CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

Extreme value theory and dependence
by
M.R. Leadbetter

Technical Report No. 109
July 1985
The purpose of this paper is to give a very brief account of some of the essential ideas underlying classical extreme value theory, and to see how these are used (modified as necessary) for dependent cases. In particular it will be shown how the classical theory still applies for moderately dependent stationary sequences, but that under higher local dependence, clustering of high values occurs, requiring modifications of the theory especially as it involves order statistics other than the maximum. Underlying concepts (especially point process convergence results) are emphasized.
EXTREME VALUE THEORY AND DEPENDENCE

by M.R. Leadbetter

Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

Abstract. The purpose of this paper is to give a very brief account of some of the essential ideas underlying classical extreme value theory, and to see how these are used (modified as necessary) for dependent cases. In particular it will be shown how the classical theory still applies for "moderately dependent" stationary sequences, but that under higher local dependence, clustering of high values occurs, requiring modifications of the theory especially as it involves order statistics other than the maximum. Underlying concepts (especially point process convergence results) are emphasized.

1. Notation, and the Classical Theory.

Throughout we write \( M_n = \max(\xi_1, \xi_2, \ldots, \xi_n) \) for any sequence \((\xi_n)\) of random variables. If the \( \xi_i \) are i.i.d. with marginal d.f. \( F \) then \( M_n \) has the d.f. \( P(M_n < x) = F^n(x) \) so that the distribution of \( M_n \) is precisely known if \( F \) is known. However in practice \( F \) is not necessarily known precisely, and approximations less dependent on the exact form of \( F \) are useful. In this vein a central contribution of the classical theory is the following result which restricts the "type" of limiting d.f. which \( M_n \) may have (under linear normalizations):

Theorem 1.1 (Extremal Types Theorem). Let \( M_n = \max(\xi_1, \xi_2, \ldots, \xi_n) \), where \( \xi_i \) are i.i.d. random variables. If for some constants \( a_n > 0, b_n \), we have

\[
P(a_n(M_n - b_n) < x) \downarrow G(x)
\]

for some non-degenerate d.f. \( G \), then \( G \) is one of the "three extreme value types":

Type I: \( G(x) = \exp(-e^{-x}), -\infty < x < \infty; \)

Type II: \( G(x) = \begin{cases} a & \text{for some } a > 0, \\ \exp(-x), & \text{for some } x < 0 \end{cases} \)

(family) according as \( x < 0 \) or \( x > 0 \)

Type III: \( G(x) = \begin{cases} a & \text{for some } a > 0, \\ \exp(-(-x)^a), & \text{for some } x < 0 \end{cases} \)

(family) according as \( x < 0 \) or \( x > 0 \)

In these d.f.'s \( x \) may be replaced by \( ax + b \) for any \( a > 0, b \). This is consistent with the terminology we use that two d.f.'s \( G, H \) are of the same type if \( G(x) = H(ax + b) \) for some \( a > 0, b \).

The classical theory provides domain of attraction criteria determining the type of limit in terms of the general behaviour of the tail \( 1-F \) of the d.f. \( F \) (see [5]). Thus Equation (1.1) may often be used to provide an approximation for the distribution of \( M_n \) when \( n \) is large. Obviously relevant questions of rates of convergence and estimation of the normalizing constants in practice are not part of this paper (see [5] and references therein for discussions of these topics).

While no detailed proof of Theorem 1.1 can be given here, it is useful to see the principal ideas involved. These can be grouped into the following two packages:

Package I: The class of d.f.'s \( G \) which occur as limits in Theorem 1.1 are precisely the max stable d.f.'s i.e. d.f.'s \( G \) such that \( G^k \) is of the same type as \( G \) \( (G^k(x) = G(a_kx + b_k) \) for some \( a_k > 0, b_k \), for each \( k = 1, 2, \ldots).\)

Package II: A d.f. \( G \) is max stable if and only if it is one of the three extreme value types.

The result of Package I follows trivially for the i.i.d. sequence considered by noting that
Lemma 1.2 Let \( \{u_n\} \) be any real sequence, and 
\[ 0 \leq \tau \leq \infty. \]
Then (1.3) and (1.4) are equivalent for the i.i.d. sequence \( \{\xi_n\} \) with d.f. \( F \).

We shall use the notation \( u_n(\tau) \) to denote any sequence satisfying (1.4). It is of interest (and makes the notation more compelling) to note that if \( u_n(\tau) \) exists for one \( \tau \) with 
\[ 0 < \tau < \infty, \]
then it exists for all such \( \tau \) and this happens if and only if 
\[ (1 - F(x))/(1 - F(x_\tau)) \equiv 1 \]
as \( x \rightarrow x_\tau \) (see [5, Theorem 1.7.13]), where

\( x_\tau \) is the upper endpoint of the d.f. \( F \).

Lemma 1.2 is central in the consideration of domains of attraction. But it is also basic in showing how the asymptotic distributions of extreme order statistics are determined by that for the maximum. To see this let \( S_n \) denote the number of exceedances of the level \( u_n \) by \( \xi_1, \ldots, \xi_n \), i.e. the number of \( i \) such that 
\[ \xi_i > u_n, \quad 1 \leq i \leq n. \]
\( S_n \) is clearly a binomial r.v. with parameters \( (n, p = 1 - F(u_n)) \) so that if (1.4) holds, \( n p_n \tau \) and \( S_n \) has a Poisson limit with mean \( \tau \), i.e. 
\[ P(S_n = r) = e^{-\tau_\nu}/r! \]
in particular if (1.1) holds (so that \( M_n \) has the asymptotic distribution \( G \)) then the identification \( u_n = x/a_n + b_n \) \( \tau = -\log G(x) \) and Lemma 1.2 show that 
\[ (1.5) \quad P(a_n(M_n - b_n) \leq x) = G(x) \sum_{r=0}^{k-1} (-\log G(x))^{r}/r! \]
Thus the asymptotic distribution for \( M_n \) determines that for each \( M_n(k) \) (using the same normalizing constants as \( M_n \)).

The Poisson property of exceedances may be pursued further with advantage. Specifically let \( u_n = u_n(\tau) \) satisfy (1.4) for some \( \tau > 0 \).

Write \( N_n \) for the point process on the interval \([0,1]\) having an atom (event) at each point \( i/n \) for which \( \xi_i > u_n \). That is a time scale change by \( 1/n \) is made, and converts an exceedance at \( i(\xi_i > u_n) \) to a point at \( i/n \) in \([0,1]\). \( N_n \) may be called the exceedance point process for the level \( u_n \). If \( N_n(B) \) denotes the number of atoms of \( N_n \) in the subset \( B \) of \([0,1]\), and (1.4) holds, it is readily shown that if \( B \) is an interval then \( N_n(B) \) is asymptotically Poisson with mean \( \tau m(B) \) where \( m(B) \) is the length of \( B \). Further \( N_n(B_1) \) and \( N_n(B_2) \) are clearly independent when \( B_1 \) and \( B_2 \) are disjoint. This suggests that the
point process \( N_n \) is taking on a Poisson character as \( n \) increases and indeed it may be shown by such consideration that if (1.4) holds then full weak convergence holds,

\[
(1.6) \quad N_n \xrightarrow{d} \mathcal{N} \quad \text{as} \quad n \to \infty
\]

where \( \mathcal{N} \) is a Poisson Process on \([0,1]\) with intensity \( \tau \).

A harvest of corollaries may be reaped from this and related results involving more than one level. In particular the asymptotic joint distribution of any group of extreme order statistics and their locations may be obtained. In the i.i.d. case these results may be also obtained directly but the point process approach is illuminating, and most useful in dependent cases to be considered next.

2. Stationary Sequences and the Extremal Types Theorem.

Turning now to dependent situations we consider the case of a \( \{\xi_n\} \) which is strictly stationary in the standard sense that its finite dimensional distributions \( \mathcal{F}_{i_1} \ldots \mathcal{F}_{i_p} \) where

\[
P\{\xi_i \leq x \ldots \xi_i \leq x_p \} \quad \text{are such that} \quad \mathcal{F}_{i_1 \ldots i_p} \quad \text{for any choice of} \quad p, \ i_1 \ldots i_p,
\]

Other cases (e.g. Markov-cf. [7], [8]) may be treated, but stationary sequences are adequate to illustrate the effects of dependence.

As far as the Extremal Types Theorem is concerned, Loynes ([6]) took the first and most significant step away from independence by showing that the result remains true under strong mixing assumptions. That is if \( \{\xi_n\} \) is stationary and strongly mixing and (1.1) holds for some non-degenerate \( G \), then \( G \) must be of extreme value type. It is obvious from Loynes' proof that the full force of strong mixing is not required and weaker assumptions will suffice. The following condition "\( D(u_n) \)" is defined with respect to a given sequence \( \{u_n\} \) and is convenient and useful. (It is clearly possible to weaken \( D(u_n) \) very slightly for the present purpose - as has been made explicit in [7], but \( D(u_n) \) is useful in other contexts also. In the following statement \( \mathcal{F}_{i_1 \ldots i_p} \) is written for \( \mathcal{F}_{i_1 \ldots i_p} \).

**Definition.** If \( \{u_n\} \) is any real sequence, the condition \( D(u_n) \) is said to hold if for any

\[
l \leq i_1 \ldots \leq i \quad j_1 < j_2 \ldots < j_p \quad n, \quad j_i - i > \epsilon
\]

we have

\[
\left| \mathcal{F}_{i_1 \ldots i_p} \mathcal{F}_{j_1 \ldots j_p} (u_n) \right| \leq \alpha_n, \quad \epsilon
\]

\( \alpha_n, \epsilon \) are such that

\[
\sup_n \mathcal{F}_{i_1 \ldots i_p} (u_n) < \infty
\]

The Extremal Types Theorem then says that if \( \{\xi_n\} \) is strictly stationary and satisfies (1.1) and if \( D(u_n) \) holds for all sequences of the form \( u_n = x/a_n + b_n \) (\( -\infty < x < \infty \)) then \( G \) is again of extreme value type. The method of proof is simply to show that such a \( G \) must be max stable and hence of extreme value type as in the i.i.d. case. The proof of max stability rests on the following basic result:

**Lemma 2.1** If \( D(u_n) \) holds for the stationary sequence \( \{\xi_n\} \) (and a given real sequence \( \{u_n\} \)) then

\[
P\{M_n < u_n \} - P^k[M_{n/k} < u_n] \to 0 \quad \text{for any} \quad k = 1, 2, \ldots
\]

It follows simply from (1.1) by using this lemma (with \( n/k \) replacing \( n \) and identifying \( u_n \) with \( x/a_n + b_n \)) that (1.2) holds and hence \( G \) is max stable exactly as in the i.i.d. case. The proof of the lemma is achieved by a standard type of argument used to "reduce dependent to independent cases (cf. [6] and earlier papers in dependent central limit theory). Very roughly in this case the integers 1...n are divided into \( k \) consecutive groups and approximate independence of the maxima on each is used. This approximate independence is established via \( D(u_n) \) by "snipping" an expanding but relatively
small piece from each group to give the separation required for $D(u_n)$, (cf. [5, Section 3.2] for details).

3. The Effect of Dependence on the Asymptotic Distribution of $M_n$

Again let $\{\xi_j\}$ be stationary with marginal d.f. $F$. Following Loynes, define the associated independent sequence to be an i.i.d. sequence $\{\xi_n\}$ with the same d.f. $F$. Write $\tilde{M}_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$. It is natural to ask whether $M_n$ has an asymptotic distribution if $\tilde{M}_n$ does, and conversely, and, if both do, whether they are of the same extremal type or related in some specific way. This matter can be resolved in unexpectedly explicit ways for most stationary sequences, in terms of a single parameter.

As before let $u(\tau)$ satisfy (1.4) for each $\tau > 0$ so that $P(M_n < u(\tau)) + e^{-\tau}$ by Lemma 1.2. Now Loynes ([6]) showed that for strongly mixing sequences, if $\lim P(M_n < u(\tau))$ exists for all $\tau$ it must have the form $e^{-\theta \tau}$ for some $\theta, 0 < \theta < 1$. This may also be shown under $D(u_n)$ - assumptions and indeed if the limit exists for one $\tau$ it exists for all $\tau$, and is $e^{-\theta \tau}$ for some $\theta$. This $\theta$ may be called the extremal index of the process and exists under wide conditions. For i.i.d. (and many stationary) sequences $\theta = 1$, but all values of $\theta$ in $[0,1]$ are possible, though $\theta = 0$ is rather pathological. $\theta = 1$ for a stationary normal process satisfying the covariance condition $\tau_n \log n \to 0$. An example of a case with $0 < \theta < 1$ is the following.

Example 3.1 Let $\eta_1, \eta_2, \ldots$ be i.i.d. with d.f. $H$ and write $\xi_j = \max(\eta_j, \eta_{j+1})$. Then $\{\xi_n\}$ is stationary with d.f. $F = H^2$ and easy calculation shows that if $u_n(\tau)$ satisfies (1.4) then

$$n [1 - H(u_n(\tau))] + \tau/2$$

and

$$P[M_n < u_n(\tau)] = P[\max(\eta_1, \ldots, \eta_n) < u_n(\tau)] P(n + 1 < u_n(\tau)) + e^{-\tau/2}$$

so that $\{\xi_n\}$ has extremal index $\theta = 1/2$.

Modifications of this example yield other values of $\theta$.

The following result from [4] illustrates the use of the index $\theta$.

**Theorem 3.2** Suppose that the stationary sequence $\{\xi_n\}$ has extremal index $\theta > 0$. Let $\{v_n\}$ be any constants and $0 < \rho < 1$. Then $P(M_n < v_n) + \rho$ if and only if $P(M_n < v_n) + \rho^n$.

This clearly exhibits the relations between $M_n$ and $\tilde{M}_n$. For example if there is an "i.i.d. limit" $G$ $P(a_n(M_n - b_n) < x) \to G(x)$, then we can write $v_n = x/\tilde{M}_n$ and obtain $P(a_n(M_n - b_n) < x) \to G(\theta)(x)$. Thus the limiting d.f. for $M_n$ is just $G(\theta)$. This is of the same type as $G(\theta) = G(ax + b)$ as is easily verified for each extremal distribution. The converse also holds so that if $\theta > 0, M_n$ has an asymptotic distribution if and only if $\tilde{M}_n$ does, they have the same normalizing constants, and the limits are of the same type (indeed powers of each other). If $\theta = 1$ the limits are identical (see [4] for a complete discussion).

4. Poisson Results and Extreme Order Statistics when $\theta = 1$.

When $\theta = 1$, the Poisson results for exceedances go through as in the i.i.d. case. In particular, the number $S_n$ of exceedances of $u_n$ satisfying (1.4) by $\xi_1, \ldots, \xi_n$ is asymptotically Poisson, leading to the same asymptotic distribution (1.5) for $M_n(\theta)$ as in the i.i.d. case. This Poisson limit is best shown as a corollary of Poisson convergence of the exceedance point process $N_n$ defined on $[0,1]$ as in Section 1, with points at those $i/n$ for which $\xi_i = u_n$.

Specifically if $u_i = u_i(\tau)$ satisfies (1.4), $P(M_i < u_i) + e^{-\tau}$ and hence $P(M_i < u_i) + e^{-\tau}$ by Theorem 3.2 ($\theta = 1$). But the events $\{M_i < u_i\}$, $N_n([0,1]) = 0$ are clearly equivalent so that $P(N_n([0,1]) = 0) = e^{-\tau}$. The same argument in fact shows easily that $P(N_n(B) = 0) = e^{-m(B)}$ where $m$ is Lebesgue measure, for any finite union $B$ of subintervals of $[0,1]$, so that $P(N_n(B) = 0) + P(N(B) = 0)$ for such sets $B$, where $N$ is a Poisson Process on $[0,1]$ with intensity...
$\tau$. This and the easily proved fact that $E N_n^d(B) = \tau m(B) = E N(B)$ are sufficient to show that $N_n^d \Rightarrow N$ in the full weak convergence sense, by a theorem of Kallenberg [3, Theorem 4.7].

In particular $S_n = N_n((0,1])$ is asymptotically Poisson with mean $\tau$. If $M_n$ has the asymptotic distribution $G$ as in (1.1) it follows readily by the usual identification $u_n = x/a_n + b_n$, $\tau = -\log G(x)$ that the $k$th largest values $M_n(k)$ again have the classical forms (1.5). Indeed, it may be shown by considering exceedances of more than one level that joint distributions of the $M_n(k)$ (and their locations) have their i.i.d. forms when $\theta = 1$.

5. Clustering and its effect on $M_n(k)$ when $0 < \theta < 1$.

The above discussion shows that when $\theta = 1$ the classical results hold without change. The less usual - but still non-pathological - case is where $0 < \theta < 1$, which we now consider. In this case, which typically involves higher local dependence of the sequence terms, clustering of exceedances may occur. For instance, it is readily seen that in Example 3.1, exceedances of $u_n(t)$ by $\xi_i$ occur in (at least) pairs. Correspondingly it is found that the limiting point process has double points, occurring at positions which form a Poisson process with intensity $\tau/2$.

In general when $0 < \theta < 1$ exceedances tend to occur in groups or clusters of stochastic size and the limiting point process is a "Compound Poisson" consisting of multiple events of random size occurring at points of $[0,1]$ which form a Poisson process with intensity $\tau/2$.

As noted already the only effect on $M_n$ taking $(0 <) \theta < 1$ is to replace the "i.i.d. limit $G$" by $G^\theta$. However, the distribution of other order statistics $M_n(k)$ are more radically altered. This may be understood intuitively by noting that the maximum $M_n$ is simply the maximum of the greatest values in each cluster. However e.g. $M_n(2)$ may be the second largest value in the cluster when $M_n$ occurs, or the largest in some other cluster, so that cluster structure becomes important.

In fact if the cluster size in the limiting compound Poisson process has distribution $\pi(i)$, $i = 1,2,...$ and if $M_n$ has the asymptotic distribution $G$ as in (1.1), then the asymptotic distribution for the $k$th largest $M_n(k)$ is

$$P(a_n(M_n(k) - b_n) \leq x) = G(x)[1 + \sum_{j=1}^{k-1} \sum_{i=j}^{k-1} (-\log G(x))^{j} \pi^j(i)]$$

where $\pi^j(i)$ is the $j$-fold convolution of the cluster size distribution $\pi(i)$. (These latter may be written as limits of probabilities associated with the $\xi_i$'s).

For completeness we state the main theorem (due to Hsing) on which this is based. This requires a modest strengthening of the $D(u_n)$ condition to a form $\Delta(u_n(t))$ which will not be specified here (but is still of similar type, and widely applicable - see [1] or [2] for the details and proof of the theorem).

**Theorem 5.1** Suppose $\tau > 0$ is constant and $\Delta(u_n(t))$ holds. If the exceedance point process $N_n$ converges in distribution to some point process $N$, the latter must be a Poisson process with Laplace transform

$$\xi e^{-\tau \int_0^t [1 - L(f(t))] dt}$$

where $L$ is the Laplace Transform of some probability distribution $\pi$ on $(1,2,...)$ and $0 < \theta < 1$.

Throughout this paper we have considered point processes of exceedance of one level. However as noted, joint distributions of order statistics may be discussed by consideration of "vector valued" point processes involving exceedances of e.g. $k$ levels $u_n(\tau_1) ... u_n(\tau)$. Alternatively a so called "complete convergence result" may be used under slightly more restrictive conditions) to obtain these results. Such results typically involve point processes
in the plane consisting of points at \((j/n, u_n^{-1}(c_j))\) where \(u_n^{-1}\) is the inverse function of \(u_n(T)\). In simple cases a Poisson limit holds (in the plane) leading to all the joint asymptotic distributions of order statistics (cf. [5, Sec. 5.7]). In cases with more local dependence \(0 < 1\) the limit involves a point process with clustering which can be explicitly defined cf.([1]) in a manner analogous to Theorem 5.1.


