Chapter 4

Declustering & the Extremal Index

4.0 Introduction

In stationary sequences, extreme values can occur in clusters, for which asymptotic characterisations and methods of inference were reviewed in Chapter 3. The first step in making inferences is identifying clusters in the data, a procedure known as declustering. Declustering is discussed in this chapter, first for univariate sequences in Section 4.1. The unsatisfactory nature of declustering schemes currently in use is noted and an alternative that does not suffer from the same failings is proposed. The new scheme relies on a novel interpretation of the extremal index (3.1) as a measure of over-dispersion of times between extreme observations. This also leads to new estimators for the extremal index and the cluster-size distribution

(3.6) that, unlike other estimators, do not require data to be declustered first. A natural extension of these ideas to address declustering in multivariate sequences is presented in Section 4.2.

4.1 Univariate Sequences

Let $\{X_i\}_{i=1}^n$ be a sample from a stationary sequence of random variables with marginal distribution function F, survival function $\bar{F} = 1-F$ and upper end-point $\omega = \sup\{x : F(x) < 1\}$. The aim is to identify within this sample approximately independent clusters of extreme observations. First a definition is required for 'extreme'. In the univariate case, this typically corresponds to an observation exceeding a chosen threshold u. All that remains is to group the exceedances into clusters.

Two popular schemes are blocks and runs declustering. Blocks declustering consists in choosing a block length, b, and partitioning the sequence $\{1, \ldots, n\}$ into $k = \lfloor n/b \rfloor$ blocks, $\{(i-1)b+1, \ldots, ib\}$ for $1 \leq i \leq k$. The scheme stipulates that any extreme observations lying within the same block belong to the same cluster. Runs declustering consists in choosing a run length, r, and stipulates that any extreme observations separated by fewer than r non-extreme observations belong to the same cluster.

An application of these two declustering schemes is to estimation of the extremal index (3.1). The relationship (3.7) shows that the extremal index can be interpreted as the reciprocal of the mean number of exceedances in a cluster. Therefore, an estimator is the ratio of the number of clusters found by a declustering scheme to the total number of extreme observations in the sequence. Let

$$M_{i,j} = \max\{X_{i+1}, \dots, X_j\}.$$

The estimators for the extremal index that correspond to the blocks and runs declustering schemes are

$$\bar{\theta}_n^B(u\,;\,b) = \frac{\sum_{i=1}^{k} I\{M_{(i-1)b,ib} > u\}}{\sum_{i=1}^{kb} I(X_i > u)}$$

and

$$\bar{\theta}_n^R(u\,;\,r) = \frac{\sum_{i=1}^{n} I(X_i > u, M_{i,i+r} \leq u)}{\sum_{i=1}^{n} I(X_i > u)}, \tag{4.1}$$

both of which are described by Smith and Weissman (1994). Hsing (1991) points out that a problem with these estimators, intrinsic to the nature of the declustering schemes involved, is that selection of values for the declustering parameters, b and r, is largely arbitrary. Furthermore, the choice can have a significant impact on the estimate of the extremal index, as illustrated in Figure 4.1.

Estimators for other cluster functionals (3.8) that are based on these declustering schemes are similarly sensitive to the declustering parameter. One example is the runs estimator for the cluster-size distribution, which estimates $\pi(j)$ by the ratio of the number of clusters with j exceedances to the total number of clusters.

Some attempts have been made to guide the choice of declustering parameter, for example Ledford and Tawn (2003), but the issue remains problematic. In Sec-

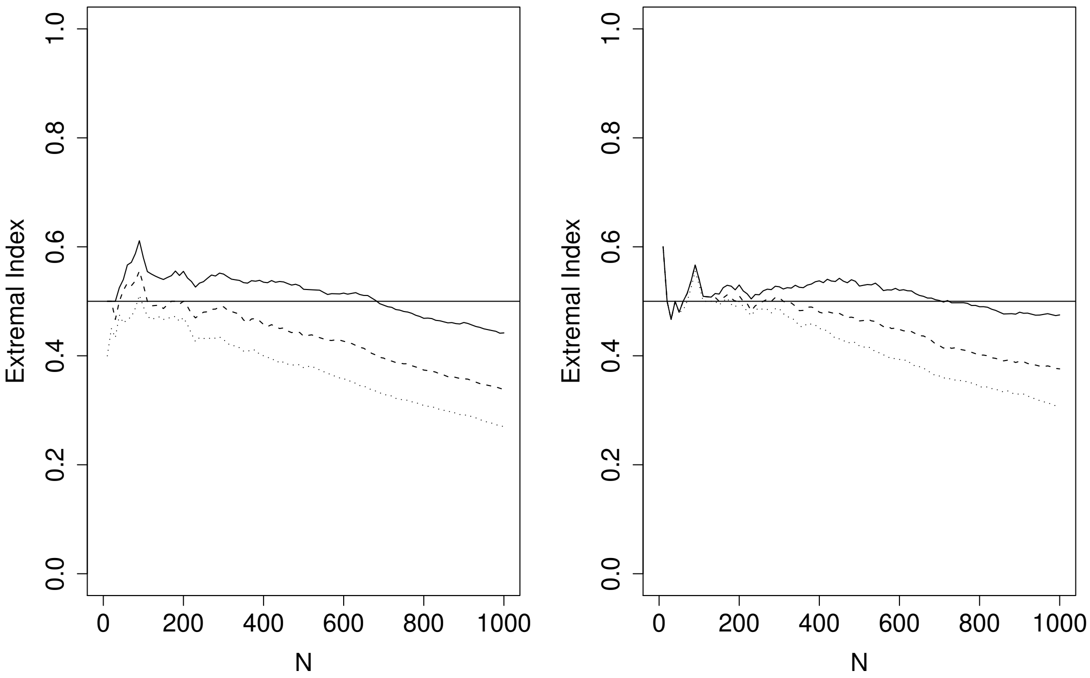


Figure 4.1: Estimates of the extremal index against number of threshold exceedances for a max-autoregressive process of length 10000 with extremal index $\theta = 0.5$. Blocks estimates (left) with $b = 10$ (——), 20 (- - -) and 30 ($\cdots$), and runs estimates (right) with $r = 1$ (——), 5 (- - -) and 9 ($\cdots$).

tion 4.1.2, new estimators for the extremal index and the cluster-size distribution are introduced that, unlike the runs, blocks and other estimators in the literature, do not depend on a declustering scheme. Two more estimation approaches are discussed in Sections 4.1.3 and 4.1.4. It is shown in Section 4.1.5 how the new estimators for the extremal index lead to a declustering scheme that is completely prescribed by the data. The performance of the estimators is evaluated with a simulation study in Section 4.1.6 and a data example in Section 4.1.7. The estimators are motivated by limit results for distributions of times between threshold exceedances, presented in the following section.

4.1.1 Inter-exceedance times

Let $T(u)$ be a random variable representing the time between consecutive exceedances of a threshold u by the process $\{X_i\}_{i\geq 1}$, that is

$$T(u) \stackrel{d}{=} \min\{n \geq 1 : \ X_{n+1} > u\} \text{ given } X_1 > u.$$

The distribution of the inter-exceedance time $T(u)$ is defined by

$$P\{T(u) > n\} = P(M_{1,n+1} \leq u \mid X_1 > u) \quad \text{for } n \geq 1.$$

The asymptotic distribution of $T(u)$ can be found easily when the sequence $\{X_i\}_{i\geq 1}$ is independent. In that case

$$P\{T(u) > n\} = F(u)^n \quad \text{for } n \geq 1,$$

so that, for $t > 0$,

$$P\{\bar{F}(u)T(u) > t\} = P\{T(u) > \lfloor t/\bar{F}(u)\rfloor\} = \exp\{\lfloor t/\bar{F}(u)\rfloor \log F(u)\}.$$

If F has no atom at its upper end-point ω then, since $\log(1+\epsilon) \sim \epsilon$ as $\epsilon \to 0$,

$$\lim_{u\to\omega} P\{\bar{F}(u)T(u) > t\} = \exp(-t) \quad \text{for } t > 0.$$

So $\bar{F}(u)T(u)$ is asymptotically standard exponentially distributed. This agrees with the result (Theorem 2.2) that the point process of exceedances has a Poisson process limit.

Now consider the general case where the sequence $\{X_i\}_{i\geq 1}$ has extremal index $\theta \in [0, 1]$. The corresponding point-process limit for the exceedance times is compound Poisson (Theorem 3.4). This anticipates that the limit distribution of the inter-exceedance times will be a mixture of an exponential distribution and a point-mass on zero. This is indeed the case, as described in Theorem 4.1 below under the following mixing condition, which is stated in slightly more general terms than those introduced in Chapter 3. Let $\mathcal{F}^*_{j,k}(u) = \sigma(\{X_i \leq u\} : j \leq i \leq k)$.

Condition $\Delta^*(u_n)$. For all $A \in \mathcal{F}^*_{1,k}(u_n)$ with $P(A) > 0$, $B \in \mathcal{F}^*_{k+l,cr_n}(u_n)$ and $1 \leq k \leq cr_n - l$,

$$|P(B \mid A) - P(B)| \leq \alpha^*(cr_n, l) \quad \text{for each } c > 0$$

and there exists $l_n = o(r_n)$ such that $\alpha^*(cr_n, l_n) \to 0$ as $n \to \infty$ for all $c > 0$.

Theorem 4.1. *Let the positive integers r_n, $n \geq 1$, and the thresholds u_n, $n \geq 1$, be such that*

$$r_n \to \infty, \qquad r_n \bar{F}(u_n) \to \tau, \quad \textit{and} \quad P(M_{r_n} \leq u_n) \to \exp(-\theta\tau),$$

for some $\tau \in (0, \infty)$ and $\theta \in [0, 1]$. If $\Delta^*(u_n)$ holds, then

$$P\{\bar{F}(u_n)T(u_n) > t\} \to \theta \exp(-\theta t) \quad \textit{for } t > 0$$

as $n \to \infty$.

A proof is contained in Appendix 4.A. Note that the mixing condition used is similar to that of Weissman and Novak (1998).

Theorem 4.1 says that if $\{X_i\}_{i\geq 1}$ has extremal index θ then

$$\bar{F}(u)T(u) \xrightarrow{d} T_\theta \quad \text{as } u \to \omega,$$

where T_θ is a random variable distributed according to the mixture distribution

$$(1-\theta)\epsilon_0 + \theta\mu_\theta, \tag{4.2}$$

ϵ_0 is the degenerate probability distribution at zero and μ_θ is the exponential distribution with mean $1/\theta$. Note the dual role of the extremal index: θ is both the proportion of non-zero inter-exceedance times and the reciprocal of the mean of the non-zero inter-exceedance times. These features are evident from quantile-quantile plots such as that in Figure 4.2.

In subsequent sections estimators for the extremal index are obtained by fitting the limiting model (4.2) to observed inter-exceedance times. First, however, the joint distributions of inter-exceedance times are investigated. Recall that the reciprocal of the extremal index is the limit of the mean of the cluster-size distribution (3.6). Therefore, Theorem 4.1 shows that the marginal distribution of inter-exceedance times is informative about the mean of the cluster-size distribution. In the remainder of this section, it is shown that limiting joint distributions of inter-exceedance

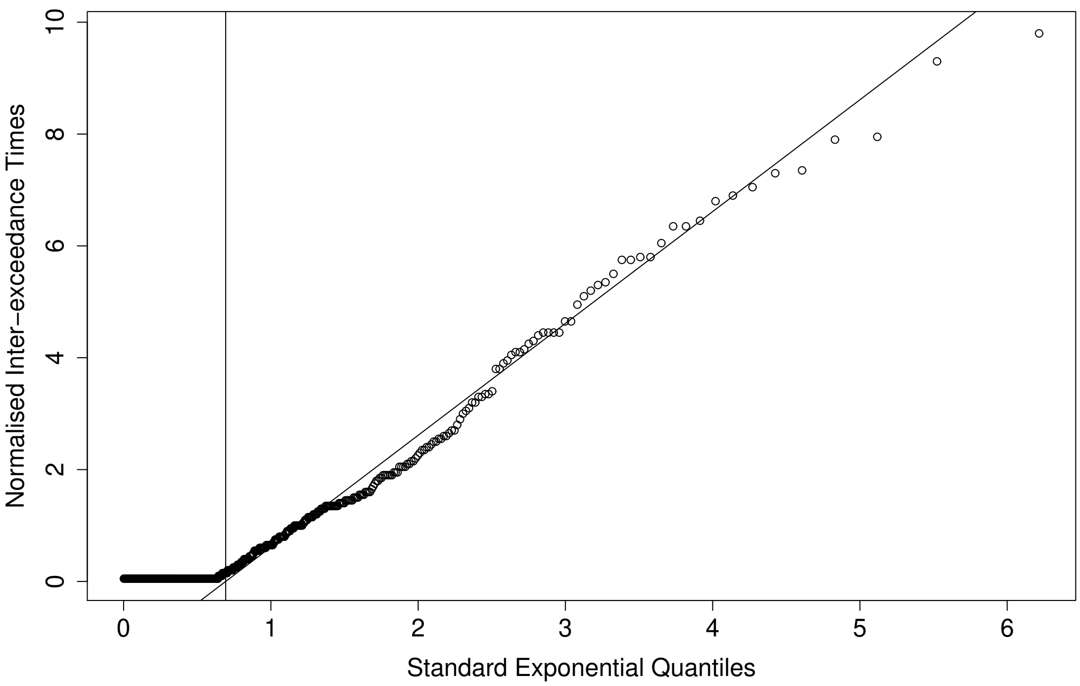


Figure 4.2: Quantile-quantile plot of 500 normalised inter-exceedance times against standard exponential quantiles for a max-autoregressive process of length 10000 with extremal index $\theta = 0.5$. The vertical line indicates the $(1-\theta)$-quantile and the sloping line has gradient $1/\theta$.

times are characterised by the limiting cluster-size distribution itself.

Let $T_1(u)$ and $T_{1+j}(u)$ be random variables representing inter-exceedance times separated by $j-1$ other inter-exceedance times. For example, the joint distribution of $T_1(u)$ and $T_2(u)$ is

$$P\{T_1(u) > s, T_2(u) > t\} = \sum_{r>s} P(M_{1,r} \leq u, X_{r+1} > u, M_{r+1,r+1+t} \leq u \mid X_1 > u)$$

for positive integers s and t.

The limit of this joint distribution will be computed under a stronger mixing condition than Theorem 4.1, namely m-dependence. For a positive integer m, the process $\{X_i\}_{i\geq 1}$ is m-dependent if, for all positive integers k, $\sigma(X_i : 1 \leq i \leq k)$ and $\sigma(X_i : i \geq k+m)$ are independent.

Theorem 4.2. *Let the thresholds u_n, $n \geq 1$, be such that*

$$n\bar{F}(u_n) \to \tau \quad \textit{and} \quad P(M_n \leq u_n) \to \exp(-\theta\tau)$$

for some $\tau \in (0,\infty)$ and $\theta \in (0,1]$. Let also $r_n = o(n)$ be such that $\pi_n(1) \to \pi(1)$, where

$$\pi_n(1) = P\{N_{r_n}(u_n) = 1 \mid N_{r_n}(u_n) > 0\}$$

and $N_r(u) = \sum_{i=1}^{r} I(X_i > u)$. If $\{X_i\}_{i\geq 1}$ is m-dependent then

$$P\{\bar{F}(u_n)T_1(u_n) > s, \bar{F}(u_n)T_2(u_n) > t\} \to \theta\pi(1)e^{-\theta(s+t)} \quad \textit{for } s,t > 0.$$

A proof is presented in Appendix 4.B. Theorem 4.2 says that

$$\big(\bar{F}(u)T_1(u), \bar{F}(u)T_2(u)\big) \xrightarrow{d} (\tilde{T}_1, \tilde{T}_2) \quad \text{as } u \to \omega,$$

where

$$P(\tilde{T}_1 > s, \tilde{T}_2 > t) = \theta\pi(1)e^{-\theta(s+t)} \quad \text{for } s,t \geq 0.$$

This distribution has mass $1-\theta$ on each of the axes $\tilde{T}_1 = 0$ and $\tilde{T}_2 = 0$, and mass $1 - \theta\{2 - \pi(1)\}$ at $(0,0)$.

The following result extends Theorem 4.2 to arbitrary lag.

Theorem 4.3. *Let the thresholds u_n, $n \geq 1$, be such that*

$$n\bar{F}(u_n) \to \tau \quad \text{and} \quad P(M_n \leq u_n) \to \exp(-\theta\tau)$$

for some $\tau \in (0, \infty)$ and $\theta \in (0, 1]$. Let also $r_n = o(n)$ be such that $\pi_n(j) \to \pi(j)$ for all positive integers j. If $\{X_i\}_{i \geq 1}$ is m-dependent then

$$\big(\bar{F}(u_n)T_1(u_n), \bar{F}(u_n)T_{1+j}(u_n)\big) \xrightarrow{d} (\tilde{T}_1, \tilde{T}_{1+j}) \quad \text{as } n \to \infty,$$

for all positive integers j, where

$$P(\tilde{T}_1 > s, \tilde{T}_{1+j} > t) = \theta e_j e^{-\theta(s+t)} \quad \text{for } s, t \geq 0 \tag{4.3}$$

and

$$\begin{aligned} e_1 &= \pi(1), \\ e_j &= \pi(1)e_{j-1} + \ldots + \pi(j-1)e_1 + \pi(j) \quad \text{for } j \geq 2. \end{aligned} \tag{4.4}$$

The proof of this result is discussed in Appendix 4.C.

The results of this section have shown that the limiting distributions (4.2) and (4.3) of normalised inter-exceedance times are characterised by the extremal index and the cluster-size distribution. In the following section, novel, moment-based estimators are presented for these quantities that do not require declustering, but are justified by the limiting distributions. In Sections 4.1.3 and 4.1.4, least-squares and maximum-likelihood estimators are discussed.

4.1.2 Moment estimators

Estimators for the extremal index and the cluster-size distribution can be constructed by equating theoretical moments of the limiting distributions (4.2) and (4.3) to their empirical counterparts.

Extremal index

For a high threshold u, let

$$N = N_n(u) = \sum_{i=1}^{n} I(X_i > u)$$

be the number of observations exceeding u and let $1 \leq S_1 < \ldots < S_N \leq n$ be the exceedance times. The inter-exceedance times are

$$T_i = S_{i+1} - S_i, \quad \text{for } 1 \leq i \leq N-1.$$

The second moment of T_θ with distribution (4.2) is $E(T_\theta^2) = 2/\theta$, which can be estimated to obtain a first estimator for θ. Let $\bar{F}_n(u)$ be an estimator for $\bar{F}(u)$. Then

$$\bar{\theta}_n(u) = \frac{2(N-1)}{\bar{F}_n(u)^2 \sum_{i=1}^{N-1} T_i^2}.$$

For example, if $\bar{F}_n(u) = N/n$ then

$$\bar{\theta}_n(u) = \frac{2n^2(N-1)}{N^2 \sum_{i=1}^{N-1} T_i^2}.$$

This can be improved. The first moment of T_θ is one so that θ is related to the coefficient of variation, ν, of the inter-exceedance times by

$$1 + \nu^2 = E(T_\theta^2)/E(T_\theta)^2 = 2/\theta.$$

In particular, the exceedance times are over-dispersed, with clustering in the limit, if and only if the extremal index is less than one. This relationship motivates another estimator for θ,

$$\hat{\theta}_n(u) = \frac{2\left(\sum_{i=1}^{N-1} T_i\right)^2}{(N-1)\sum_{i=1}^{N-1} T_i^2}, \tag{4.5}$$

where $\bar{F}(u)$ need not be estimated.

A further improvement is possible by considering a penultimate approximation to the limiting mixture distribution (4.2). Setting $r_n = n$ in the proof of Theorem 4.1 (see Appendix 4.A), the distribution of the inter-exceedance times can be seen to satisfy

$$P\{T(u_n) > n\} = \theta F(u_n)^{\theta n} + o(1).$$

An estimator for θ can be based on this relationship. Let T denote a random variable on the positive integers whose distribution is given by

$$P(T > n) = \theta p^{\theta n} \quad \text{for } n \geq 1, \tag{4.6}$$

where $\theta \in (0,1]$ and $p \in (0,1)$ can be thought of as $F(u_n)$. Note that

$$\begin{aligned}
\frac{2E(T)^2}{E(T^2)} &= \frac{2\{1-(1-\theta)p^\theta\}^2}{2\theta p^\theta + \theta p^\theta(1-p^\theta)+(1-p^\theta)^2} \\
&= \theta + \theta(2-3\theta/2)(1-p) + O\left\{(1-p)^2\right\} \quad \text{as } p \to 1,
\end{aligned}$$

which follows from

$$\begin{aligned}
E(T-1) &= \sum_{n=1}^{\infty} P(T>n) = \theta p^\theta (1-p^\theta)^{-1}, \\
E\left\{\frac{T(T-1)}{2}\right\} &= \sum_{n=1}^{\infty} nP(T>n) = \theta p^\theta(1-p^\theta)^{-2}.
\end{aligned}$$

Therefore, the first-order bias of $\hat{\theta}_n(u)$ is approximately $\theta(2-3\theta/2)\bar{F}(u)$. In contrast, the relationship

$$\frac{2\{E(T-1)\}^2}{E\{(T-1)(T-2)\}} = \theta \tag{4.7}$$

motivates the estimator

$$\hat{\theta}_n^*(u) = \frac{2\left\{\sum_{i=1}^{N-1}(T_i-1)\right\}^2}{(N-1)\sum_{i=1}^{N-1}(T_i-1)(T_i-2)}, \tag{4.8}$$

whose first-order bias is approximately zero. Compared to estimator (4.5), the estimators for the first and second moments from which $\hat{\theta}_n^*(u)$ is constructed have been shrunk towards zero. This is intuitively appealing. The smallest observed inter-exceedance times are positive, whereas the limiting distribution (4.2) models them as zero. Estimator $\hat{\theta}_n^*(u)$ ensures that the contributions from the smallest inter-exceedance times are indeed zero; the larger inter-exceedance times are relatively unaffected.

Finally, note that $\hat{\theta}^*_n(u)$ can take values that are greater than one and is undefined if the largest inter-exceedance time is less than three. To avoid these cases, define

$$\tilde{\theta}_n(u) = \begin{cases} 1 \wedge \hat{\theta}_n(u) & \text{if } \max\{T_i :\ 1 \le i \le N-1\} \le 2, \\ 1 \wedge \hat{\theta}^*_n(u) & \text{if } \max\{T_i :\ 1 \le i \le N-1\} > 2, \end{cases} \tag{4.9}$$

and call this the intervals estimator for the extremal index.

Consistency of the intervals estimator (4.9) for m-dependent processes is stated in the following theorem, a proof of which is in Appendix 4.D. Denote convergence in probability by $\xrightarrow{p}$.

Theorem 4.4. *Let the positive integers r_n and the thresholds u_n, for $n \ge 1$, be such that*

$$r_n = o(n), \quad r_n \to \infty, \quad r_n \bar{F}(u_n) \to \tau \quad \text{and} \quad P(M_{r_n} \le u_n) \to \exp(-\theta\tau),$$

for some $\tau \in (0, \infty)$ and $\theta \in (0, 1]$. If $\{X_i\}_{i \ge 1}$ is m-dependent, then $\tilde{\theta}_n(u_n) \xrightarrow{p} \theta$.

Cluster-size distribution

Just as the limiting distribution (4.2) has been used to derive a moment estimator for the extremal index, so the limiting distributions (4.3) can be used to derive moment estimators for the cluster-size distribution. Rewriting the recurrence relation

(4.4) yields

$$\begin{aligned} \pi(1) &= e_1, \\ \pi(j) &= e_j - \pi(1)e_{j-1} - \ldots - \pi(j-1)e_1 \quad \text{for } j \geq 2. \end{aligned} \tag{4.10}$$

Two of many possible expressions for e_j that can be derived from the joint distribution (4.3) are

$$e_j = \theta E(\tilde{T}_1 \tilde{T}_{1+j}) \quad \text{and} \quad e_j = 2E\{\min(\tilde{T}_1, \tilde{T}_{1+j})\}.$$

Recalling $E(\tilde{T}_1) = 1$ and $E(\tilde{T}_1^2) = 2/\theta$, it follows that

$$e_j = 2E(\tilde{T}_1 \tilde{T}_{1+j})/E(\tilde{T}_1^2) \quad \text{and} \quad e_j = 2E\{\min(\tilde{T}_1, \tilde{T}_{1+j})\}/E(\tilde{T}_1).$$

The advantage of these expressions is that the estimators defined by substituting empirical versions of the expectations are independent of the normalisation used for the inter-exceedance times. In particular, e_j can be estimated for each $1 \leq j \leq N-2$ by

$$\hat{e}_j = \frac{2(N-1)\sum_{i=1}^{N-1-j}(T_i - 1)(T_{i+j} - 1)}{(N-1-j)\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)}$$

or

$$\tilde{e}_j = \frac{2(N-1)\sum_{i=1}^{N-1-j}\min(T_i - 1, T_{i+j} - 1)}{(N-1-j)\sum_{i=1}^{N-1}(T_i - 1)}, \tag{4.11}$$

where a penultimate approximation is again used to shrink the expectations. Simulations indicate that estimators based on $\tilde{e}_j$ are superior to those based on $\hat{e}_j$, which are therefore discarded. The estimators, $\tilde{\pi}_n(j)$, for the cluster-size distribution obtained recursively for $j = 2,\ldots,N-2$ by substituting expression (4.11) into relation (4.10) can be constrained by

$$\tilde{\pi}_n(j) = \min\left\{\max\{0, \tilde{e}_j - \tilde{\pi}_n(1)\tilde{e}_{j-1} - \ldots - \tilde{\pi}_n(j-1)\tilde{e}_1\}, 1 - \sum_{i=1}^{j-1}\tilde{\pi}_n(i)\right\} \quad (4.12)$$

to ensure that they form a proper distribution function. These will be referred to as the direct estimators for $\pi(j)$.

4.1.3 Least-squares estimators

In this section, an approach to estimating the extremal index and the cluster-size distribution based on quantile-quantile plots like Figure 4.2 is described. The limiting mixture distribution (4.2) determines that plots of normalised inter-exceedance times against standard exponential quantiles should be piecewise-linear with a breakpoint at the $(1-\theta)$-quantile, $-\log\theta$. The segment on the left should be constant with value zero; the segment on the right should have gradient $1/\theta$. The intervals estimator (4.9) is constructed from the moments of the limiting distribution, so it ignores this structure in the distribution. Here, an attempt is made to exploit the structure evident in the quantile-quantile plot.

Extremal index

Let $\{t_i\}_{i=1}^{N-1}$ be a collection of normalised inter-exceedance times, $NT_i/(n+1)$ say, ordered so that $t_1 \leq \ldots \leq t_{N-1}$. Let $\{x_i\}_{i=1}^{N-1}$ be a collection of standard exponential quantiles, where $x_i = -\log(1 - i/N)$. Figure 4.2 is an example of t_i plotted against x_i. If the t_i were a sample from the limiting distribution (4.2) then the broken-stick model

$$t = \beta(x - \log\beta)_+ \quad \text{for } \beta \geq 1$$

would have $\beta = 1/\theta$. In fact, it is better to fit this model when the t_i are defined by $N(T_i - 1)/(n + 1)$ since then the smallest inter-exceedance times are indeed zero.

Dependence between the ordered inter-exceedance times will be ignored, but heteroskedasticity accounted for by using weighted least squares to reduce the influence of the larger inter-exceedance times. The model is fitted by minimising

$$S(\beta) = \sum_{i=1}^{N-1} w_i \{t_i - \beta(x_i - \log\beta)_+\}^2,$$

where the weights are $w_i = 1/\text{var}(t_i)$. These weights could be computed from the mixture distribution (4.2) but, since θ is unknown and $\theta < 1$ can yield infinite weights, it is assumed for the purpose of choosing the weights that the t_i form an ordered sample of size $N - 1$ from a standard exponential distribution. In this case,

$$w_i^{-1} = \sum_{j=N-i}^{N-1} j^{-2};$$

see David (1981, page 49) for example. The least-squares estimator for the extremal index is

$$\tilde{\theta}_n^{LS}(u) = \{\text{argmin } S(\beta)\}^{-1}, \tag{4.13}$$

where it is necessary to search over β in the interval $[1, N]$ only since $S(\beta)$ takes the same value for all $\beta \geq \exp(x_{N-1}) = N$. More sophisticated choices of weights, using the intervals estimator to obtain a preliminary estimate of θ for example, were found to make little difference to the performance of the least-squares estimator.

Cluster-size distribution

Least-squares estimators can be defined for the cluster-size distribution as well. The joint distribution (4.3) implies

$$P\{\min(\tilde{T}_1, \tilde{T}_{1+j}) > t\} = \theta e_j \exp(-2\theta t),$$

and e_j can be estimated by fitting a broken-stick model to the corresponding quantile-quantile plot. The extremal index can be estimated simultaneously, or a prior estimate such as $\tilde{\theta}_n^{LS}(u)$ can be plugged in. Estimates of the cluster-size distribution follow from the recurrence relation (4.10).

4.1.4 Maximum-likelihood estimators

Two models for the marginal distribution of the inter-exceedance times have been described: the limiting form (4.2) and the penultimate approximation (4.6). In this section, the possibility of using these models to define maximum-likelihood estimators for the extremal index is examined. The joint distribution of all inter-exceedance times, which can be found by applying arguments similar to those used to prove Theorem 4.3, could be used, but likelihoods will be constructed from the marginal distribution assuming independent inter-exceedance times. The joint likelihood is more complicated and suffers from the same problems that are outlined below for the marginal likelihood.

With $t_i = NT_i/n$, the log-likelihood from the limiting model (4.2) is

$$\sum_{i=1}^{N-1} \log \left\{ (1-\theta)^{I(t_i=0)} \left(\theta^2 e^{-\theta t_i}\right)^{I(t_i>0)} \right\} = 2(N-1)\log\theta - \theta \sum_{i=1}^{N-1} t_i,$$

since the observed inter-exceedance times are always strictly positive. The resulting maximum-likelihood estimator for θ is $\min(1, 2/\bar{t})$, where $\bar{t}$ is the mean of the t_i. This estimator tends in distribution to one as n increases, a failure arising from the model assigning all of the inter-exceedance times to the exponential component of the mixture distribution. It is possible to circumvent this problem by grouping a certain number of the smallest inter-exceedance times and treating them as though they were equal to zero. This solution requires the choice of an auxiliary parameter however. The choice is arbitrary and can have a significant impact on estimates, which is no improvement on runs or blocks declustering.

The log-likelihood from the penultimate model (4.6) is

$$m_1 \log(1-\theta p^\theta) + \{\log\theta + \log(1-p^\theta)\}(N-1-m_1) + \theta(\log p)\sum_{i=1}^{N-1}(T_i - 1),$$

where m_1 is the number of inter-exceedance times that are equal to one and $p = F(u)$. Maximum-likelihood estimates can be found by numerical optimisation. Unfortunately, a combination of two features can cause them to perform poorly. First, the distribution (4.6), from which the likelihood is constructed, is in general not a good model for the smallest inter-exceedance times. In particular, $P(T > 1) = \theta p^\theta$ converges to θ as $p \to 1$, where θ arises as the limit of $P(M_{1,m} \leq u \mid X_1 > u)$ as $u \to \omega$ and $m \to \infty$ at an appropriate rate (O'Brien, 1987). On the other hand, $P\{T(u) > 1\} = P(X_2 \leq u \mid X_1 > u)$ is larger than $P(M_{1,m} \leq u \mid X_1 > u)$ and need not converge to θ. For example, the doubly stochastic process of Smith and Weissman (1994) with parameters $0 < \psi < 1$ and $0 < \eta < 1$ has $\lim_{u\to\omega} P\{T(u) > 1\} = 1 - \eta\psi \geq (1-\psi)/(1-\psi+\psi\eta) = \theta$. The second, troublesome feature is the sensitivity of the log-likelihood to m_1, which is clearer after noting that $\sum_{i=1}^{N-1}(T_i - 1) = S_N - S_1 - (N-1)$ is asymptotically equivalent to n as $N \to \infty$ and $N/n \to 0$. Previously made comments warn that the observed value of m_1 may well be far from the value expected under the model, in which case poor estimates result. Grouping small inter-exceedance times and modelling only the larger times is again a possible but unattractive solution.

In contrast to the problems encountered with likelihood estimation, the intervals estimator (4.9) is rather robust to deviations of the smallest inter-exceedance times from the postulated model. This is due to relationship (4.7) in which the numerator and denominator both tend to infinity as $p = F(u) \to 1$, so that the smallest inter-exceedance times play no role asymptotically and the estimator depends most heavily on the well-modelled, large inter-exceedance times.

4.1.5 Declustering and bootstrapping

Although the extremal index and the cluster-size distribution can be estimated without recourse to declustering, estimating other cluster characteristics still may require clusters to be identified. All declustering schemes proposed in the literature require an auxiliary parameter, the choice of which is largely arbitrary. This section explains how the limiting distribution (4.2) can be used to identify clusters without making an arbitrary choice. Declustering will also be shown to support a bootstrap procedure for assessing estimation uncertainty. Importantly, this procedure accounts for the uncertainty in the choice of declustering scheme.

As the limiting process of exceedance times is a compound Poisson process, inter-exceedance times can be categorised as either independent inter-cluster times (between clusters) or independent sets of intra-cluster times (within clusters). As mentioned at the end of Section 4.1.1, the extremal index is the proportion of inter-exceedance times that can be regarded as inter-cluster times. Suppose that N exceedance times, $S_1 < \ldots < S_N$, are observed, with inter-exceedance times $T_i = S_{i+1} - S_i$ for $1 \leq i \leq N-1$, and let $\hat{\theta}$ be an estimate of the extremal index. The largest $C - 1 = \lfloor \hat{\theta}(N-1) \rfloor$ inter-exceedance times can be assumed to be approximately independent inter-cluster times that divide the remainder into approximately independent sets of intra-cluster times. To be precise, if $T_{(C)}$ is the C-th largest inter-exceedance time and T_{i_j} is the j-th inter-exceedance time to exceed $T_{(C)}$, then $\{T_{i_j}\}_{j=1}^{C-1}$ is a set of approximately independent inter-cluster times. If $C = N$, set $T_{(C)} = 0$; in the case of ties, decrease C until $T_{(C-1)}$ is strictly greater than $T_{(C)}$. Let $\mathcal{T}_j = \{T_{i_{j-1}+1}, \ldots, T_{i_j-1}\}$, where $i_0 = 0$, $i_C = N$ and $\mathcal{T}_j = \emptyset$ if $i_j = i_{j-1} + 1$. Then $\{\mathcal{T}_j\}_{j=1}^{C}$ is a collection of approximately independent sets of intra-cluster times. Furthermore, each set $\mathcal{T}_j$ has associated with it a set of threshold exceedances $\mathcal{C}_j = \{X_k : k \in \mathcal{S}_j\}$, where $\mathcal{S}_j = \{S_{i_{j-1}+1}, \ldots, S_{i_j}\}$.

This interpretation justifies a decomposition of the observed process into C clusters, where the j-th cluster comprises the exceedances $\mathcal{C}_j$. This is equivalent to runs declustering with run length $r = T_{(C)}$, which is estimated directly from the data. If $\hat{\theta}$ is the intervals estimate (4.9) then the declustering procedure is entirely automatic and justified by the limiting theory.

Suppose now that inferences are required for a cluster functional $c(\cdot)$ such as those listed in (3.8). The functional can be evaluated for each cluster to obtain values $CF_i = c\{(X_j - u) : j \in \mathcal{S}_i\}$, $1 \leq i \leq C$, that can be used to find estimates, $\widehat{CF}$, of properties of c. For example, the expectation could be estimated by $\widehat{CF} = (CF_1 + \ldots + CF_C)/C$. Assuming independence of clusters and inter-cluster times, the bootstrap can be employed as follows to obtain confidence limits.

1. Resample with replacement $C - 1$ inter-cluster times from $\{T_{i_j}\}_{j=1}^{C-1}$.

2. Resample with replacement C sets of intra-cluster times (some of which may be empty) and associated exceedances from $\{(\mathcal{T}_j, \mathcal{C}_j)\}_{j=1}^{C}$.

3. Intercalate these inter-exceedance times and clusters to form a bootstrap replication of the process above the threshold.

4. Compute N for the bootstrap process, estimate θ, and decluster accordingly.

5. Compute $\widehat{CF}$ from the declustered bootstrap process.

Forming B such bootstrap processes yields collections of estimates, $\{\hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(B)}\}$ and $\{\widehat{CF}_{(1)}, \ldots, \widehat{CF}_{(B)}\}$, that can be used to approximate the distributions of the original point estimates, $\hat{\theta}$ and $\widehat{CF}$. In particular, the empirical α- and $(1-\alpha)$-quantiles of each sample define $(1-2\alpha)$-confidence intervals. This procedure is demonstrated for the data-set in Section 4.1.7, and used in the simulation study of the following section. Note first that if $CF_i = I(\#\mathcal{S}_i = j)$ then $\widehat{CF}$ estimates $\pi(j)$. The estimator

$$\bar{\pi}_n(j) = C^{-1} \sum_{i=1}^{C} I(\#\mathcal{S}_i = j) \quad \text{for } j \geq 1 \tag{4.14}$$

will be referred to as the indirect estimator for the cluster-size distribution.

4.1.6 Simulation study

In this section a simulation study is conducted to investigate the performances of the various estimators introduced in this chapter. For the extremal index comparisons are made between the runs estimator (4.1), which can be written in terms of the inter-exceedance times as

$$\bar{\theta}_n^R(u\,;\,r) = N^{-1} \left\{ \sum_{i=1}^{N-1} I(T_i > r) + 1 \right\}$$

when the run length is r, the intervals estimator (4.9) and the least-squares estimator (4.13). For the cluster-size distribution only estimators for $\pi(1)$ are compared:

estimators for other elements of the distribution were found to have similar relative performance. The estimators used are the direct estimator (4.12), the indirect estimator (4.14) when both intervals declustering and runs declustering are used, and the least-squares estimator described in Section 4.1.3.

Data are simulated from two stationary processes: a max-autoregressive process (3.10) with dependence parameter α, and a first-order Markov chain with extreme-value logistic transition distribution (2.9) and dependence parameter β. The extremal index of the Markov chain for specific values of β can be found by simulation (Smith *et al.*, 1997). One-thousand sequences of length 5000 are simulated from both of these processes for each of three extremal indices, $\theta = 0.25$, 0.5 and 0.75, corresponding to $\alpha = 0.75$, 0.5 and 0.25 for the max-autoregressive process, and $\beta = 0.43$, 0.64 and 0.82 for the Markov chain. For the max-autoregressive process, $\pi(1) = \theta$; for the three Markov chains, the values of $\pi(1)$ are 0.35, 0.56 and 0.77.

Extremal index

First, the intervals estimator is compared to three runs estimators, with run lengths 1, 5 and 9, at a range of thresholds chosen so that there are N exceedances, with N ranging from 10 to 1000. Root-mean-square errors of the point estimates and the coverage probability of bootstrapped confidence intervals, computed with $B = 1000$, are reported in each case.

Figure 4.3 shows the empirical root-mean-square errors and coverage probabilities for the four estimators applied to the three max-autoregressive processes. The sensitivity of the runs estimator to run length, threshold and the true extremal index is clear; the intervals estimator is robust to these factors. This robustness is particularly evident at lower thresholds, where the coverage probability for the intervals estimator remains close to the nominal value. The cause of the poor coverage for the runs estimator can be traced to step (d) of the bootstrap algorithm: the run length is fixed so that any uncertainty in the declustering scheme is unaccounted for, resulting in confidence intervals that are too narrow. The apparently uniform decrease of root-mean-square error with threshold for the intervals estimator is only indicative of the fact that the optimal threshold for the intervals estimator is typically lower than that for the runs estimator. The reason for this is the greater sensitivity of the runs estimator to disproportionately large numbers of small inter-exceedance times, which enhances the bias at low thresholds. The intervals estimator is robust to such discrepancies. Nevertheless, the runs estimators appear to outperform the intervals estimator for high thresholds. Figure 4.4 presents the results for the Markov chains. Similar comments can be made here with regard to the robustness of the intervals estimator. Note, however, that not every choice of run length for the runs estimator leads to a performance that is superior at high thresholds to that of the intervals estimator.

In Figure 4.5, root-mean-square error is plotted for the least-squares estimator, the intervals estimator and the runs estimators of the extremal index. The least-squares estimator often performs similarly to the best runs estimator, improving on the intervals estimator at high thresholds but worsening as the threshold decreases. This behaviour is due to the break-point in the quantile-quantile plot being poorly

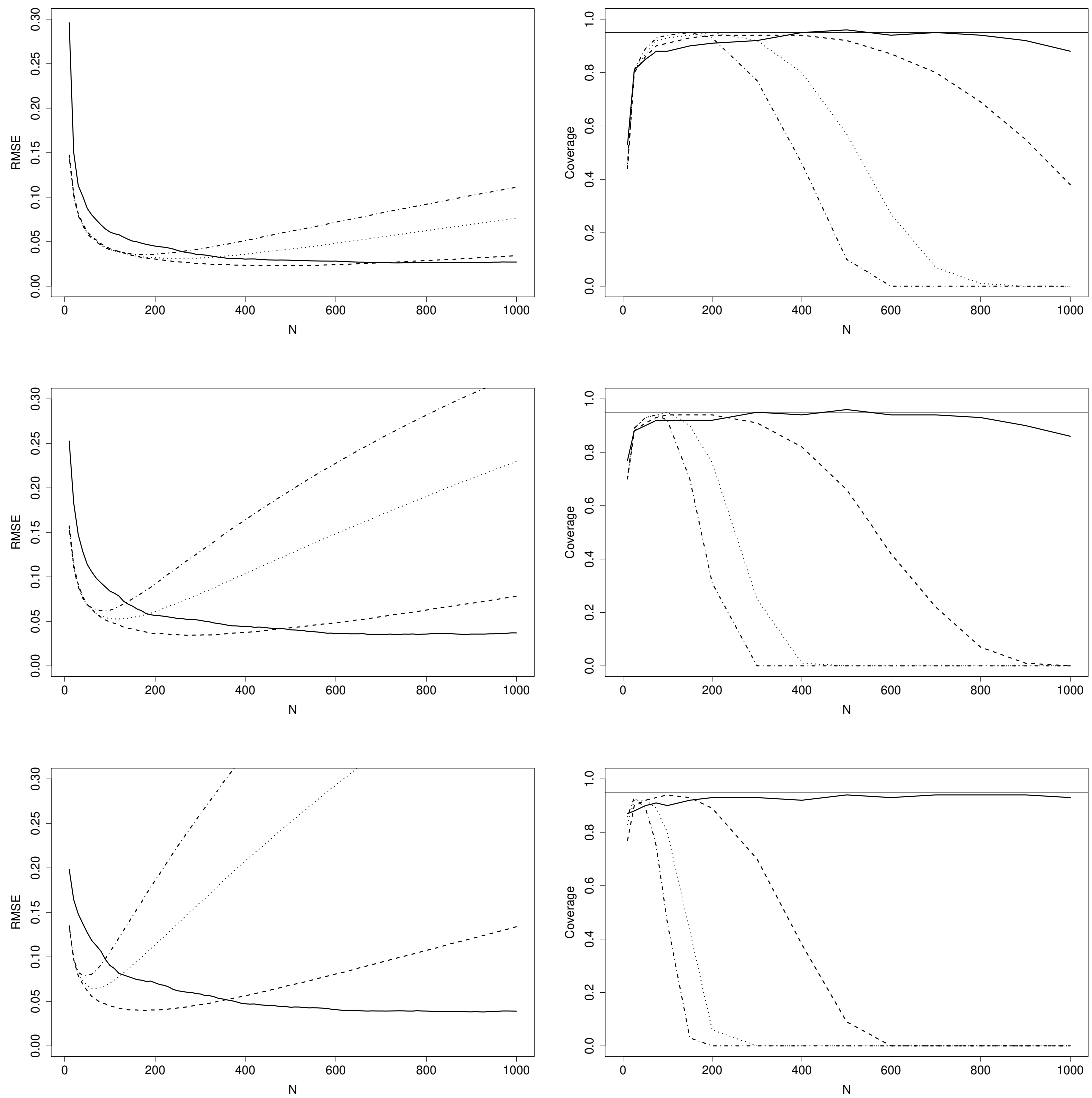


Figure 4.3: Root-mean-square errors (left) and coverage probabilities (right) for the intervals estimator (——) and runs estimators with run lengths 1 (- - -), 5 (· · ·) and 9 (-·-·-) applied to max-autoregressive sequences of length 5000 with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom). The nominal coverage probability (0.95) is indicated by a horizontal line.

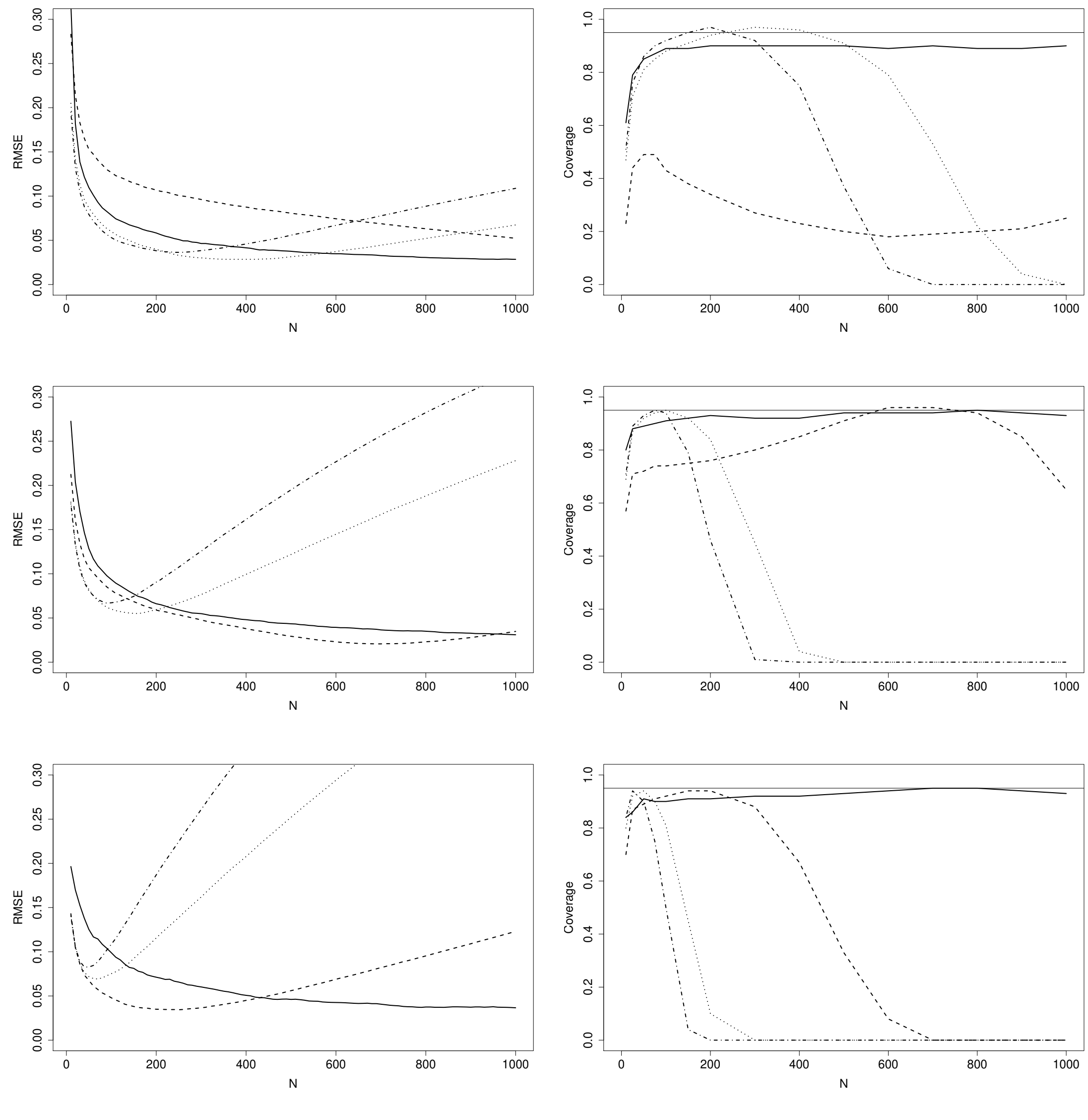


Figure 4.4: Root-mean-square errors (left) and coverage probabilities (right) for the intervals estimator (——) and runs estimators with run lengths 1 (- - -), 5 (· · ·) and 9 (-·-·-) applied to Markov chain sequences of length 5000 with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom). The nominal coverage probability (0.95) is indicated by a horizontal line.

defined at lower thresholds. As with the intervals estimator, the least-squares estimator does not require the choice of an auxiliary parameter, so its performance is very favourable.

Cluster-size distribution

Finally, a comparison is made of the estimators for $\pi(1)$, with N in the smaller range 10 to 500. The root-mean-square errors for the indirect estimators with both runs and intervals declustering are shown in Figure 4.6. Intervals declustering is seen to provide a robust alternative to runs declustering. The root-mean-square error for the direct estimator, the least-squares estimator and the indirect estimator with intervals declustering are compared in Figure 4.7. The least-squares estimator performs similarly to the indirect estimator, with the direct estimator performing slightly worse. The least-squares estimator for $\pi(1)$ is therefore competitive, but, as it is more expensive to compute than the indirect estimator, the cluster-size distribution and other cluster functionals are perhaps best estimated by declustering.

4.1.7 Data example

To conclude, the intervals estimator and intervals declustering are applied to a sequence of daily minimum temperatures recorded to the nearest degree Fahrenheit at Wooster, Ohio, between 1 June 1893 and 31 December 1987. The raw data exhibit strong seasonality, which is removed here by using the data from only

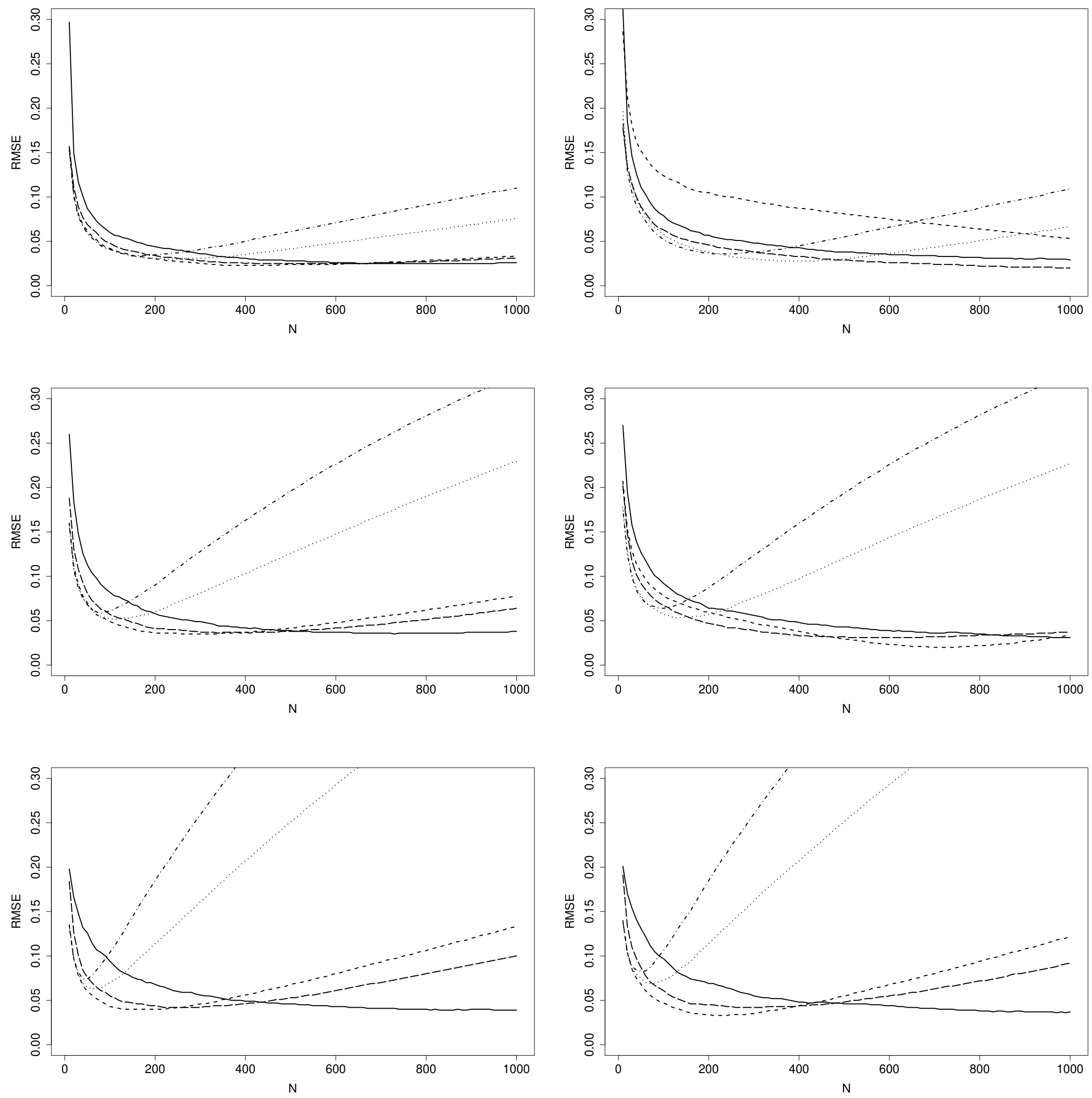


Figure 4.5: Root mean-square-error for the intervals estimator (——), the least-squares estimator (– – –) and the runs estimator with run lengths 1 (- - -), 5 (· · ·) and 9 (-·-·-) applied to max-autoregressive (left) and Markov chain (right) sequences of length 5000 with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom).

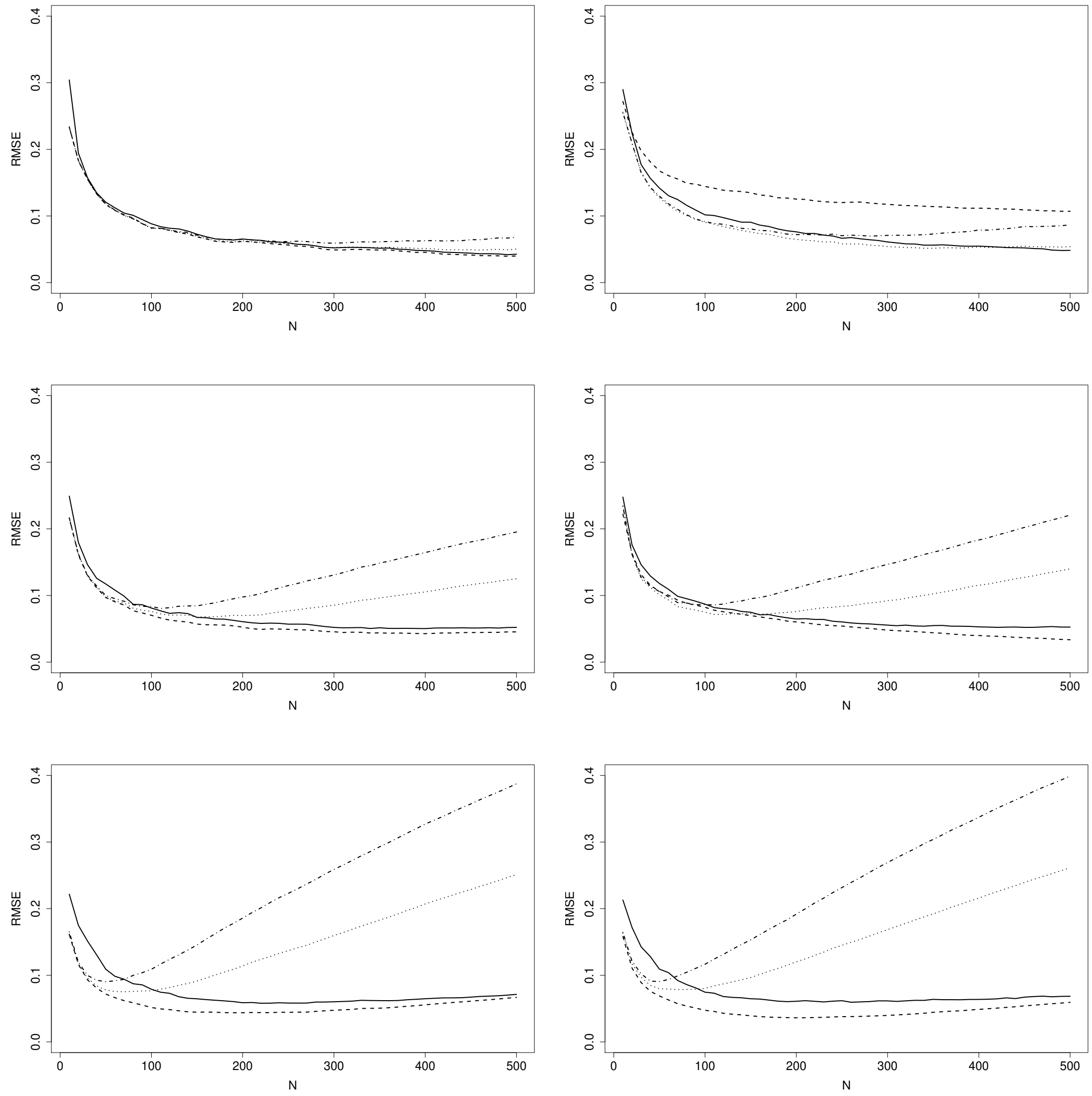


Figure 4.6: Root mean-square-error for the indirect estimator of $\pi(1)$ with intervals declustering (——) and runs declustering with run lengths 1 (- - -), 5 ($\cdots$) and 9 (-·-·-) applied to max-autoregressive (left) and Markov chain (right) sequences of length 5000 with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom).

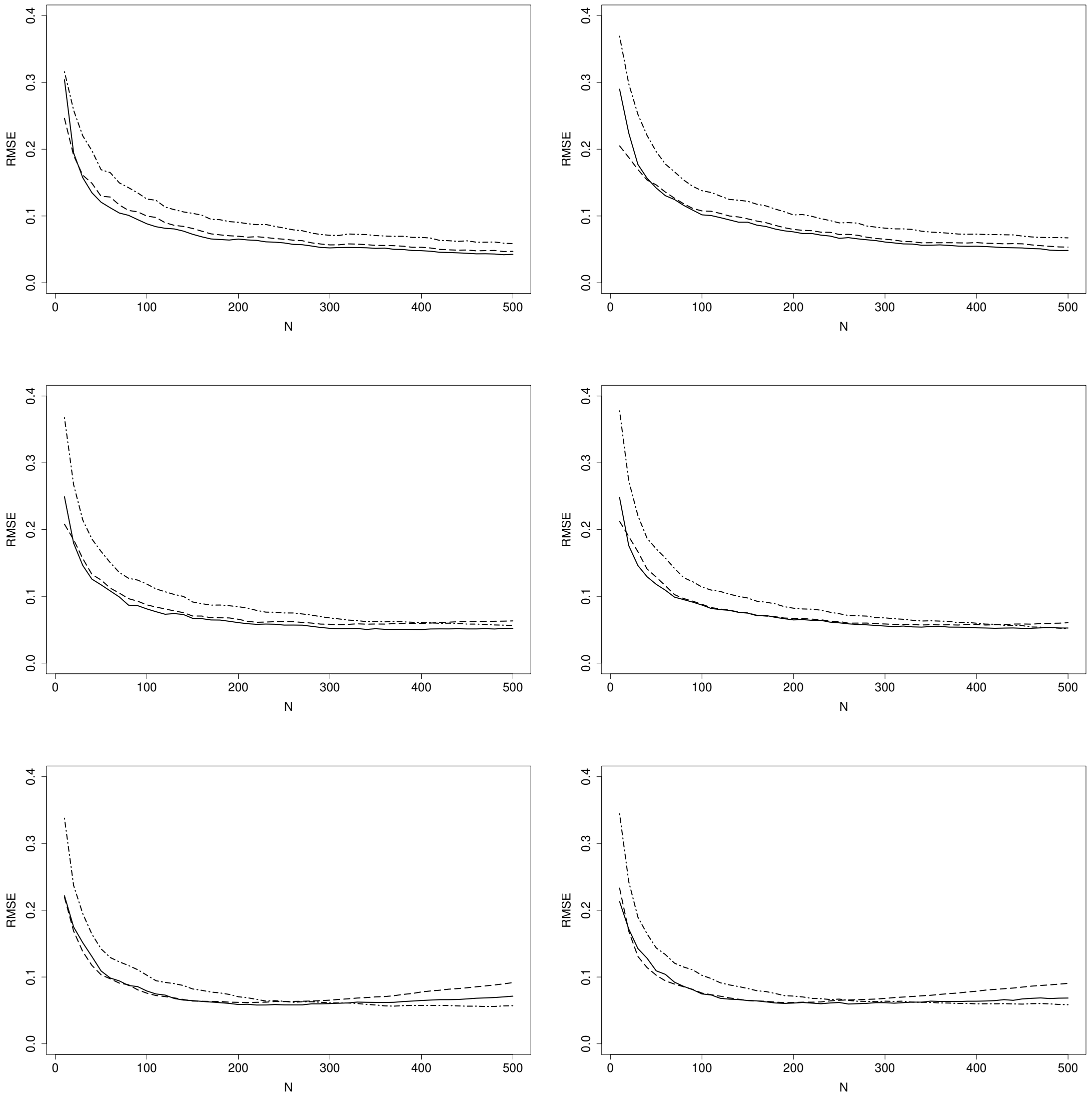


Figure 4.7: Root mean-square-error for the indirect estimator of $\pi(1)$ with intervals declustering (——), the least-squares estimator (– – –) and the direct estimator (-·-·-) applied to max-autoregressive (left) and Markov chain (right) sequences of length 5000 with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom).

the months December, January and February. Two missing values on 1 February 1954 and 3 January 1981 are also removed. The concatenated data are plotted in Figure 4.8. The extremal index, cluster-size distribution and mean cluster excess (3.8) of the series will be estimated at thresholds $u = 0.5, 1.5, \ldots, 14.5$. See Smith *et al.* (1997) for additional information and analysis.

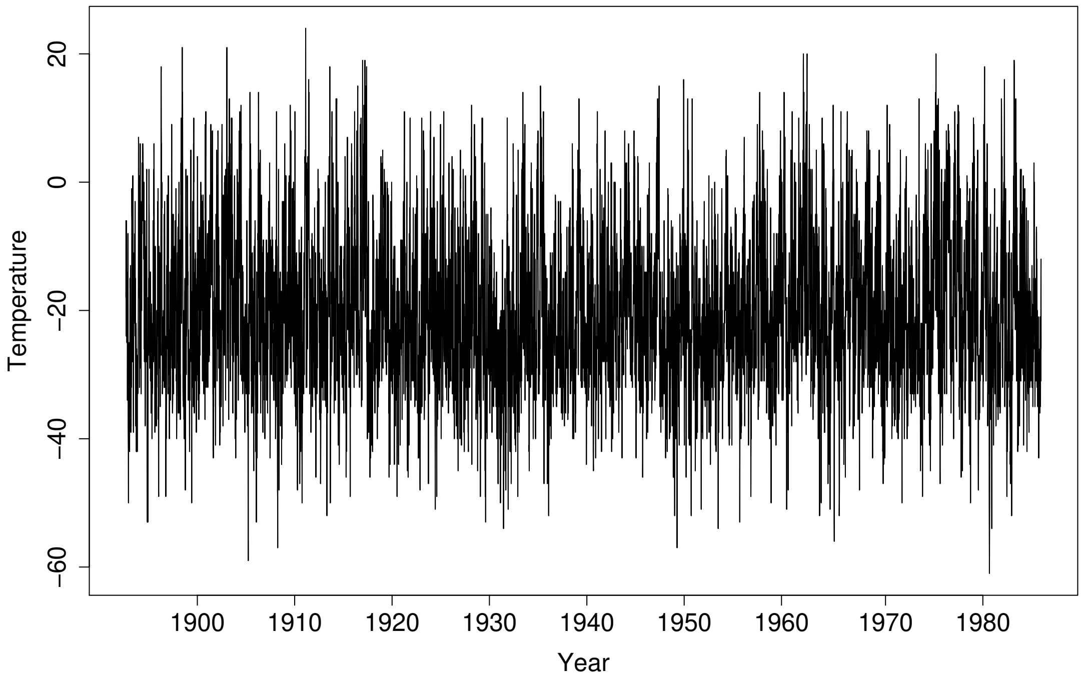
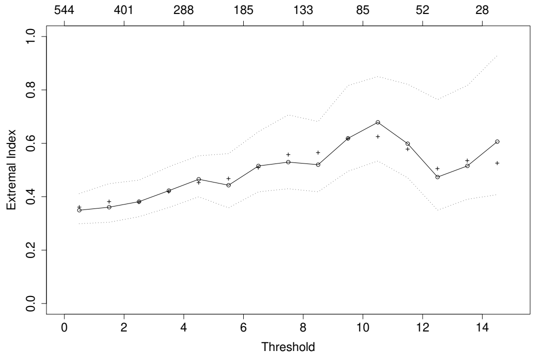


Figure 4.8: Daily minimum temperature recorded in degrees Fahrenheit below zero at Wooster. The first observation in each decade is labelled.

The intervals estimates of the extremal index are shown in Figure 4.9 with bootstrap confidence limits. For comparison, least-squares estimates and runs estimates with run length five days are also shown. The point estimates from the three methods are similar. As expected, the variance of the intervals estimator is greater than that of the runs estimator, which does not account for uncertainty



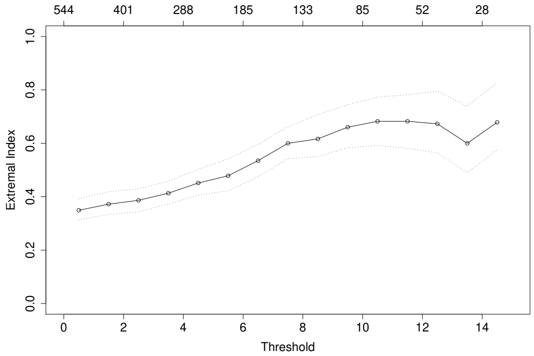


Figure 4.9: Estimates (–•–) of the extremal index for the Wooster temperature series. The intervals estimator (left) and the runs estimator with run length five (right) are shown with bootstrapped 95% confidence limits (- - -). The least-squares estimator (+) is superimposed and the number of exceedances at each threshold is indicated on the upper axis.

in the run length. Both plots support an extremal index of about 0.6, which is similar to the values found by Smith *et al.* (1997).

Estimates of the cluster-size distribution, obtained by declustering according to the intervals scheme, are in Figure 4.10. The estimates of $\pi(1)$ are similar to those of the extremal index; the estimates of $\pi(2)$ are around 0.25. Approximately 4% of clusters have more than four exceedances. Again, these values are similar to those found by Smith *et al.* (1997).

The estimates of the mean cluster excess are presented in the left-hand plot of Figure 4.11. These decrease approximately linearly with unit slope as the threshold increases before levelling out at the ten-degree threshold with mean cluster excess about six degrees. The pattern is the inverse of the extremal index, suggesting that the mean cluster excess follows the mean cluster size, with the distribution of excesses approximately stable over thresholds.

Finally, returning to the bootstrap procedure of Section 4.1.5, notice that each

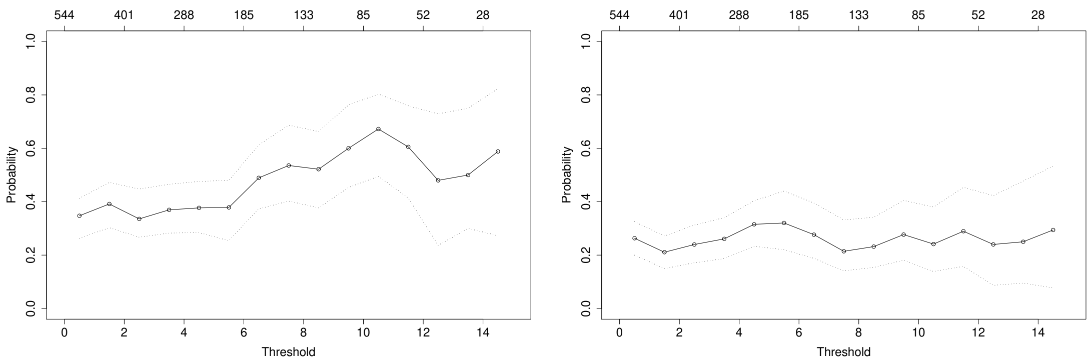


Figure 4.10: Estimates (–•–) based on intervals declustering of the cluster-size distribution, size 1 (left) and 2 (right), for the Wooster temperature series. Bootstrapped 95% confidence limits (- - -) are also shown. The number of exceedances at each threshold is indicated on the upper axis.

bootstrapped process has an associated run length, $r_{(b)} = T_{(C_{(b)})}$, where $C_{(b)} = \lceil \hat{\theta}_{(b)}(N_{(b)} - 1) \rceil$ and $N_{(b)}$ is the number of exceedances. At each threshold, the collection $\{r_{(b)}\}_{b=1}^{B}$ describes the uncertainty in the declustering scheme and indicates the extent of the uncertainty ignored by using runs declustering with an arbitrary run length. The right-hand plot in Figure 4.11 summarises the run-length distribution. A run length of about one week is suggested with the interpretation that cold periods separated by more than one week can be considered independent.

As well as stability of cluster estimates, the goodness-of-fit of the limiting model (4.2) to the observed inter-exceedance times can help to identify thresholds above which the asymptotic models are reasonable approximations. The fit can be assessed using quantile-quantile plots. For the Wooster data, all of the thresholds considered in the analysis gave good fits.

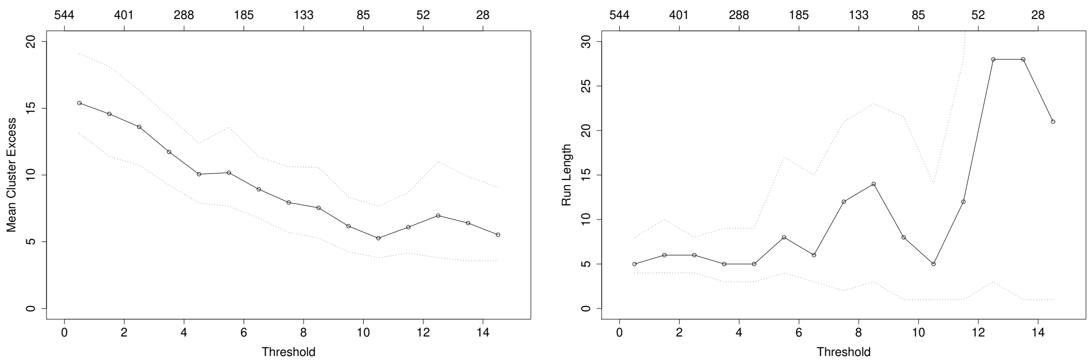


Figure 4.11: Estimates (–•–) based on intervals declustering of mean cluster excess (left) and run length (right) for the Wooster temperature series. Bootstrapped 95% confidence limits (- - -) are also shown. The number of exceedances at each threshold is indicated on the upper axis.

4.2 Multivariate Sequences

A few declustering schemes have been proposed for multivariate sequences: for example, Coles and Tawn (1991) and Nadarajah (2001). These schemes are designed to extract independent observations from a multivariate, stationary sequence: clusters are identified and then summarised by a single value, such as the componentwise maximum of the observations in the cluster. The approach of Coles and Tawn (1991) is a multivariate version of blocks declustering; that of Nadarajah (2001) is an extension of runs declustering. Both methods require the choice of one or more declustering parameters. A multivariate extension of the scheme outlined in the previous section that does not require an arbitrary choice of declustering parameter is proposed below.

Before declustering, the subset of observations considered to be extreme must be chosen. For univariate sequences this amounts to choosing a threshold; for multivariate data the decision is less straightforward. In Section 2.2, defining an observation to be extreme if and only if it falls in a set $A = \{x \in \mathbb{R}^D : x \nleq u\}$ for some choice of high thresholds $u = (u_1, \ldots, u_D)$ allowed the marginal and dependence features of the point process model to be estimated simultaneously. Adopting this definition, the problem of multivariate declustering reduces to the problem of declustering the univariate sequence, $\{I(X_i \in A)\}_{i=1}^n$, of indicator variables. The approach of the previous section can be applied to this sequence to accomplish the task as the next result shows.

Let $A_n = \{x \in \mathbb{R}^D : x \nleq u_n\}$ for $u_n = (u_{n1}, \ldots, u_{nD})$ and define $p(A_n) = P(X_1 \in A_n)$. Let $T(A_n)$ be the random variable representing the return time to the set A_n, with distribution

$$P\{T(A_n) > n\} = P(X_i \notin A_n \text{ for } 2 \leq i \leq n \mid X_1 \in A_n).$$

The following theorem uses a multivariate version of the $\Delta^*(u_n)$ condition of Section 4.1.1, in which the only difference is that the random variables and the thresholds have dimension D.

Theorem 4.5. *Let the thresholds u_n, $n \geq 1$, be such that*

$$nP(X_{1d} > u_{nd}) \to \tau_d \quad \textit{for } 1 \leq d \leq D$$

and $\{X_i\}_{i=1}^n$ has multivariate extremal index $\theta(\tau)$, that is

$$np(A_n) \to \nu \quad \textit{and} \quad P(M_n \leq u_n) \to \exp\{-\theta(\tau)\nu\}$$

for some $\nu \in (0, \infty)$ *and where* $\tau = (\tau_1, \ldots, \tau_D)$. *If the multivariate version of* $\Delta^*(u_n)$ *holds, then*

$$P\{p(A_n)T(A_n) > t\} \to \theta(\tau)\exp\{-\theta(\tau)t\} \quad \textit{for all } t > 0$$

as $n \to \infty$.

Proof. Let $\tilde{X} = \Psi(X)$ have standard Fréchet component distributions and let $\tilde{u}_n = n/\tau \sim \Psi(u_n)$. Following Proposition 2.1 of Smith and Weissman (1996), let

$$Z_i(\tau) = \max_{1 \le d \le D} \{\tau_d \tilde{X}_{id}\}.$$

Then,

$$P\{Z_1(\tau) \le n\} = P(\tilde{X}_1 \le \tilde{u}_n) \sim P(X_1 \le u_n)$$

so that the multivariate condition $\Delta^*(u_n)$ for the process $\{X_i\}_{i=1}^n$ is equivalent to the univariate analogue for the process $\{Z_i(\tau)\}_{i=1}^n$ with threshold n. Define $\bar{F}_Z(z) = P\{Z_1(\tau) > z\}$, $M_n^Z = \max_{1 \le i \le n} Z_i(\tau)$ and let $T_Z(z)$ be the random variable representing the time between exceedances of z by $\{Z_i(\tau)\}_{i=1}^n$. Now,

$$n\bar{F}_Z(n) = nP(\tilde{X}_1 \not\le \tilde{u}_n) \sim nP(X_1 \not\le u_n) \to \nu$$

and

$$P(M_n^Z \le n) = P(\tilde{M}_n \le \tilde{u}_n) \sim P(M_n \le u_n) \to \exp\{-\theta(\tau)\nu\},$$

that is $\{Z_i(\tau)\}_{i=1}^n$ has extremal index $\theta(\tau)$. Therefore, Theorem 4.1 applies and yields

$$P\{p(A_n)T(A_n) > t\} \sim P\{\bar{F}_Z(n)T_Z(n) > t\} \to \theta(\tau)\exp\{-\theta(\tau)t\} \quad \text{for } t > 0$$

as $n \to \infty$. □

The application of this result is demonstrated with a simple example due to Smith and Weissman (1996). Consider the bivariate M_4 process (3.11) with $\alpha_{101} = 1$, $\alpha_{102} = \alpha_{112} = 1/2$, that is $X_i = (X_{i1}, X_{i2})$, where

$$X_{i1} = Z_i \quad \text{and} \quad X_{i2} = \max(Z_{i-1}, Z_i)/2$$

and the Z_i are independent, standard Fréchet random variables. The multivariate extremal index satistfies $\theta(c\tau) = \theta(\tau)$ for all positive c (Nandagopalan, 1994), so the extremal index of the bivariate process has the form $\theta(\tau_1, \tau_2) = \theta(w, 1-w)$, where $w = \tau_1/(\tau_1 + \tau_2) \in [0, 1]$. It is straightforward to derive

$$\theta(w, 1-w) = \begin{cases} 1/2 & \text{if } 0 \le w \le 1/3, \\ 2w/(1+w) & \text{if } 1/3 \le w \le 1. \end{cases}$$

The limiting cluster-size distribution, where clusters contain values in the set A_n above, satisfies

$$\pi(1) = \begin{cases} 0 & \text{if } 0 \le w \le 1/3, \\ (3w-1)/2w & \text{if } 1/3 \le w \le 1 \end{cases}$$

and $\pi(2) = 1 - \pi(1)$ (Smith and Weissman, 1996).

By the theorem, an estimator for $\theta(c\tau)$ is the intervals estimator evaluated for return times to a set $A = \{x \in \mathbb{R}^2 : x \not\leq u\}$, where $u = (u_1, u_2)$ satisfies

$$n\bar{F}_1(u_1) = cw \quad \text{and} \quad n\bar{F}_2(u_2) = c(1-w).$$

In practice, thresholds are estimated using the component empirical distribution functions, and the constant c can be chosen so that the number of points in A is the same for each w. This method, called Method I, can be adapted to any definition of the extreme set A. A simpler but less general approach, Method II, is to apply the intervals estimator to the univariate sequence $Z_i(\tau) = \max\{wX_{i1}, (1-w)X_{i2}\}$, $1 \leq i \leq n$, identified in the proof of the theorem as having extremal index $\theta(\tau)$. Other estimators are described by Smith and Weissman (1996).

The multivariate process can be declustered by applying intervals declustering to either the return times to A, or the inter-exceedance times of $\{Z_i(\tau)\}_{i=1}^n$. The cluster-size distribution and other cluster functionals can then be estimated empirically.

A single realisation of length 5000 is simulated from the M_4 process, and the extremal index and cluster-size distribution are estimated. For estimates based on return times to the extreme set A, c is chosen so that A contains 250 observations for each w. Estimates based on $\{Z_i(\tau)\}_{i=1}^n$ are evaluated at a threshold equal to the empirical 95% quantile. The estimates of the multivariate extremal index and the cluster-size distribution from the two methods are displayed in Figure 4.12.

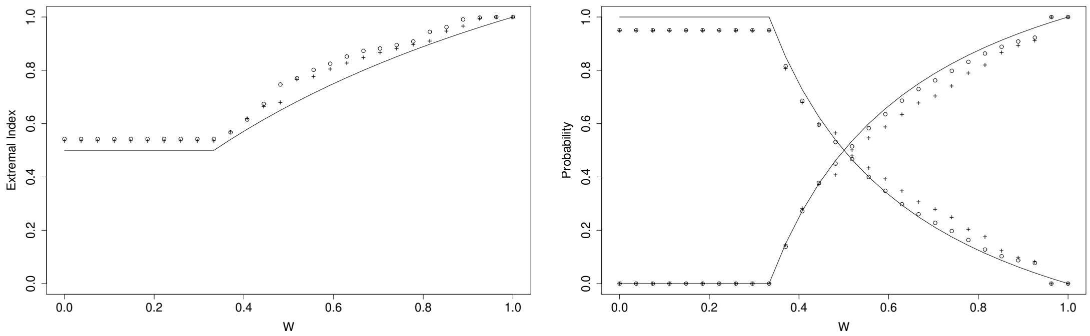


Figure 4.12: Estimates of the extremal index (left) and the first two elements of the cluster-size distribution (right) using Method I (∘) and Method II (+). In both plots, the solid lines correspond to the truth. For the cluster-size distribution, $\pi(1)$ increases and $\pi(2)$ decreases as w increases.

4.A Proof of Theorem 4.1

(The proofs in Appendices 4.A and 4.D were constructed by Johan Segers, co-author to Ferro and Segers (2003a).)

Since both $l_n = o(r_n)$ and $\alpha_n = \alpha_{r_n,l_n}(u_n) = o(1)$, there exist positive integers p_n such that

$$p_n = o(r_n), \quad r_n\alpha_n = o(p_n) \quad \text{and} \quad l_n = o(p_n).$$

(Take for instance $p_n = \lfloor \{r_n \max(l_n, r_n\alpha_n)\}^{1/2} \rfloor$.) Since $F(u_n)^{r_n} \to e^{-\tau}$, all conditions of Corollary 2.3 of O'Brien (1987) are fulfilled, and

$$P(M_{r_n} \le u_n) - F(u_n)^{r_n P(M_{1,p_n} \le u_n | X_1 > u_n)} \to 0.$$

But $P(M_{r_n} \leq u_n) \to e^{-\theta\tau}$ by assumption, so necessarily

$$P(M_{1,p_n} \leq u_n \mid X_1 > u_n) \to \theta.$$

Now put $t_n = \lfloor t/\bar{F}(u_n) \rfloor$. Then

$$\begin{aligned} P\{\bar{F}(u_n)T(u_n) > t\} &= P\{T(u_n) > t_n\} \\ &= P(M_{1,1+t_n} \leq u_n \mid X_1 > u_n). \end{aligned}$$

For n large enough so that $l_n < p_n$,

$$\begin{aligned} P(M_{p_n,p_n+l_n} > u_n \mid X_1 > u_n) &\leq P(M_{p_n,p_n+l_n} > u_n) + \alpha_n \\ &\leq l_n\bar{F}(u_n) + \alpha_n \to 0, \end{aligned}$$

so that

$$\begin{aligned} &P(M_{1,1+t_n} \leq u_n \mid X_1 > u_n) \\ &\quad = P(M_{1,p_n} \leq u_n,\, M_{p_n+l_n,t_n} \leq u_n \mid X_1 > u_n) + o(1). \end{aligned}$$

But $t_n \sim tr_n/\tau = O(r_n)$, so that by assumption

$$\begin{aligned} &P(M_{1,p_n} \leq u_n,\, M_{p_n+l_n,t_n} \leq u_n \mid X_1 > u_n) \\ &\quad = P(M_{p_n+l_n,t_n} \leq u_n \mid M_{1,p_n} \leq u_n,\, X_1 > u_n)P(M_{1,p_n} \leq u_n \mid X_1 > u_n) \\ &\quad = \{P(M_{p_n+l_n,t_n} \leq u_n) + o(1)\}\{\theta + o(1)\}. \end{aligned}$$

Now also $s_n = k_n - (p_n + q_n) \sim t\tau^{-1} r_n = O(r_n)$. Apply Corollary 2.3 of O'Brien (1987) again to deduce that

$$\begin{aligned} P(M_{p_n+l_n,t_n} \leq u_n) &= P(M_{s_n} \leq u_n) \\ &= F(u_n)^{s_n P(M_{1,p_n} \leq u_n | X_1 > u_n)} + o(1) \to e^{-\theta t}. \end{aligned}$$

$\square$

4.B Proof of Theorem 4.2

Lemma 4.1. *Let $\{X_i\}_{i\geq 1}$ be an m-dependent, strictly stationary sequence and let the thresholds u_n, $n \geq 1$, be such that*

$$n\bar{F}(u_n) \to \tau \quad \textit{and} \quad P(M_n \leq u_n) \to \exp(-\theta\tau)$$

for some $\tau \in (0, \infty)$ and $\theta \in (0, 1]$. Define also

$$\pi_n(1) = P\{N_{r_n}(u_n) = 1 \mid N_{r_n}(u_n) > 0\},$$

where $r_n = o(n)$ and $r_n \to \infty$. Then, there exists $l_n = o(r_n)$ such that $\pi_n^*(1) = \pi_n(1) + o(1)$, where

$$\pi_n^*(1) = P(M_{l_n - 1} \leq u_n \mid X_{l_n} > u_n, M_{l_n, 2l_n - 1} \leq u_n).$$

Proof. By Theorem 2.1 of O'Brien (1987) there exists $l_n = o(r_n)$ such that $l_n \to \infty$ and

$$P(M_{1,l_n} \leq u_n \mid X_1 > u_n) \to \theta.$$

Note that $P\{N_{r_n}(u_n) > 0\} = 1 - P(M_{r_n} \leq u_n) \sim \theta\tau/k_n$, where $k_n = \lfloor n/r_n \rfloor$. Therefore,

$$\pi_n(1) = \frac{k_n}{\theta\tau} P\{N_{r_n}(u_n) = 1\}\{1 + o(1)\}.$$

Now,

$$P\{N_{r_n}(u_n) = 1\} = \sum_{i=1}^{r_n} P(X_i > u_n, X_j \leq u_n \text{ for } j = 1, \ldots, r_n, j \neq i)$$

and

$$\sum_{i=1}^{l_n} P(X_i > u_n, X_j \leq u_n \text{ for } j = 1, \ldots, r_n, j \neq i) \leq l_n \bar{F}(u_n) \sim \tau l_n/n$$

so that

$$\pi_n(1) = \frac{k_n}{\theta\tau} \sum_{i=l_n+1}^{r_n-l_n} P(X_i > u_n, X_j \leq u_n \text{ for } j = 1, \ldots, r_n, j \neq i)\{1 + o(1)\} + o(1).$$

For $1 + l_n \leq i \leq r_n - l_n$ and n large enough so that $l_n \geq m$,

$$\begin{aligned}
0 \;&\leq\; P(M_{i-l_n,i-1} \leq u_n, X_i > u_n, M_{i,i+l_n-1} \leq u_n) \\
&\qquad - P(M_{i-1} \leq u_n, X_i > u_n, M_{i,r_n} \leq u_n) \\
&= P(M_{i-l_n} > u_n, M_{i-l_n,i-1} \leq u_n, X_i > u_n, M_{i,i+l_n-1} \leq u_n, M_{i+l_n-1,r_n} > u_n) \\
&\leq P(M_{i-l_n} > u_n, X_i > u_n) \\
&\leq \bar{F}(u_n)^2 = o\left(n^{-1}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\pi_n(1) &= \frac{k_n}{\theta\tau} \sum_{i=l_n+1}^{r_n-l_n} P(M_{i-l_n,i-1} \leq u_n, X_i > u_n, M_{i,i+l_n-1} \leq u_n)\{1+o(1)\} + o(1) \\
&= \frac{k_n}{\theta\tau}(r_n - 2l_n) P(M_{l_n-1} \leq u_n, X_{l_n} > u_n, M_{l_n,2l_n-1} \leq u_n)\{1+o(1)\} + o(1) \\
&= \frac{k_n}{\theta\tau}(r_n - 2l_n)\pi_n^*(1) P(M_{1,l_n-1} \leq u_n \mid X_1 > u_n)\bar{F}(u_n)\{1+o(1)\} + o(1) \\
&= \frac{k_n r_n}{\theta\tau}\pi_n^*(1)\theta\bar{F}(u_n)\{1+o(1)\} + o(1) \\
&= \pi_n^*(1) + o(1).
\end{aligned}$$

□

Remark The definition of $\pi_n^*(1)$ is based on runs declustering rather than the usual blocks declustering. It can be read as the probability that X_{l_n} is the first exceedance in a cluster given that it is the last exceedance in a cluster, where clusters are defined by runs declustering with run length $l_n - 1$.

Turn now to the theorem itself. The conditions of Lemma 4.1 are satisfied. Let

$s_n = \lfloor s/\bar{F}(u_n) \rfloor \sim sn/\tau$, $t_n = \lfloor t/\bar{F}(u_n) \rfloor \sim tn/\tau$ and denote $\lceil wn \rceil$ by w_n. Then

$$
\begin{aligned}
& P\{\bar{F}(u_n)T_1(u_n) > s, \bar{F}(u_n)T_2(u_n) > t\} \\
&= \int_{s_n}^{\infty} P(M_{1,\lceil w \rceil} \le u_n, X_{\lceil w \rceil + 1} > u_n, M_{\lceil w \rceil + 1, \lceil w \rceil + 1 + t_n} \le u_n \mid X_1 > u_n)\, \mathrm{d}w \\
&= \int_{s_n/n}^{\infty} nP(M_{1,w_n} \le u_n, X_{w_n+1} > u_n, M_{w_n+1, w_n+1+t_n} \le u_n \mid X_1 > u_n)\, \mathrm{d}w.
\end{aligned}
$$

By arguments similar to those in the proof of Lemma 4.1 and for sufficiently large n, the integrand is

$$
\begin{aligned}
& n\bar{F}(u_n)P(M_{1,w_n} \le u_n, M_{w_n+1,w_n+1+t_n} \le u_n \mid X_1 > u_n, X_{w_n+1} > u_n) \\
&= n\bar{F}(u_n)P(M_{1,l_n} \le u_n, M_{l_n+m, w_n-l_n-m} \le u_n, M_{w_n-l_n, w_n} \le u_n, \\
&\qquad M_{w_n+1, w_n+l_n} \le u_n, M_{w_n+l_n+m, w_n+1+t_n} \le u_n \mid X_1 > u_n, X_{w_n+1} > u_n) + o(1) \\
&= n\bar{F}(u_n)P(M_{1,l_n} \le u_n \mid X_1 > u_n)P(M_{w_n-2l_n-2m} \le u_n)P(M_{t_n-l_n-m+1} \le u_n) \\
&\qquad \times P(M_{w_n-l_n, w_n} \le u_n, M_{w_n+1, w_n+l_n} \le u_n \mid X_{w_n+1} > u_n) + o(1) \\
&= \tau\theta^2\pi(1)e^{-\theta(t+w\tau)} + o(1).
\end{aligned}
$$

Finally, the integrand is dominated by $C\gamma^w$ when $n \ge n_0$ for some $C < \infty$, $\gamma \in (0,1)$ and positive integer n_0 since

$$
\begin{aligned}
& nP(M_{1,w_n} \le u_n, X_{w_n+1} > u_n, M_{w_n+1, w_n+1+t_n} \le u_n \mid X_1 > u_n) \\
&\le nP(M_{m, w_n+1-m} \le u_n, X_{w_n+1} > u_n \mid X_1 > u_n) \\
&\le n\bar{F}(u_n)P(M_{w_n-2m} \le u_n),
\end{aligned}
$$

$n\bar{F}(u_n) \to \tau \in (0, \infty)$ and $P(M_{w_n - 2m} \leq u_n) \to \exp(-\theta\tau w)$. The result follows by dominated convergence. $\square$

4.C Proof of Theorem 4.3

A proof of Theorem 4.3 follows the same lines as that of Theorem 4.2; the reasoning is outlined here. There are j exceedances between the two inter-exceedance times under consideration so that the joint tail function can be written as

$$\sum_{i=1}^{j} P\{T_1(u_n) \geq s_n, i \text{ intervening clusters}, T_{1+j}(u_n) \geq t_n\},$$

where $s_n = \lceil s/\bar{F}(u_n) \rceil$ and $t_n = \lceil t/\bar{F}(u_n) \rceil$. The summands factorise approximately into terms that behave similarly to

$$\begin{gathered} P(M_{1,l_n} \leq u_n \mid X_1 > u_n) P(M_{s_n} \leq u_n) P(M_{t_n} \leq u_n) \\ \times P(j \text{ consecutive exceedances form } i \text{ clusters}). \end{gathered}$$

This has limit $\theta e^{-\theta(s+t)} p_{i,j}$, where

$$\begin{aligned} p_{1,j} &= \pi(j) \\ p_{i+1,j} &= \sum_{k=1}^{j-i} \pi(k) p_{i,j-k} \quad \text{for } i = 1, \ldots, j-1 \end{aligned}$$

by independence of clusters. Finally, define $e_j = \sum_{i=1}^{j} p_{i,j}$ for $j \geq 1$ and observe that

$$\begin{aligned} e_j &= \sum_{i=1}^{j-1} p_{i+1,j} + \pi(j) = \sum_{i=1}^{j-1} \sum_{k=1}^{j-i} \pi(k) p_{i,j-k} + \pi(j) \\ &= \sum_{k=1}^{j-1} \pi(k) \sum_{i=1}^{j-k} p_{i,j-k} + \pi(j) = \sum_{k=1}^{j-1} \pi(k) e_{j-k} + \pi(j). \end{aligned}$$

□

4.D Proof of Theorem 4.4

Denote $\bar{F}(u_n)$ by p_n. To avoid ambiguities, let $\tilde{\theta}_n(u_n)$ be an arbitrary number, say 1, if $N \in \{0, 1\}$.

Lemma 4.2. *There exist $C > 0$, $0 < \gamma < 1$ and positive integer n_0 such that*

$$\begin{aligned} &(i) & P(M_k \leq u_n) &\leq C\gamma^{k/r_n} \\ &(ii) & P(M_{1,k} \leq u_n \mid X_1 > u_n) &\leq C\gamma^{k/r_n} \end{aligned}$$

for all integer $k \geq 1$ and $n \geq n_0$.

Proof. (i) Fix positive integer k and write $k = qm + j$ where $q = 0, 1, 2, \ldots$ and $j = 0, 1, \ldots, m-1$. Divide the sample of size k into q blocks of size m and one block of size j. By m-dependence,

$$P(M_k \leq u_n) \leq (1 - p_n)^q \leq (1 - p_n)^{-j/m} \left\{ (1 - p_n)^{r_n/m} \right\}^{k/r_n}.$$

Since $1-p_n \to 1$ and $(1-p_n)^{r_n} \to e^{-\tau} < 1$, there exists $0 < \gamma < 1$ and a positive integer n_0 such that for all integer $n \geq n_0$,

$$\begin{aligned} (1-p_n)^{-j/m} &\leq (1-p_n)^{-1} \leq 2 \quad \text{for } j = 0, 1, \ldots, m-1, \\ (1-p_n)^{r_n/m} &\leq \gamma. \end{aligned}$$

Hence for all integer $n \geq n_0$, $P(M_k \leq u_n) \leq 2\gamma^{k/r_n}$.

(ii) For integer $k \geq m+1$, m-dependence and stationarity give

$$P(M_{1,k} \leq u_n \mid X_1 > u_n) \leq P(M_{m,k} \leq u_n) = P(M_{k-m} \leq u_n).$$

Thanks to (i), the latter is bounded by $2\gamma^{-m}\gamma^{k/r_n}$. For $k \leq m$, $2\gamma^{-m}\gamma^{k/r_n} \geq 1$. Put $C = 2\gamma^{-m}$. □

Lemma 4.3. *If the positive integers s_n satisfy $s_n \sim \lambda r_n$ for some $0 < \lambda < \infty$, then*

$$P(M_{s_n} \leq u_n) \to e^{-\theta\lambda\tau} \quad \textit{and} \quad P(M_{1,s_n} \leq u_n \mid X_1 > u_n) \to \theta e^{-\theta\lambda\tau}.$$

Proof. Similar to the proof of Theorem 4.1. □

By convention, set $\sum_{i=1}^{N-1}(T_i - 1) = 0$ if $N \in \{0, 1\}$.

Lemma 4.4. $N \overset{p}{\sim} np_n$ *and* $\sum_{i=1}^{N-1}(T_i - 1) \overset{p}{\sim} n$.

Proof. Clearly $E(N) = np_n$. By stationarity and m-dependence, also $\operatorname{var}(N) \leq$

$2mnp_n$. Together, $E[\{(np_n)^{-1}N - 1\}^2] \leq 2m(np_n)^{-1} \to 0$, which implies $N \overset{p}{\sim} np_n$. In particular, $N \xrightarrow{p} \infty$, so that $P(N \geq 2) \to 1$.

Next, $\sum_{i=1}^{N-1}(T_i - 1) = S_N - S_1 - (N-1) = n - (n - S_N) - S_1 - (N-1)$ if $N \geq 2$. For $c > 0$, Lemma 4.2 gives

$$P(n - S_N \geq cr_n \text{ and } N \geq 2) \leq P(M_{\lfloor cr_n \rfloor} \leq u_n) \leq C\gamma^{\lfloor cr_n \rfloor / r_n}$$

so that $n - S_N = O_p(r_n) = o_p(n)$. Similarly, $S_1 = O_p(r_n) = o_p(n)$. From the first part, it is already known that $N = o_p(n)$. □

It remains to deal with the term $\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)$, which is also set equal to zero if $N \in \{0, 1\}$. Define

$$Y_n = \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} (j-1) I(X_i > u_n \geq M_{i,j}).$$

Observe that if $N \geq 2$, then

$$\begin{aligned} Y_n &= \sum_{i=1}^{N-1} \sum_{j=1}^{T_i - 1} (j-1) + \sum_{j=1}^{n-S_N} (j-1) \\ &= \frac{1}{2} \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2) + \frac{1}{2}(n - S_N)(n - S_N - 1). \end{aligned}$$

In particular,

$$Y_n = \frac{1}{2} \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2) + O_p(r_n^2).$$

Hence, to prove the Theorem, it is sufficient to show that $n^{-1} p_n Y_n \xrightarrow{p} \theta^{-1}$. To this end, it is sufficient to show that

$$n^{-1}p_nE(Y_n) \to \theta^{-1} \quad \text{and} \quad n^{-2}p_n^2E(Y_n^2) \to \theta^{-2}.$$

Lemma 4.5. $n^{-1}p_nE(Y_n) \to \theta^{-1}$

Proof. By stationarity of the process $\{X_n\}_{n\geq 1}$,

$$\begin{aligned} E(Y_n) &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}(j-1)P(X_i > u_n \geq M_{i,i+j}) \\ &= p_n\sum_{j=1}^{n}(j-1)(n-j)P(M_{1,1+j} \leq u_n \mid X_1 > u_n). \end{aligned}$$

Rewrite the sum as an integral to find

$$\begin{aligned} E(Y_n) &= p_n\int_0^n(\lceil s\rceil - 1)(n-\lceil s\rceil)P(M_{1,1+\lceil s\rceil} \leq u_n \mid X_1 > u_n)\,\mathrm{d}s \\ &= p_nr_n\int_0^{n/r_n}(\lceil r_ns\rceil - 1)(n-\lceil r_ns\rceil)P(M_{1,1+\lceil r_ns\rceil} \leq u_n \mid X_1 > u_n)\,\mathrm{d}s. \end{aligned}$$

Hence

$$\begin{aligned} &n^{-1}p_nE(Y_n) \\ &\quad = p_nr_n\int_0^{n/r_n}p_n(\lceil r_ns\rceil - 1)(1-\lceil r_ns\rceil/n)P(M_{1,1+\lceil r_ns\rceil} \leq u_n \mid X_1 > u_n)\,\mathrm{d}s. \end{aligned}$$

By Lemmas 4.2 and 4.3, the dominated convergence theorem yields

$$n^{-1}p_nE(Y_n) \to \tau^2\int_0^\infty s\theta e^{-\theta\tau s}\,\mathrm{d}s = \theta^{-1},$$

as required. □

Lemma 4.6. $n^{-2}p_n^2 E(Y_n^2) \to \theta^{-2}$

Proof. For integers $n \geq 1$, $1 \leq i \leq n-1$ and $1 \leq j \leq n-i$, abbreviate

$$I_{i,j} = I_{i,j}^{(n)} = I(X_i > u_n \geq M_{i,i+j}).$$

Then

$$E(Y_n^2) = \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=1}^{n-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{i,j}I_{k,l}).$$

The expectations can be computed as follows:

$$\begin{aligned}
E(I_{i,j}) &= E(I_{1,j}), \\
E(I_{i,j}I_{i,l}) &= E(I_{1,j\vee l}), \\
E(I_{i,j}I_{k,l}) &= 0, \quad \text{for } i < k \leq i+j, \\
E(I_{i,j}I_{k,l}) &= E(I_{1,j})E(I_{1,l}), \quad \text{for } k \geq i+j+m.
\end{aligned}$$

Hence $E(Y_n^2) = A_n + B_n + C_n$ where

$$\begin{aligned}
A_n &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{l=1}^{n-i}(j-1)(l-1)E(I_{1,j\vee l}), \\
B_n &= 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=i+j+1}^{i+j+m-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{1,j}I_{1,l}), \\
C_n &= 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=i+j+m}^{n-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{1,j})E(I_{1,l}).
\end{aligned}$$

For the terms A_n and B_n,

$$\begin{aligned} A_n &\leq n \sum_{j=1}^{n-1} \sum_{l=1}^{n-1} (j-1)(l-1) E(I_{1,j \vee l}), \\ B_n &\leq 2mn \sum_{j=1}^{n-1} \sum_{l=1}^{n-1} (j-1)(l-1) E(I_{1,j \vee l}). \end{aligned}$$

Changing sums gives

$$\sum_{j=1}^{n-1} \sum_{l=1}^{n-1} (j-1)(l-1) E(I_{1,j \vee l}) \leq 2 \sum_{j=1}^{n-1} (j-1) \sum_{l=j}^{n-1} (l-1) E(I_{1,l}) = \sum_{l=1}^{n-1} (l-1)^2 l E(I_{1,l}).$$

By Lemma 4.2, $E(I_{1,l}) = P(X_1 > u_n \geq M_{1,1+l}) \leq p_n C \gamma^{l/r_n}$. Hence both $n^{-1} A_n$ and $n^{-1} B_n$ are bounded by some positive constant times $p_n \sum_{l=1}^{\infty} (l-1)^2 l \gamma^{l/r_n}$. As $n \to \infty$, this expression is $O\left\{p_n (1-\gamma^{1/r_n})^{-4}\right\} = O(p_n^{-3})$. Therefore, both A_n and B_n are $O(np_n^{-3}) = o(n^2 p_n^{-2})$.

For the term C_n, change the order of summation to obtain

$$\begin{aligned} C_n &= 2 \sum_{j=1}^{n} \sum_{l=1}^{n} (j-1)(l-1) E(I_{1,j}) E(I_{1,l}) \sum_{i=1}^{n-j} \sum_{k=i+j+m}^{n-l} 1 \\ &= \sum_{j=1}^{n} \sum_{l=1}^{n} (j-1)(l-1)(n-l-j-m+1)(n-l-j-m)_+ E(I_{1,j}) E(I_{1,l}). \end{aligned}$$

Write the sum as an integral:

$$\begin{aligned}
C_n &= \int_0^n \int_0^n (\lceil s \rceil - 1)(\lceil t \rceil - 1)(n - \lceil s \rceil - \lceil t \rceil - m + 1) \\
&\qquad \times (n - \lceil s \rceil - \lceil t \rceil - m)_+ E(I_{1,\lceil s \rceil}) E(I_{1,\lceil t \rceil}) \,\mathrm{d}s\,\mathrm{d}t \\
&= r_n^2 \int_0^{n/r_n} \int_0^{n/r_n} (\lceil r_n s \rceil - 1)(\lceil r_n t \rceil - 1)(n - \lceil r_n s \rceil - \lceil r_n t \rceil - m + 1) \\
&\qquad \times (n - \lceil r_n s \rceil - \lceil r_n t \rceil - m)_+ E(I_{1,\lceil r_n s \rceil}) E(I_{1,\lceil r_n t \rceil}) \,\mathrm{d}s\,\mathrm{d}t.
\end{aligned}$$

Since $E(I_{1,\lceil r_n s \rceil}) = p_n P(M_{1,1+\lceil r_n s \rceil} \leq u_n \mid X_1 > u_n)$, Lemmas 4.2 and 4.3 apply, giving

$$n^{-2} p_n^2 C_n \to \tau^4 \int_0^\infty \int_0^\infty st\theta e^{-\theta\tau s} \theta e^{-\theta\tau t} \,\mathrm{d}s\,\mathrm{d}t = \theta^{-2}$$

by dominated convergence. □