ON EXTREME VALUES OF SAMPLED AND
CONTINUOUS STOCHASTIC DATA

BY

M. R. LEADBETTER

TECHNICAL REPORT NO. 10
AUGUST 1, 1977

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH

PREPARED UNDER SUPPORT TO SIMS FROM
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION (ERDA)
ROCKEFELLER FOUNDATION
SLOAN FOUNDATION
ENVIRONMENTAL PROTECTION AGENCY (EPA)
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED ^{EB}

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ON EXTREME VALUES OF SAMPLED AND
CONTINUOUS STOCHASTIC DATA

M. R. Leadbetter
University of North Carolina
and
Stanford University

I. INTRODUCTION

The maximum concentration of a given material in a specific time period is of clear interest in the study of problems of pollution. In particular, one would like to obtain the statistical distribution of such a maximum. There is a great deal of useful theory available in the "classical" case - where the concentrations are measured at discrete time points, and may be regarded as independent and identically distributed (i.i.d.) random variables (cf. [2]). In particular, rather complete answers are known for the following two questions:

- (a) What statistical laws are possible for long-term maxima?
- (b) Which of the possible such laws applies in a given situation when specific assumptions are made concerning the statistical nature of the individual observations?

As regards (a), it is known that essentially only three different laws are possible - these being the three so-called "extreme value distributions" or "extreme value types", having the specific distribution function (d.f.) forms given in the next section. In answer to (b), classical extreme value theory provides detailed conditions on

the statistical distribution of each observation, to determine which of the three extreme value forms arises in a given situation.

A more detailed, though brief, statement of these i.i.d. results is given in the next section.

While this classical theory is indeed most useful, it is clear that the necessary independence assumptions will not be universally satisfied in all cases of interest - especially if measurements are made with a reasonably short sampling interval. It turns out that much of the classical theory still applies provided the dependence is not "too strong", and indeed that for certain cases of stronger dependence alternative results are known ([3], [7]). This is discussed in Section III.

The main purpose of this paper is to explore the corresponding situation for maxima over continuous intervals of time, rather than for sampled data. The main (new) results of Section IV show that again under conditions of "not too strong dependence", some important conclusions of the classical theory still apply (along the same lines as for the case of dependent sequences described in Section III). In this case there are, however, some important differences in deciding which extreme value distribution applies. We shall describe (and attempt to motivate) these results in Sections IV - VI - a full and detailed theory will be found in [4].

The practical implication of these results lies in the confidence they give to the use of the extreme value distributions, even in a continuous context. Of course, there are cautions to be observed (as well as unsolved problems) - these will be indicated in the later discussion.

II. CLASSICAL EXTREME VALUE THEORY - TWO RESULTS

In classical Extreme Value Theory (EVT) one is concerned with the maximum M_n of n i.i.d. random variables $X_1, \dots, X_n$, as n becomes large. The central result - a celebrated theorem of Gnedenko - states that if there are normalizing constants $a_n > 0, b_n$ such that $a_n(M_n - b_n)$ has a non-degenerate limiting distribution (function G , then G must be one of the three following general types:

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

Type II: $G(x) = \exp(-x^\alpha) \quad \alpha > 0, x > 0 \quad (0 \text{ for } x < 0)$

Type III: $G(x) = \exp(-x)^\alpha \quad \alpha > 0, x \leq 0 \quad (1 \text{ for } x > 0)$

(In these, x may be replaced by $ax+b$ for any $a > 0, b$).

Further, classical EVT provides necessary and sufficient conditions on the common d.f. F of the individual observations X_i to determine which (if any) limit law holds.

It is not always possible to find normalizing constants a_n, b_n to obtain a non-degenerate limiting distribution for $a_n(M_n - b_n)$, i.e. such that $P\{M_n \leq x/a_n + b_n\}$ converges for each x . However, even for such cases it may be possible to find constants $u_n = u_n(x)$ such that $P\{M_n \leq u_n\}$ converges for each x but $u_n(x)$ is not restricted to have the form $x/a_n + b_n$. In fact, the following result is proved almost trivially in the i.i.d. case.

If u_n is such that

$$(2.1) \quad n(1 - F(u_n)) \rightarrow \tau > 0$$

then

$$(2.2) \quad P\{M_n \leq u_n\} \rightarrow e^{-\tau}$$

and conversely. Hence if $G(x)$ is a d.f. and if $u_n(x)$ may be chosen so that $n(1 - F(u_n(x))) \rightarrow \tau(x) = -\log G(x)$, then $P\{M_n \leq u_n(x)\} \rightarrow G(x)$, or $P\{u_n^{-1}(M_n) \leq x\} \rightarrow G(x)$ if $u_n(x)$ has an inverse function u_n^{-1} . Thus under quite general circumstances $u_n^{-1}(M_n)$ will have a limiting distribution. One may thus regard Gnedenko's Theorem as answering the further question, "Under what conditions may we take $u_n(x) = x/a_n + b_n$?"

In addition, the equivalence of (2.1) and (2.2) can be useful in determining which extreme value distribution applies in a particular case, and the form of the normalizing constants a_n, b_n . For if we can find u_n of the form $x/a_n + b_n$ to satisfy (2.1) with $\tau = \tau(x)$, then we have the constants, and $G(x) = \exp -\tau(x)$. In particular, this procedure is useful when the X_i are (standard) normal, leading to the values

$$a_n = (2 \log n)^{\frac{1}{2}}, \quad b_n = (2 \log n)^{\frac{1}{2}} - \frac{1}{2}(2 \log n)^{-\frac{1}{2}} [\log 4\pi + \log \log n]$$

and $G(x) = \exp(-e^{-x})$, i.e. Type I.

There are, of course, many other results of classical EVT, but we focus on the two described above.

III. DEPENDENT SEQUENCES

Much of classical EVT may be extended to include dependent (stationary) sequences provided the dependence is "not too strong". For example, Gnedenko's Theorem still applies to so-called strongly mixing stationary sequences ([6]). However, the assumption of strong mixing can be changed to a weaker form which may be called "distributional mixing", and is defined as follows

Let $F_{i_1, i_2, \dots, i_n}$ denote the joint d.f. of the members $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ taken from the sequence $X_1, X_2, \dots$ and write $F_{i_1, \dots, i_n}(u) = F_{i_1, \dots, i_n}(u, u, \dots, u)$. If $\{u_n\}$ is a real sequence, then the condition $D(u_n)$ will be said to hold if for any choice of integers $n, p, p', p+p' \leq n, i_1 < i_2 < \dots < i_p < j_1 < \dots < j_{p'}, j_1 - i_p \geq \ell$,

$$(3.1) \quad |F_{i_1, \dots, i_p, j_1, \dots, j_{p'}}(u_n) - F_{i_1, \dots, i_p}(u_n) F_{j_1, \dots, j_{p'}}(u_n)| < \alpha_{n, \ell}$$

where $\alpha_{n, \ell} \rightarrow 0$ for some sequence $\ell_n \rightarrow \infty, \ell_n = o(n)$. This restriction says that for certain (equal) values of its arguments the joint d.f. of two separated groups of the X_i factors approximately into the product of the two joint d.f.'s for each group. This is a rather weak form of "dependence decay" and leads to the following form of Gnedenko's Theorem ([3]).

Theorem 3.1. Let $\{X_n\}$ be a stationary sequence, $M_n = \max\{X_1, X_2, \dots, X_n\}$ and suppose that for some constant $a_n > 0, b_n$ we have

$$P\{a_n(M_n - b_n) \leq x\} \rightarrow G(x)$$

for some non-degenerate d.f. G . If $D(u_n)$ holds with $u_n = x/a_n + b_n$, for each x , then G is one of the three extreme value forms. #

For later reference we state the following lemma which is proved in [3], and which is basic to the proof of Theorem 3.1.

Lemma 3.2. Suppose that $D(u_n)$ holds for the stationary sequence $\{X_i\}$. Then, for any fixed $k = 1, 2, 3, \dots$, $P\{M_{nk} \leq u_{nk}\} - P^k\{M_n \leq u_{nk}\} \rightarrow 0$ as $n \rightarrow \infty$. #

In certain cases, when the X_i are normal, the $D(u_n)$ condition is relatively easy to verify, and it turns out to be a very weak condition

indeed in such cases, restricting the dependence in a rather mild way. Hence even though it may be difficult to check $D(u_n)$ precisely in other (non-normal) cases, this result does give considerable confidence for the applicability of the classical extreme value distributions for dependent sequences.

As pointed out in the last section, the equivalence of (2.1) and (2.2) is rather useful, and is almost trivial to show in the i.i.d. case. For dependent sequences it is equally useful, but requires a further condition for its validity. Specifically, we shall say that the (stationary) sequence X_i satisfies the condition $D'(u_n)$ (where again u_n is a given sequence) if

$$(2.3) \quad \limsup_{n \rightarrow \infty} n \sum_{j=2}^n P[X_1 > u_{nk}, X_j > u_{nk}] = o(1/k) \quad \text{as } k \rightarrow \infty.$$

The following result then holds.

Theorem 3.3. Suppose that $\{X_n\}$ is a stationary sequence, and that $D(u_n)$, $D'(u_n)$ hold for a sequence $\{u_n\}$ satisfying (2.1). Then (2.2) holds, i.e. $P\{M_n \leq u_n\} \rightarrow e^{-\tau}$.#

This theorem provides an important connection between the independent and dependent cases. In the following M_n will denote (as usual) the maximum of the first n members $X_1, X_2, \dots, X_n$ of the stationary sequence $\{X_i\}$. We shall write also $\hat{M}_n = \max(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n)$ where $\hat{X}_i$ are i.i.d. random variables with the same marginal d.f. F as the X_i . Then the following corollary to Theorem 3.3 is easily shown.

Corollary. Suppose that $D(u_n)$, $D'(u_n)$ hold for the stationary sequence $\{X_n\}$ and that $P\{\hat{M}_n \leq u_n\} \rightarrow \rho > 0$. Then also $P\{M_n \leq u_n\} \rightarrow \rho$.

In particular, if for some constants $a_n > 0$, b_n

$$P\{a_n(\hat{M}_n - b_n) \leq x\} \rightarrow G(x)$$

for some non-degenerate G and if $D(u_n)$, $D'(u_n)$ hold with $u_n = x/a_n + b_n$ for each x , then

$$P\{a_n(M_n - b_n) \leq x\} \rightarrow G(x) . \#$$

This corollary shows that (under D , D' conditions) the maximum of the stationary sequence $\{X_n\}$ has the same asymptotic distribution as it would if the individual random variables were independent, with the same marginal d.f. In particular, the same criteria may be used as for independent random variables to determine which extreme value distribution applies in a given case.

As for $D(u_n)$, $D'(u_n)$ is quite easily checked for normal sequences, and its wide validity there again suggests that commonly the same extremal distribution will apply in a stationary case as under independence, for non-normal situations - even though D' may then be harder to verify. It should be pointed out that there are normal cases where these conditions are not satisfied (involving a higher degree of dependence). Some such situations are described in [7].

IV. CONTINUOUS PARAMETER PROCESSES (GENERAL COMMENTS)

The main object of this paper is to compare the situation for extrema of processes which occur continuously in time with that for sequences described in the previous section. The results of this section are new and, as already indicated, will be treated in greater detail in [4].

Clearly the maximum of a process over a continuous period of time is at least as large as the maximum of its sampled values. It is also clearly important to know whether this can be significantly larger, i.e. whether measurement only at sampled time points can lead to a distribution quite different from that which really holds for the continuous maximum.

One may argue intuitively that a sufficiently fine sampling scheme should guarantee approximate equality of the continuous and sampled maxima. While this may be essentially true in many practical cases, it may also be false in others - especially if we wish to apply the asymptotic theory. For then so many samples may be necessary to ensure that the asymptotic theory applies, that the process has very many (however small) intervals between samples in which excursions above sampled values are possible. Even though high excursions in any one interval may be very rare, the large number of these intervals can lead to significant differences for the extrema in the continuous and sampled cases. In fact, it may even be the case that if one's data is such that there is clearly no difference between the continuous and sampled cases, then the sampled values must be very highly correlated, and hence one may need very many more of them than one has in order to apply asymptotic extremal theory with confidence!

In any event, it is clearly important to consider the possible asymptotic laws for the maximum in the continuous case, and the comforting result is that the same three laws still apply under rather general circumstances. At the same time there are some significant differences (e.g. in the normalizing constants). Further, it is no longer necessarily true that the same limit law holds as would apply

to an i.i.d. sequence with the same marginal distribution. There is an appropriate corresponding i.i.d. sequence, but it may have a different marginal distribution from the original process. In fact, this "associated i.i.d. sequence" will have a tail distribution proportional to the mean number of upcrossings of a high level per unit time by the original process, as we shall see.

The idea involved in dealing with the continuous case is a very simple one indeed - merely to reduce it to a sequence case by writing such a maximum as the greatest of maxima over smaller intervals of fixed length. Specifically, denote the stationary process by $\{X_t : t \geq 0\}$ and write $M(T) = \max\{X_t : 0 \leq t \leq T\}$. (We should perhaps write "sup" for "max", but will assume without comment that X_t has basic regularity properties, such as continuous sample functions.) Write also $Z_i = \max\{X_t : i-1 \leq t \leq i\}$. Then (for $T=n$) we have

$$(4.1) \quad M(n) = \max(Z_1, Z_2, \dots, Z_n) .$$

Hence it is reasonable to expect (e.g. taking n to be the integer part of T) that $M(T)$ will have asymptotic properties close to that of the maximum of n terms of the Z_i -sequence.

We can capitalize at once on this observation by giving a simple form of Gnedenko's Theorem applicable to strongly mixing processes X_t . (Strong mixing is a widely used dependence restriction - see e.g. [3] for a precise definition.)

Theorem 4.1. Let the stationary process X_t be also strongly mixing. Suppose that

$$P\{a_T[M(T)-b_T] \leq x\} \rightarrow G(x)$$

for some constants a_T, b_T and some non-degenerate G , as $T \rightarrow \infty$.

Then G has one of the three extreme value forms.

Proof. The limit holds in particular if T tends to infinity through integral values. But from the representation (4.1), since the Z_i are clearly strongly mixing, we may apply the form of Gnedenko's Theorem for such sequences (a corollary of Theorem 3.1, or directly from [6]) to see that G has an extreme value form. #

The foregoing result already indicates the potential usefulness of the extreme value distributions in continuous cases. We shall see that in fact their applicability is even wider than for strongly mixing processes. Specifically, conditions on X_t will be found such that the Z_i essentially satisfy $D(u_n)$ and $D'(u_n)$ for appropriate u_n , so that the sequence theory will apply in toto. In particular, the last part of the corollary to Theorem 3.3 will then hold if M_n is now the maximum of n i.i.d. random variables distributed not as the X_i , but as the Z_i .

Further, Theorem 3.3 itself will hold if (2.1) holds, when F is replaced by the distribution of the Z_i , i.e. if

$$(4.2) \quad nP\{Z_1 > u_n\} \rightarrow \tau.$$

Finally, in the sequence case, it is the form of the tail of the d.f. F which determines which extreme value distribution applies.

Here we simply replace F by the distribution of Z_1 .

For these reasons, the behavior of $P\{Z_1 > u\}$ for large u plays a crucial role in the continuous case (replacing $P\{X_1 > u\}$ for sequences). We therefore study the asymptotic behavior of $P\{Z_1 > u\}$ in the next section, and then state the main results in Section VI.

V. UPCROSSINGS OF A LEVEL, AND THE TAIL OF THE DISTRIBUTION OF Z_1

The tail of the distribution of $Z_1 = \sup\{X_t: 0 \leq t \leq 1\}$ depends on properties of the upcrossings of a level u by X_t , i.e. the points where the sample path of X_t crosses u in an upwards direction (cf. [1] for a precise definition). Let $N_u(t)$ denote the number of such upcrossings of the level u by X in the time interval $[0, t]$ and write $\mu (= \mu(u)) = \mathcal{E}N_u(0, 1)$. μ is thus the mean number of upcrossings of u per unit time (or the "intensity" of the upcrossings). It is easily shown that, in general, $\mu = \lim_{t \rightarrow 0} I_t(u)$ where

$$(5.1) \quad I_t(u) = P\{X_0 < u < X_t\}/t.$$

Then under appropriate conditions ([5]) it may be shown that μ is obtained explicitly as

$$(5.2) \quad \mu = \int_0^\infty zp(u, z) dz$$

where p is the joint density for $X(0)$ and its derivative $X'(0)$. If X_t is a normal (Gaussian) process with zero mean and covariance function r , this leads at once to the celebrated formula of Rice, viz

$$\mu = (\lambda_2^{1/2}/2\pi) e^{-u^2/2} \quad (\lambda_2 = -r''(0))$$

(cf. [1], [5]). We shall assume without comment that μ is finite (which holds under general conditions, though not universally). We shall also need the following restriction on the second moment of upcrossings:

$$(5.3) \quad \mathcal{E}N_u(h)[N_u(h) - 1] = o(\mu)$$

as $u \rightarrow \infty$, h being fixed. (This may typically be verified from formulae analogous to (5.2). The following result then holds.

Theorem 5.1. Suppose that (5.3) holds with $h = 1$ for the stationary process X_t and that $P(X_0 > u) = o(\mu)$. Then

$$P\{Z_1 > u\} \sim \mu$$

as $u \rightarrow \infty$ (where $Z_1 = M(1)$ and $\mu = \mu(u)$).

Sketch of Proof: Clearly (writing $N_u = N_u(1)$),

$$P\{N_u \geq 1\} \leq P\{Z_1 > u\} \leq P\{X_0 > u\} + P\{N_u \geq 1\} \leq \mu + o(\mu)$$

by assumption, and since $P\{N_u \geq 1\} \leq EN_u = \mu$. Now, writing $p_j = P\{N_u = j\}$,

$$EN_u(N_u - 1) \geq \sum_{j=2}^{\infty} jp_j = \mu - P\{N_u = 1\}$$

giving

$$P\{N_u \geq 1\} \geq P\{N_u = 1\} \geq \mu - EN_u(N_u - 1) = \mu - o(\mu)$$

from (5.3). Hence by combining results we have

$$\mu - o(\mu) \leq P\{Z_1 > u\} \leq \mu + o(\mu)$$

from which the desired conclusion follows. #

VI. MAIN THEOREMS

Again X_t will denote our basic stationary process and $Z_i = \max\{X_t : i-1 \leq t \leq i\}$. We may now give results concerning $M(T) = \max\{X_t : 0 \leq t \leq T\}$ based on the assumptions (to be stated) that the Z_i -sequence satisfies appropriate $D(u_n)$, $D'(u_n)$ conditions. Obviously, in

practice the basic assumptions should be given in terms of X_t , not the Z_i . However, for clarity we state them in terms of the Z_i , and in the next section will give conditions on X_t which can replace the D and D' conditions assumed here for the Z_i -sequence.

Theorem 6.1. Let X_t be stationary and let $M(T) = \max\{X_t : 0 \leq t \leq 1\}$ satisfy

$$P\{a_T\{M(T) - b_T\} \leq x\} \rightarrow G(x)$$

(non-degenerate) as $T \rightarrow \infty$ for some families of constants $a_T > 0$, b_T . Suppose that the condition $D(u_n)$ is satisfied for the sequence $Z_i = \max\{X_t : i-1 \leq t \leq i\}$ for each u_n of the form $x/a_n + b_n$. Then G has one of the three extreme value forms.

This result follows at once from Theorem 3.1. #

As noted above, restrictions on X_t to guarantee $D(u_n)$ for the Z_i are given later (Theorem 7.2). Theorems 6.1 and 7.2 indicate the wide applicability of the extreme value distributions - even for continuous maxima.

Theorems 3.3 and 5.1 also lead at once to the following result.

Theorem 6.2. Suppose that X_t is stationary, and $\{u_T\}$ is a family of constants chosen so that

$$(6.1) \quad \mu_T = \mu(u_T) \sim \tau/T$$

for some fixed $\tau > 0$. Suppose that the conditions of Theorem 5.1 hold, and that the sequence $Z_i = \max\{X_t : i-1 \leq t \leq i\}$ satisfies the conditions $D(u_n)$, $D'(u_n)$. Then

$$P\{M(T) \leq u_T\} \rightarrow e^{-\tau} \text{ as } T \rightarrow \infty.$$

Proof. From the representation (4.1) and Theorems 3.3 and 5.1 we have that $P\{M(n) \leq u_n\} \rightarrow e^{-\tau}$ (and hence also $P\{M(n+1) \leq u_{n+1}\} \rightarrow e^{-\tau}$) as $n \rightarrow \infty$. Now if n denotes the integer part $[T]$ of T , $n \leq T < n+1$, we have

$$P\{M(T) \leq u_T\} \leq P\{M(n) \leq u_n\} + P\{M(T) \leq u_T, M(n) > u_n\}.$$

The last term is zero if $u_T < u_n$ since $M(T) \geq M(n)$, and if $u_T > u_n$ it does not exceed

$$\begin{aligned} P\{u_n < M(n) \leq u_T\} &= P\left\{\bigcup_{i=1}^n (u_n < Z_i \leq u_T)\right\} \leq nP\{u_n < Z_1 \leq u_T\} \\ &= n[P(Z_1 > u_n) - P(Z_1 > u_T)] \\ &= n[\mu_n - \mu_T + o(\mu_n) + o(\mu_T)] \end{aligned}$$

by Theorem 5.1. This tends to zero since $\mu_n \sim \tau/n \sim \tau/T$, $\mu_T \sim \tau/T$.

Hence $\limsup_T P\{M(T) \leq u_T\} \leq e^{-\tau}$. Similarly, a comparison of the events $\{M(n+1) \leq u_{n+1}\}$, $\{M(T) \leq u_T\}$ gives the opposite inequality for $\liminf P\{M(T) \leq u_T\}$ and the result follows. #

Corollary. Suppose that $\{X_t\}$ is stationary and that $\{\hat{Z}_i\}$ is a sequence of i.i.d. random variables such that $P\{\hat{Z}_1 > u\} \sim \mu(= \mathcal{E}N_u(1))$, and satisfying

$$P\{a_n(\hat{M}_n - b_n) \leq x\} \rightarrow G(x),$$

for some non-degenerate G , constants $a_n > 0$, b_n , ($\hat{M}_n = \max(\hat{Z}_1 \dots \hat{Z}_n)$).

If the sequence $Z_i = \max\{X_t : i-1 \leq t \leq i\}$ satisfies $D(u_n)$, $D'(u_n)$ for all u_n of the form $x/a_n + b_n$, then

$$P\{a_T(M(T) - b_T) \leq x\} \rightarrow G(x)$$

where $a_T > 0$, b_T are constants, equal to a_n, b_n when $T = n$.

Proof. It follows from the equivalence of (2.1) and (2.2) for i.i.d. random variables (here the $\hat{Z}_i$) that writing $u_n = x/a_n + b_n$, $\tau = -\log G(x)$, we have $\mu_n = \mu(u_n) \sim \tau/n$.

Let $a_T = a_{[T]}$, $b_T = b_{[T]}$, $u_T = x/a_T + b_T$, so that

$$\mu_T = \mu(u_T) = \mu(u_{[T]}) \sim \tau/[T] \sim \tau/T$$

i.e. (6.1) holds. Hence by the theorem

$$P\{M(T) \leq u_T\} \rightarrow e^{-\tau} = G(x)$$

which gives the desired conclusion. #

Thus we see that the central results quoted from classical E.V.T. have counterparts not only for dependent sequences, but also for continuous time processes. The maximum of a dependent sequence $\{X_i\}$ will (under appropriate conditions) have the same asymptotic distribution as the i.i.d. sequence $\hat{X}_i$ with the same constants a_n, b_n . However, in the continuous case the "corresponding i.i.d. sequence" is the sequence of $\hat{Z}_i$ with tail distribution $P\{\hat{Z}_i > u\} \sim \mu = \mathcal{E}N_u(1)$. Hence the constants can be different from the classical ones for the given marginal d.f. of X_t (and it is even possible that the limit law is different though this question is an open one).

As an example, consider the (standardized) stationary normal case. For such sequences with covariance function $r_n, D(u_n), D'(u_n)$ are always satisfied if $r_n \log n \rightarrow 0$. Hence the limiting distribution for the maximum is the double exponential one with the classical constants $a_n = (2 \log n)^{\frac{1}{2}}, b_n = (2 \log n)^{\frac{1}{2}} - \frac{1}{2}(2 \log n)^{-\frac{1}{2}} [\log 4\pi + \log \log n]$ given before.

For a continuous parameter stationary normal process $\{X_t\}$, a corresponding covariance condition ensures that D and D' conditions are satisfied by the sequence $\{Z_i\}$. However $P\{Z_1 > u\} \sim \mu = (\lambda_2^2/2\pi)e^{-u^2/2}$ whereas $P(X_t > u) \sim (2\pi)^{-1/2} e^{-u^2/2}/u$ (the asymptotic form of the tail of a standard normal d.f.). The asymptotic distribution for the maximum of i.i.d. random variables $\hat{Z}_i$ with the same tail distribution as the Z_i is still Type I. However, the coefficients are now (when X_t is suitably standardized) $a_n = (2 \log n)^{1/2}$ and $b_n = (2 \log n)^{1/2} - (2 \log n)^{-1/2} \log 2\pi$, so that a_n and b_n are obtained by putting T for n , i.e.

$$a_T = (2 \log T)^{1/2}, \quad b_T = (2 \log T)^{1/2} - (2 \log T)^{-1/2} \log 2\pi.$$

VII. DEPENDENCE CONDITIONS FOR X_t

The main theorems of Section VI contained, as assumptions, that the sequence $Z_i = \max\{X_t : i-1 \leq t \leq i\}$ should satisfy D - and D' -type conditions. As noted, these were so stated for purposes of clarity. In practice one would want to use conditions on the X_t -process itself. We now give such conditions which will imply, or may replace, those we need for the Z_i sequence. These results will be sketched only - full details being given in [4].

It will be convenient to define continuous analogs $D_c(u_n)$, $D'_c(u_n)$ of the conditions $D(u_n)$, $D'(u_n)$. The condition D_c will be simply a developed version of D and we deal with it first. The D'_c condition involves some changes which give further insight into the relationship of the discrete and continuous cases.

D_c will be defined in terms of a sampling of the X_t -process by points spaced a certain distance q apart. We noted that the mean number of upcrossings of the level u per unit time by X_t is generally given as $\mu = \mu(u) = \lim_{t \rightarrow 0} I_t(u)$ where $I_t(u)$ is defined by (5.1). Here we require the further assumption that (for certain u_n) constants q may be chosen (converging to zero) so that, in fact

$$(7.1) \quad \mu_n = \mu(u_n) \sim I_{q_n}(u_n) \text{ as } n \rightarrow \infty.$$

(That is, the fixed u in the limit of $I_t(u)$ is replaced by a changing level u). It may be shown that such a choice is possible in certain cases of interest, e.g. stationary normal processes).

Write, further, $F_{t_1 \dots t_n}$ to denote the joint d.f. of $X_{t_1} \dots X_{t_n}$ and $F_{t_1 \dots t_n}(u) = F_{t_1 \dots t_n}(u, u \dots u)$. Then $\{X_t\}$ will be said to satisfy the condition $D_c(u_n)$ for a given u_n -sequence, if there is a sequence q_n such that (7.1) holds for the sequences u_n, q_n , and if for any choice of integers $n, i_1 < i_2 \dots < i_p < j_1 \dots < j_p, \leq n/q_n, j_1 - i_p \geq \ell/q_n$ we have (writing $q_n = q$)

$$(7.2) \quad |F_{i_1 q \dots i_p q, j_1 q \dots j_p q}(u_n) - F_{i_1 q \dots i_p q}(u_n) F_{j_1 q \dots j_p q}(u_n)| < \alpha_{n, \ell}$$

where $\alpha_{n, \ell} \rightarrow 0$ for some sequence $\ell_n = o(n)$.

The procedure for showing that $D(u_n)$ holds for the sequence $\{Z_i\}$ is based on the following lemma, the proof of which we briefly indicate only. (The full proof is straightforward and given in detail in [4].)

Lemma 7.1. Suppose u_n is a sequence, q_n chosen so that (7.1) holds, and I an interval of fixed length h . Assume also that $P\{X_0 > u\} = o(\mu)$ as $u \rightarrow \infty$. Then, writing $M(I) = \max\{X_t; t \in I\}$, we have

$$0 \leq P\left\{\bigcap_{jq \in I} \{X_{jq} \leq u_n\}\right\} - P\{M(I) \leq u_n\} = o(\mu_n).$$

This lemma may be proved by dominating the difference in probabilities by $\mathcal{E}\{N_u - N_u^{(q)}\}$ where N_u and $N_u^{(q)}$ are the numbers of upcrossings of u in I by X_t itself, and by the "sampled" process $\{X_{jq}\}$ respectively. (It may be shown that $\mathcal{E}N_u^{(q)} = \mu h + o(\mu)$.)#

Theorem 7.2. Let $\{X_t\}$ satisfy $D_c(u_n)$, for a given sequence $\{u_n\}$. Suppose that $P\{X_0 > u\} = o(\mu)$ and $\limsup_n n\mu(u_n) < \infty$. Then $\{Z_i\}$ satisfies $D(u_n)$.

Indication of proof. If $i_1 \dots i_r$ are any integers and $I_s = [s-1, s]$ we may show simply that

$$\begin{aligned} 0 &\leq P\left\{\bigcap_{jq \in UI_{i_s}} \{X_{jq} \leq u\}\right\} - P\{Z_{i_1} \leq u, \dots, Z_{i_r} \leq u\} \\ &\leq r \left[P\left\{\bigcap_{jq \in I_1} \{X_{jq} \leq u\}\right\} - P\{Z_1 \leq u\} \right] = ro(\mu) \end{aligned}$$

by Lemma 7.1. By applying this approximation to $i_1 \dots i_p, j_1 \dots j_p$, and the combined set of these integers, it follows in a straightforward way that

$$\begin{aligned} P\left\{\bigcap_1^p (Z_{i_r} \leq u_n) \bigcap_1^{p'} \{Z_{j_s} \leq u_n\}\right\} &- P\left\{\bigcap_1^p (Z_{i_r} \leq u_n)\right\} P\left\{\bigcap_1^{p'} (Z_{j_s} \leq u_n)\right\} \\ &\leq \alpha_{n,\ell} + o(n\mu) \end{aligned}$$

if $i_1 < i_2 \dots < i_p < j_1 \dots < j_p, \leq n, j_1 - i_p \geq l$, by using $D_c(u_n)$.
 $D(u_n)$ for $\{Z_i\}$ follows at once. #

The appropriate D' condition for X_t is more interesting, and its motivation will be given in the next section. Here we just note the definition. Specifically, we shall say that $\{X_t\}$ satisfies $D'_c(u_n)$ if

$$\limsup_{n \rightarrow \infty} \mathcal{E}\{N_{u_{nk}}(n) [N_{u_{nk}}(n) - 1]\} = o(1/k) \text{ as } k \rightarrow \infty.$$

We may then cast Theorem 6.2 in the following form.

Theorem 7.3. The conclusion of Theorem 6.2 holds if the requirements that $\{Z_i\}$ satisfy $D'(u_n)$ are replaced by the condition that $\{X_t\}$ should satisfy $D'_c(u_n)$. (Of course, $D(u_n)$ may also be replaced by the sufficient conditions of Theorem 7.2.)

Indication of proof. Let k be fixed. Then a straightforward calculation shows that

$$P\{X_0 > u_{nk}\} + P\{M(n) \leq u_{nk}\} \geq P\{N_{u_{nk}}(n) = 0\} \geq 1 - n\mu_{nk}$$

($\mu_{nk} = \mu(u_{nk})$), and, letting $n \rightarrow \infty$, we obtain

$$\liminf_n P\{M(n) \leq u_{nk}\} \geq 1 - \tau/k.$$

A similar calculation gives ($N = N_{u_{nk}}(n)$)

$$P\{M(n) \leq u_{nk}\} \leq 1 - \mu_{nk}n + \mathcal{E}N(N-1)$$

from which, letting $n \rightarrow \infty$, we obtain

$$\limsup_n P\{M(n) \leq u_{nk}\} \leq 1 - \tau/k + o(1/k).$$

By Lemma 3.2 applied to the sequence $\{Z_i\}$ we at once obtain

$$\begin{aligned} (1 - \tau/k)^k &\leq \liminf_n P\{M(nk) \leq u_{nk}\} \leq \limsup_n P\{M(nk) \leq u_{nk}\} \\ &\leq (1 - \tau/k + o(1/k))^k. \end{aligned}$$

Now by an argument similar to that used in the proof of Theorem 6.2 (choosing r dependent on n so that $rk \leq n \leq (r+1)k$, we may replace nk by n in the above inequality giving

$$\begin{aligned} (1 - \tau/k)^k &\leq \liminf_n P\{M(n) \leq u_n\} \leq \limsup_n P\{M(n) \leq u_n\} \\ &\leq (1 - \tau/k + o(1/k))^k \end{aligned}$$

from which the desired result follows by letting $k \rightarrow \infty$. #

VIII. A COMPARISON, AND DISCUSSION

One may ask why it is that for a sequence $\{X_n\}$ it is the tail distribution $P\{X_n > u\} = 1 - F(u)$ which plays a central role in determining which extreme value distribution applies, whereas in the continuous case the same role is played by the mean number of upcrossings of u by X_t per unit time ($\mathcal{E}N_u(1)$). In fact, even in the discrete case $1 - F(u)$ may be interpreted as the mean number of upcrossings for large u . Indeed, if X_i are i.i.d., and we say that an upcrossing of u occurs at i if $X_i \leq u < X_{i+1}$, then since zero or one such upcrossing occurs in unit time, the expected number per unit time is just

$$P(X_i \leq u < X_{i+1}) = F(u)[1 - F(u)].$$

Since $F(u) \rightarrow 1$ as $u \rightarrow \infty$, this is asymptotically the same as $1 - F(u)$. Hence the central quantity may be regarded as the mean number of upcrossings per unit time, in both discrete and continuous cases.

Along these lines one may wish to compare the conditions $D'(u_n)$, $D'_c(u_n)$, the former being based on (bivariate) probabilities viz $\sum P(X_i > u, X_j > u)$ and the latter on the second factorial moment $\sum N_u(N_u - 1)$. One may see here also that, in the sequence case, this factorial moment is expressible (for large u) as a sum of bivariate probabilities $P(X_i > u, X_j > u)$. (The calculation is again trivial when the X_i are i.i.d., but holds more generally.) Hence these seemingly different conditions are really of the same character.

The main conclusion for application to practical cases, such as maximum levels of concentrations of some substance over time, is that the extreme value distributions will still have wide validity. The normalizing constants, however, will not usually be the same for a continuous interval of measurement, as for sampled data. In some cases (e.g. normal) the constants may be obtained theoretically - or by theory combined with some simple estimation. For other non-normal cases the theoretical evaluation of the constants may be difficult or impossible. However, the knowledge that the distribution of the maximum is likely to be one of the three extreme value types suggests fitting these experimentally either by parameter estimation from repeated measurements of maxima, or from complete single records by some form of function or parameter estimation for the mean numbers of upcrossings of the varying level u . Such questions as these require further investigation in order to apply classical EVT in

practice in the continuous context - the main intent of this paper has been to extend the theoretical framework for this.

Finally, we list some open questions (some mentioned above) which would lead to further extension of the theory, or which are relevant to its application:

1. Is it possible for one extreme value type to apply to a process X_t , and for another to apply to a sampled version of X_t ? (We conjecture that this is not possible, even though the normalizing constants may differ markedly.)

2. In a given case, how may the correct one of the three types be determined and what procedures may be used to estimate the constants?

3. Can the varied techniques of function estimation be used profitably to estimate the function $\mu(x)$ and hence provide an alternative method to determine the extreme value type?

4. Can the theory be extended to deal with cases where the mean number of upcrossings is infinite? This is by no means a pathological situation, and perhaps the use of so-called " ϵ -upcrossings" would be appropriate.

5. The stationarity assumption used throughout may need modification, due e.g. to periodic meteorological factors. It may be profitable to consider a number of shorter series of observations - assumed independent - obtained from comparable periods of the day or season (and omitting values between these periods).

BIBLIOGRAPHY

- [1] Cramér, Harald, and Leadbetter, M. R. (1967). Stationary and Related Stochastic Processes, John Wiley & Sons, New York.
- [2] Gumbell, E. J. (1958). Statistics of Extremes, Columbia University Press, New York.
- [3] Leadbetter, M. R. (1974). "On extreme values in stationary sequences," Z. Wahr. verw. Geb. 28, pp. 289-303.
- [4] _____ (1977). "Continuous parameter extreme value theory," University of North Carolina Mimeo Series, in preparation.
- [5] _____ (1966). "On crossings of levels and curves by a wide class of stochastic processes," Ann. Math. Statist. 37, pp. 260-267.
- [6] Loynes, R. M. (1965). "Extreme values in uniformly mixing stationary stochastic processes," Ann. Math. Statist. 36, pp. 993-999.
- [7] Mittal, Y., and Ylvisaker, D. (1975). "Limit distributions for maxima in strongly correlated stationary Gaussian sequences," J. Stoch. Proc. & Appl. 3, pp. 1-18.

ABSTRACT

Statistical extreme value theory is obviously a potentially useful tool in the study of concentrations of materials associated with atmospheric pollution, when maximum levels are of primary interest.

The classical theory requires the observed values to be statistically independent. Within the last decade there has been considerable interest in allowing a dependence structure and an extensive extremal theory now exists for a wide class of statistically dependent values, measured as a discrete sequence, e.g. at specific sampling points.

Of at least equal interest, however, are the statistical properties of the maxima of such quantities over continuous periods of time. For it is such a maximum - not necessarily occurring at a sampling point - which may lead to concern for effects on health, and so on.

In the case of normal (Gaussian) data there has been considerable activity, and many explicit results are known. In this report we present a general theory which applies to maxima in the continuous time context, for a wide class of basic distributions (normal and non-normal). It is found that the same basic results hold as for the sampled case - with certain important practical differences, which are described in detail.