The Annals of Statistics
1993, Vol. 21, No. 4, 2043–2071

EXTREMAL INDEX ESTIMATION FOR A WEAKLY DEPENDENT STATIONARY SEQUENCE[1]

BY TAILEN HSING

Texas A & M University

Under stationarity and weak dependence, the statistical significance and the estimation of the extremal index are considered. It is shown that the distribution of the sample maximum can be uniformly approximated given the extremal index and the marginal distribution as the sample size increases. An adaptive procedure is proposed for estimating the extremal index. The procedure is shown to be asymptotically optimal in a class of estimators.

1. Introduction. The problem of estimating the tail distribution of the maximum of a large number of observations has its roots in many important practical situations and has long been studied by statisticians. See Gumbel (1958), Galambos (1978) and Leadbetter, Lindgren and Rootzén (1983). In this paper we consider the problem in the context of stationarity, for which the notion of an extremal index provides a powerful modeling tool. Roughly speaking, the extremal index together with the marginal distribution paints a very vivid picture of the distribution of the maximum. Our goals are to further motivate the use of the extremal index and consider how it can be estimated.

Throughout this paper, we assume that $\{\xi_j\}$ is a strictly stationary sequence of random variables with a continuous marginal distribution F. Write $\overline{F} = 1 - F$ and $x_0 = \sup\{x: F(x) < 1\}$. Also write

$$M_{i,j} = \max_{i \le k \le j} \xi_k \quad \text{and} \quad M_n = M_{1,n}.$$

We say that the extremal index of $\{\xi_j\}$ is equal to θ if

$$\lim_{n \to \infty} P\left[n\overline{F}(M_n) \ge x\right] = e^{-\theta x}, \qquad x > 0. \tag{1.1}$$

This is essentially equivalent to saying that the distribution of $n\overline{F}(M_n)$ converges weakly to an exponential distribution with mean θ^{-1}. The extremal index θ exists for most weakly dependent stationary sequences with continuous marginal distributions. The notion of the extremal index was implicitly mentioned in Theorem 2 of Loynes (1965). However, the formal definition and the terminology first appeared in Leadbetter (1983).

By Boole's inequality, regardless of the dependence structure,

$$\liminf_{n \to \infty} P\left[n\overline{F}(M_n) \ge x\right] \ge 1 - x, \qquad x > 0.$$

Comparing this with (1.1) for small x, it is readily concluded that $\theta \in [0, 1]$.

Received July 1990; revised December 1991.
[1]Research supported by ONR Grant N000014-92-J-1007 and NSF Grant 9107507.
AMS 1991 *subject classification*. 62F12.
Key words and phrases. Adaptive estimation, extremal index, stationary sequence.

Examples show that θ can be of any value in $[0, 1]$. The extremal index measures the strength of the dependence of $\{\xi_j\}$. It is plausible to say that $\theta = 0$ corresponds to a long memory sequence, $0 < \theta < 1$ a short memory sequence, and $\theta = 1$ a no memory sequence (insofar as the tail dependence structure is concerned). If $\theta > 0$, then the dependence is weak so that M_n can be normalized with the normalization which is appropriate for the maximum of n i.i.d. random variables from F, whereas if $\theta = 0$, then the dependence is so strong that a different normalization is called for. Thus the nature of long memory sequences is very different from that of short or no memory sequences. In this paper we focus on the case $\theta > 0$, which we assume henceforth.

It is interesting to note in passing a parallel situation. If, instead, one is interested in the asymptotic distribution of $\sum_{j=1}^{n} \xi_j$, the role of θ will be in a sense played by the value of the spectral density at 0 [cf. Grenander and Szegö (1958)].

Some theoretical results are available for the purpose of understanding and computing the extremal index. Here and hereafter, $\bar{F}^{-1}$ denotes the right continuous inverse of $\bar{F}$. Leadbetter (1983) showed that under a mixing condition called D, for any sequence r_n which tends to ∞ but is $o(n)$ (and constrained in some way by D),

$$E\left(\sum_{j=1}^{r_n} I\left(\xi_j > \bar{F}^{-1}(x/n)\right) \middle| M_{r_n} > \bar{F}^{-1}(x/n) \right) = \frac{r_n P\left[\xi_1 > \bar{F}^{-1}(x/n)\right]}{P\left[M_{r_n} > \bar{F}^{-1}(x/n)\right]} \to \frac{1}{\theta} \quad \text{as } n \to \infty, \tag{1.2}$$

from which he obtained the interpretation that θ^{-1} is the mean number of exceedances of a high level in a "cluster" of large observations. See also Hsing, Hüsler and Leadbetter (1988). O'Brien (1987) showed that if $\{\xi_j\}$ satisfies a mixing condition called AIM, for some $\{r_n\}$ satisfying constraints similar to those mentioned previously,

$$P\left[M_{2, r_n} \le \bar{F}^{-1}(x/n) | \xi_1 > \bar{F}^{-1}(x/n)\right] \to \theta \quad \text{as } n \to \infty, \tag{1.3}$$

which, when given the structure of $\{\xi_j\}$, is very useful for computing the extremal index. See also Hsing (1989), Chernick, Hsing and McCormick (1991) and Rootzén (1988) for related results. Examples illustrating the extremal index can also be found in Chernick (1981), Davis and Resnick (1985), Denzel and O'Brien (1975), de Haan, Resnick, Rootzén and de Vries (1989) and Leadbetter, Lindgren and Rootzén (1983), to name a few. Recently, a number of papers discussed the estimation of θ. These include Hsing (1990, 1991a), Joe (1991), Leadbetter, Weissman, de Haan and Rootzén (1989), Nandagopalan (1990), Smith (1989) and Smith and Weissman (1993).

An important statistical issue in this context is to understand what information the extremal index provides. The definition (1.1) shows that, in general terms, the distribution of M_n can be approximated by $F^{n\theta}(x)$. In practice where the distribution of the maximum is of interest, the tail portion of that distribution is almost always the focus. For example, when a study is done to determine how high a sea dike should be, the criterion might be that the dike should be high enough so that the probability of floods in 1000 years is less than a small probability. Equation (1.1) itself does not provide information of how good the approximation is toward the tail. We show in Section 3 that under a very general condition,

$$\lim_{n\to\infty} \sup_x \left| \frac{P[M_n > x]}{1 - F^{n\theta}(x)} - 1 \right| \to 0.$$

This shows that to estimate the tail distribution of M_n, it suffices to estimate θ and the marginal tail distribution. Much effort has been put into the problem of estimating the tail marginal distribution in the i.i.d. case. See Gumbel (1958), Pickands (1975), Smith (1987), Dekkers and de Haan (1989) and Dekkers, Einmahl and de Haan (1989), among others. Hsing (1991b) partially confirmed that the same problem under a weakly dependent stationary setting can be treated in similar ways.

We consider the problem of estimating the extremal index in Section 2. There we suggest a model motivated by (1.3) and consider the problem in that setting. While the model puts restriction on the class of processes from which the observations are generated, it is nevertheless quite rich and is suitable for most situations. The estimator that we propose to use contains a tuning constant which partially determines the quality of the estimate. We show that it is possible to choose from data the optimal tuning constant, relative to the mean squared error criterion. To keep the presentation concise, parts of the proofs are given in simplified settings. Extensions of those proofs to more general settings are intuitively obvious but technically complicated.

In Section 4 we consider the m-dependent case in detail. There is it possible to consider specifically how $P[M_r > x]$, $P[M_{2,r} \le x | \xi_1 > x]$ and $\bar{F}(x)$ are related. As a result, we obtain refined versions of (1.2) and (1.3). These results can be used to suggest models which contain θ and hence different ways of estimating θ. In particular, we will mention an approach motivated by (1.2).

To consider the distributions of other extreme order statistics, additional parameters will have to be included. This is best seen from the viewpoint of a functional limit theorem. See Hsing (1993) for details.

2. Statistical estimation. Consider the problem of estimating the extremal index in the following setting. As mentioned in Section 1, $\{\xi_j\}$ is a stationary sequence of random variables with a continuous marginal distribution function F and a nonzero extremal index θ. Assume also that there exists a finite positive integer $r \ge 2$ such that

$$P[M_{2,r} \le x | \xi_1 > x] = \theta + R(\bar{F}(x)), \tag{2.1}$$

where $R(p) \to 0$ as $p \to 0$. We assume that r can be chosen in some way, but R is unknown. The role of r and how it is determined will be explained in more detail below. While, by (1.3) [Theorem 2.1 of O'Brien (1987)], (2.1) obviously holds if $\{\xi_j\}$ is m-dependent where $m \le r$ (cf. Theorem 4.2), interestingly enough such a representation is appropriate for many weakly dependent sequences whose ranges of dependence are infinite. This property is studied by Chernick, Hsing and McCormick (1991) [cf. Leadbetter and Nandagopolan (1989)]. This class of models is very rich so that (2.1) is a rather reasonable model for most situations. Compare also with (3.3). We first give some examples of $\{\xi_j\}$ for which (2.1) holds.

Example A. Let $\{Z_j\}$ be i.i.d. with distribution G, where

$$\overline{G}(z) = 1 - G(z) = \eta z^{-\alpha}(1 + \gamma z^{-\beta} + o(z^{-\beta})) \quad \text{as } z \to \infty,$$

where $\alpha, \eta, \beta > 0$ and $\gamma \in \mathbb{R}$. Define

$$\xi_j = \max(Z_j, \rho Z_{j+1}),$$

where $\rho \ge 1$ is a constant. Since $\{\xi_j\}$ is 2-dependent, (2.1) holds with $r = 2$ where straightforward computations show that

$$P[\xi_2 \le x | \xi_1 > x] = \frac{\rho^\alpha}{1 + \rho^\alpha} - \frac{\eta\rho^\alpha + \eta\rho^{2\alpha}(1 + \rho^\alpha)}{(1 + \rho^\alpha)^2} x^{-\alpha} + \frac{\gamma\rho^\alpha(\rho^\beta - 1)}{(1 + \rho^\alpha)^2} x^{-\beta} + o(x^{-\alpha}) + o(x^{-\beta}).$$

The form of the remainder R depends on whether $\alpha > \beta$ or $\alpha \le \beta$.

Example B. Consider the following sequence defined in Smith and Weissman (1993). Let $\{I_i\}$ be i.i.d. Bernoulli with $P[I_1 = 1] = \rho \in (0,1)$ and $\{Z_i\}$ be i.i.d. continuous also independent of $\{I_i\}$. Define

$$\xi_1 = Z_1,$$

and ξ_j, $j \ge 2$, recursively by

$$\xi_j = \xi_{j-1} I_j + Z_j(I_j - 1).$$

It is easy to show that $\{\xi_j\}$ has extremal index $\theta = 1 - \rho$, and (2.1) holds with $r = 2$ and $R(p) \sim (1 - \rho)p$.

Example C. Consider the AR(1) Cauchy sequence $\xi_j = \rho\xi_{j-1} + Z_j$, where $\rho \in (0,1)$ and $\{Z_i\}$ is i.i.d. standard Cauchy. Assume also that ξ_0 has the same distribution as $1 - \rho$ times the standard Cauchy random variable so that $\{\xi_j\}$ is strictly stationary with this marginal distribution. By Chernick, Hsing and McCormick (1991), the extremal index is $1 - \rho$. It is shown in the Appendix that

$$P[\xi_2 \le x | \xi_1 > x] = 1 - \rho + c\overline{F}(x)(1 + o(1)), \tag{2.2}$$

where

$$c = -2(1 - \rho) + (1 \quad \rho)^3 \int_{-\infty}^{-1} \frac{dy}{(1 - \rho y)y^2} + 2\rho^2(1 - \rho)\int_0^1 \frac{dy}{1 - \rho^2 y^2}.$$

Define

$$I_{i,1}(x) = I(\xi_i > x \geq M_{i+1,i+r-1}) \quad \text{and} \quad I_{i,2}(x) = I(\xi_i > x).$$

The estimator of θ that we propose to use under (2.1) is

$$\hat{\theta}_n(x) = \frac{\sum_{i=1}^n I_{i,1}(x)}{\sum_{i=1}^n I_{i,2}(x)},$$

where x is a tuning constant which determines the quality of the estimate. Generally speaking, in order to estimate θ well, x should be large enough so that the relevant tail characteristics of $\{\xi_j\}$ are captured by the quantities included in $\hat{\theta}_n(x)$. In reality, the choice of x typically has to be made from data. Thus we will consider the properties of $\hat{\theta}_n(x)$ for $x = \overline{F}_n^{-1}(p) :=$ $[np] + 1$th largest value of the sample $\xi_1, \dots, \xi_n$. Note that although the procedure $\hat{\theta}_n$ requires observing $\xi_1, \dots, \xi_{n+r-1}$, without any loss of generality we will regard the sample size as n for convenience.

Under a mixing condition such as Leadbetter's condition D [Leadbetter (1974)] or any stronger condition, it is easy to see [cf. Chernick, Hsing and McCormick (1991), Theorem 1.1] that (2.1) implies

$$P\left[M_{2,r} \leq x, \xi_j > x | \xi_1 > x\right] \to 0 \quad \text{as } x \to x_0, \text{ all } j > r. \tag{2.3}$$

Then one simply observes that if (2.1) holds for some $r = r_0$, then it holds for all $r > r_0$, where the remainder R depends on r. Thus one has some flexibility in choosing r. Note that the effectiveness of the estimator $\hat{\theta}_n$ is partly affected by the choice of r. For example, using the mean squared error criterion, typically the best choice of r tends to be the smallest integer r for which (2.1) holds, and when a larger integer r is used, the limiting mean squared error will generally have the same order of magnitude but possibly a larger coefficient. For situations where the estimation of θ is called for, the theoretical structure of the sequence is normally unknown. Thus we recommend selecting r conservatively; namely r should be as small as possible, but should reasonably cover the range of dependence. Using model selection techniques, this can be achieved in a number of ways. Smith and Weissman (1993) consider some issues in a related context.

Our first result shows that $\hat{\theta}_n(x)$ is weakly consistent for a wide range of x values.

THEOREM 2.1. *Let* $0 < \lambda_1 < \lambda_2 < \infty$ *be constants, and let* p_n *be such that* $p_n \to 0$ *and* $np_n \to \infty$. *Assume that* (2.1) *holds, and for* $x_n = \overline{F}^{-1}(\lambda p_n)$, $\lambda \in [\lambda_1, \lambda_2]$,

$$\limsup_{n\to\infty} \frac{\left|\sum_{i=1}^n \operatorname{Cov}(I_{1,s}(x_n), I_{i,s}(x_n))\right|}{\overline{F}(x_n)} < \infty, \qquad s = 1, 2. \tag{2.4}$$

Then

$$\text{(2.5)} \qquad \sup_{\bar F^{-1}(\lambda_2 p_n) \le x \le \bar F^{-1}(\lambda_1 p_n)} \left|\hat\theta_n(x) - \theta\right| \to_p 0$$

and

$$\text{(2.6)} \qquad \sup_{\bar F_n^{-1}(\lambda'_2 p_n) \le x \le \bar F_n^{-1}(\lambda'_1 p_n)} \left|\hat\theta_n(x) - \theta\right| \to_p 0,$$

for all λ'_1, λ'_2 *satisfying* $\lambda_1 < \lambda'_1 \le \lambda'_2 < \lambda_2$.

PROOF. For a nonrandom x, write

$$\begin{aligned}\hat\theta_n(x) - \theta = &\left(\left(n\bar F(x)\right)^{-1} \sum_{i=1}^n I_{i,2}(x)\right)^{-1} \\ &\times\left\{\left(n\bar F(x)\right)^{-1} \sum_{i=1}^n \left(I_{i,1}(x) - EI_{i,1}(x)\right)\right. \\ &\left. -\theta\left(n\bar F(x)\right)^{-1} \sum_{i=1}^n \left(I_{i,2}(x) - EI_{i,2}(x)\right) + \bar F(x)^{-1}\left(EI_{1,1}(x) - \theta EI_{1,2}(x)\right)\right\}.\end{aligned}$$

In order to show (2.5) it therefore suffices to show

$$\text{(2.7)} \qquad \sup_{\bar F^{-1}(\lambda_2 p_n) \le x \le \bar F^{-1}(\lambda_1 p_n)} \frac{\left|\sum_{i=1}^n (I_{i,1}(x) - EI_{i,1}(x))\right|}{n\bar F(x)} \to_p 0,$$

$$\text{(2.8)} \qquad \sup_{\bar F^{-1}(\lambda_2 p_n) \le x \le \bar F^{-1}(\lambda_1 p_n)} \frac{\left|\sum_{i=1}^n (I_{i,2}(x) - EI_{i,2}(x))\right|}{n\bar F(x)} \to_p 0,$$

and

$$\text{(2.9)} \qquad \sup_{\bar F^{-1}(\lambda_2 p_n) \le x \le \bar F^{-1}(\lambda_1 p_n)} \frac{\left|EI_{1,1}(x) - \theta EI_{1,2}(x)\right|}{\bar F(x)} \to 0.$$

Formula (2.9) follows readily from (2.1). We next prove (2.8). For any positive integer k, write

$$x_j^{(k)} = \bar F^{-1}\left(p_n(\lambda_1 + (\lambda_2 - \lambda_1)(k-j)/k)\right), \qquad 0 \le j \le k.$$

It suffices to show

$$\text{(2.10)} \qquad \sup_{0 \le j \le k} \frac{\left|\sum_{i=1}^n \left(I_{i,2}\left(x_j^{(k)}\right) - EI_{i,2}\left(x_j^{(k)}\right)\right)\right|}{np_n} \to_p 0, \qquad k \ge 1,$$

$$\text{(2.11)} \qquad \lim_{k\to\infty} \limsup_{n\to\infty} P\left[\sup_{1 \le j \le k} \sup_{x_{j-1}^{(k)} \le x \le x_j^{(k)}} \frac{\left|\sum_{i=1}^n \left(I_{i,2}\left(x_j^{(k)}\right) - I_{i,2}(x)\right)\right|}{np_n} > \varepsilon\right] = 0,$$

$$\varepsilon > 0,$$

$$\text{(2.12)} \qquad \lim_{k\to\infty} \limsup_{n\to\infty} \sup_{1 \le j \le k} \sup_{x_{j-1}^{(k)} \le x \le x_j^{(k)}} \frac{\left|\bar F\left(x_j^{(k)}\right) - \bar F(x)\right|}{p_n} = 0.$$

Formula (2.10) follows from (2.4) using Chebyshev's inequality and (2.12) follows from the definition of $x_j^{(k)}$. To show (2.11), fix $\varepsilon > 0$ and observe that, by (2.4),

$$\sup_{1 \le j \le k} \sup_{x_{j-1}^{(k)} \le x \le x_j^{(k)}} \frac{\left|\sum_{i=1}^n \left(I_{i,2}\left(x_j^{(k)}\right) - I_{i,2}(x)\right)\right|}{np_n}$$

$$= \sup_{1 \le j \le k} \frac{\sum_{i=1}^n \left(I_{i,2}\left(x_{j-1}^{(k)}\right) - I_{i,2}\left(x_j^{(k)}\right)\right)}{np_n} \to_p \frac{\lambda_2 - \lambda_1}{k},$$

which is less than ε for large k. This proves (2.11) and hence (2.8). We omit the proof of (2.7) which follows from essentially the same argument. Thus (2.5) is proved. To prove (2.6), note that, by (2.4),

$$P\left[\overline{F}_n^{-1}(\lambda'_2 p_n) \le \overline{F}^{-1}(\lambda_2 p_n)\right] = P\left[\sum_{i=1}^n I_{i,2}\left(\overline{F}^{-1}(\lambda_2 p_n)\right) \le \lambda'_2 n p_n\right] \to 0.$$

Similarly,

$$P\left[\overline{F}_n^{-1}(\lambda'_1 p_n) \ge \overline{F}^{-1}(\lambda_1 p_n)\right] \to 0.$$

Thus (2.6) follows from (2.5). □

We now address the issue of selecting the optimal threshold x, or equivalently, the optimal p in $\hat{\theta}_n(\overline{F}_n^{-1}(p))$, using the information given by a sample. The criterion that we will use is the mean squared error criterion. Much of the difficulty in selecting the optimal threshold arises from the fact that the unknown function R in (2.1) partly determines the mean squared error of $\hat{\theta}_n(\overline{F}_n^{-1}(p))$. Therefore, to decide the optimal p, it is crucial to have some information on R. If R can be estimated accurately enough from the data, then a $\hat{p}$ can be picked to minimize the asymptotic mean squared error of $\hat{\theta}_n(\overline{F}_n^{-1}(p))$. It is indeed very difficult to solve this problem in its entire generality, because we do not yet have a complete understanding of the nature of the function R. However, there is evidence to believe that the model

$$R(p) = L(p)p^{\beta},$$

where L is nonzero and slowly varying at 0 and $\beta > 0$, is satisfied by a large class of processes. See Feller (1971) for some general information on the notion of slow variation. In fact, for most of the processes for which we could specifically compute R, R has the very simple form

$$R(p) \sim \eta p^{\beta} \quad \text{as } p \to 0, \tag{2.13}$$

where $\eta \ne 0$ and $\beta > 0$. Three such examples are given by Examples A through C. See the paragraph following Theorem 4.2 for a discussion on the value of β. As a first step in solving this problem, let us assume henceforth in this section that (2.13) holds. To illustrate how to proceed in this situation, we make a technical simplification at this point. If we assume that $\{\xi_j\}$ has an infinite range of dependence but satisfies some mixing condition, the technical details involved in deciding the optimal p will be quite lengthy and indeed rather overwhelming so that the clarity of the presentation will be less than ideal. It seems feasible in this case to compromise details and clarity by presenting the proofs in the very simple m-dependence setting, whereby rigor is preserved and the reader will be able to see what is involved in making extensions. Thus, in the rest of this section, we assume that $\{\xi_j\}$ is m-dependent for some finite m, keeping in mind that the results hold under more general assumptions of dependence. To be specific, recall that $\{\xi_j\}$ is m-dependent if $\sigma\{\xi_j: j \le k\}$ is independent of $\sigma\{\xi_j: j \ge k+m\}$ for all k. Whether $m \le r$ or $m > r$, where r is the constant appearing in (2.1), is irrelevant.

Let p_n be such that $p_n \to 0$ and $np_n \to \infty$. First note that we can write

$$\begin{aligned}\hat{\theta}_n\big(\bar{F}_n^{-1}(p_n)\big) - \theta &= \frac{1}{[np_n]}\sum_{i=1}^{n}\Big(I_{i,1}\big(\bar{F}_n^{-1}(p_n)\big) - EI_{i,1}\big(\bar{F}_n^{-1}(p_n)\big)\Big)\\ &\quad+\left(\frac{1}{[np_n]}\sum_{i=1}^{n} EI_{i,1}\big(\bar{F}_n^{-1}(p_n)\big) - \theta\right) =: A_n + B_n.\end{aligned} \tag{2.14}$$

It is clear that the variance and bias of $\hat{\theta}_n(\bar{F}_n^{-1}(p_n))$ are contributed by A_n and B_n, respectively. We first analyze A_n.

LEMMA 2.2.

$$(np_n)^{1/2}\big(p_n^{-1}\bar{F}\big(\bar{F}_n^{-1}(p_n)\big) - 1\big) = O_P(1).$$

PROOF. It is easy to see that, for $x > 0$,

$$\begin{aligned}&P\Big[(np_n)^{1/2}\big(p_n^{-1}\bar{F}\big(\bar{F}_n^{-1}(p_n)\big) - 1\big) \ge x\Big]\\ &\quad= P\Big[\bar{F}_n^{-1}(p_n) \le \bar{F}^{-1}\Big(p_n\big(1 + x/\sqrt{np_n}\,\big)\Big)\Big]\\ &\quad= P\left[\sum_{j=1}^{n} I\Big(\xi_j > \bar{F}^{-1}\Big(p_n\big(1 + x/\sqrt{np_n}\,\big)\Big)\Big) \le np_n\right] = P[Y_n(x) \le -x],\end{aligned}$$

where

$$Y_n(x) = \frac{\sum_{j=1}^{n} I\Big(\xi_j > \bar{F}^{-1}\Big(p_n\big(1 + x/\sqrt{np_n}\,\big)\Big)\Big) - np_n\big(1 + x/\sqrt{np_n}\,\big)}{\sqrt{np_n}},$$

and similarly, for $x < 0$,

$$P\Big[(np_n)^{1/2}\big(p_n^{-1}\bar{F}\big(\bar{F}_n^{-1}(p_n)\big) - 1\big) \le x\Big] = P[Y_n(x) \ge -x + o(1)].$$

By Lemma 4 of Deo (1973), following the arguments on page 874 in that paper, it is easy to show that for any fixed x, the limiting distribution of $Y_n(x)$ is the same as that of $Y_n(0)$. Since the latter is clearly stochastically bounded, the result follows. □

The stochastic equicontinuity argument in the preceding proof will be used in two other places below. See also Billingsley (1968) and Withers (1975).

Define

$$(2.15)\quad X_n(u) = (np_n)^{-1/2} \sum_{i=1}^{n} \left[I_{i,1}(\bar{F}^{-1}(up_n)) - EI_{i,1}(\bar{F}^{-1}(up_n))\right],$$

for u in a neighborhood of 1.

LEMMA 2.3.

$$(np_n)^{1/2} A_n \to_d \text{normal}(0, \theta).$$

PROOF. By Lemma 4 of Deo (1973), arguing as on page 874 in that paper, it can be seen that for every sequence of nonnegative constants ε_n tending to 0, we have

$$(2.16)\quad \sup_{1-\varepsilon_n \le u \le 1+\varepsilon_n} |X_n(u) - X_n(1)| \to_p 0,$$

where $X_n(u)$ is defined by (2.15). Note that

$$\bar{F}_n^{-1}(p_n) = \bar{F}^{-1}\left(p_n^{-1}\bar{F}(\bar{F}_n^{-1}(p_n))p_n\right).$$

Hence, by (2.16) and Lemma 2.2, $(np_n)^{1/2}A_n$ has the same asymptotic distribution as $X_n(1)$. Thus the only issue in question is the variance of the limit. Note that (2.1) together with m-dependence implies (2.3), since (2.3) obviously holds if $r \ge m$, and for $r < m$, Theorem 2.1 of O'Brien (1987) implies that $P[M_{2,m} \le x | \xi_1 > x] \to \theta$ [cf. Hsing (1989)] which in turn implies (2.3). Note that

$$I_{1,1}(\bar{F}^{-1}(p_n)) I_{i,1}(\bar{F}^{-1}(p_n)) = 0 \quad \text{for } i = 2, \dots, r.$$

By this, (2.1) and (2.3),

$$EI_{1,1}(\bar{F}^{-1}(p_n)) I_{i,1}(\bar{F}^{-1}(p_n)) = \begin{cases} \theta p_n(1 + o(1)), & i = 1, \\ 0, & 2 \le i \le r, \\ o(p_n), & i \ge r + 1. \end{cases}$$

Thus it follows from m-dependence that the asymptotic variance of $X_n(1)$ is θ. □

To handle B_n, we need the following two lemmas.

LEMMA 2.4. *There exist subsets* $\mathscr{I}_1, \dots, \mathscr{I}_n$ *of* $\{1, 2, \dots, n\}$ *such that* $\sup_{1 \le i \le n}(n - \#(\mathscr{I}_i)) = o(\sqrt{n/p_n})$ *and*

$$(2.17)\quad \begin{aligned} \sup_{1 \le i \le n} \Big| & P\left[\xi_i > \bar{F}_n^{-1}(p_n) \ge M_{i+1, i+r-1}\right] \\ & - P\left[\hat{\xi}_1 > \bar{F}_{n,i}^{-1}(p_n) \ge \hat{M}_{2,r}\right]\Big| = o\left(\sqrt{\frac{p_n}{n}}\right), \end{aligned}$$

where $\bar{F}_n^{-1}(p_n)$ *is the* $[np] + 1$*th largest value of* $\xi_1, \ldots, \xi_n$, $\bar{F}_{n,i}^{-1}(p_n)$ *is the* $[np] + 1$*th largest value of* ξ_j, $j \in \mathscr{I}_i$, *and* $(\hat{\xi}_1, \hat{M}_{2,r})$ *has the same distribution as* $(\xi_1, M_{2,r})$ *but is independent of* $\bar{F}_{n,i}^{-1}(p_n)$, $i = 1, \ldots, n$.

PROOF. Let $\mathscr{I}_i = \{1, 2, \ldots, i - m, i + r - 1 + m, \ldots, n\}$. By m-dependence $(\xi_i, M_{i,i+r-1})$ is independent of $\bar{F}_{n,i}^{-1}(p_n)$. Observe that

$$\begin{aligned}
&\left|P\left[\xi_i > \bar{F}_n^{-1}(p_n) \ge M_{i+1,i+r-1}\right] - P\left[\hat{\xi}_1 > \bar{F}_{n,i}^{-1}(p_n) \ge \hat{M}_{2,r}\right]\right| \\
&\quad = \left|P\left[\xi_i > \bar{F}_n^{-1}(p_n) \ge M_{i+1,i+r-1}\right] - P\left[\xi_i > \bar{F}_{n,i}^{-1}(p_n) \ge M_{i+1,i+r-1}\right]\right| \\
&\quad \le P\left[\bar{F}_{n,i}^{-1}(p_n) < \xi_k \le \bar{F}_n^{-1}(p_n) \text{ for some } k = i, \ldots, i + r - 1\right].
\end{aligned}$$

Since $n - \#(\mathscr{I}_i) = 2m + r - 2$, the number of $\xi_1, \ldots, \xi_n$ that are in the interval $(\bar{F}_{n,i}^{-1}(p_n), \bar{F}_n^{-1}(p_n)]$ can at most be $2m + r - 1$. Thus the last probability is easily shown to be bounded by

$$\begin{aligned}
&\sum_{k=i}^{i+r-1} \sum_{j=1}^{2m+r-2} P\left[\xi_k \text{ is the } [np_n] + j\text{th largest value of } \xi_1, \ldots, \xi_n\right] \\
&\quad \le Cn^{-1} = o\left(\sqrt{\frac{p_n}{n}}\right),
\end{aligned}$$

where C is a constant independent of i. This completes the proof. □

LEMMA 2.5. *Let* $F_{n,i}$ *be as defined in Lemma* 2.4. *Then uniformly in* i,

$$E\left(\bar{F}\left(\bar{F}_{n,i}^{-1}(p_n)\right)\right) = p_n\left(1 + o\left((np_n)^{-1/2}\right)\right)$$

and

$$E\left(\bar{F}^{\beta+1}\left(\bar{F}_{n,i}^{-1}(p_n)\right)\right) \sim p_n^{\beta+1}.$$

PROOF. Both asymptotic statements are proved using similar arguments. Thus we illustrate by proving the first statement, whose proof is the more difficult of the two. Let $n' = \#(\mathscr{I}_i)$ and define

$$Z_{n,i} = (np_n)^{1/2}\left\{p_n^{-1}\bar{F}\left(\bar{F}_{n,i}^{-1}(p_n)\right) - 1\right\},$$

$$Y_{n,i}(x) = \frac{\sum_{j \in \mathscr{I}_i} I\left(\xi_j > \bar{F}^{-1}\left(p_n\left(1 + x(np_n)^{-1/2}\right)\right)\right) - n'p_n\left(1 + x(np_n)^{-1/2}\right)}{\sqrt{n'p_n\left(1 + x(np_n)^{-1/2}\right)}}$$

and

$$Y_{n,i} = Y_{n,i}(0).$$

By the proof of Lemma 2.2, for $x > 0$,

$$\begin{aligned} P[Z_{n,i} > x] &= P\Big[Y_{n,i}(x) \le \big(np_n - n'p_n\big(1 + x(np_n)^{-1/2}\big)\big) \\ &\qquad \times \Big(\sqrt{n'p_n\big(1 + x(np_n)^{-1/2}\big)}\Big)^{-1}\Big] \qquad (2.18) \\ &= P\Big[Y_{n,i}(x) < -x/\sqrt{1 + x(np_n)^{-1/2}}\,(1 + o(1))\Big], \end{aligned}$$

where the error of approximation $o(1)$ is uniform in i and $x > 0$. Similarly, uniformly for all i and $x < 0$,

$$P[Z_{n,i} < x] = P\Big[Y_{n,i}(x) > -x/\sqrt{1 + x(np_n)^{-1/2}}\,(1 + o(1))\Big]. \qquad (2.19)$$

It is clear that, for a fixed $\Delta > 0$,

$$\begin{aligned} \sup_{0<x<\Delta} &|P[Z_{n,i} > x] - P[Y_{n,i} < -x]| \\ &\le \sup_{0<x<\Delta} |P[Z_{n,i} > x] - P[Y_{n,i}(x) < -x]| \\ &\quad + \sup_{0<x<\Delta} |P[Y_{n,i}(x) < -x] - P[Y_{n,i} < -x]|, \end{aligned}$$

where the first term on the right tends to 0 by (2.18) and the second term tends to 0 by the standard arguments for proving sample path equicontinuity. Thus

$$\sup_{0<x<\Delta} |P[Z_{n,i} > x] - P[Y_{n,i} < -x]| \to 0, \qquad \Delta > 0. \qquad (2.20)$$

Similarly,

$$\sup_{0<x<\Delta} |P[Z_{n,i} < -x] - P[Y_{n,i} > x]| \to 0, \qquad \Delta > 0. \qquad (2.21)$$

By m-dependence and the fact that $Y_{n,i}$ is a normalized partial sum of indicators, $E(Y_{n,i}(x))^4$ is bounded in n, i and x [cf. Deo (1973) and Ibragimov (1962)]. Thus by (2.18), (2.19) and the Schwarz inequality, for any fixed $\Delta > 0$ there exists a constant C such that

$$\int_\Delta^\infty P[|Z_{n,i}| > x]\,dx \le C\int_\Delta^\infty \left(\frac{1}{x^4} + \frac{2}{\sqrt{np_n}}\frac{1}{x^3} + \frac{1}{np_n}\frac{1}{x^2}\right) dx.$$

Thus we have

$$\lim_{\Delta\to\infty}\lim_{n\to\infty}\int_\Delta^\infty P[|Z_{n,i}| > x]\,dx = 0. \qquad (2.22)$$

Also

$$\lim_{\Delta\to\infty}\lim_{n\to\infty}\int_\Delta^\infty P[|Y_{n,i}| > x]\,dx = 0. \qquad (2.23)$$

Since $EY_{n,i} = 0$, by (2.20) through (2.23), we have

$$0 = \lim_{n\to\infty} EY_{n,i} = \lim_{n\to\infty} \int_0^\infty (P[Y_{n,i} > x] - P[Y_{n,i} < -x])\,dx$$
$$= \lim_{\Delta\to\infty} \lim_{n\to\infty} \int_0^\Delta (P[Y_{n,i} > x] - P[Y_{n,i} < -x])\,dx$$
$$= \lim_{\Delta\to\infty} \lim_{n\to\infty} \int_0^\Delta (P[Z_{n,i} < -x] - P[Z_{n,i} > x])\,dx$$
$$= \lim_{n\to\infty} \int_0^\infty (P[Z_{n,i} < -x] - P[Z_{n,i} > x])\,dx = -\lim_{n\to\infty} EZ_{n,i}.$$

This concludes the proof of the first assertion of the lemma. □

The asymptotic behavior of B_n is now derived as follows.

LEMMA 2.6.

$$B_n = o\big((np_n)^{-1/2}\big) + \eta p_n^\beta(1 + o(1)).$$

PROOF. Using the notation of Lemma 2.4, by (2.14) and (2.17) we have

$$B_n = o\big((np_n)^{-1/2}\big) + (np_n)^{-1} \sum_{i=1}^n P\Big[\hat{\xi}_1 > \bar{F}_{n,i}^{-1}(p_n) \ge \hat{M}_{2,r}\Big] - \theta. \tag{2.24}$$

By independence and Lemma 2.5, uniformly in $i = 1, \dots, n$,

$$\begin{aligned} P\Big[\hat{\xi}_1 > \bar{F}_{n,i}^{-1}(p_n) \ge \hat{M}_{2,r}\Big] \\ &= \theta E\big(\bar{F}(\bar{F}_{n,i}^{-1}(p_n))\big) + \eta E\big(\bar{F}^{\beta+1}(\bar{F}_{n,i}^{-1}(p_n))\big)(1 + o(1)) \\ &= \theta p_n\big(1 + o\big((np_n)^{-1/2}\big)\big) + \eta p_n^{\beta+1}(1 + o(1)). \end{aligned} \tag{2.25}$$

The result follows from (2.24) and (2.25). □

Combining the conclusions of Lemmas 2.3 and 2.6, we have

$$\hat{\theta}_n\big(\bar{F}_n^{-1}(p_n)\big) - \theta = \eta p_n^\beta(1 + o(1)) + (np_n)^{-1/2} Z_n(1 + o(1)), \tag{2.26}$$

where the limiting distribution of Z_n is normal$(0, \theta)$. Thus one can loosely write

$$\mathrm{MSE}\big(\hat{\theta}_n\big(\bar{F}_n^{-1}(p_n)\big)\big) = \eta^2 p_n^{2\beta} + \frac{\theta}{np_n}.$$

Thus the p_n which minimizes this expression is

$$p_{n,0} = \lambda_0 n^{-[1/(2\beta+1)]}, \tag{2.27}$$

where

$$\lambda_0 = \left(\frac{\theta}{2\beta\eta^2}\right)^{1/(2\beta+1)}.$$

It is clear that

$$n^{\beta/(2\beta+1)}\Big(\hat{\theta}_n\big(\bar{F}_n^{-1}(p_{n,0})\big) - \theta\Big) \to_d \text{normal}\left(\eta\lambda_0^\beta, \frac{\theta}{\lambda_0}\right).$$

The goal now is to find preliminary estimates of θ, β and η, from which $p_{n,0}$ can be estimated.

For $p_n = n^{-\delta}$, it follows from (2.26) that

$$\hat{\theta}_n\big(\bar{F}_n^{-1}(n^{-\delta})\big) = \begin{cases} \theta + O_P(n^{-[(1-\delta)/2]}), & \text{if } \dfrac{1}{2\beta+1} < \delta < 1, \\ \theta + \eta n^{-\delta\beta}(1 + o_P(1)), & \text{if } 0 < \delta < \dfrac{1}{2\beta+1} \end{cases}. \tag{2.28}$$

To estimate α, β and γ, we consider a method inspired by Hall and Welsh (1985). Suppose we can pick two positive constants β_1 and β_2 such that $\beta_1 < \beta < \beta_2$. While the method will be particularly effective if the bounds are chosen accurately, we do not assess the precise effect of the choice of the bounds. Next choose δ, δ_1 and δ_2 such that

$$\frac{1}{2\beta_1 + 1} < \delta < 1 \quad \text{and} \quad 0 < \delta_1 < \delta_2 < \frac{1-\delta}{2\beta_2}. \tag{2.29}$$

Define

$$\begin{aligned} \tilde{\theta} &= \hat{\theta}_n\big(\bar{F}_n^{-1}(n^{-\delta})\big), \\ \tilde{\beta} &= \frac{\log\left(\left(\dfrac{\big\{\hat{\theta}_n\big(\bar{F}_n^{-1}(n^{-\delta_1})\big) - \tilde{\theta}\big\}}{\big\{\hat{\theta}_n\big(\bar{F}_n^{-1}(n^{-\delta_2})\big) - \tilde{\theta}\big\}} \wedge n^{\beta_2(\delta_2-\delta_1)}\right) \vee n^{\beta_1(\delta_2-\delta_1)}\right)}{\log(n^{\delta_2-\delta_1})}, \\ \tilde{\eta} &= n^{\tilde{\beta}\delta_1}\big[\hat{\theta}_n\big(\bar{F}_n^{-1}(n^{-\delta_1})\big) - \tilde{\theta}\big]. \end{aligned} \tag{2.30}$$

LEMMA 2.7. *Under the assumptions stated previously, as $n \to \infty$,*

$$\tilde{\eta} \to_p \eta, \qquad \tilde{\theta} \to_p \theta, \qquad \tilde{\beta} = \beta + o_P(1/\log n).$$

PROOF. Now (2.29) implies that

$$\frac{1-\delta}{2\beta_2} < \frac{1}{2\beta_2+1} < \frac{1}{2\beta+1} < \frac{1}{2\beta_1+1} < \delta,$$

where the leftmost inequality is equivalent to $\delta > 1/(2\beta_2 + 1)$. Consequently, it follows from (2.29) that

$$0 < \delta_1 < \delta_2 < \frac{1}{2\beta+1} < \delta < 1 \tag{2.31}$$

and

$$\beta\delta_1 < \beta\delta_2 < \frac{1-\delta}{2}. \tag{2.32}$$

By (2.31) and (2.28),

$$\begin{aligned} \hat{\theta}_n\big(\overline{F}_n^{-1}(n^{-\delta_1})\big) &= \theta + \eta n^{-\delta_1\beta}(1 + o_P(1)),\\ \hat{\theta}_n\big(\overline{F}_n^{-1}(n^{-\delta_2})\big) &= \theta + \eta n^{-\delta_2\beta}(1 + o_P(1)),\\ \tilde{\theta} &= \theta + O_P\big(n^{-[(1-\delta)/2]}\big). \end{aligned} \tag{2.33}$$

It is clear that $\tilde{\theta} \to_p \theta$. By (2.32) and (2.33),

$$\frac{\hat{\theta}_n\big(\overline{F}_n^{-1}(n^{-\delta_1})\big) - \tilde{\theta}}{\hat{\theta}_n\big(\overline{F}_n^{-1}(n^{-\delta_2})\big) + \tilde{\theta}} = n^{\beta(\delta_2-\delta_1)}\big(1 + o_P(1)\big).$$

Since $\beta_1 < \beta < \beta_2$,

$$\tilde{\beta} = \beta + o_P\left(\frac{1}{\log n}\right). \tag{2.34}$$

Similarly,

$$\hat{\theta}_n\big(\overline{F}_n^{-1}(n^{-\delta_1})\big) - \tilde{\theta} = \eta n^{-\beta\delta_1}\big(1 + o_P(1)\big).$$

Therefore, by (2.34), $\tilde{\eta} \to_p \eta$. This completes the proof. □

We now summarize the preceding derivation in the following result.

THEOREM 2.8. *Suppose that* $\{\xi_j\}$ *is m-dependent,* (2.1) *holds with* $R(p) \sim \eta p^{\beta}$ *for some* $r \geq 2$, $\eta \neq 0$ *and* $\beta > 0$. *Then the asymptotic mean squared error of* $\hat{\theta}_n(\overline{F}_n^{-1}(p_n))$ *is minimized with*

$$p_n = p_{n,0} = \left(\frac{\theta}{2\beta\eta^2 n}\right)^{1/(2\beta+1)}.$$

Furthermore, for $\tilde{\theta}$, $\tilde{\eta}$ *and* $\tilde{\beta}$ *defined by* (2.30), *and writing*

$$\hat{p}_{n,0} = \left(\frac{\tilde{\theta}}{2\tilde{\beta}\tilde{\eta}^2 n}\right)^{1/(2\tilde{\beta}+1)},$$

we have

$$\frac{\hat{p}_{n,0}}{p_{n,0}} \to_p 1 \quad \text{as } n \to \infty.$$

Under the conditions of the preceding theorem, θ can be estimated by the adaptive estimate $\hat{\theta}_n(\overline{F}_n(\hat{p}_{n,0}))$. Now that we have a good estimate of θ, a marginal improvement can be made by applying again the preceding procedure, with $\tilde{\theta}$ replaced by $\hat{\theta}_n(\overline{F}_n^{-1}(\hat{p}_{n,0}))$.

3. Approximating the distribution of the maximum. Let $\{\xi_j\}$ be stationary and have a nonzero extremal index θ. The main result of this section, Theorem 3.1, shows that the distribution of M_n is determined by the dependence structure of $\{\xi_j\}$ through θ if n is large. For convenience, we work under the following mixing condition. For $c > 0$ and $1 \le l \le n - 1$, define

$$\alpha_n(l;c) = \sup\Big\{\big|P\big[M_{j+l+1,j+l+k} \le x | M_{1,j} > x\big] - P[M_k \le x]\big|: \\ j,k \ge 1,\ j+l+k \le n,\ x \ge \bar{F}^{-1}(c/n)\Big\}.$$

We assume in this section that, for each $c > 0$,

$$\alpha(l;c) := \limsup_{n\to\infty} \alpha_n(l;c) \to 0 \quad \text{as } l \to \infty. \tag{3.1}$$

Note that this implies that for each $c > 0$, there exists a sequence $\{l_n\}$ such that

$$l_n/n \to 0 \quad \text{and} \quad \alpha_n(l_n;c) \to 0 \quad \text{as } n \to \infty. \tag{3.2}$$

Observe that

$$P\big[M_{j+l+1,j+l+k} \le x | M_{1,j} \le x\big] - P[M_k \le x] \\ = \frac{P\big[M_{1,j} > x\big]}{P\big[M_{1,j} \le x\big]}\Big(P[M_k \le x] - P\big[M_{j+l+1,j+l+k} \le x | M_{1,j} > x\big]\Big),$$

where

$$\liminf_{n\to\infty} \inf_{x \ge \bar{F}^{-1}(c/n)} \inf_{1\le j\le n} P\big[M_{1,j} \le x\big] \ge \liminf_{n\to\infty} P\big[M_{1,n} \le \bar{F}^{-1}(c/n)\big] \\ = e^{-\theta c} > 0.$$

Thus (3.1) and (3.2) hold with α_n replaced by $\tilde{\alpha}_n$ defined by

$$\tilde{\alpha}_n(l;c) = \sup\Big\{\big|P\big[M_{j+l+1,j+l+k} \le x | M_{1,j} \le x\big] \\ -P[M_k \le x]\big|: j,k \ge 1,\ j+l+k \le n,\ x \ge \bar{F}^{-1}(c/n)\Big\}.$$

Thus (3.1) implies O'Brien's AIM($\bar{F}^{-1}(c/n)$) [cf. O'Brien (1987)] and it follows readily from O'Brien (1987), Theorem 2.1, that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \big|P\big[M_{2,m} \le \bar{F}^{-1}(c/n) | \xi_1 > \bar{F}^{-1}(c/n)\big] - \theta\big| = 0,$$

which implies that

$$\lim_{m\to\infty} \limsup_{x\to x_0} \big|P\big[M_{2,m} \le x | \xi_1 > x\big] - \theta\big| = 0. \tag{3.3}$$

The main result of this section is the following.

THEOREM 3.1. *Under the mixing condition* (3.1),

$$\lim_{n\to\infty} \sup_{\text{all } x} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| = 0,$$

where $0/0$ *is interpreted as* 1.

PROOF. First fix $m \in \{1, 2, \dots\}$ and $0 < c_0 < c_1 < \infty$. By (3.2) we can pick r_n such that

$$(3.4) \qquad r_n/n \to 0, \qquad l_n/r_n \to 0, \qquad k_n \alpha_n(l_n; c_1) \to 0 \quad \text{as } n \to \infty,$$

where $k_n = [n/r_n]$. Write

$$\tilde{M}_n = \max\left(\xi_j \colon j \in \{1, \dots, k_n r_n\} - \bigcup_{i=1}^{k_n} \{ir_n - l_n, \dots, ir_n\} \right).$$

By the triangle inequality,

$$\begin{aligned}
\left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \le{}& \frac{\left|P[M_n > x] - P\left[\tilde{M}_n > x\right]\right|}{1 - F^{\theta n}(x)} \\
&+ \frac{\left|P\left[\tilde{M}_n > x\right] - 1 + P^{k_n}\left[M_{r_n - l_n} \le x\right]\right|}{1 - F^{\theta n}(x)} \\
&+ \frac{\left|P^{k_n}\left[M_{r_n - l_n} \le x\right] - F^{\theta n}(x)\right|}{1 - F^{\theta n}(x)} \\
=:{}& A_n(x) + B_n(x) + C_n(x).
\end{aligned}$$

Making use of Lemmas 3.2 through 3.5 (which follows), we get

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\text{all } x} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \\
&\quad\le \limsup_{n\to\infty} \sup_{x \le \bar{F}^{-1}(c_1/n)} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \\
&\qquad + \limsup_{n\to\infty} \sup_{x > \bar{F}^{-1}(c_1/n)} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \\
&\quad\le \limsup_{n\to\infty} \sup_{x \le \bar{F}^{-1}(c_1/n)} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| + \limsup_{n\to\infty} \sup_{x > \bar{F}^{-1}(c_1/n)} A_n(x) \\
&\qquad + \limsup_{n\to\infty} \sup_{\bar{F}^{-1}(c_1/n) < x \le \bar{F}^{-1}(c_0/n)} B_n(x) + \limsup_{n\to\infty} \sup_{x > \bar{F}^{-1}(c_0/n)} B_n(x) \\
&\qquad + \limsup_{n\to\infty} \sup_{x > \bar{F}^{-1}(c_1/n)} C_n(x) \\
&\quad\le \frac{2e^{-\theta c_1}}{1 - e^{-\theta c_1}} + \frac{c_0^2}{1 - e^{-\theta c_0}} + \frac{c_1}{1 - e^{-\theta c_1}} \varepsilon(m, c_1),
\end{aligned}$$

where $\varepsilon(m, c_1) \to 0$ as $m \to \infty$. Thus the result follows from first letting $m \to \infty$, and then $c_0 \to 0$ and $c_1 \to \infty$. □

We remark that it is also possible to assess the rate of convergence of the convergence statement in subsection 3.1, if additional information on the dependence structure is given. See Hsing (1990).

In the following we will apply the simple but useful facts that

$$\frac{1}{1 - F^{\theta n}(x)} \le \frac{1}{1 - e^{-\theta n \overline{F}(x)}} \tag{3.5}$$

and

$$\frac{v^a}{1 - e^{-\theta n v}}, \ v > 0, \text{ is increasing in } v \text{ for } a \ge 1. \tag{3.6}$$

LEMMA 3.2.

$$\limsup_{n\to\infty} \sup_{x \le \overline{F}^{-1}(c_1/n)} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \le \frac{2e^{-\theta c_1}}{1 - e^{-\theta c_1}}.$$

PROOF. Straightforward arguments show

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{x \le \overline{F}^{-1}(c_1/n)} \left| \frac{P[M_n > x]}{1 - F^{\theta n}(x)} - 1 \right| \\
&\quad \le \limsup_{n\to\infty} \sup_{x \le \overline{F}^{-1}(c_1/n)} \left(\frac{P[M_n \le x]}{1 - F^{\theta n}(x)} + \frac{F^{\theta n}(x)}{1 - F^{\theta n}(x)} \right) \\
&\quad \le \limsup_{n\to\infty} \left(\frac{P\big[M_n \le \overline{F}^{-1}(c_1/n)\big]}{1 - F^{\theta n}\big(\overline{F}^{-1}(c_1/n)\big)} + \frac{F^{\theta n}\big(\overline{F}^{-1}(c_1/n)\big)}{1 - F^{\theta n}\big(\overline{F}^{-1}(c_1/n)\big)} \right) \\
&\quad \le \frac{2e^{-\theta c_1}}{1 - e^{-\theta c_1}}.
\end{aligned}$$

□

LEMMA 3.3.

$$\lim_{n\to\infty} \sup_{x > \overline{F}^{-1}(c_1/n))} A_n(x) = 0.$$

PROOF. By Boole's inequality,

$$A_n(x) \le \frac{(k_n l_n + r_n)\overline{F}(x)}{1 - F^{\theta n}(x)}.$$

Thus, by (3.5), (3.6) and (3.4),

$$\sup_{x > \overline{F}^{-1}(c_1/n)} A_n(x) \le \frac{k_n l_n + r_n}{n} \frac{c_1}{1 - e^{-\theta c_1}} \to 0 \quad \text{as } n \to \infty.$$

□

LEMMA 3.4.

$$\lim_{n\to\infty} \sup_{\bar{F}^{-1}(c_1/n)<x\le\bar{F}^{-1}(c_0/n)} B_n(x) = 0$$

and

$$\limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_0/n)} B_n(x) \le \frac{c_0^2}{1-e^{-\theta c_0}}.$$

PROOF. It suffices to show that, for $x > \bar{F}^{-1}(c_1/n)$,

$$B_n(x) \le \frac{k_n\tilde{\alpha}_n(l_n; c_1)}{1-F^{\theta n}(x)} \wedge \left[\frac{n^2\bar{F}^2(x)}{1-F^{\theta n}(x)} + \frac{k_n n\bar{F}(x)\alpha_n(l_n; c_1)}{2(1-F^{\theta n}(x))}\right]. \tag{3.7}$$

This is so since the first bound of (3.7) together with (3.5) and (3.4) implies that

$$\sup_{\bar{F}^{-1}(c_1/n)<x\le\bar{F}^{-1}(c_0/n)} B_n(x) \le \frac{k_n\tilde{\alpha}_n(l_n; c_1)}{1-e^{-\theta c_0}} \to 0 \quad \text{as } n\to\infty,$$

and the second bound of (3.7) plus (3.5), (3.6) and (3.4) imply that

$$\sup_{x>\bar{F}^{-1}(c_0/n)} B_n(x) \le \frac{c_0^2}{1-e^{-\theta c_0}} + \frac{c_0 k_n \alpha_n(l_n; c_1)}{2(1-e^{-\theta c_0})} \to \frac{c_0^2}{1-e^{-\theta c_0}} \quad \text{as } n\to\infty.$$

The first term on the right of (3.7) comes from repeatedly applying

$$\left|P\left[\bigvee_{i=j}^{k_n} M_{(i-1)r_n+1,\, ir_n-l_n} \le x\right] - P\left[M_{1,r_n-l_n} \le x\right] P\left[\bigvee_{i=j+1}^{k_n} M_{(i-1)r_n+1,\, ir_n-l_n} \le x\right]\right|$$
$$\le \tilde{\alpha}_n(l_n; c_1), \qquad 1 \le j \le k_n,$$

and the triangle inequality. A different argument is needed to derive the second term on the right of (3.7). By the triangle inequality, we get

$$\begin{aligned}
&\left|P\left[\tilde{M}_n > x\right] - 1 + P^{k_n}\left[M_{r_n-l_n} \le x\right]\right| \\
&\quad \le \left|P\left[\tilde{M}_n > x\right] - k_n P\left[M_{r_n-l_n} > x\right]\right| \\
&\qquad + \left|k_n P\left[M_{r_n-l_n} > x\right] - \left(1 - P^{k_n}\left[M_{r_n-l_n} \le x\right]\right)\right|.
\end{aligned}$$

By Bonferroni's inequality and Boole's inequality,

$$\begin{aligned}
0 &\le k_n P\left[M_{r_n-l_n} > x\right] - \left(1 - P^{k_n}\left[M_{r_n-l_n} \le x\right]\right) \\
&\le \frac{k_n(k_n-1)}{2} P^2\left[M_{r_n-l_n} > x\right] \le \frac{k_n(k_n-1)(r_n-l_n)^2}{2} \bar{F}^2(x).
\end{aligned}$$

By the same token plus the assumed mixing condition,

$$0 \le k_n P\left[M_{r_n - l_n} > x\right] - P\left[\tilde{M}_n > x\right]$$
$$\le \frac{k_n(k_n-1)}{2} P\left[M_{r_n-l_n} > x\right]\left(P\left[M_{r_n-l_n} > x\right] + \alpha_n(l_n; c_1)\right)$$
$$\le \frac{k_n(k_n-1)(r_n-l_n)^2}{2}\overline{F}^2(x) + \frac{k_n(k_n-1)(r_n-l_n)}{2}\overline{F}(x)\alpha_n(l_n;c_1).$$

Inequality (3.7) follows from this. □

LEMMA 3.5.

$$\limsup_{n\to\infty} \sup_{x > \overline{F}^{-1}(c_1/n)} C_n(x) \le \frac{c_1}{1 - e^{-\theta c_1}}\varepsilon(m, c_1),$$

where $\varepsilon(m, c_1) \to 0$ *as* $m \to \infty$ *for fixed* $0 < c_1 < \infty$.

PROOF. By Bonferroni's inequality, there exists a constant $B > 0$ such that

$$1 - F^{\theta n/k_n}(x) - r_n\theta\overline{F}(x) \le Br_n^2\overline{F}^2(x), \tag{3.8}$$

for all $x > \overline{F}^{-1}(c_1/n)$. Next write

$$P\left[M_{r_n-l_n} > x\right] - r_n\theta\overline{F}(x) = \sum_{i=1}^{r_n-l_n} P\left[\xi_i > x \ge M_{i+1, r_n-l_n}\right] - r_n\theta\overline{F}(x)$$
$$= \sum_{i=1}^{r_n-l_n} P\left[\xi_1 > x \ge M_{2, r_n-l_n-i+1}\right] - r_n\theta\overline{F}(x),$$

which can be seen to equal

$$\left(\sum_{i=1}^{r_n} P[\xi_1 > x \ge M_{2,m}] - r_n\theta\overline{F}(x)\right) - \sum_{i=r_n-l_n-m+2}^{r_n} P[\xi_1 > x \ge M_{2,m}]$$
$$- \sum_{i=1}^{r_n-l_n-m+1} P\left[\xi_1 > x \ge M_{2,m}, M_{m+1, r_n-l_n-i+1} > x\right] \tag{3.9}$$
$$+ \sum_{i=r_n-l_n-m+2}^{r_n-l_n} P\left[\xi_1 > x \ge M_{2, r_n-l_n-i+1}\right]$$
$$=: C_{n,m,1}(x) + C_{n,m,2}(x) + C_{n,m,3}(x) + C_{n,m,4}(x).$$

Now using the elementary inequality,

$$C_n(x) \le \frac{k_n\left|P\left[M_{r_n-l} \le x\right] - F^{\theta n/k_n}(x)\right|}{1 - F^{\theta n}(x)}$$

$$\le \frac{k_n\left|P\left[M_{r_n-l} > x\right] - r_n\theta\bar{F}(x)\right|}{1 - F^{\theta n}(x)} + \frac{\left|1 - F^{\theta n/k_n}(x) - r_n\theta\bar{F}(x)\right|}{1 - F^{\theta n}(x)},$$

it follows from (3.8) and (3.9) that

$$C_n(x) \le k_n \frac{Br_n^2\bar{F}^2(x) + \Sigma_{i=1}^4|C_{n,m,i}(x)|}{1 - F^{\theta n}(x)}.$$

Clearly, by Boole's inequality,

$$\limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} k_n \frac{Br_n^2\bar{F}^2(x) + |C_{n,m,2}(x)| + |C_{n,m,4}(x)|}{1 - F^{\theta n}(x)}$$

$$\le \limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} k_n \frac{Br_n^2\bar{F}^2(x) + (l_n + 2m - 2)\bar{F}(x)}{1 - F^{\theta n}(x)},$$

and, by (3.5), (3.6) and (3.4), is bounded by

$$\limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} k_n \frac{Br_n^2c_1^2/n^2 + (l_n + 2m - 2)c_1/n}{1 - e^{-\theta c_1}} = 0.$$

Write

$$\varepsilon_1(m, c_1) := \limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} \frac{k_n|C_{n,m,1}(x)|}{1 - F^{\theta n}(x)},$$

which, by (3.3), (3.5) and (3.6), is bounded by

$$\limsup_{n\to\infty} \frac{k_n r_n c_1/n}{1 - e^{-\theta c_1}} \sup_{x>\bar{F}^{-1}(c_1/n)} \left|P\left[M_{2,m} \le x|\xi_1 > x\right] - \theta\right|$$

$$= \frac{c_1}{1 - e^{-\theta c_1}} \limsup_{x\to x_0} \left|P\left[M_{2,m} \le x|\xi_1 > x\right] - \theta\right|.$$

Finally, define

$$\varepsilon_3(m, c_1) := \limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} \frac{k_n|C_{n,m,3}(x)|}{1 - F^{\theta n}(x)},$$

which is bounded by

$$\limsup_{n\to\infty} \sup_{x>\bar{F}^{-1}(c_1/n)} \frac{k_n r_n P\left[\xi_1 > x,\, M_{m+1,r_n} > x\right]}{1 - F^{\theta n}(x)}.$$

Thus, by the assumed mixing condition, (3.5) and (3.6),

$$\varepsilon_3(m, c_1) \le \limsup_{n \to \infty} \frac{k_n r_n c_1/n(r_n c_1/n + \alpha_n(m; c_1))}{1 - e^{-\theta c_1}} = \frac{c_1}{1 - e^{-\theta c_1}} \alpha_n(m; c_1).$$

It is easy to see that

$$\begin{aligned}\varepsilon(m, c_1) &:= \varepsilon_1(m, c_1) + \varepsilon_3(m, c_1) \\ &\le \frac{c_1}{1 - e^{-\theta c_1}} \Big(\limsup_{x \to x_0} |P[M_{2,m} \le x | \xi_1 > x] - \theta| + \alpha_n(m; c_1) \Big),\end{aligned}$$

which approaches 0 as $m \to \infty$ by (3.1) and (3.3). This concludes the proof. □

4. The extremal index of an *m*-dependent sequence. Assume in this section that $\{\xi_j\}$ is an m-dependent stationary sequence of random variables for some finite m. This section is devoted to finding the asymptotic relationship between $\bar{F}(x)$ and $P[M_r > x]$ and between $\bar{F}(x)$ and $P[M_{2,r} \le x | \xi_1 > x]$. The following results are useful in a number of ways. First, they provide a better understanding of how the extremal index relates to local dependence and explain the sources of the approximation errors in (1.2), (1.3) and (2.1). More importantly, they are useful in suggesting alternative ways of modeling dependence in the context of estimating the extremal index.

We could derive the results in this section in a more general setting of dependence. But the amount of extra work outweighs what would be gained by doing so.

For convenience, define

$$\alpha_k(x) = \sum P\Big[\bigcap_{1 \le j \le k} (\xi_{i_j} > x) \Big],$$

$$\beta_k(x) = \sum (i_k - 1) P\Big[\bigcap_{1 \le j \le k} (\xi_{i_j} > x) \Big],$$

where the summations are taken over the set

$$\{(i_1, \dots, i_k) \colon 1 = i_1 < \cdots < i_k \le m\}.$$

Define

$$\theta(x) = \frac{\sum_{k=1}^{m} (-1)^{k+1} \alpha_k(x)}{\bar{F}(x)},$$

$$\theta'(x) = \frac{\sum_{k=2}^{m} (-1)^{k} \beta_k(x)}{\bar{F}(x)}. \tag{4.1}$$

THEOREM 4.1. *Suppose $r = r_n$ and $x = x_n$ are such that $r \to \infty$ and $r\bar{F}(x) \to 0$ as $n \to \infty$. Then*

$$\frac{P(M_r > x)}{r\bar{F}(x)} = \theta(x) + \frac{\theta'(x)}{r} - \frac{r\bar{F}(x)}{2}\theta(x)^2 + o\big(r\bar{F}(x)\big). \tag{4.2}$$

PROOF. First write

$$P[M_r > x] = \sum_{k=1}^{r} (-1)^{k+1} p_k(x), \tag{4.3}$$

where

$$p_k(x) = \sum_{1 \le i_1 < \cdots < i_k \le r} P\left[\bigcap_{1 \le j \le k} (\xi_{i_j} > x)\right].$$

For $k \le r$ define

$$B_{k,r} = \{\mathbf{i}: 1 \le i_1 < \cdots < i_k \le r\},$$

$$B_{k,r}^{(1)} = \{\mathbf{i} \in B_{k,r}: 0 \le i_k - i_1 \le m - 1\},$$

$$B_{k,r}^{(2)} = \{\mathbf{i} \in B_{k,r}: \text{there exists } j \text{ such that } 0 \le i_k - i_j \le m - 1, \; 0 \le i_{j-1} - i_1 \le m - 1 \text{ and } i_j - i_{j-1} \ge m\},$$

$$B_{k,r}^{(3)} = \{\mathbf{i} \in B_{k,r}: m \le i_k - i_1 \le 3m - 3\},$$

$$B_{k,r}^{(4)} = \{\mathbf{i} \in B_{k,r}: \text{there exists } j \text{ such that } i_k - i_j \ge m \text{ and } i_j - i_1 \ge m\}.$$

If $k > r$, $B_{k,r}$ and $B_{k,r}^{(i)}$ denote null sets. Note that

$$\big(B_{k,r}^{(4)}\big)^c = \{\mathbf{i} \in B_{k,r}: \text{for each } j, (i_j - i_1) \wedge (i_k - i_j) \le m - 1\}.$$

Thus

$$B_{k,r}^{(2)} = \big(B_{k,r}^{(4)}\big)^c \cap \big(B_{k,r}^{(1)} \cup B_{k,r}^{(3)}\big)^c.$$

Consequently, $B_{k,r} = \bigcup_{j=1}^{4} B_{k,r}^{(j)}$, and hence

$$p_k(x) = \sum_{B_{k,r}^{(1)} \cup B_{k,r}^{(2)} \cup B_{k,r}^{(3)} \cup B_{k,r}^{(4)}} P\left[\bigcap_{1 \le j \le k} (\xi_{i_j} > x)\right].$$

Here and hereafter, we use the convention that any summation over the null set is 0. It is easy to conclude from m-dependence that

$$\sum_{\mathbf{i} \in B_{k,r}^{(3)}} P\left[\bigcap_{1 \le j \le k} (\xi_{i_j} > x)\right] \le \binom{3m-3}{k} r\bar{F}^2(x) = o\big(r^2\bar{F}^2(x)\big)$$

and

$$\sum_{\mathbf{i} \in B_{k,r}^{(4)}} P\left[\bigcap_{1 \le j \le k} (\xi_{i_j} > x)\right] \le \binom{r}{3} \bar{F}^3(x) = o\big(r^2\bar{F}^2(x)\big).$$

Thus

$$p_k(x) = \begin{cases} \sum_{B^{(1)}_{k,r} \cup B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] + o(r^2\overline{F}^2(x)), & 1 \le k \le m, \\ \sum_{B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] + o(r^2\overline{F}^2(x)), & m+1 \le k \le 2m, \\ o(r^2\overline{F}^2(x)), & k \ge 2m+1. \end{cases}$$

It follows from (4.3) that

$$\begin{aligned} P[M_r > x] = {} & \sum_{k=1}^{m} (-1)^{k+1} \sum_{B^{(1)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] \\ & + \sum_{k=1}^{2m} (-1)^{k+1} \sum_{B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] + o(r^2\overline{F}(x)). \end{aligned} \tag{4.4}$$

Observe that

$$\begin{aligned} & \sum_{B^{(1)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] \\ & \quad = \sum_{1=i_1<\cdots<i_k\le m} (r - i_k + 1)_+ P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right], \end{aligned} \tag{4.5}$$

where $(\cdot)_+ = \cdot \vee 0$, and

$$\begin{aligned} & \sum_{B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] \\ & \quad = \sum_{\substack{s+t=k \\ 1\le s\le m \\ 1\le t\le m}} \sum_{\substack{1=i_1<\cdots<i_s\le m \\ 1=j_1<\cdots<j_t\le m}} I(r - i_s - j_t \ge m) \\ & \qquad \times P\left[\bigcap_{1\le l\le s} (\xi_{i_l} > x)\right] P\left[\bigcap_{1\le l\le t} (\xi_{j_l} > x)\right] \sum_{u=m}^{r-i_s-j_t} (r - i_s - j_t - u + 1). \end{aligned} \tag{4.6}$$

Notice that up to this point we have not made use of the fact that $r \to \infty$, and the arguments will also be applicable in the proof of Lemma 4.2. However, we assume that $r \to \infty$ from this point on. For large r, it follows from (4.5) that

$$\sum_{B^{(1)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] = r\alpha_k(x) - \beta_k(x), \qquad 1 \le k \le m, \tag{4.7}$$

and from (4.6) that

$$\sum_{B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j} > x)\right] = \frac{r^2}{2} \sum_{\substack{s+t=k \\ 1\le s\le m \\ 1\le t\le m}} \alpha_s(x)\alpha_t(x) + o(r^2\overline{F}^2(x)), \qquad 1 \le k \le 2m. \tag{4.8}$$

By (4.4), (4.7) and (4.8),

$$P[M_r > x] = r \sum_{k=1}^{m} (-1)^{k+1} \alpha_k(x) + \sum_{k=1}^{m} (-1)^k \beta_k(x)$$
$$-\frac{r^2}{2}\left(\sum_{k=1}^{m} (-1)^{k+1} \alpha_k(x)\right)^2 + o\big(r^2 \bar{F}^2(x)\big).$$

This concludes the proof. □

Comparing (4.2) with Leadbetter's result (1.2), it is readily seen that the extremal index of $\{\xi_j\}$ exists and is equal to θ if and only if $\theta(x)$ converges to θ as $x \to x_0$. In this case,

$$\text{(4.9)} \qquad \frac{P[M_r > x]}{r\bar{F}(x)} - \theta = (\theta(x) - \theta) + \frac{\theta'(x)}{r} - \frac{r\bar{F}(x)\theta^2}{2} + o\big(r\bar{F}(x)\big).$$

This motivates the following estimator which has a different form from $\hat{\theta}_n$ in Section 2. For given constants x and r, define

$$\tilde{\theta}_n(x, r) = \frac{\sum_{i=1}^{k} I(M(i) > x)}{\sum_{j=1}^{n} I(\xi_j > x)},$$

where

$$M(i) = \max_{(i-1)r+1 \le j \le ir} \xi_j \quad \text{and} \quad k = [n/r].$$

Under fairly general conditions $\tilde{\theta}_n(x_n, r_n)$ estimates θ consistently [cf. Hsing (1991a)]. Equation (4.9) is useful in suggesting models for the bias of this estimator. Although we do not investigate this approach in the present paper, we mention the following simple observation. It can be seen from the proof of Theorem 4.1 that for any fixed r,

$$\frac{P[M_r > x]}{r\bar{F}(x)} \nrightarrow \theta \quad \text{as } x \to x_0,$$

despite the fact that the range of dependence of $\{\xi_j\}$ is finite. Compare this with (2.1), where r can be chosen fixed for a large class of processes whose ranges of dependence may possibly be infinite. Thus both x and r in $\tilde{\theta}_n(x, r)$ will always have to be chosen to ensure a good rate of convergence even if $\{\xi_j\}$ is m-dependent. This fact alone makes $\tilde{\theta}_n$ more difficult to implement.

The following result is a refinement of (1.3) under m-dependence.

THEOREM 4.2. *Suppose $r = r_n \ge m$ and $x = x_n$ are such that $r\bar{F}(x) \to 0$ as $n \to \infty$. Then*

$$\text{(4.10)} \qquad P[M_{2,r} \le x | \xi_1 > x] = \theta(x) + \varepsilon(x) + o\big(r^2 \bar{F}(x)\big),$$

where $\theta(x)$ *is defined by* (4.1), *and*

$$\varepsilon(x) = \left(\bar{F}(x)\right)^{-1} \sum_{k=1}^{2m} (-1)^{k+1} \sum_{\substack{s+t=k \\ 1\le s\le m \\ 1\le t\le m}} \sum_{\substack{1=i_1<\cdots<i_s\le m \\ 1=j_1<\cdots<j_t\le m}} \times I(r-i_s-j_t\ge m) P\left[\bigcap_{1\le l\le s} (\xi_{i_l}>x)\right] P\left[\bigcap_{1\le l\le t} (\xi_{j_l}>x)\right].$$

Along any infinite subsequence $x = x_n$ *for which* $r \ge 3m$, $\varepsilon(x) = -\bar{F}(x)\theta^2(x)$.

PROOF. By stationarity and (4.4),

$$\begin{aligned} P[\xi_1 > x > M_{2,r}] &= P[M_{1,r} > x] - P[M_{1,r-1} > x] \\ &= A + B + o\left(r^2\bar{F}^2(x)\right), \end{aligned} \tag{4.11}$$

where

$$A = \sum_{k=1}^{m} (-1)^{k+1}\left(\sum_{B^{(1)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right] - \sum_{B^{(1)}_{k,r-1}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right]\right)$$

and

$$B = \sum_{k=1}^{2m} (-1)^{k+1}\left(\sum_{B^{(2)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right] - \sum_{B^{(2)}_{k,r-1}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right]\right).$$

By (4.7),

$$\begin{aligned} \sum_{B^{(1)}_{k,r}} P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right] &= \sum_{1=i_1<\cdots<i_k\le m} (r-i_k+1) P\left[\bigcap_{1\le j\le k} (\xi_{i_j}>x)\right] \\ &= r\alpha_k(x) - \beta_k(x), \end{aligned}$$

for $1 \le k \le m$ if $r \ge m$, and the identity also holds for $1 \le k \le m-1$ if $r = m-1$. Thus, for $r \ge m+1$,

$$A = \sum_{k=1}^{m} (-1)^{k+1}\alpha_k(x), \tag{4.12}$$

and for $r = m$,

$$\begin{aligned} A &= m\sum_{k=1}^{m} (-1)^{k+1}\alpha_k(x) + \sum_{k=1}^{m} (-1)^{k}\beta_k(x) \\ &\quad -(m-1)\sum_{k=1}^{m-1} (-1)^{k+1}\alpha_k(x) - \sum_{k=1}^{m-1} (-1)^{k}\beta_k(x) \\ &= \sum_{k=1}^{m} (-1)^{k+1}\alpha_k(x), \end{aligned} \tag{4.13}$$

since $(m-1)\alpha_m(x) = \beta_m(x)$. By (4.6), whether $r \geq m+1$ or $r = m$,

$$\sum_{B_{k,r}^{(2)}} P\left[\bigcap_{1\leq j\leq k} (\xi_{i_j} > x)\right] - \sum_{B_{k,r-1}^{(2)}} P\left[\bigcap_{1\leq j\leq k} (\xi_{i_j} > x)\right]$$

$$(4.14) \qquad = \sum_{\substack{s+t=k\\ 1\leq s\leq m\\ 1\leq t\leq m}} \sum_{\substack{1=i_1<\cdots<i_s\leq m\\ 1=j_1<\cdots<j_t\leq m}} I(r-i_s-j_t\geq m)$$

$$\times P\left[\bigcap_{1\leq l\leq s} (\xi_{i_l} > x)\right] P\left[\bigcap_{1\leq l\leq t} (\xi_{j_l} > x)\right],$$

and thus $B = \overline{F}(x)\varepsilon(x)$, where $\varepsilon(x)$ is given by the statement of the lemma. Hence (4.10) follows from (4.11) through (4.14). If $r \geq 3m$, then the indicator in (4.14) is equal to 1, and it is easily seen that $\varepsilon(x) = -\overline{F}(x)\theta^2(x)$. This completes the proof. □

By Theorem 4.2, if the extremal index θ exists and $r \geq 3m$, then

$$P[M_{2,r} \leq x|\xi_1 > x] - \theta = (\theta(x) - \theta) - \theta^2\overline{F}(x)\big(1 + o(\overline{F}(x))\big).$$

The expression on the right, which is denoted by $R(\overline{F}(x))$ in (2.1), can be regarded as the sum of two components. The first component $\theta - \theta(x)$ depends on the local dependence structure, whereas the (fixed) order of magnitude of the second component $-\overline{F}(x)\theta^2$ merely reflects short-range dependence. Thus the value of β in (2.13) is most likely to be less than or equal to 1. But it is theoretically possible to have

$$(\theta(x) - \theta) - \theta^2\overline{F}(x) = o(\overline{F}(x)),$$

however unlikely this may be. For instance, in Example A of Section 2, for fixed ρ and $\alpha = \beta$, if η, γ are constrained by

$$\eta(1 + \rho^\alpha(1 + \rho^\alpha)) = \gamma(\rho^\alpha - 1),$$

then $R(p) = o(p)$. When this occurs, the optimal rate of convergence of $\hat{\theta}_n(\overline{F}_n^{-1}(p_n))$ is faster than $n^{-1/3}$.

APPENDIX

We now derive (2.2) in Example C of Section 2. Let G be the standard Cauchy c.d.f., and use the notation of Example C. By stationarity,

$$P[\xi_1 > x \geq \xi_2] = P[\xi_1 \leq x < \xi_2] = P[\xi_1 \leq x, \rho\xi_1 + Z_2 > x]$$

$$= \int_{-\infty}^{x} \overline{G}(x - \rho z)\, dF(z).$$

Let

$$A(x) = P[\xi_1 > x \geq \xi_2] - \frac{1}{\pi}\int_{-\infty}^{x}\left(\frac{1}{x-\rho z} - \frac{1}{3(x-\rho z)^3}\right) dF(z).$$

Also write

$$\int_{-\infty}^{x} \frac{1}{\pi}\left(\frac{1}{x-\rho z} - \frac{1}{3(x-\rho z)^3}\right) dF(z)$$
$$= \frac{1}{\pi}\int_{-\infty}^{-x}\frac{dF(z)}{x-\rho z} + \frac{1}{\pi}\int_{-x}^{x}\frac{dF(z)}{x-\rho z} - \frac{1}{3\pi}\int_{-\infty}^{x}\frac{dF(z)}{(x-\rho z)^3}$$
$$=: B(x) + C(x) + D(x).$$

By the expansion

$$\bar{G}(u) = \frac{1}{\pi}\left(\frac{1}{u} - \frac{1}{3u^3} + \frac{1}{5u^5} - \cdots\right) \quad \text{as } u \to \infty, \tag{A.1}$$

we get

$$A(x) + D(x) = O(\bar{F}^3(x)).$$

Letting $y = z/x$, by (A.1) and dominated convergence,

$$B(x) = \frac{1-\rho}{\pi^2}\int_{-\infty}^{-1}\frac{dy}{(1-\rho y)(1+(1-\rho)x^2y^2)}$$
$$\sim \frac{1-\rho}{\pi^2 x^2}\int_{-\infty}^{-1}\frac{dy}{(1-\rho y)y^2} \sim (1-\rho)^3\int_{-\infty}^{-1}\frac{dy}{(1-\rho y)y^2}\bar{F}^2(x).$$

Next write

$$C(x) = C_1(x) + C_2(x),$$

where

$$C_1(x) := \frac{1}{\pi x}\int_{-x}^{x} dF(z) = \frac{1}{\pi x}(1 - 2\bar{F}(x))$$
$$= (1-\rho)\bar{F}(x) - 2(1-\rho)\bar{F}^2(x)(1+o(1))$$

and

$$C_2(x) := \frac{\rho}{\pi x}\int_{-x}^{x}\frac{z}{x-\rho z}\, dF(z) = \frac{2\rho^2}{\pi x^2}x\int_0^x \frac{z^2}{x^2-\rho^2 z^2}\, dF(z).$$

By Karamata's theorem [cf. Feller (1971)],

$$\lim_{a\downarrow 0}\lim_{x\to\infty} x\int_0^{ax}\frac{z^2}{x^2-\rho^2z^2}\, dF(z) = 0.$$

Thus

$$\lim_{x\to\infty} x\int_0^x \frac{z^2}{x^2-\rho^2z^2}\, dF(z)$$
$$= \lim_{a\downarrow 0}\lim_{x\to\infty} x\int_{ax}^{x}\frac{z^2}{x^2-\rho^2z^2}\, dF(z)$$
$$= \lim_{a\downarrow 0}\lim_{y\to\infty}\int_a^1 \frac{y^2}{1-\rho^2y^2}\frac{(1-\rho)x^2\, dy}{\pi\left[1+(1-\rho)^2x^2y^2\right]}$$
$$= \lim_{a\downarrow 0}\frac{1}{\pi(1-\rho)}\int_a^1\frac{dy}{1-\rho^2y^2} = \frac{1}{\pi(1-\rho)}\int_0^1\frac{dy}{1-\rho^2y^2}.$$

Thus

$$C_2(x) \sim \frac{2\rho^2}{\pi x^2}\frac{1}{\pi(1-\rho)}\int_0^1 \frac{dy}{1-\rho^2y^2}$$
$$\sim 2\rho^2(1-\rho)\int_0^1 \frac{dy}{1-\rho^2y^2}\bar{F}^2(x).$$

Summarizing,

$$\frac{P[\xi_1 > x \geq \xi_2]}{\bar{F}(x)} = 1-\rho + c\bar{F}(x)(1+o(1)),$$

where

$$c = -2(1-\rho) + (1-\rho)^3\int_{-\infty}^{-1}\frac{dy}{(1-\rho y)y^2} + 2\rho^2(1-\rho)\int_0^1 \frac{dy}{1-\rho^2y^2}.$$

This proves (2.2).

Acknowledgments. The comments of the referees and an Associate Editor resulted in a significant improvement over a previous version of the paper. The author gratefully acknowledges their assistance.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

CHERNICK, M. (1981). A limit theorem for the maximum term of autoregressive processes with uniform marginal distributions. *Ann. Probab.* **9** 145–149.

CHERNICK, M. R., HSING, T. and MCCORMICK, W. (1991). Calculating the extremal index of a class of stationary sequences. *Adv. in Appl. Probab.* **23** 835–850.

DAVIS, R. A. and RESNICK, S. I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195.

DE HAAN, L., RESNICK, S. I., ROOTZÉN, H. and DE VRIES, C. G. (1989). Extremal behavior of solutions to a stochastic differential equation with applications to ARCH processes. *Stochastic Process. Appl.* **32** 213–224.

DEKKERS, A. L. M. and DE HAAN, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795–1832.

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855.

DENZEL, G. E. and O'BRIEN, G. L. (1975). Limit theorems for extreme values of chain dependent processes. *Ann. Probab.* **3** 773–779.

DEO, C. M. (1973). A note on empirical processes for strong mixing sequences. *Ann. Probab.* **1** 870–875.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Univ. California Press, Berkeley.

GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press.

HALL, P. and WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331–341.

HSING, T. (1989). On a loss of memory property of the maximum. *Statist. Probab. Lett.* **8** 493–496.

HSING, T. (1990). Estimating the extremal index under m-dependence and tail balancing. Technical Report, Dept. Statistics, Texas A&M Univ.

HSING, T. (1991a). Estimating the parameters of rare events. *Stochastic Process. Appl.* **37** 117–139.

HSING, T. (1991b). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569.

HSING, T. (1993). On some estimates based on sample behavior near high level excursions. *Probab. Theory Related Fields* **95** 331–356.

HSING, T., HÜSLER, J. and LEADBETTER, M. R. (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78** 97–112.

IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382.

JOE, H. (1991). Estimation of the extremal index function and the distribution of the maximum of a stationary dependent sequence. Preprint.

LEADBETTER, M. R. (1974). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **28** 289–303.

LEADBETTER, M. R. (1983). Extremes and local dependence in a stationary sequence. *Z. Wahrsch. Verw. Gebiete* **65** 291–306.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.

LEADBETTER, M. R. and NANDAGOPOLAN, L. (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. In *Extreme Value Theory: Proceedings, Oberwolfach 1987. Lecture Notes in Statist.* **51** 69–80. Springer, Berlin.

LEADBETTER, M. R., WEISSMAN, I., DE HAAN, L. and ROOTZÉN, H. (1989). On clustering of high values in statistically stationary series. Technical Report 253, Center for Stochastic Processes, Dept. Statistics, Univ. North Carolina.

LOYNES, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36** 993–999.

NANDAGOPOLAN, L. (1990). Multivariate extreme value theory and extremal index estimation. Ph.D. dissertation, Dept. Statistics, Univ. North Carolina, Chapel Hill.

O'BRIEN, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.* **15** 281–289.

PICKANDS, J., III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131.

ROOTZÉN, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. in Appl. Probab.* **20** 371–390.

SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207.

SMITH, R. L. (1989). Extreme value analysis of environmental time series. *Statist. Sci.* **4** 367–393.

SMITH, R. L. and WEISSMAN, I. (1993). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B*. To appear.

WITHERS, C. S. (1975). Convergence of empirical processes of mixing random variables on [0, 1]. *Ann. Statist.* **3** 1101–1108.

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143