Chapter 3

Background Theory & Methods

II – Dependent Sequences

3.0 Introduction

Serial independence was assumed throughout Chapter 2 but is rarely a valid assumption for applications. This chapter reviews key results for strongly stationary processes. The presence of serial dependence causes quantitative and qualitative changes in the behaviour of extremes.

3.1 Univariate Sequences

Let $\{X_i\}_{i\geq 1}$ be a stationary sequence of random variables with marginal distribution function F. Without any restriction on the strength of dependence, the limiting distribution of the block maximum M_n can remain obscure. For example, if $X_i = X_1$ almost surely for all $i \geq 1$ then the distribution of M_n is F. The serial dependence need be restricted at only extreme levels, however, to obtain a non-trivial limit. One such restriction is the following mixing condition introduced by Leadbetter (1983). Define $M(I) = \max\{X_i : i \in I\}$ and, for a sequence of thresholds u_n, let $I_{j,k}(u_n) = \{\{M(I) \leq u_n\} : I \subseteq \{j, \ldots, k\}\}$ be the set of all intersections of the events $\{X_i \leq u_n\}$, $j \leq i \leq k$.

Condition $D(u_n)$. For all $A \in I_{1,k}(u_n)$, $B \in I_{k+l,n}(u_n)$ and $1 \leq k \leq n - l$,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n, l)$$

and $\alpha(n, l_n) \to 0$ as $n \to \infty$ for some $l_n = o(n)$.

Condition $D(u_n)$ ensures that any two events of the form $\{M(I_1) \leq u_n\}$ and $\{M(I_2) \leq u_n\}$ can become approximately independent as n increases when the index sets $I_i \subset \{1, \ldots, n\}$ are separated by a relatively short interval of length $l_n = o(n)$. Hence, $D(u_n)$ limits the long-range dependence between such events. It is a weak condition relative to many imposed in other areas of statistics. For example, $D(u_n)$ is satisfied for linear Gaussian processes if the correlation decays faster than the logarithm of the lag (Berman, 1964), and this is weaker than the geometric decay assumed by autoregressive models. With the restriction of $D(u_n)$, Leadbetter (1974) proves the following analogue of Theorem 2.1.

Theorem 3.1. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G, and if $D(u_n)$ holds with $u_n = a_n x + b_n$ for each x such that $G(x) > 0$, then G is a generalised extreme-value distribution function (2.1).

The possible limit distributions for block maxima from stationary processes are therefore the same as in the independent case, and can be used as models in the same way. Dependence does influence the particular limit that is achieved, however, as made clear by the next theorem, due to Leadbetter (1983). Define $\{\hat{X}_i\}_{i\geq 1}$ to be the sequence of independent random variables also with marginal distribution F and let $\hat{M}_n = \max\{\hat{X}_1, \ldots, \hat{X}_n\}$.

Theorem 3.2. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{\hat{M}_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G, if $D(u_n)$ holds with $u_n = a_n x + b_n$ for each x such that $G(x) > 0$, and if $P\{(M_n - b_n)/a_n \leq x\}$ converges for some x then

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G^{\theta}(x) \quad \textit{as } n \to \infty \tag{3.1}$$

for some constant $\theta \in [0, 1]$.

Block maxima from stationary processes are therefore stochastically smaller than those from independent processes. The constant θ is called the extremal index and summarises the strength of dependence between extremes in a stationary sequence. It has several interpretations and will play a crucial role in Chapter 4.

The two-dimensional point process

$$Q_n(\cdot) = \sum_{i=1}^{n} \delta_{(i/n\,,\,(X_i - b_n)/a_n)}(\cdot)$$

considered in Theorem 2.2 for independent sequences has a limit for stationary processes too. The following mixing condition of Hsing (1987) is needed to ensure convergence. Let $\mathcal{F}_{j,k}(u_n\,;\,\tau_1,\ldots,\tau_s) = \sigma(\{X_i \le u_n(\tau_r)\}\ :\ j \le i \le k, 1 \le r \le s)$.

Condition $\Delta(u_n)$. For all integer $s \ge 1$, $\tau_1,\ldots,\tau_s > 0$, $A \in \mathcal{F}_{1,k}(u_n\,;\,\tau_1,\ldots,\tau_s)$, $B \in \mathcal{F}_{k+l,n}(u_n\,;\,\tau_1,\ldots,\tau_s)$ and $1 \le k \le n-l$,

$$|P(A \cap B) - P(A)P(B)| \le \alpha(n,l)$$

and $\alpha(n,l_n) \to 0$ as $n \to \infty$ for some $l_n = o(n)$.

This condition is stricter than $D(u_n)$ because the events required to become independent are more numerous and they are required to become independent at many levels simultaneously. The following result was proved by Hsing (1987).

Theorem 3.3. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G of generalised extreme-value form (2.1), if $\Delta(u_n)$ holds when $u_n(\tau) = a_n\tau + b_n$ and if $Q_n \xrightarrow{d} Q$, then Q is the point process

$$Q(\cdot) = \sum_{i=1}^{\infty}\sum_{j=1}^{K_i} \delta_{(S_i\,,\mu+\sigma\xi^{-1}\{(T_iY_{i,j})^{-\xi}-1\})}(\cdot), \tag{3.2}$$

where $\{(S_i, T_i)\}_{i\geq 1}$ are points of a homogeneous Poisson process, η, on $(0, 1] \times (0, \infty)$ with Lebesgue measure, $\{Y_{i,j}\}_{j=1}^{K_i}$ are points of a point process γ_i on $[1, \infty)$ with an atom $Y_{i,1} = 1$ for each $i \geq 1$ and $\{\gamma_i\}_{i\geq 1}$ are independent, identically distributed and independent of η.

The limit process (3.2) has the following interpretation. For each i, the points $\{(S_i, T_{i,j})\}_{j=1}^{K_i}$, where

$$T_{i,j} = \mu + \sigma\xi^{-1}\{(T_iY_{i,j})^{-\xi} - 1\},$$

all occur at the same time, S_i. Also, $T_{i,1}$, which corresponds to the atom $Y_{i,1} = 1$, is the largest of the $T_{i,j}$ occurring at time S_i. This structure describes clusters of points occurring through time: K_i points occur at time S_i with cluster maximum $T_{i,1}$; the remaining points in the cluster are defined relative to $T_{i,1}$ by the random variables

$$Y_{i,j} = \frac{\left\{1 + \xi\left(\frac{T_{i,j}-\mu}{\sigma}\right)\right\}^{-1/\xi}}{\left\{1 + \xi\left(\frac{T_{i,1}-\mu}{\sigma}\right)\right\}^{-1/\xi}}, \quad 1 \leq j \leq K_i. \tag{3.3}$$

The collection $\{Y_{i,j}\}_{j=1}^{K_i}$ is called the string associated with the i-th cluster.

The cluster maxima, $\{(S_i, T_{i,1})\}_{i\geq 1}$, are points of a non-homogeneous Poisson process on $(0,1] \times (\omega_0, \omega_1)$ with intensity measure

$$\Lambda(A) = -(b-a)\log G(x) \tag{3.4}$$

on $A = (a,b) \times [x, \omega_1)$ when $0 \leq a < b \leq 1$ and $\omega_0 < x < \omega_1$. This is the same point process found in Theorem 2.2 for independent sequences except that the intensity measure (3.4) is reduced by a factor θ due to Theorem 3.2. For each i, the number of the remaining points in the cluster and their distribution relative to the cluster maximum is determined by the point process γ_i. The distribution of γ_i depends on the particular short-range dependence structure of the process and Mori (1977) shows that γ_i can be any point process. Nevertheless, the theorem does say that clusters are independent and the distribution of points within a cluster relative to the cluster maximum is the same for all clusters.

Statistical application of Theorem 3.3 will be discussed later in this section. Another point process is investigated first, which requires the following definitions. A marked point process Q counts for each point X_i in some space S a quantity Y_i, and has the form

$$Q(A) = \sum_{i \in \mathcal{I}} Y_i \, \delta_{X_i}(A), \quad A \subseteq S,$$

where the marks $\{Y_i : i \in \mathcal{I}\}$ are identically distributed. A compound Poisson process is a marked point process for which the points occur according to a Poisson process independently of the marks, which are themselves independent and identically distributed.

For the stationary sequence $\{X_i\}_{i\geq 1}$, the point process

$$Q'_n(\cdot) = \sum_{i\in\mathcal{I}} \delta_{i/n}(\cdot), \quad \mathcal{I} = \{i : X_i > u_n, 1 \leq i \leq n\}, \tag{3.5}$$

counts the times on $(0,1]$ at which a threshold u_n is exceeded. The limit of Q'_n is found by Hsing *et al.* (1988) under the following mixing condition. Let $\mathcal{F}'_{j,k}(u_n) = \sigma(\{X_i \leq u_n\} : j \leq i \leq k)$.

Condition $\Delta'(u_n)$. For all $A \in \mathcal{F}'_{1,k}(u_n)$, $B \in \mathcal{F}'_{k+l,n}(u_n)$ and $1 \leq k \leq n-l$,

$$|P(A\cap B) - P(A)P(B)| \leq \alpha'(n,l)$$

and $\alpha'(n,l_n) \to 0$ as $n\to\infty$ for some $l_n = o(n)$.

Define also the cluster-size distribution

$$\pi_n(j) = P\{N_{r_n}(u_n) = j \,|\, N_{r_n}(u_n) > 0\} \quad \text{for integer } j \geq 1, \tag{3.6}$$

where $N_r(u) = \sum_{i=1}^r I(X_i > u)$ is the number of exceedances in a block of length r, and denote by $\lfloor x \rfloor$ the integer part of x.

Theorem 3.4. *If there exist sequences of constants $a_n > 0$ and b_n such that, for some x,*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x) > 0 \quad \text{as } n\to\infty$$

and if, for $u_n = a_n x + b_n$, $\Delta'(u_n)$ holds, there exist sequences of constants k_n and l_n such that $k_n \to \infty$, $k_n l_n = o(n)$ and $k_n\alpha'(n,l_n) \to 0$, and there exists a distribution

π such that, when $r_n = \lfloor n/k_n \rfloor$, the cluster-size distribution $\pi_n(j) \to \pi(j)$ for all positive integers j, then $Q'_n \xrightarrow{d} Q'$, where Q' is a compound Poisson process on $(0, 1]$ with mark distribution π and intensity measure

$$\Lambda(A) = -(b-a)\log G(x)$$

on $A = (a, b)$ when $0 \leq a < b \leq 1$.

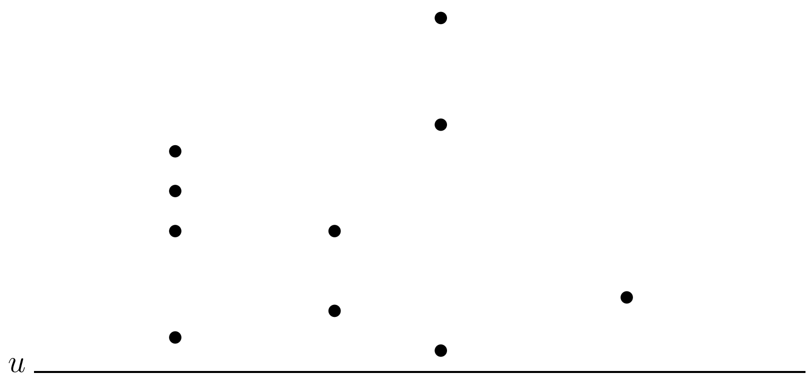
The x in the intensity measure of Q' is the particular, fixed value used to define the threshold u_n. Therefore, the result states that, in the limit, the threshold is exceeded at times that occur according to a homogeneous Poisson process, and that each event comprises multiple exceedances distributed according to the limiting cluster-size distribution, π. This agrees with the structure of the limiting two-dimensional process in Theorem 3.3.

Leadbetter (1983) notes the following connection between the cluster-size distribution and the extremal index θ:

$$\theta^{-1} = \lim_{n\to\infty} \sum_{j=1}^{\infty} j\pi_n(j), \tag{3.7}$$

that is, the extremal index is the reciprocal of the limiting mean cluster size. To see this, recall that the intensity (3.4) of clusters is smaller by a factor θ than the intensity (2.2) of exceedances in independent sequences. Since the expected number of exceedances is unaffected by the strength of the dependence, an average of $1/\theta$ exceedances must occur in each cluster.

The previous two theorems have described the asymptotic clustering behaviour of stationary processes. The following picture gives an idealised view of a stationary sequence over a high but finite threshold u.



The four clusters in the diagram might represent storms, heat waves or other extreme events. Certain features of a cluster, such as the number of exceedances, may be of particular interest. In general, cluster characteristics can be written as

$$c\{(X_i - u) : i \in \mathcal{S}\}, \quad \mathcal{S} \subset \mathbb{N},$$

for a set $\{X_i : i \in \mathcal{S}\}$ of threshold exceedances. Examples of cluster characteristics are

$$\begin{array}{ll} \text{Cluster Size} & \sum_{i \in \mathcal{S}} I\{(X_i - u) > 0\}, \\ \text{Cluster Excess} & \sum_{i \in \mathcal{S}} (X_i - u), \\ \text{Maximum Excess} & \max_{i \in \mathcal{S}} (X_i - u). \end{array} \tag{3.8}$$

The first step in making inferences about such functionals is identifying clusters in the data. Declustering has two stages: identifying those observations regarded as extreme, and grouping those observations into clusters. In the univariate case, the first stage is completed by choosing a threshold u above which the limiting behaviour described in Theorem 3.3 is believed to be a good approximation. Threshold choice involves a trade-off between bias and variance: if the threshold is too low then limiting forms may be poor models; if the threshold is too high then there will be few observations with which to make inferences. Often, if a limiting form holds at one threshold then it holds also at all higher thresholds. Therefore, one strategy is to try several thresholds before selecting u to be the lowest in the range of thresholds for which estimates appear stable and models fit the data reasonably well. Objective threshold selection procedures are described by Guillou and Hall (2001) among others; a different approach, which requires the specification of a model for the data below the threshold, has been developed by Frigessi *et al.* (2002). The second stage of declustering is not trivial: threshold exceedances do not occur at single points in time except in the limit; instead, they are slightly spread out and it may not be obvious if a group of exceedances should form one cluster or be split into two clusters. This problem is the topic of Chapter 4.

Suppose that C clusters have been identified above u, where the i-th cluster comprises exceedances $\mathcal{C}_i = \{X_j : j \in \mathcal{S}_i\}$ in some set $\mathcal{S}_i \subseteq \{1, \ldots, n\}$. Theorem 3.3 provides a model for these clusters (Leadbetter, 1991). According to the model, the cluster maxima can be thought of as points of a non-homogeneous Poisson process with intensity measure (3.4). This is the same model described in Chapter 2 and can be estimated by maximum likelihood. The theorem also justifies the assumption that clusters are independent and identically distributed. Inference for cluster functionals $c(\cdot)$ can therefore be made using the empirical values $c\{(X_j - u) : j \in \mathcal{S}_i\}$, $1 \leq i \leq C$. Estimators for cluster functionals that exploit more of the structure described by Theorem 3.3 are introduced in Chapter 6.

The statistical methods described so far have relied on point process results for quite general stochastic processes. Another approach for analysing extremes is to select a time-series model for the evolution of the process at extreme levels. For example, Smith (1992) and Smith *et al.* (1997) consider first-order Markov chains with bivariate extreme-value distributions, such as the logistic model (2.9), for consecutive elements (X_i, X_{i+1}). Such models can be fitted to threshold exceedances and then asymptotic properties of the fitted model found by simulation. Simpler models include the moving-maximum process,

$$X_i = \max_{j \geq 0}(\alpha_j Z_{i-j}) \quad \text{for } i \geq 1, \tag{3.9}$$

where the constants, $\alpha_j \geq 0$, satisfy $\sum_{j\geq 0} \alpha_j = 1$ and the Z_i are independent, standard Fréchet random variables. These processes were developed by Newell (1964) and Deheuvels (1983, 1985). A special case of the moving-maximum process, when $\alpha_j = (1-\alpha)\alpha^j$ for $0 \leq \alpha < 1$, is the max-autoregressive process (Leadbetter, 1983) defined recursively by

$$X_i = \max\{\alpha X_{i-1}, (1-\alpha)Z_i\} \quad \text{for } i \geq 1. \tag{3.10}$$

The developments in subsequent chapters will avoid making assumptions about the serial structure in the underlying process, preferring to make inferences free of any models not justified by asymptotic theory. The time-series models listed above, for which extremal properties can be obtained analytically or by simulation, will be used instead in simulation studies for assessing the performance of new estimators.

3.2 Multivariate Sequences

Let $\{X_i\}_{i\geq 1}$ be a stationary sequence of D-dimensional random variables $X_i = (X_{i1}, \ldots, X_{iD})$ with marginal distribution function F. Just as Theorem 2.1 has an analogue for stationary processes under a suitable mixing condition, so there is an extension of Theorem 2.3 for componentwise maxima. An appropriate multivariate mixing condition is introduced by Hsing (1989) and is precisely the same as the $D(u_n)$ condition except that the thresholds, u_n, are vectors. Hsing (1989) also proves the following result.

Theorem 3.5. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G and if the multivariate version of $D(u_n)$ holds with $u_n = a_n x + b_n$ for each x such that $G(x) > 0$, then G is a multivariate extreme-value distribution function (2.4).

Inference for componentwise maxima from stationary processes can therefore be conducted with the same models used in the independence case. Dependence again influences the particular limit achieved, and Nandagopalan (1994) defines a multivariate extremal index in direct analogy with Theorem 3.2 as follows. A stationary sequence has multivariate extremal index $\theta(\cdot) \in [0,1]$ if there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$P\left(\frac{\hat{M}_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \text{and} \quad P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G^{\theta(x)}(x).$$

Unlike its univariate counterpart, the multivariate extremal index depends on the point at which the distribution function is evaluated.

A multivariate extension of the point process (3.5) of exceedance times is

$$Q_n'(\cdot) = \sum_{i \in \mathcal{I}} (I(X_{i1} > u_{n1}), \ldots, I(X_{iD} > u_{nD}))\delta_{i/n}(\cdot), \quad \mathcal{I} = \{i : X_i \not\leq u_n\}.$$

This marked point process has marks on $\{0,1\}^D$ and counts points for which X_i exceeds u_n in at least one component. Define also a multivariate extension of the cluster-size distribution (3.6) as

$$\pi_n(j) = P\{N_{r_n}(u_n) = j \,|\, N_{r_n}(u_n) \neq 0\} \quad \text{for } j \in \mathbb{N}_0^D \setminus \{0\},$$

where $N_r(u) = (\sum_{i=1}^r I(X_{i1} > u_1), \ldots, \sum_{i=1}^r I(X_{iD} > u_D))$ and $\mathbb{N}_0$ is the set of non-negative integers. The limit of Q_n' is found by Nandagopalan (1994) under the multivariate version of condition $\Delta'(u_n)$, where, again, u_n is a vector.

Theorem 3.6. *If there exist sequences of constants $a_n > 0$ and b_n such that, for* *some x,*

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x) > 0 \quad \text{as } n \to \infty,$$

and if, for $u_n = a_n x + b_n$, $\Delta'(u_n)$ holds, there exist sequences of constants k_n and l_n such that $k_n \to \infty$, $k_n l_n = o(n)$ and $k_n \alpha'(n, l_n) \to 0$, and there exists a distribution π such that, when $r_n = \lfloor n/k_n \rfloor$, the multivariate cluster-size distribution $\pi_n(j) \to \pi(j)$ for all $j \in \mathbb{N}_0^D \setminus \{0\}$, then $Q'_n \xrightarrow{d} Q'$, where Q' is a compound Poisson process on $(0, 1]$ with mark distribution π and intensity measure

$$\Lambda(A) = -(b - a) \log G(x)$$

on $A = (a, b)$ when $0 \le a < b \le 1$.

The result shows how clusters (of observations extreme in at least one component) occur according to a homogeneous Poisson process in time and that each event comprises a certain number of exceedances in each component distributed according to the limiting multivariate cluster-size distribution, π.

Estimation for features of multivariate clusters has hardly been discussed in the literature. As in the univariate case, the first step is to decluster the data, but this is more complicated here because a single threshold is an insufficient demarcation. Multivariate declustering will be addressed in Chapter 4. Once clusters have been identified, however, the foregoing result justifies the assumption that they are independent. A sequence comprising single observations from each cluster is therefore an independent sequence of multivariate observations that can be modelled on regions bounded away from the origin by the Poisson process of Theorem 2.4. This is the model used by Coles and Tawn (1994), where they summarise each cluster by its componentwise maximum. A procedure for making inferences about multivariate clusters that removes the need for such summaries is introduced in Chapter 7.

Time-series models also exist for multivariate extremes. A multivariate extension of the moving-maximum process (3.9) called the M_4 process was introduced by Smith and Weissman (1996) and is defined by

$$X_{id} = \max_{l \geq 1} \max_{j \geq 0} (\alpha_{ljd} Z_{l,i-j}) \quad \text{for } i \geq 1 \text{ and } 1 \leq d \leq D, \tag{3.11}$$

where the constants, $\alpha_{ljd} \geq 0$, satisfy $\sum_{l \geq 1} \sum_{j \geq 0} \alpha_{ljd} = 1$ for $1 \leq d \leq D$ and the $Z_{l,i}$ are independent, standard Fréchet random variables. Multivariate Markov chains with extreme-value transition distributions have been studied by Perfekt (1997) but statistical methods have yet to be developed.