To deal with the case when $\lambda(\mathbb{R}^d)$ is infinite, decompose λ into a countable sum of finite measures, λ_k, that are just the restriction of λ

to a finite set F_k, where the F_k form a partition of $\mathbb{R}^d$. Then N^* is just the sum of independent $PPP(\lambda_k)$ N_k^*. □

5.4 Convergence of point processes

Before we turn to applications to extremal processes, we still have to discuss the notion of convergence of point processes. As point processes are probability distributions on the space of point measures, we will naturally think about *weak convergence*. This means that we will say that a sequence of point processes, N_n, converges weakly to a point process, N, if for all continuous functions, f, on the space of point measures,

$$\mathbb{E}f(N_n) \to \mathbb{E}f(N). \tag{5.12}$$

However, to understand what this means, we must discuss what continuous functions on the space of point measures are, i.e. we must introduce a topology on the set of point measures. The appropriate topology for our purposes will be that of *vague convergence*.

Vague convergence. We consider the space $\mathbb{R}^d$ equipped with its natural Euclidean metric. Clearly $\mathbb{R}^d$ is a complete, separable metric space. We will denote by $C_0(\mathbb{R}^d)$ the set of continuous real-valued functions on $\mathbb{R}^d$ that have compact support; $C_0^+(\mathbb{R}^n)$ denotes the subset of non-negative functions. We consider $\mathcal{M}_+(\mathbb{R}^d)$ the set of all σ-finite, positive measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We denote by $\mathcal{M}_+(\mathbb{R}^d)$ the smallest sigma-algebra of subsets of $M_+(\mathbb{R}^d)$ that makes the maps $m \to m(f)$ measurable for all $f \in C_0^+(\mathbb{R}^d)$.

We will say that a sequence of measures, $\mu_n \in M_+(\mathbb{R}^d)$ converges *vaguely* to a measure $\mu \in M_+(\mathbb{R}^d)$, if, for all $f \in C_0^+(\mathbb{R}^d)$,

$$\mu_n(f) \to \mu(f) \tag{5.13}$$

Note that for this topology, typical open neighborhoods are of the form

$$B_{f_1,\dots,f_k,\epsilon}(\mu) \equiv \{\nu \in M_+(\mathbb{R}^d) : \forall_{i=1}^k \, |\nu(f_i) - \mu(f_i)| < \epsilon\},$$

i.e. to test the closeness of two measures, we test it on their expectations on finite collections of continuous, positive functions with compact support. Given this topology, on can of course define the corresponding Borel sigma algebra, $\mathcal{B}(M_+(\mathbb{R}^d))$, which (fortunately) turns out to coincide with the sigma algebra $\mathcal{M}_+(\mathbb{R}^d)$ introduced before.

The following properties of vague convergence are useful.

Proposition 5.4.1 *Let μ_n, $n \in \mathbb{N}$ be in $M_+(\mathbb{R}^d)$. Then the following statements are equivalent:*

(i) *μ_n converges vaguely to μ, $\mu_n \stackrel{v}{\to} \mu$.*
(ii) *$\mu_n(B) \to \mu(B)$ for all relatively compact sets, B, such that $\mu(\partial B) = 0$.*
(iii) $\limsup_{n\uparrow\infty} \mu_n(K) \leq \mu(K)$ *and* $\limsup_{n\uparrow\infty} \mu_n(G) \geq \mu(G)$, *for all compact K, and all open, relatively compact G.*

In the case of point measures, we would of course like to see that the point where the sequence of vaguely convergent measures are located converge. The following proposition tells us that this is true.

Proposition 5.4.2 *Let μ_n, $n \in \mathbb{N}$, and μ be in $M_p(\mathbb{R}^d)$, and $\mu_n \stackrel{v}{\to} \mu$. Let K be a compact set with $\mu(\partial K) = 0$. Then we have a labeling of the points of μ_n, for $n \geq n(K)$ large enough, such that*

$$\mu_n(\cdot \cap K) = \sum_{i=1}^{p} \delta_{x_i^n}, \quad \mu(\cdot \cap K) = \sum_{i=1}^{p} \delta_{x_i},$$

such that $(x_1^n, \dots, x_p^n) \to (x_1, \dots, x_p)$.

Another useful and unsurprising fact is that

Proposition 5.4.3 *The set $M_p(\mathbb{R}^d)$ is vaguely closed in $M_+(\mathbb{R}^d)$.*

Thus, in particular, the limit of a sequence of point measures, will, if it exists as a σ-finite measure, be again a point measure.

Proposition 5.4.4 *The topology of vague convergence can be metrized and turns M_+ into a complete, separable metric space.*

Although we will not use the corresponding metric directly, it may be nice to see how this can be constructed. We therefore give a proof of the proposition that constructs such a metric.

Proof The idea is to first find a countable collection of functions, $h_i \in \mathbb{C}_0^+(\mathbb{R}^d)$, such that $\mu_n \stackrel{v}{\to} \mu$ if and only if, for all $i \in \mathbb{N}$, $\mu_n(h_i) \to \mu(h_i)$. The construction below is form [6]. Take a family $G_i, i \in \mathbb{N}$, that form a base of relatively compact sets, and assume it to be closed under finite unions and finite intersections. One can find (by Uryson's theorem), families of functions $f_{i,n}, g_{i,n} \in C_0^+$, such that

$$f_{i,n} \uparrow \mathbb{1}_{G_i}, \quad g_{i,n} \downarrow \mathbb{1}_{G_i}$$

Take the countable set of functions $g_{i,n}, f_{i,n}$ as the collection h_i. Now

$\mu \in M_+$ is determined by its values on the h_j. For, first of all, $\mu(G_i)$ is determined by these values, since

$$\mu(f_{i,n}) \uparrow \mu(G_i) \quad \text{and} \mu(g_{i,n}) \downarrow \mu(G_i)$$

But the family G_i is a Π-system that generates the sigma-algebra $\mathcal{B}(\mathbb{R}^d)$, and so the values $\mu(G_i)$ determine μ.

Now, $\mu_n \stackrel{v}{\to} \mu$, iff and only if, for all h_i, $\mu_n(h_i) \to c_i = \mu(h_i)$.

From here the idea is simple: Define

$$d(\mu, \nu) \equiv \sum_{i=1}^{\infty} 2^{-i} \left(1 - e^{-|\mu(h_i) - \nu(h_i)|}\right) \tag{5.14}$$

Indeed, if $D(\mu_n, \mu) \downarrow 0$, then for each ℓ, $|\mu_n(h_\ell) - \mu(h_\ell)| \downarrow 0$, and conversely. $\square$

It is not very difficult to verify that this metric is complete and separable.

Weak convergence. Having established the space of σ-finite measures as a complete, separable metric space, we can think of weak convergence of probability measures on this space just as if we were working on an Euclidean space.

One very useful fact about weak convergence is Skorohod's theorem, that relates weak convergence to *almost sure convergence.*

Theorem 5.4.5 *Let $X_n, n = 0, 1, \ldots$ be a sequence of random variables on a complete separable metric space. Then X_n converges weakly to a random variable X_0, iff and only if there exists a family of random variables X_n^*, defined on the probability space $([0,1], \mathcal{B}([0,1]), m)$, where m is the Lebesgue measure, such that*

(i) For each n, $X_n \stackrel{\mathcal{D}}{=} X_n^*$, and
(ii) $X_n^* \to X_0^*$, almost surely.

(for a proof, see [1]). While weak convergence usually means that the actual realisation of the sequence of random variables do not converge at all and oscillate widely, Skorohod's theorem says that it is possible to find an "equally likely" sequence of random variables, X_n^*, that do themselves converge, with probability one. Such a construction is easy in the case when the random variables take values in $\mathbb{R}$. In that case, we associate with the random variable X_n (whose distribution function is

F_n, that form simplicity we may assume strictly increasing), the random variable $X_n^*(t) \equiv F_n^{-1}(t)$. It is easy to see that

$$m(X_n^* \leq x) = \int_0^1 \mathbb{1}_{F_n^{-1}(t) \leq x} dt = F_n(x) = \mathbb{P}(X_n \leq x)$$

On the other hand, if $\mathbb{P}[X_n \leq x] \to \mathbb{P}[X_0 \leq x]$, for all points of continuity of F_0, that means that for Lebesgue almost all t, $F_n^{-1}(t) \to F_0^{-1}(t)$, i.e. $X_n^* \to X_0^*$, m-almost surely.

Skorohod's theorem is very useful to extract important consequences from weak convergence. In particular,it allows to prove the convergence of certain functionals of sequences of weakly convergent random variables, which otherwise would not be obvious.

A particularly useful criterion for convergence of point processes is provided by *Kallenberg's theorem* [6].

Theorem 5.4.6 *Assume that ξ is a simple point process on a metric space E, and $\mathcal{T}$ is a Π-system of relatively compact open sets, and that for $I \in \mathcal{T}$,*

$$\mathbb{P}\left[\xi(\partial I) = 0\right] = 1.$$

If ξ_n, $n \in \mathbb{N}$ are point processes on E, and for all $I \in \mathcal{T}$,

$$\lim_{n\uparrow\infty} \mathbb{P}\left[\xi_n(I) = 0\right] = \mathbb{P}\left[\xi(I) = 0\right], \tag{5.15}$$

and

$$\lim_{n\uparrow\infty} \mathbb{E}\xi_n(I) = \mathbb{E}\xi(I) < \infty, \tag{5.16}$$

then

$$\xi_n \stackrel{w}{\to} \xi \tag{5.17}$$

Remark 5.4.1 The Π-system, $\mathcal{T}$, can be chosen, in the case $E = \mathbb{R}^d$, as finite unions of bounded rectangles.

Proof The key observation needed to prove the theorem is that simple point processes are uniquely determined by their avoidance function. This seems rather intuitive, in particular in the case $E = \mathbb{R}$: if we know the probability that in an interval there is no point, we know the distribution of the gape between points, and thus the distribution of the points.

Let us note that we can write a point measure, μ, as

$$\mu = \sum_{y \in S} c_y \delta_y,$$

where S is the support of the point measure and c_y are integers. We can associate to μ the simple point measure

$$T^*\mu = \mu^* = \sum_{y\in S} \delta_y,$$

Then it is true that the map T^* is measurable, and that, if ξ_1 and ξ_2 are point measures such that, for all $I \in \mathcal{T}$,

$$\mathbb{P}\left[\xi_1(I) = 0\right] = \mathbb{P}\left[\xi_2(I) = 0\right], \tag{5.18}$$

then

$$\xi_1^* \stackrel{\mathcal{D}}{=} \xi_2^*.$$

To see this, let

$$\mathcal{C} \equiv \left\{\{\mu \in M_p(E) : \mu(I) = 0\}, I \in \mathcal{T}\right\}.$$

The set $\mathcal{C}$ is easily seen to be a Π-system. Thus, since by assumption the laws, $\mathbb{P}_i$, of the point processes ξ_i coincide on this Π-system, they coincide on the sigma-algebra generated by it. We must now check that T^* is measurable as a map from $(M_p, \sigma(\mathcal{C}))$ to $(M_p, \mathcal{M}_p)$, which will hold, if for each I, the map $T_1^* : \mu \to \mu^*(I)$ is measurable form $(M_p, \sigma(\mathcal{C})) \to \{0, 1, 2, \dots\}$. Now introduce a family of finite coverings of (the relatively compact set) I, $A_{n,j}$, with $A_{n,j}$'s whose diameter is less than $1/n$. We will chose the family such that for each j, $A_{n+1,j} \subset A_{n,i}$, for some i. Then

$$T_1^*\mu = \mu^*(I) = \lim_{n\uparrow\infty} \sum_{j=1}^{k_n} \mu(A_{n,j}) \wedge 1,$$

since eventually, no $A_{n,j}$ will contain more than one point of μ. Now set $T_2^*\mu = (\mu(A_{n,j}) \wedge 1)$. Clearly,

$$(T_2^*)^{-1}\{0\} = \{\mu : \mu(A_{n,j}) = 0\} \subset \sigma(\mathcal{C}),$$

and so T_2^* is measurable as desired, and so is T_1^*, being a monotone limit of finite sums of measurable maps. But now

$$\mathbb{P}\left[\xi_1^* \in B\right] = \mathbb{P}\left[T^*\xi_1 \in B\right] = \mathbb{P}\left[\xi_1 \in (T^*)^{-1}(B)\right] = \mathbb{P}_1\left[(T^*)^{-1}(B)\right].$$

But since $(T^*)^{-1}(B) \in \sigma(\mathcal{C})$, by hypothesis, $\mathbb{P}_1\left[(T^*)^{-1}(B)\right] = \mathbb{P}_2\left[(T^*)^{-1}(B)\right]$, which is also equal to $\mathbb{P}\left[\xi_1^* \in B\right]$, which proves (5.18).

Now, as we have already mentioned, (5.16) implies uniform tightness of the sequence ξ_n. Thus, for any subsequence n', there exist a sub-sub-sequence, n'', such that $\xi_{n''}$ converges weakly to a limit, η. By compactness of M_p, this is a point process. Let us assume for a moment that (a) η is simple, and (b), for any relatively compact A,

$$\mathbb{P}\left[\xi(\partial A)=0\right] \quad \Rightarrow \quad \mathbb{P}\left[\eta(\partial A)=0\right]. \tag{5.19}$$

Then, the map $\mu \to \mu(I)$ is a.s. continuous with respect to η, and therefore, if $\xi_{n'} \stackrel{w}{\to} \eta$, then

$$\mathbb{P}\left[\xi_{n'}(I)=0\right] \to \mathbb{P}\left[\eta(I)=0\right].$$

But we assumed that

$$\mathbb{P}\left[\xi_n(I)=0\right] \to \mathbb{P}\left[\xi(I)=0\right],$$

so that, by the foregoing observation, and the fact that both η and ξ are simple, $\xi=\eta$.

It remains to check simplicity of η and (5.19).

To verify the latter, we will show that for any compact set, K,

$$\mathbb{P}\left[\eta(K)=0\right] \geq \mathbb{P}\left[\xi(K)=0\right]. \tag{5.20}$$

We use that for any such K, there exist sequences of functions, $f_j \in C_0^+(\mathbb{R}^d)$, and compact sets, K_j, such that

$$\mathbb{1}_K \leq f_j \leq 1_{K_j},$$

and $\mathbb{1}_{K_j} \downarrow \mathbb{1}_K$. Thus,

$$\mathbb{P}\left[\eta(K)=0\right] \geq \mathbb{P}\left[\eta(f_j)=0\right] = \mathbb{P}\left[\eta(f_j)\leq 0\right]$$

But $\xi_{n'}(f_j)$ converges to $\eta(f_j)$, and so

$$\mathbb{P}\left[\eta(f_j)\leq 0\right] \geq \limsup_{n'} \mathbb{P}\left[\xi_{n'}(f_j)\leq 0\right] \geq \mathbb{P}\left[\xi_{n'}(K_j)\leq 0\right].$$

Finally, we can approximate K_j by elements $I_j \in \mathcal{T}$, such that $K_j \subset I_j \downarrow K$, so that

$$\mathbb{P}\left[\xi_{n'}(K_j)\leq 0\right] \geq \limsup_{n'} \mathbb{P}\left[\xi_{n'}(I_j)\leq 0\right] = \mathbb{P}\left[\xi(K_j)\leq 0\right],$$

so that (5.20) follows.

Finally, to show simplicity, we take $I \in \mathcal{T}$ and show that the η has multiple points in I with zero probability. Now

$$\mathbb{P}\left[\eta(I)>\eta^*(I)\right] = \mathbb{P}\left[\eta(I)-\eta^*(I)<1/2\right] \leq 2\left(\mathbb{E}\eta(I)-\mathbb{E}\eta^*(I)\right))$$

□

The latter, however, is zero, due to the assumption of convergence of the intensity measures.

Remark 5.4.2 The main requirement in the theorem is the convergence of the so-called *avoidance function*, $\mathbb{P}[\xi_n(I) = 0]$, (5.16). The convergence of the mean (the intensity measure) provides tightness, and ensures that all limit points are simple. It is only a sufficient, but not a necessary condition. It may be replaced by the tightness criterion that for all $I \in \mathcal{T}$, and any $\epsilon > 0$, one can find $R \in \mathbb{N}$, such that, for all n large enough,

$$\mathbb{P}[\xi_n(I) > R] \leq \epsilon, \tag{5.21}$$

if one can show that all limit points are simple (see [4]). Note that, by Chebeychev's inequality, (5.16) implies, of course, (5.21), but but vice versa. There are examples when (5.15) and (5.21) hold, but (5.16) fails.

5.5 Point processes of extremes

We are now ready to describe the structure of extremes of random sequences in terms of point processes. There are several aspect of these processes that we may want to capture:

(i) the distribution of the values largest values of the process; if $u_n(x)$ is the scaling function such that $\mathbb{P}[M_n \leq u_n(x)] \stackrel{w}{\to} G(x)$, it would be natural to look at the point process

$$N_n \equiv \sum_{i=1}^{n} \delta_{u_n^{-1}(X_i)}. \tag{5.22}$$

As n tends to infinity, most of the points $u_n(X_i)$ will disappear to minus infinity, but we may hope that as a point process, this object will converge.

(ii) the "spatial" structure of the large values: we may fix an extreme level, u_n, and ask for the distribution of the values i for which X_i exceeds this level. Again, only a finite number of exceedances will be expected. To represent the exceedances as point process, it will be convenient to embed $1 \leq i \leq n$ in the unit interval $(0, 1]$, via the map $i \to i/n$. This leads us to consider the point process of exceedances on $(0, 1]$,

$$N_n \equiv \sum_{i=1}^{n} \delta_{i/n} \mathbb{1}_{X_i > u_n}. \tag{5.23}$$

(iii) we may consider the two aspects together and consider the point process on $\mathbb{R} \times (0, 1]$,

$$N_n \equiv \sum_{i=1}^{n} \delta_{(u_n^{-1}(X_i), i/n)} \tag{5.24}$$

a restriction of this point process to $(-5, \infty) \times (0, 1]$ is detected in Figure 5.5 for three values of n in the case of iid exponential random variables.

The point process of exceedances. We begin with the simplest, object, the process N_n of exceedances of an extremal level u_n.

Theorem 5.5.1 *Let X_i be a stationary sequence of random variables with marginal distribution function F.*

(i) Let $\tau > 0$, and assume that $D(u_n), D'(u_n)$ hold with $u_n \equiv u_n(\tau)$ such that $n(1 - F(u_n(\tau))) \to \tau$. Let N_n be defined in (5.23). Then N_n converges weakly to a Poisson point process, N, on $(0, 1]$ with intensity measure τdx.

(ii) If the assumptions under (i) hold for all $\tau > 0$, then the point process

$$\widetilde{N}_n \equiv \sum_{i=-\infty}^{\infty} \delta_{i/n} \mathbb{1}_{X_i > u_n(\tau)} \tag{5.25}$$

converges weakly to the Poisson point process $\widetilde{N}$ on $\mathbb{R}$ with intensity measure τdx.

Proof We will use Kallenberg's theorem. First note that trivially,

$$\begin{aligned} \mathbb{E}N_n((c, d]) &= \sum_{i=1}^{n} \mathbb{P}\left[X_i > u_n(\tau)\right] \mathbb{1}_{i/n \in (c,d]} \\ &= n(d - c)(1 - F(u_n(\tau))) \to \tau(d - c) \end{aligned}$$

so that the intensity measure converges to the desired one.

Next we need to show that

$$\mathbb{P}\left[N_n(I) = 0\right] \to e^{-\tau |I|}$$

for I any finite union of disjoint intervals. Consider first I a single interval. Then

$$\mathbb{P}\left[N_n(I) = 0\right] = \mathbb{P}\left[\forall_{i/n \in I} X_i \leq u_n\right] \to e^{-\tau |I|}$$

by Proposition 2.3.3. For finite unions of disjoint intervals the same result follows using that the distances of the intervals when mapped

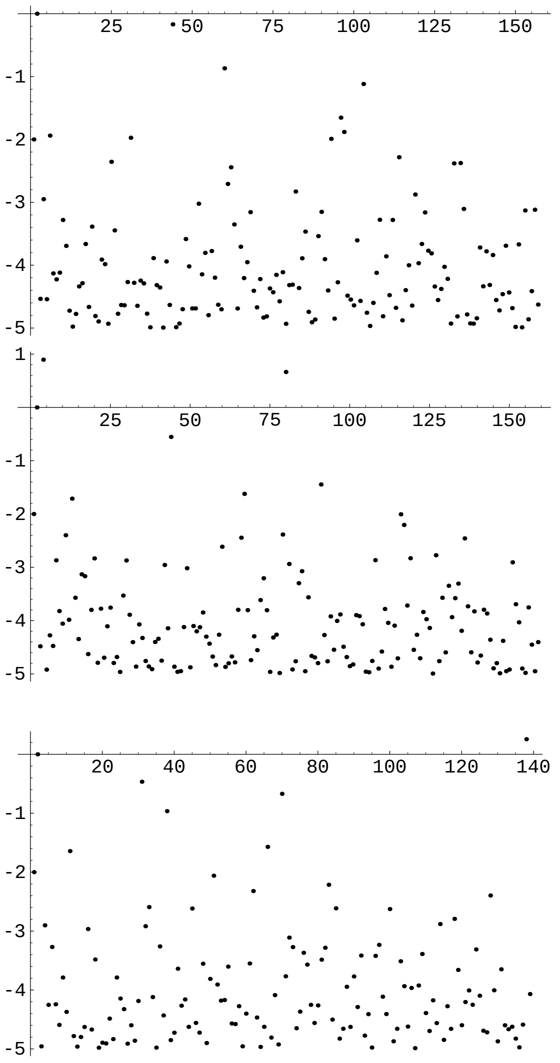

Fig. 5.1. The process P_n for values $n = 1000, 10000, 100000$, truncated at -5. The expected number of points in each picture is 148.41

back to the integers scale like n, so that we can use Proposition 2.1.1 to show that, if $I = \cup_k I_k$, then

$$\mathbb{P}\left[N_n(I) = 0\right] \sim \prod_k \mathbb{P}\left[N_n(I_k) = 0\right].$$

In case (i) all intervals are smaller than one, so it is enough to have conditions $D(u_n), D'(u_n)$ with $u_n = u_n(\tau)$ for the given τ. In case (ii), we must also consider larger intervals, too, which requires to have the conditions with any $\tau > 0$. □

The point process of extreme values. Let us now turn to an alternative point process, that of the values of the largest maxima. In the iid case we have already shown enough in Theorem 1.3.2 to get the following result.

Theorem 5.5.2 *Let X_i be a stationary sequence of random variables with marginal distribution function F, and assume that $D(u_n)$ and $D'(u_n)$ hold for all $u_n = u_n(\tau)$ s.t. $n(1-F(u_n(\tau))) \to \tau$. Then the point process*

$$\mathcal{P}_n \equiv \sum_{i=1}^{n} \delta_{u_n^{-1}(X_i)} \tag{5.26}$$

converges weakly to the Poisson Point process on $\mathbb{R}_+$ with intensity measure the Lebesgue measure.

Proof We will just as before use Kallenberg's theorem. Thus, note first that for any interval $(a, b] \subset \mathbb{R}_+$,

$$\mathbb{E}\mathcal{P}_n((a,b]) = n\mathbb{P}\left[u_n^{-1}(X_1) \in (a,b]\right] = n(F(u_n(a)) - F(u_n(b))) \to b - a.$$

Next we consider the probability that $\mathcal{P}_n((a,b]) = 0$. Clearly,

$$\begin{aligned} &\mathbb{P}\left[\mathcal{P}_n((a,b]) = 0\right] \\ &= \sum_{k=0}^{\infty} \mathbb{P}\left[\#\{i : X_i > u_n(a) = k\} \wedge \#\{i : X_i \le u_n(b)\} = n-k\right] \end{aligned} \tag{5.27}$$

Now we use the decomposition of the interval $(1, \dots, n)$ into disjoint subintervals I_ℓ and I'_ℓ, $\ell =, \dots, r$, as in the proof of Theorem 2.1.2 to estimate the terms appearing in the sum. Then we have that

$$\mathbb{P}\left[\exists_\ell : M(I'_\ell) \ge u_n(b)\right] \le rm(1 - F(u_n(b))) \le \frac{mrb}{n} \downarrow 0.$$

and hence

$$\begin{aligned} &\mathbb{P}\left[\#\{i : X_i > u_n(a) = k\} \wedge \#\{i : X_i \le u_n(b)\} = n-k\right] \\ &\le \mathbb{P}\left[\#\{i \in \cup_\ell I_\ell : X_i > u_n(a) = k\} \wedge \#\{i \in \cup_\ell I_\ell : X_i \le u_n(b)\} = n-k\right] \\ &+ \mathbb{P}\left[\exists_\ell : M(I'_\ell) \ge u_n(b)\right], \end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left[\#\{i : X_i > u_n(a) = k\} \wedge \#\{i : X_i \leq u_n(b)\} = n-k\right] \\
&\geq \mathbb{P}\left[\#\{i \in \cup_\ell I_\ell : X_i > u_n(a) = k\} \wedge \#\{i \in \cup_\ell I_\ell : X_i \leq u_n(b)\} = n-k\right] \\
&- \mathbb{P}\left[\exists_\ell : M(I'_\ell) \geq u_n(b)\right].
\end{aligned}$$

Moreover, due to condition $D'(u_n)$,

$$\begin{aligned}
&\mathbb{P}\left[\exists_\ell : \#\{i \in I_\ell : X_i > u_n(a)\} \geq 2\right] \\
&\leq r \sum_{1 \neq j \in I_1} \mathbb{P}\left[X_1 > u_n(a), X_j > u_n(b)\right] \leq n \sum_{1 \neq j \in I_1} \mathbb{P}\left[X_1 > u_n(b), X_j > u_n(b)\right] \downarrow 0.
\end{aligned}$$

Thus, we can safely restrict our attention to the event that none of the intervals I_ℓ contains more than one exceedance of the level $u_n(a)$. Therefore, the only way to realise the event that k of the X_i exceed $u_n(a)$, while all others remain below $u_n(b)$ is to ask that in k of the intervals I_ℓ the maximum exceeds $u_n(a)$ while in all others it remains below $u_n(b)$. Using stationarity, this gives

$$\begin{aligned}
&\mathbb{P}\left[\#\{i : X_i > u_n(a) = k\} \wedge \#\{i : X_i \leq u_n(b)\} = n-k\right] \\
&= \binom{r}{k} \left(\mathbb{P}\left[M(I_1) > u_n(b)\right]\right)^k \left(\mathbb{P}\left[M(I_1) \leq u_n(b)\right]\right)^{n-k} + o(1).
\end{aligned}$$

Finally,

$$\mathbb{P}\left[M(I_1) > u_n(a)\right] \sim \frac{n}{r}(1 - F(u_n(a)) \sim a/r$$

and

$$\mathbb{P}\left[M(I_1) \leq u_n(b)\right] \sim 1 - \frac{n}{r}(1 - F(u_n(b)) \sim 1 - b/r$$

and so

$$\binom{r}{k} \left(\mathbb{P}\left[M(I_1) > u_n(a)\right]\right)^k \left(\mathbb{P}\left[M(I_1) \leq u_n\right]\right)^{n-k} \sim \frac{1}{k!} a^k e^{-b}.$$

Inserting and summing over k yields, as desired,

$$\mathbb{P}\left[\mathbb{P}_n((a,b]) = 0\right] = \sum_{k=0}^{\infty} \frac{1}{k!} a^k e^{-b}(1 + o(1)) = e^{a-b}(1 + o(1)) \tag{5.28}$$

where $o(1) \downarrow 0$, when $n \uparrow \infty$, $r \uparrow \infty, m \uparrow \infty$, $r/n \downarrow 0$, $\frac{mr}{n} \downarrow 0$.

□

Independence of maxima in disjoint intervals. Theorem 5.5.1 says that the number of exceedances of a given extremal level in disjoint intervals are asymptotically independent. We want to sharpen this in the sense that the same is true if different levels are considered in each interval, and that in fact the processes of extremal values over disjoint intervals are independent. To go in this direction, let us fix $r \in \mathbb{N}$, and consider for $k = 1, \dots, r$, sequences u_n^k such that

$$n(1 - F(u_n^k)) \to \tau_k \tag{5.29}$$

with $u_n^1 \geq \dots u_n^r$. We need to introduce a condition $D_r(\mathbf{u}_n)$ that sharpens condition $D(u_n)$:

Definition 5.5.1 A stationary random sequence satisfies condition $D_r(\mathbf{u}_n)$, for sequences u_n^k as above, if and only if

$$|F_{\mathbf{ij}}(\mathbf{v}, \mathbf{w}) - F_{\mathbf{i}}(\mathbf{v})F_{\mathbf{j}}(\mathbf{w})| \leq \alpha_{n,\ell} \tag{5.30}$$

where $\mathbf{i}, \mathbf{j}$ are as in the definition of $D(u_n)$, and $\mathbf{v} = (v_1, \dots, v_p)$, $\mathbf{w} = (w_1, \dots, w_q)$, where the entries are taken arbitrarily form the levels u_n^k, $v_1, \dots, w_q \in \{u_n^1, \dots, u_n^r\}$. $\alpha_{n,\ell}$ is as in the definition of $D(u_n)$.

Condition $D_r(\mathbf{u}_n)$ has of course the obvious implication:

Lemma 5.5.3 *Under condition $D_r(\mathbf{u}_n)$, for $E_1, \dots, E_s \subset \{1, \dots, n\}$ with $\mathrm{dist(E_j, E_{j'})} \geq \ell$,*

$$\left| \mathbb{P}\left[\forall_{j=1}^s M(E_j) \leq u_{n,j}\right] - \prod_{j=1}^s \mathbb{P}\left[M(E_j) \leq u_{n,j}\right] \right| \leq \alpha_{n,\ell}(s-1) \tag{5.31}$$

if all $u_{n_j} \in \{u_n^1, \dots, u_n^r\}$.

We skip the proof.

In the following theorem $J_1, \dots, J_s$ denote disjoint subintervals of $\{1, \dots, n\}$, such that $|J_j| \sim \theta_k n$.

Theorem 5.5.4 *Under condition $D_r(\mathbf{u}_n)$, with $J_1, \dots, J_s$ as above, and $u_{n,j}$ sequences as in Lemma 5.5.3, we have:*

(i)

$$\mathbb{P}\left[\forall_{j=1}^s M(J_j) \leq u_{n,j}\right] - \prod_{j=1}^s \mathbb{P}\left[M(J_j) \leq u_{n,j}\right] \to 0 \tag{5.32}$$

as $n \uparrow \infty$.

(ii) *Let m be fixed, and $J_1, \dots, J_s$ be disjoint subintervals of $\{1, \dots, nm\}$, with $|J_j| \sim n\theta_j$ and $\sum_j \theta_j \leq m$. Then, if $D_r(\mathbf{u}_n)$ holds with $\mathbf{u}_n = (u_n(\tau_1 m), \dots, u_n(\tau_s m))$, where $n(1 - F(u_n(\tau))) \to \tau$, for each $\tau > 0$, then (5.32) holds.*

The proof of this Theorem should by now be fairly straightforward and I will skip it.

If we add suitable conditions D', we get the following nice corollary:

Corollary 5.5.5(i) *Under the conditions of (i) of Theorem 5.5.4, assume that conditions $D'(u_{n,k})$ hold for $k = 1, \dots, r$. Then with $\sum_{k=1}^r \theta_k \leq 1$,*

$$\mathbb{P}\left[\forall_{k=1}^r M(J_k) \leq u_{n,k}\right] \to \exp\left(-\sum_{k=1}^r \theta_k \tau_k\right) \tag{5.33}$$

where τ_k is such that

$$\tau_k \equiv \lim_{n \uparrow \infty} n(1 - F(u_{n,k})).$$

(ii) *The same conclusion holds, if the conditions of (ii) of the theorem and in addition $D'(v_n)$ holds with $v_n = u_n(\theta_k \tau_k)$, for $k = 1, \dots, r$.*

Exceedances of multiple levels. We are now ready to consider the exceedances of several extremal levels at the same time. We choose sequences u_n^k as before, and define the point process on $\mathbb{R}^2$

$$N_n \equiv \sum_{i=1}^n \sum_{k=1}^r \mathbb{1}_{X_i > u_n^k} \delta_{(\ell_k, i/n)}. \tag{5.34}$$

We can think of the values $\ell_1 < \dots < \ell_r$ naturally as $\ell_k = e^{-\tau_k}$, with $\ell_k = \lim_{n \uparrow \infty} n(1 - F(u_n^k))$, but this is not be necessary.

To formulate a convergence result, we first define a candidate limit point process. Consider a Poisson point process, $\mathcal{P}_r$, with intensity τ_r on $\mathbb{R}$, and let x_j be the positions of the atoms of this process, i.e. $\mathcal{P}_r = \sum_j \delta_{x_{1,j}}$. Let $\beta_j, j \in \mathbb{N}$ be iid random variables, independent of $\mathcal{P}_r$, that take the values $1, \dots, r$ with probabilities

$$\mathbb{P}[\beta_j = s] = \begin{cases} (\tau_{r-s+1} - \tau_{r-s})/\tau_r, & \text{if } s = 1, \dots, r-1 \\ \tau_1/\tau_r, & \text{if } s = r \end{cases} \tag{5.35}$$

Now define the point process

$$N = \sum_{j=1}^\infty \sum_{k=1}^{\beta_j} \delta_{(\ell_k, x_j)}. \tag{5.36}$$

Note that the restriction of this process to any of the lines $x = \ell_k$ is a Poisson point process with intensity τ_k. The construction above constructs r such processes that are successive "thinnings" of each other. Thus, it appears rather natural that N will be the limit of N_n (under the suitable mixing conditions); we already know that the marginals of N on the lines ℓ_k converge to those of N, and by construction, the consecutive processes on the lines must be thinnings of each other.

Theorem 5.5.6 *(i) Assume that $D_r(\mathbf{u}_n)$ and $D'(u_n^k)$ for all $1 \leq k \leq r$ hold. Then N_n converges weakly in distribution to N as a point process on $(0,1] \times \mathbb{R}$.*

(ii) If, in addition, for $0 \leq \tau < \infty$, $u_n(\tau)$ satisfies $n(1 - F(u_n(\tau))) \to \tau$, and $D_r(\mathbf{u}_n)$ holds with $\mathbf{u} = (u_n(m\tau_1), \dots, u_n(m\tau_t))$, for all $m \geq 1$, and $D'(u_n(\tau))$ holds for all $\tau > 0$, then N_n converges to N as a point process on $\mathbb{R}_+ \times \mathbb{R}$.

Proof The proof again uses Kallenberg's criteria. To verify that $\mathbb{E}N_N(B) \to \mathbb{E}N(B)$ is very easy. The main task is to show that

$$\mathbb{P}\left[N_n(B) = 0\right] \to \mathbb{P}\left[N(B) = 0\right],$$

B a collection of disjoint, half-open rectangles, $B = \cup_k (c_k, d_k] \times (\gamma_k, \delta_k] \equiv \cup_k F_k$. By considering differences and intersections, we may write B in the form $B = \cup_j (c_j, d_j] \times E_j \equiv \cup_j F_j$, where E_j is a finite union of semi-closed intervals. Now the nice thing is that $N_n(F_j) = 0$, if and only if the lowest line ℓ_k that intersects F_j is empty. Let this line be numbered m_j. Then $\{N_n(F_j) = 0\} = \{\mathcal{P}_{m_j}((c_j, d_j]) = 0\}$. Since this is just the event that $\{M([c_j n, d_j n]) \leq u_{n,m_j}\}$: thus, we can use Corollary 5.5.5 to deduce that

$$\mathbb{P}\left[N_n(B) = 0\right] \to \exp\left(-\sum_{j=1}^{s} (d_j - c_j)\tau_{m_j}\right), \tag{5.37}$$

which is readily verified to equal $\mathbb{P}\left[N(B) = 0\right]$. This proves the theorem. □

Complete Poisson convergence. We now come to the final goal of this section, the characterisation of the space-value process of extremes as a two-dimensional point process. We consider again $u_n(\tau)$ such that $n(1 - F(u_n(\tau))) \to \tau$. Then we define

$$\mathcal{N}_n \equiv \sum_{i=1}^{\infty} \delta_{(i/n, u_n^{-1}(X_i))} \tag{5.38}$$

as a point process on $\mathbb{R}^2$ (or more precisely, on $(0,1] \times \mathbb{R}_+$.

Theorem 5.5.7 *Let $u_n(\tau)$ be as above; assume that, for any $\tau > 0$, $D'(u_n(\tau)$ holds, and, for any $r \in \mathbb{N}$, and any $\mathbf{u}_n \equiv (u_n(\tau_1), \ldots, u_N(\tau_r))$, $D_r(\mathbf{u}_n)$ holds. Then the point process $\mathcal{N}_n$ converges to the Poisson point process, $\mathcal{N}$, on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measure given by the Lebesgue measure.*

Proof Again we use the criteria of Kallenberg's lemma. It is straightforward to see that, if $B = (c,d] \times (\gamma, \delta]$, then

$$\begin{aligned}\mathbb{E}N(B) =&([nd]-[nc])\mathbb{P}\left[\gamma < u_n^{-1}(X_1) \leq \delta\right] \\ \sim& n(d-c)\mathbb{P}\left[u_n(\gamma) < X_1 \leq u_n(\delta)\right] \qquad (5.39)\\ =& n(d-c)(F(u_n(c)) - F(u_n(d))) \to (d-c)(\delta-\gamma) = \mathbb{E}\mathcal{N}(B). \qquad (5.40)\end{aligned}$$

To prove the convergence of the avoidance function, the basic idea is to express the probabilities that appear for a given set B in terms of processes N_n that were considered in the preceding theorem, and use the convergence of those. E.g., if B is a single rectangle, it is clear that B is free of points of $\mathcal{N}_n$, if and only if the number of exceedances of the levels $u_N(\gamma)$ and $u_n(\delta)$ in $(c,d]$ are the same, which can be expressed in terms of the process N_n corresponding to the two levels $u_N(\gamma)$ and $u_n(\delta)$. But this process converges weakly, and thus the corresponding probabilities converge to those with respect to the process N. But the latter probability is easy to compute: any number of points in the lower process are allowed, provided that all the β_j concerned take the value 2. This yields

$$N(B) = \sum_{l=0}^{\infty} e^{-(d-c)\delta} \frac{[(d-c)\delta]^l}{l!} \left(\frac{\gamma}{\delta}\right)^l = e^{-(d-c)(\delta-\gamma)}, \qquad (5.41)$$

as desired. $\square$

6

Processes with continuous time

6.1 Stationary processes.

We will now consider extremal properties of stochastic processes $\{X_t\}_{t\in\mathbb{R}_+}$ whose index set are the positive real numbers. We will mostly be concerned with stationary processes. In this context, stationarity means that, for all $k \in \mathbb{N}$, and all $t_1, \dots, t_k \in \mathbb{R}_+$, and all $s \in \mathbb{R}_+$, the random vectors $(X_{t_1}, \dots, X_{t_k})$ and $(X_{t_1+s}, \dots, X_{t_k+s})$ have the same distribution.

We will further restrict our attention to processes whose *sample paths*, $X_t(\omega)$, are *continuous functions* for almost all ω, and that the marginal distribution of $X(t)$, for any given $t \in \mathbb{R}_+$, is continuous. Most of our discussion will moreover concern Gaussian processes.

A crucial notion in the extremal theory of such processes are the notion of *up-crossings* and *down-crossings* of a given level.

Let us define, for $u \in \mathbb{R}$, the set, G_u, of function that are not equal to the constant u on any interval, i.e.

$$G_u \equiv \{f \in C_0(\mathbb{R}) : \forall_{I\in\mathbb{R}} f_I \neq u\} \tag{6.1}$$

Note that the processes we will consider enjoy this property.

Lemma 6.1.1 *If X_t is a continuous time random process with the property that all its marginals have a continuous distribution. Then, for any $u \in \mathbb{R}$,*

$$\mathbb{P}[X_t \in G_u] = 0 \tag{6.2}$$

Proof If $X_t = u$ for all $u \in I$ for some interval I then, it must be true that for some rational number s, $X_s = u$. Thus,

$$\mathbb{P}[X_t \in G_u] \leq \mathbb{P}[\exists s \in \mathbb{Q} : X_s = u] \leq \sum_{s\in\mathbb{Q}} \mathbb{P}[X_s = u] = 0 \tag{6.3}$$

since each term in the sum is zero by the assumption that the distribution of X_s is continuous. □

We can now define the notion of up-crossings.

Definition 6.1.1 A function $f \in G_u$ has a strict up-crossing of u at t_0, if there are $\eta > 0, \epsilon > 0$, such that for all $t \in (t_o - \epsilon, t_0)$ $f(t) \leq u$, and for all $t \in (t_0, t_0 + \eta)$, $f(t) \geq u$.

A function n $f \in G_u$ has a up-crossing of u at t_0, if there are $\eta > 0$ such that for all $t \in (t_o - \epsilon, t_0)$ $f(t) \leq u$, and for all $\eta > 0$, there exists $t \in (t_0, t_0 + \eta)$, such that $f(t) > u$.

Remark 6.1.1 Down-crossings of a level u are defined in a completely analogous way.

The following lemma collects some elementary facts on up-crossings.

Lemma 6.1.2 *Let $f \in G_u$. Then*

(i) If for $0 \leq t_1 < t_2$ $f(t_1) < u < f(t_2)$, then f has an up-crossing in (t_1, t_2).
(ii) If f has a non-strict up-crossing of u at t_0, then for all $\epsilon > 0$, there are infinitely many up-crossings of u in $(t_0, t_0 + \epsilon)$.

We leave the proof as an exercise.

For a given function f we will denote, for $I \subset \mathbb{R}$, by $N_u(I)$ the number of up-crossings of u in I. In particular we set $N_u(t) \equiv N_u((0, t])$. For a stochastic process X_t we define

$$J_q(u) \equiv q^{-1}\mathbb{P}\left[X_0 < u < X_q\right] \tag{6.4}$$

Lemma 6.1.3 *Consider a continuous stochastic process X_t as discussed above. Let $I \subset \mathbb{R}$ be an interval. Let $q_n \downarrow 0$ be a decreasing sequence of positive real numbers. Let N_n denote the number of points jq_n, $j \in \mathbb{N}$, such that both $(j-1)q_n \in I$ and $jq_n \in I$, and $X_{(j-1)q_n} < u < X_{jq_n}$. Then*

(i) $N_n \leq N_u(I)$.
(ii) $N_n \uparrow N_u(I)$ almost surely. This implies that $N_u(I)$ is a random variable.
(iii) $\mathbb{E}N_n \uparrow \mathbb{E}N_u(I)$, and whence $\mathbb{E}N_u(t) = t \lim_{q \downarrow 0} J_q(u)$.

Proof Assertion (i) is trivial. To prove (ii), we may use that

$$\mathbb{P}\left[\exists_{k,n} : X_{kq_n} = u\right] = 0.$$

If $N_u(I) \geq m$, then we can select m up-crossings $t_1, \dots, t_t$, such that in intervals $(t_i - e, t_i)$, $X_t \leq u$, and in each interval $(t_i, t_i + \eta)$, there is τ s.t. $X_\tau > u$. By continuity of X_t, there is an interval around each of these t s.t. $X_t > u$ in the entire interval.

Thus, for sufficiently large values of n, each of these intervals will contain one of the points kq_n, and hence there are at least m pairs $k_i q_n, \ell_i q_n$, such that $X_{k_i q_n} < u < X_{\ell_i q_n}$, i.e. $N_n \geq m$. Thus $\liminf_{n\uparrow\infty} N_n \geq N_u(I)$. Because of (i), the limsup of N_n is less than $N_u(I)$, and (ii) follows.

To prove (iii), note that by Fatou's lemma, (ii) implies that $\liminf_n \mathbb{E}N_n \geq \mathbb{E}N_u(I)$. If $\mathbb{E}N_u(I) = \infty$, this proves $\mathbb{E}N_n \to \infty$, as desired. Otherwise, we can use dominated convergence due to (i) to show that $\mathbb{E}N_n \to \mathbb{E}N_u(I)$.

Now if $I = (0, t]$, then there are $\nu_n \sim t/q_n$ points jq_n in I, and so

$$\mathbb{E}N_n \sim (\nu_n - 1)\mathbb{P}\left[X_0 < u < X_q\right] \sim tJ_{q_n}(u).$$

Hence $\lim_{n\uparrow\infty} \mathbb{E}N_n = tJ_{q_n}(u)$ for any sequence $q_n \downarrow 0$, which implies the second assertion of (iii). $\square$

An obvious corollary of this lemma provides a criterion for up-crossings to be strict:

Corollary 6.1.4 *If $\mathbb{E}N_u(t) < \infty$, resp. if $\lim_{q\downarrow 0} J_q(u) < \infty$, then all up-crossings of u are strict, a.s.*

We now turn to the first non-trivial result of this section, an integral representation of the function $J_q(u)$ that will prove particularly useful in the case of Gaussian processes. This is generally known as Rice's theorem, although the version we give was obtained later by Leadbetter.

Theorem 6.1.5 *Assume that X_0 and $\zeta_q \equiv q^{-1}(X_q - X_0)$ have a joint density, $g_q(u, z)$ that is continuous in u for all z and all q small enough, and that there exists $p(u, z)$, such that $g_q(u, z) \to p(u, z)$, uniformly in u for fixed z, as $q \downarrow 0$. Assume moreover that, for some function $h(z)$ such that $\int_0^\infty dz z h(z) < \infty$, $g_q(u, z) \leq h(z)$. Then*

$$\mathbb{E}N_u(1) = \lim_{q\downarrow 0} J_q(u) = \int_0^\infty z p(u, z) dz \tag{6.5}$$

Proof Note that

$$\{X_0 < u < X_q\} = \{X_0 < u < X_0 + q\zeta_q\} = \{X_0 < u < X_0\} \cap \{\zeta_q > q^{-1}(u - X_0)\}.$$

Thus

$$J_q(u) = q^{-1} \int_{-\infty}^{u} dx \int_{q^{-1}(u-x)}^{\infty} dy g_q(x,y). \tag{6.6}$$

Now change to variables v, z, via

$$x = u - qzv, \quad y = z,$$

to get that

$$J_q(u) = \int_0^\infty z dz \int_0^1 dv g_u(u - qzv, z) \tag{6.7}$$

By our assumptions, Lebesgue's dominate convergence theorem implies that

$$\lim_{q \downarrow 0} \int_0^\infty z dz \int_0^1 dv g_u(u-qzv, z) = \int_0^\infty z dz \int_0^1 dv p(u,z) = \int_0^\infty z dz p(u,z), \tag{6.8}$$

as claimed. □

Remark 6.1.2 We can see $p(u, z)$ as the joint density of X_t, X_t', if the process is differentiable. Considering the process together with its derivative will be an important tool in the analysis. Equation (6.5) then can be interested as saying that the average number of up-crossings of u equals the mean derivative of X_t, conditioned on $X_t = u$.

We conclude the general discussion with two simple observations that follow from continuity.

Theorem 6.1.6 *Suppose that the function* $\mathbb{E}N_u(1)$ *is continuous in* u *at* u_0*, and that* $\mathbb{P}[X_t = u] = 0$ *for all* u*. Then,*

(i) with probability one, all points, t*, such that* $X_t = u_0$ *are either (strict) up-crossings or down-crossings.*
(ii) If $M(T)$ *denotes the maximum of* X_t *in* $[0,T]$*, then* $\mathbb{P}[M(T) = u_0] = 0$*.*

Proof If $X(t) = u_0$, but neither a strict up-crossing nor a strict down-crossing, there are either infinitely many crossings of the level u_0, or X_t is tangent to the line $x = u_0$. The former is impossible since by assumption $\mathbb{E}N_{u_0}(t)$ is finite. We need to show that that the probability of a tangency to any fixed level u is zero. Let b_u denote the number of tangencies of u in the interval $(0, t]$. Assume that $N_u + B_u \geq m$, and let t_i be the points where these occur. Since X_t has no plateaus, there must be at least one up-crossing of the level $u - 1/n$ next to each t_i for

n large enough. Thus, $\liminf_{n\downarrow 0} N_{u-1/n} \geq N_u + B_u$. Thus, by Fatou's lemma and continuity of $\mathbb{E}N_u$,

$$\mathbb{E}N_u + \mathbb{E}B_u \leq \liminf_{n\downarrow 0} \mathbb{E}N_{u-1/n} = \mathbb{E}N_u,$$

hence $\mathbb{E}B_u = 0$, which proofs that tangencies have probability zero. Now to prove (ii), note that $M(T) = u$ either because the maximum of X_t in reached in the interior of $(0,T)$, and then X_t must be tangent to u, or $X_0 = u$ or $X_T = u$. All three events have zero probability. □

6.2 Normal processes.

We now turn to the most amenable class of processes, stationary Gaussian processes. A process $\{X_t\}_{t\in\mathbb{R}_+}$ is called *stationary normal (Gaussian) process*, if X_t is a stationary random process, and if for any collection, $t_1, \ldots, t_k \in \mathbb{R}_t$, the vector $(X_{t_1}, \ldots, X_{t_k})$ is a multivariate normal vector. In particular, $\mathbb{E}X_t = 0$ and $\mathbb{E}X_t^2 = 1$, for all $t \in \mathbb{R}_+$. We denote by $r(\tau)$ the covariance function

$$r(\tau) = \mathbb{E}X_t X_{t+\tau} \tag{6.9}$$

Clearly, $r(0) = 1$, and if r is differentiable (resp. twice differentiable) at zero, we set $\lambda_1 = r'(0) = 0$ and $\lambda_2 \equiv -r''(0)$.

We say that X_t is differentiable in square-mean, if there is X_t', such that

$$\lim_{h\to\infty} \mathbb{E}\left(h^{-1}(X_{t+h} - X_t) - X_t'\right)^2 = 0. \tag{6.10}$$

Lemma 6.2.1 *X_t is differentiable in square mean, if and only if*

$$\lambda_2 < \infty.$$

Proof Let $\lambda_2 < \infty$. Define the Gaussian process X_t' with

$$\mathbb{E}X_t' = 0, \quad \mathbb{E}\left(X_t'\right)^2 = \lambda_2, \quad \mathbb{E}X_t'X_{t+\tau}' = -r''(\tau).$$

Then

$$\mathbb{E}\left(h^{-1}(X_{t+h} - X_t) - X_t'\right)^2 = h^{-2}\mathbb{E}\left(2 - 2r(h)\right) - r''(0),$$

The first term converges to $r''(0)$ by hypothesis, and so X_t is differentiable in quadratic mean. On the other hand, if X_t is differentiable in mean, then

$$h^{-2}(r(h) - 2) = \mathbb{E}\left(h^{-1}(X_{t+h} - X_t)\right) \to \mathbb{E}\left(X_t'\right)^2,$$

and so $r''(0) = \lim_{h\to 0} h^{-2}(2 - r(h))$ exists and is finite. □

Moreover, one sees that in this case,

$$\mathbb{E}(h^{-1}(X_{t+h} - X_t)X_t) = h^{-1}(r(h) - r(0)) \to r'(0) = 0,$$

so that $X'(t)$ and $X(t)$ are independent for each t.

$$\begin{aligned}&\mathbb{E}\left((h^{-1}(X_{t+h} - X_t) - X'_t)((h^{-1}(X_{s+h} - X_s) - X'_s)\right)\\&= h^{-2}\left(2r(t-s) - r(t-s-h) - r(t-s+h)\right) + \mathbb{E}X'_tX'_s,\end{aligned}$$

and so the covariance of X'_t is as given above. In this case the joint density, $p(u,z)$, of the pair X_t, X'_t is given explicitly by

$$p(u,z) = \frac{1}{2\pi\sqrt{\lambda_2}} \exp\left(-\frac{1}{2}(u^2 + z^2/\lambda_2)\right). \tag{6.11}$$

Note also that X_0, ζ_q are bivariate Gaussian with covariance matrix

$$\begin{pmatrix} 1 & q^{-1}(r(q) - r(0)) \\ q^{-1}(r(q) - r(0)) & 2q^{-2}(r(0) - r(q)) \end{pmatrix}.$$

Thus, in this context, we can apply Theorem 6.1.5 to get an explicit formula, called *Rice's formula*, for the mean number of up-crossings,

$$\mathbb{E}N_u(1) = \int_0^\infty zp(u,z)dz = \frac{1}{2\pi}\sqrt{\lambda_2}\exp\left(-\frac{u^2}{2}\right). \tag{6.12}$$

6.3 The cosine process and Slepians lemma.

Comparison between processes is also an important tool in the case of continuous time Gaussian processes. There is a particularly simple stationary normal process, called the cosine process, where everything can be computed explicitly. Let η, ζ be two independent normal random variables, and define

$$X^*_t \equiv \eta\cos\omega t + \zeta\sin\omega t. \tag{6.13}$$

A simple computation shows that this is a normal process with covariance

$$\mathbb{E}X^*_{t+\tau}X^*_t = \cos\omega\tau.$$

Another way to realise this process as

$$X^*_t = A\cos\left(\omega t - \phi\right), \tag{6.14}$$

where $\eta = A\cos\phi$ and $\zeta = A\cos\phi$. It is easy to check that the random variables A and ϕ are independent, ϕ is uniformly distributed on $[0, 2\pi)$, and A is Raighley-distributed on $\mathbb{R}_+$, i.e. its density is given by

$$xe^{-x^2/2}.$$

We will now derive the probability distribution of the maximum of this random process.

Lemma 6.3.1 *If $M^*(T)$ denotes the maximum of the cosine process on $[0,T]$, then*

$$\mathbb{P}\left[M^*(T) \leq u\right] = \Phi(u) - \frac{\omega T}{2\pi} e^{-u^2/2}, \tag{6.15}$$

for $u > 0$ and $0 < \omega T < 2\pi$.

Proof We have $\lambda_2 = \omega^2$, and so by Rice's formula, the mean number of up-crossings in this process satisfies

$$\mathbb{E}N_u(T) = \frac{\omega T}{2\pi} e^{-u^2/2}.$$

Now clearly,

$$\mathbb{P}\left[M^*(T) > u\right] = \mathbb{P}[X_0^* > 0] + \mathbb{P}[X_0^* \leq u \wedge N_u(T) \geq 1].$$

Consider $\omega T < \pi$. Then, if $X_0^* > u$, then the next up-crossing of u cannot occur before time $t = \pi/\omega$, i.e. not before T. Thus, $N_u(T) \geq 1$ only if $X_0^* \leq u$, and thus

$$\mathbb{P}[X_0^* \leq u \wedge N_u(T) \geq 1] = \mathbb{P}[N_u(T) \geq 1].$$

Also, the number of up-crossings of u before T is bounded by one, so that

$$\mathbb{P}[N_u(T) \geq 1] = \mathbb{E}N_u(T) = \frac{\omega T}{2\pi} e^{-u^2/2}.$$

Hence,

$$\mathbb{P}\left[M^*(T) > u\right] = 1 - \Phi(u) + \frac{\omega T}{2\pi} e^{-u^2/2}, \tag{6.16}$$

which is the same as formula (6.15). □

Using the periodicity of the cosine, we see that the restriction to $T < 2\pi/\omega$ gives already all information on the process.

Let us note that from this formula it follows that

$$\frac{\mathbb{P}[M^*(h) > u]}{h\varphi(u)} \to \left(\frac{\lambda_2}{2\pi}\right)^{1/2}, \tag{6.17}$$

as $u \uparrow \infty$, where φ denotes the normal density. This will later be shown to be a general fact. The point is that the cosine process will turn out to be a good model for more general processes *locally*, i.e. it will reflect well the effect of short range correlations on the behaviour of maxima.

We conclude this section with a statement of Slepian's lemma for continuous time normal processes.

Theorem 6.3.2 *Let X_t, Y_t be independent standard normal processes with almost surely continuous sample paths. Let r_1, r_2 denote their covariance function. Assume that, for all s, t, $r_1(s,t) \geq r_2(s,t)$, then*

$$\mathbb{P}\left[M_(T) \leq u\right] \geq \mathbb{P}\left[M_2(T) \leq u\right],$$

if M_1, M_2 denote the maxima of the processes X and Y, respectively.

The proof follows from the analogous statement for discrete time processes and the fact the sample paths are continuous. Details are left as an exercise.

6.4 Maxima of mean square differentiable normal processes

We will now consider stationary normal processes whose covariance function is twice differentiable at zero, i.e. that

$$r(\tau) = 1 - \frac{\lambda_2 \tau^2}{2} + o(\tau^2). \tag{6.18}$$

We will first derive the behaviour of the maximum of the process X_t on a time interval $[0,T]$.

The basic idea is to use a discretization of the process and Gaussian comparison results. We cut $(0,T)$ into pieces of length $h > 0$ and set $n = [T/h]$. We call $M(T)$ the maximum of X_t on $(0,T)$; we set $M(nh) \equiv \max_{i=1}^{n} M(((i-1)h, ih))$.

Lemma 6.4.1 *Let X_t be as described above. Then*

(i) for all $h > 0$, $\mathbb{P}[M(h) > u] \leq 1 - \Phi(u) + \mu h$, and so

$$\limsup_{u \uparrow \infty} \mathbb{P}[M(h) > u]/(\mu h) \leq 1.$$

(ii) Given $\theta < 1$, there exists h_0, such that for all $h < h_0$,

$$\mathbb{P}[M(h) > u] \geq 1 - \Phi(u) + \theta \mu h. \tag{6.19}$$

Proof To prove (i), note that $M(h)$ exceeds one either because $X_0 \geq u$, or because there is an up-crossing of u in $(0,h)$. Thus

$$\mathbb{P}[M(h) > u] \leq \mathbb{P}[X_0 > u] + \mathbb{P}[N_u(h) \geq 1] \leq 1 - \Phi(u) + \mathbb{E}N_u(h).$$

(ii) follows from Slepian's lemma by comparing with the cosine process.

$\square$

Next we compare maxima to maxima of discretizations. We fix $q > 0$, and let N_u and N_u^q be the number of up-crossing of X_t, respectively the sequence X_{nq} in an interval I of length h.

Lemma 6.4.2 *With q, u such that $qu \downarrow 0$ as $u \uparrow \infty$,*

(i) $\mathbb{E}N_u^q = h\mu + o(\mu)$,
(ii) $\mathbb{P}[M(I) \le u] = \mathbb{P}[\max_{kq\in I} X_{kq} \le u] + o(\mu)$,

with $o(\mu)$ uniform in I with $|I| \le h_0$.

Proof (i) follows from

$$\mathbb{E}N_u^q \sim h/q\mathbb{P}[X_0 < u < X_q] = hJ_q(u) = h\mu(1 + o(1)).$$

(ii) follows since

$$\begin{aligned}
&\mathbb{P}[\max_{kq\in I} X_{kq} \le u] - \mathbb{P}[M(I) \le u] \le \mathbb{P}(X_I > u] + \mathbb{P}[X_a < u, N_u \ge 1, N_u^q = 0] \\
&\le \mathbb{P}(X_I > u] + \mathbb{P}[N_u - N_u^q \ge 1] \\
&\le 1 - \Phi(u) + \mathbb{E}(N_u - N_u^q) = o(\mu)
\end{aligned}$$

□

We now return to intervals of length $T = nh$, where $T \uparrow \infty$ and $T\mu \to \tau > 0$. We fix $0 < \epsilon < h$, and divide each sub-interval of length h into two intervals I_i, I_i^*, of length $h - \epsilon$ and ϵ.

Lemma 6.4.3 *With the notation above, and q such that $qu \to 0$,*

(i)

$$\limsup_{T\uparrow\infty} |\mathbb{P}\left[M\left(\cup_i I_i\right) \le u\right] - \mathbb{P}\left[M(nh) \le u\right]| \le \tau\epsilon/h \tag{6.20}$$

(ii)

$$\mathbb{P}\left[X_{kq} \le u, \forall kq \in \cup I_i\right] - |\mathbb{P}\left[M\left(\cup_i I_i\right) \le u\right]| \to 0 \tag{6.21}$$

Proof For (i):

$$\begin{aligned}
0 &\le |\mathbb{P}\left[M\left(\cup_i I_i\right) \le u\right] - \mathbb{P}\left[M(nh) \le u\right]| \\
&\le n|\mathbb{P}\left[M\left(I_i^*\right) > u\right] \sim \frac{\tau\epsilon}{h}\frac{\mathbb{P}\left[M\left(I_i^*\right) > u\right]}{\mu\epsilon}
\end{aligned}$$

Because of Lemma 6.4.1 (i), (i) follows. (ii) follows from Lemma 6.4.2 (ii) and the fact that the right-hand side is bounded by

$$\sum_{i=1} n\left(\mathbb{P}\left[X_{qk} \le u, \forall qk \in I_i\right] - \mathbb{P}[M(I_j) \le u]\right).$$

□

It is now quite straightforward to prove the asymptotic independence of maxima as in the discrete case.

Lemma 6.4.4 *Assume that $r(\tau) \downarrow 0$, as $\tau \uparrow \infty$, and that as $T \uparrow \infty$ and $u \uparrow \infty$,*

$$\frac{T}{q} \sum_{\epsilon \leq qk \leq T} |r(kq)| \exp\left(-\frac{u^2}{1+|r(qk)|}\right) \downarrow 0, \tag{6.22}$$

for any $\epsilon > 0$, and some q such that $qu \to 0$. Then, it $T\mu \to \tau > 0$,

(i)

$$\mathbb{P}\left[X_{kq} \leq u, \forall kq \in \cup_i I_i\right] - \prod_{i=1}^{n} \mathbb{P}\left[X_{kq} \leq u, \forall kq \in I_i\right] \to 0, \tag{6.23}$$

(ii)

$$x \limsup_{n} \left| \prod_{i=1}^{n} \mathbb{P}\left[X_{kq} \leq u, \forall kq \in I_i\right] - \left(\mathbb{P}[M(h) \leq u]\right)^n \right| \leq \frac{2\tau\epsilon}{h} \tag{6.24}$$

Proof The details of the proof are a bit long and boring. I just give the idea: (i) is proven as the Gaussian comparison lemma, comparing the variance of the sequence X_{kn} with the one where the covariances between those variables for which $kq, k'q$ are not in the same I_i are set to zero.

(ii) uses Lemmata 6.4.1 and 6.4.2. □

From independence it is only a small step to the exponential distribution.

Theorem 6.4.5 *Let U, T tend to infinity in such a way that $T\mu(u) = (T/2\pi)\lambda_2^{1/2} \exp(-u^2/2) \to \tau \geq 0$. Suppose that $r(t)$ satisfies (6.18) and either $\rho(t) \ln t \downarrow 0$, as $t \downarrow 0$, or the weaker condition (6.22) for some q s.t. $qu \downarrow 0$. Then*

$$\lim_{T \uparrow \infty} \mathbb{P}\left[M(T) \leq u\right] = e^{-\tau} \tag{6.25}$$

Proof If $\tau = 0$, i.e. $T\mu \downarrow 0$, $\mathbb{P}\left[M(T) > u\right] \leq 1 - \Phi(u) + T\mu(u) \to 0$. If $\tau > 0$, we are under the assumptions of Lemma 6.4.4, and from our earlier Lemmata 6.1.2 and 6.1.3,

$$\limsup_{T \uparrow \infty} \left|\mathbb{P}\left[M(nh) \leq u\right] - \mathbb{P}\left[M(n) \leq u\right]^n\right| \leq \frac{3\tau}{h}\epsilon \tag{6.26}$$

for any $\epsilon >$. Also, since $nh \leq t < (n+1)h$,

$$0 \leq \mathbb{P}\left[M(T) \leq u\right] - \mathbb{P}\left[M(nh) \leq u\right] \leq \mathbb{P}\left[N_u(h) \geq 1\right] \leq \mu h$$

which tends to zero. Now we choose $0 < h < h_0(\theta)$ with $h_0(\theta)$ as in (ii) of Lemma 6.1.2. Then

$$\mathbb{P}\left[M(h) > u\right)] \geq \theta\mu h(1 + o(1)) = \frac{\theta\tau}{n}(1 + o(1))$$

and thus

$$\begin{aligned}\mathbb{P}\left[M(T) \leq u\right] &\leq (1 - \mathbb{P}\left[M(h) > u\right])^n + o(1)\\ \left(1 - \frac{\theta\tau}{n}(1 + o(1))\right)^n &+ o(1/n) = e^{-\theta\tau + o(1)} + o(1)\end{aligned}$$

and so for all $\theta < 1$,

$$\limsup_{T\uparrow\infty} \mathbb{P}\left[M(T) \leq u\right] \leq e^{-\theta\tau}$$

Using (i) of Lemma 6.1.2, one sees immediately that

$$\liminf_{T\uparrow\infty} \mathbb{P}\left[M(T) \leq u\right] \geq e^{-\tau}$$

from which the claimed result follows. $\square$

6.5 Poisson convergence

6.5.1 Point processes of up-crossings

In this section we will show that the up-crossings of levels u such that $T\mu(u) \to \tau$ converge to a Poisson point process. We denote this process by N_T^*, which can be simply defined by setting, for all Borel-subsets, B, of $\mathbb{R}_+$,

$$N_T^*(B) = N_u(TB) \tag{6.27}$$

It will not be a big surprise to find that this process converges to a Poisson point process:

Theorem 6.5.6 *Consider a stationary normal process as in the previous section, and assume that T, u tend to infinity such that $T\mu(u) \to \tau$. Then, the point-processes of u-up-crossings, N_T^*, converge to the Poisson point process with intensity τ on $\mathbb{R}_+$.*

The proof of this result follows exactly as in the discrete case from the independence of maxima over disjoint intervals. In fact, one proves, using the same ideas as in the proceeding section the following key lemma (which is stronger than needed here).

Lemma 6.5.7 *Let* $0 < c = c_1 < d_1 \leq c_2 < d_2 \leq \cdots \leq c_r < d_r = d$ *be given. Set* $E_i = (Tc_i, Td_i]$. *Let* $\tau_1 \ldots, \tau_i$ *be positive numbers, and let* $u_{T,i}$ *be such that* $T\mu(u_{T,i}) \to \tau_i$. *Then, under the assumptions of the theorem,*

$$\mathbb{P}\left[\cap_{i=1}^{r} M(E_i) \leq u_{T_i}\right] - \prod_{i=1}^{r} \mathbb{P}\left[M(E_i) \leq u\right] \to 0 \tag{6.28}$$

6.5.2 Location of maxima

We will denote by $L(T)$ the values, $t \leq T$, where X_t attains its maximum in $[0, T]$ for the first time. $L(T)$ is a random variable, and $\mathbb{P}[L(T) \leq t] = \mathbb{P}[M(0, t]) \geq M((t, T]$. Moreover, under mild regularity assumptions, the distribution of $L(T)$ is continuous except possibly at 0 and at T.

Lemma 6.5.8 *Suppose that* X_t *has a derivative in probability at* t *for* $0 < t < T$, *and that the distribution of the derivative is continuous at zero. Then* $\mathbb{P}[L(T) = t] = 0$.

One may be tempted to think that the distribution of $L(T)$ is uniform on $[0, T]$, for stationary sequences.

Bibliography

[1] P. Billingsley. *Weak convergence of measures: Applications in probability.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

[2] A. Bovier and I. Kurkova. Poisson convergence in the restricted k-partioning problem. preprint 964, WIAS, 2004.

[3] M. R. Chernick. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Ann. Probab.*, 9(1):145–149, 1981.

[4] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.* Springer Series in Statistics. Springer-Verlag, New York, 1988.

[5] Jean-Pierre Kahane. Une inégalité du type de Slepian et Gordon sur les processus gaussiens. *Israel J. Math.*, 55(1):109–110, 1986.

[6] O. Kallenberg. *Random measures.* Akademie-Verlag, Berlin, 1983.

[7] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983.

[8] St. Mertens. Phase transition in the number partitioning problem. *Phys. Rev. Lett.*, 81(20):4281–4284, 1998.

[9] St. Mertens. A physicist's approach to number partitioning. *Theoret. Comput. Sci.*, 265(1-2):79–108, 2001.

[10] S.I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York, 1987.

[11] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.*, 41:463–501, 1962.

[12] M. Talagrand. *Spin glasses: a challenge for mathematicians*, volume 46 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. (3). [Results in Mathematics and Related Areas. (3)].* Springer-Verlag, Berlin, 2003.