The Annals of Probability
1988, Vol. 16, No. 2, 431–478

SPECIAL INVITED PAPER

EXTREMAL THEORY FOR STOCHASTIC PROCESSES[1]

By M. R. Leadbetter and Holger Rootzén

University of North Carolina at Chapel Hill and University of Copenhagen

The purpose of this paper is to provide an overview of the asymptotic distributional theory of extreme values for a wide class of dependent stochastic sequences and continuous parameter processes. The theory contains the standard classical extreme value results for maxima and extreme order statistics as special cases but is richer on account of the diverse behavior possible under dependence in both discrete and continuous time contexts. Emphasis is placed on stationary cases but some departures from stationarity are considered. Significant ideas and methods are described rather than details, and, in particular, the nature and role of important underlying point processes (such as exceedances and upcrossings) are emphasized. Applications are given to particular classes of processes (e.g., normal, moving average) and connections with related theory (such as convergence of sums) are indicated.

Table of contents

Received July 1986; revised July 1987.

[1]Research supported by Air Force Office of Scientific Research Contract AFOSR 85-C-0144 and by the Danish National Science Research Council.

AMS 1980 *subject classifications*. Primary 60G10; secondary 60-02, 60G15, 60G17, 60G55, 60F05.

Key words and phrases. Extreme values, stationary processes, point processes.

1. Introduction: The classical theory of extremes

1.1. *Scope and content of the paper.* The purpose of this paper is to give a "motivated overview" of the principal results in and related to the distributional theory of extremes of stationary sequences and processes. In particular, we shall be concerned with distributional properties of the maximum $M_n = \max(\xi_1, \xi_2, \dots, \xi_n)$ and other order statistics from stationary sequences $\{\xi_i\}$ as $n \to \infty$ and with corresponding results for continuous parameter processes. The emphasis throughout will be on the motivation for and significant methods used in obtaining the results. Full proofs will not generally be given—in many cases the details of such proofs may be found in [66] or from the references cited therein.

The results to be described may, in part, be regarded as extensions of the classical theory of extremes of sequences of independent, identically distributed (i.i.d.) random variables (r.v.'s), (cf. [51] and [49]). However, they constitute more than just such an extension of the classical theory, since the dependent framework provides a natural setting for the theory and one in which its essential ideas and methods may be clearly exposed. In particular, it will be seen that the central results may often be regarded as special cases of the convergence of certain point processes—a view which may, of course, be taken in the classical case but which is less needed there in view of the detailed i.i.d. assumptions. Our discussion will emphasize the centrality of these underlying point process convergence results.

As indicated in the table of contents, this paper is organized in three main parts. This first introductory part contains central distributional results of the classical i.i.d. theory and, in particular, the "extremal types theorem," which restricts the possible limiting distributions for maxima to essentially three "different types." We shall indicate only the general organization and main features of the most recently available derivations of these results.

The second part of the paper concerns extremes of sequences—primarily (but not always) assumed stationary and is largely based on point process methods. It will be seen that the classical theory may be regarded as a special case of the more general theory for dependent sequences—some results being identical and others generalizing in interesting and nontrivial ways. For example, under weak dependence restrictions, the general "type" of limiting distribution for the maximum is the same as for an i.i.d. sequence with the same marginal d.f. (though the normalizing constants may change). However, the limiting distributions for other order statistics can be quite different from those under i.i.d. assumptions.

Some particular cases of special interest (e.g., normal sequences, moving averages, Markov sequences) will be discussed in Part 2. Other aspects of the theory (e.g., rates of convergence, multivariate extremes) are also briefly described along with some interesting connections with convergence of sums.

In Part 3 attention is turned to continuous parameter processes. The theory here may be made to rest on the sequence case by the simple device of regarding the maximum of a process $\xi(t)$ up to, say time $T = n$, as the maximum of the values of the sequence $\zeta_i = \sup\{\xi(t): i - 1 \le t \le i\}$ for $1 \le i \le n$. Whereas this is simple and obvious in principle, the details are more complicated and require analogous but somewhat more intricate assumptions regarding the dependence structure of the process. The point process approach is also very valuable here—considering, for example, upcrossings of high levels in lieu of exceedances. Again, a rather full and satisfying theory results and is applied, in particular, to special cases such as normal and χ^2-processes. Properties of point processes of *local maxima* may also be obtained, as will be briefly indicated.

It may be noted that the *stationarity* assumption, where made, primarily provides for convenience and clarity, and that some departures from this will either not alter the result, or will alter it in an interesting way that can be determined. This will be evident, e.g., in discussion of normal sequences, where extensions to useful nonstationary cases will be briefly mentioned. Finally, this paper is not by any means intended as a complete review of all aspects of extremal theory—a number of important topics are not referred to at all. Rather it is our purpose to provide an overview of much of a developing area, which includes but is more widely applicable than the classical theory, and is based on the interplay of interesting mathematical techniques. In particular, we emphasize recent results—especially those obtained since the publication of [66].

1.2. *Classical extreme value theory.* The principal concern of classical extreme value theory is with asymptotic distributional properties of the maximum $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ from an i.i.d. sequence $\{\xi_i\}$ as $n \to \infty$. Whereas the distribution function (d.f.) of M_n may be written down exactly $[P\{M_n \le x\} = F^n(x)$, where F is the d.f. of each $\xi_i]$, there is nevertheless virtue in obtaining asymptotic distributions, which are less dependent on the precise form of F, i.e., relations of the form

$$(1.2.1) \qquad P\{a_n(M_n - b_n) \le x\} \to_d G(x), \quad \text{as } n \to \infty,$$

where G is a nondegenerate d.f. and $a_n > 0$, b_n, are normalizing constants.

The central result of classical extreme value theory, due in varying degrees of generality to Fréchet [47], Fisher and Tippett [46] and Gnedenko [50], restricts the class of possible limiting d.f.'s G in (1.2.1) to essentially three different types as follows.

THEOREM 1.2.1 (Extremal types theorem). *Let $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$, where ξ_i are i.i.d. If (1.2.1) holds for some constants $a_n > 0$, b_n and some nondegenerate G, then G must have one of the following forms (in which x may* *be replaced by $ax + b$ for any $a > 0$, b):*

$$\text{type } I: \quad G(x) = \exp(-e^{-x}), \qquad -\infty < x < \infty,$$

$$\text{type } II: \quad G(x) = \begin{cases} 0, & x \le 0, \\ \exp(-x^{-\alpha}), \quad \text{for some } \alpha > 0, & x > 0, \end{cases}$$

$$\text{type } III: \quad G(x) = \begin{cases} \exp(-(-x)^{\alpha}), \quad \text{for some } \alpha > 0, & x \le 0, \\ 1, & x > 0. \end{cases}$$

Conversely, any such d.f. G may appear as a limit in (1.2.1) *and in fact does so when G is itself the d.f. of each ξ_i.*

It will be convenient to say that two nondegenerate d.f.'s G_1 and G_2 are of the *same type* if $G_1(x) = G_2(ax + b)$ for some $a > 0$, b, and to refer to the equivalence classes so determined as "types." The use of "type" in the above theorem is a slight abuse of this since types II and III really represent *families* of types —one corresponding to each $\alpha > 0$. However, this abuse is convenient and it is conventional to refer to "the three types" of limit. It should also be noted that the three types may be incorporated into a single family, for example by writing $G_\alpha(x) = \exp\{-(1 - \alpha x)^{1/\alpha}\}$, $-\infty < \alpha < \infty$, $\alpha x < 1$, G_0 being interpreted as $\lim_{\alpha \to 0} G_\alpha(x) = \exp(-e^{-x})$. (Such a parametrization was introduced by von Mises.)

A straightforward proof of Theorem 1.2.1 is given in [66], Theorem 1.4.2, and here we note only the fact that this consists of two parts—a division which is most useful for later forms of the result. The first part is to show that the class of limit laws G in (1.2.1) is precisely the class of *max-stable* d.f.'s. Specifically, a d.f. G is called *max-stable* if for each $n = 1, 2, \dots,$ the d.f. G^n is of the same type as G, i.e., if there exist constants $a_n > 0$, b_n such that $G^n(a_n x + b_n) = G(x)$.

The second part of the proof of Theorem 1.2.1 is to identify the class of max-stable d.f.'s with the type I, II and III extreme value d.f.'s. This is a purely function-analytic (nonprobabilistic) procedure and will apply verbatim in dependent cases. A smooth proof due to de Haan (using inverse functions) may be found in [66].

It is, of course, important to know which (if any) of the three types of limit law applies when ξ_n has a given d.f. F. Necessary and sufficient conditions are known, involving only the behavior of the tail $1 - F(x)$ as x increases, for each possible limit. For example, the criterion for a type II limit is simply that $1 - F(x)$ should be regularly varying with index $-\alpha$, $\alpha > 0$, as $x \to \infty$. The conditions for all three types may be found in [66], Theorem 1.6.2, together with simple proofs of their sufficiency. The necessity is more complicated (though perhaps also less important) but may be achieved by using methods of regular variation (cf. [38] for a recent smooth treatment).

The following almost trivially proved result is also used in "domain of attraction" determinations, and has important (and less trivially proved) extensions to dependent cases.

LEMMA 1.2.2. *Let $\{u_n, n \geq 1\}$ be constants and $0 \leq \tau \leq \infty$. If $\xi_1, \xi_2, \ldots$ are i.i.d. with d.f. F, then*

$$P\{M_n \leq u_n\} \to e^{-\tau}, \tag{1.2.2}$$

if and only if

$$n(1 - F(u_n)) \to \tau. \tag{1.2.3}$$

It may be noted that (1.2.1) is a special case of (1.2.2) using a linear parametrization, by making identifications $\tau = -\log G(x)$, $u_n = a_n^{-1}x + b_n$. Thus a necessary and sufficient condition for the limit G is

$$n\big(1 - F\big(a_n^{-1}x + b_n\big)\big) \to -\log G(x), \quad \text{as } n \to \infty,$$

for each x and some $a_n > 0$, b_n. This explains the relevance of the tail $1 - F(x)$ for domain of attraction criteria. Use of Lemma 1.2.2 also enables expressions to be obtained for the normalizing constants a_n, b_n in terms of the $(1 - n^{-1})$ percentile γ_n defined to satisfy $F(\gamma_n -) \leq 1 - n^{-1} \leq F(\gamma_n)$. For example, in the type II case, a_n, b_n may be taken to be $a_n = \gamma_n^{-1}$, $b_n = 0$. Of course, whereas γ_n may be determined (and hence a_n, b_n found) when F is known, the practical problem lies in the estimation of those constants when the form of F is not precisely known.

It is readily checked that a standard normal sequence belongs to the type I domain with normalizing constants

$$\begin{aligned} a_n &= (2\log n)^{1/2}, \\ b_n &= (2\log n)^{1/2} - \tfrac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi). \end{aligned} \tag{1.2.4}$$

The exponential and log-normal distributions also have type I limits as does the d.f. $F(x) = 1 - e^{1/x}$, $x < 0$, with a finite right endpoint $x_F = 0$. The Pareto and Cauchy distributions give type II limits, whereas the uniform distribution belongs to the type III domain.

Not every d.f. F belongs to a domain of attraction at all. For example, this occurs for the Poisson and geometric distributions—for which there is no sequence $\{u_n\}$ such that (1.2.3) holds for $0 < \tau < \infty$. This typically happens in cases when the jumps of the d.f. do not decay sufficiently quickly relative to the tail ([66], Theorem 1.7.13). However, it is also possible for there to be no limit even if there is a sequence $\{u_n\}$ satisfying (1.2.3) for any τ—such as the d.f. $F(x) = 1 - e^{-x-\sin x}$, an example due to von Mises.

We turn now, in this brief tour of classical results, to other extreme order statistics, writing $M_n^{(k)}$ for the kth largest among the i.i.d. $\xi_1, \ldots, \xi_n$ with common d.f. F. Suppose that $M_n = M_n^{(1)}$ has the limiting distribution G as in (1.2.1). By identifying $u_n = a_n^{-1}x + b_n$, $\tau = -\log G(x)$, it follows that (1.2.3) holds. Let S_n be the number of *exceedances* of u_n by $\xi_1, \ldots, \xi_n$, i.e., the number of i, $1 \leq i \leq n$, such that $\xi_i > u_n$. Then S_n is binomial with parameters (n, $p_n = 1 - F(u_n)$) and $np_n \to \tau$ so that S_n has a Poisson limit with mean τ. The obvious equivalence of the events $\{M_n^{(k)} \leq u_n\}$ and $\{S_n < k\}$ leads directly to the relation

$$P\{a_n(M_n^{(k)} - b_n) \le x\} \to G(x) \sum_{s=0}^{k-1} (-\log G(x))^s/s!. \tag{1.2.5}$$

Thus if the maximum M_n has a limiting distribution G, then the kth largest $M_n^{(k)}$ has a limiting distribution given by (1.2.5) (with the *same* normalizing constants a_n, b_n as the maximum itself).

These results foreshadow a more detailed discussion of the exceedances and related point processes, which will be taken up in the next section.

Finally, topics from the classical theory not dealt with in this part include (a) rate of convergence results (considered in the dependent setting in Section 2.8), (b) asymptotic distributions of minima (obtainable by simple transformations of the results for maxima) and (c) asymptotic theory of variable rank order statistics (cf. [99]).

1.3. *Point processes associated with extremes.* The above asymptotic Poisson property of the number of exceedances of u_n satisfying (1.2.3) may be generalized by considering the actual *point process N_n of exceedances* of the level u_n. Specifically, N_n consists of the point process on $(0,1]$ formed by normalizing the actual exceedance points by the factor $1/n$, i.e., if i is the time of an exceedance ($\xi_i > u_n$), then a point of N_n is plotted at i/n. If $E \subset (0,1]$, then $N_n(E)$ denotes the number of such points in E, so that $N_n(E) = \#\{i/n \in E: \xi_i > u_n, 1 \le i \le n\} = \#\{i \in nE: \xi_i > u_n, 1 \le i \le n\}$. The actual exceedance points and the point process N_n are illustrated in Figure 1.

For u_n satisfying (1.2.3) it follows immediately as before that $N_n(I) \to_d N(I)$, where N is a Poisson process on $[0,1]$ with intensity τ for each interval $I \subset [0,1]$. By independence, corresponding convergence holds for joint distributions of $N_n(I_1), \ldots, N_n(I_k)$ for disjoint intervals $I_1, \ldots, I_k$. This is, in fact, sufficient for convergence in distribution $N_n \to_d N$ (i.e., full weak convergence of PN_n^{-1} to PN^{-1}) of the point processes N_n to N. This may be regarded as a "fountainhead" result from which the asymptotic distributions for the maximum and all extreme order statistics follow. The result may be extended ([66], Section 5.5) by considering the vector point process of exceedances of multiple levels, to give joint asymptotic distributions of finite numbers of order statistics. For example, if the maximum has the asymptotic distribution G given as in (1.2.1), consideration of two levels leads to the asymptotic joint distribution of the first

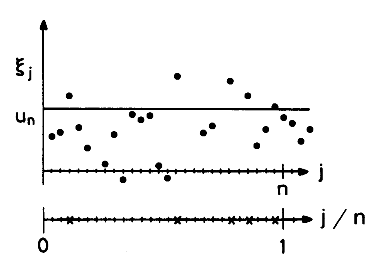

FIG. 1. *Point process of exceedances. Upper: actual exceedance points; lower: the point process N_n.*

two order statistics $M_n^{(1)}$ $(= M_n)$, $M_n^{(2)}$:

$$P\{a_n(M_n^{(1)} - b_n) \le x_1,\ a_n(M_n^{(2)} - b_n) \le x_2\} \tag{1.3.1}$$
$$\to G(x_2)(\log G(x_1) - \log G(x_2) + 1), \quad \text{as } n \to \infty,\ x_1 > x_2.$$

In general, consideration of r levels enables calculation of the asymptotic joint distribution of r extreme order statistics. On the other hand, these results for $r = 1, 2, \dots$ may be summarized simultaneously in one theorem sometimes referred to as a "complete" convergence result. This can be given quite a general form (cf. [66], Theorem 5.7.1), which reduces to the following when the maximum has a limiting distribution G [and writing $x_0 = \inf\{x: G(x) > 0\}$]. This result was first proved by Pickands [81].

THEOREM 1.3.1. *Suppose* (1.2.1) *holds for the i.i.d. sequence* $\{\xi_j\}$, *and let* N_n' *be the point process in the plane with points at* $(j/n,\ a_n(\xi_j - b_n))$. *Then* $N_n' \to N'$ *on* $(0, \infty) \times (x_0, \infty)$, *where* N' *is a Poisson process whose intensity measure is the product of Lebesgue measure and that defined by the increasing function* $\log G(y)$.

In the i.i.d. case the previous theorem may be regarded as fundamental in yielding all relevant asymptotic distributional properties. On the other hand, when dependence is introduced the "partial" r-level results require somewhat fewer assumptions than does the "complete" result. For i.i.d. sequences the proofs of both "r-level" and "complete" results rely on similar (though somewhat more complicated) arguments to those already indicated for single level exceedances.

2. Extremes of sequences.

2.1. *The extremal types theorem for stationary sequences.* Obviously, some form of dependence restriction is necessary to obtain an extremal types result in dependent cases (since, e.g., one might take all ξ_i to be equal with arbitrary d.f., so that M_n would also have this assigned d.f.). Loynes [72] first obtained such a result under strong mixing [viz., $\sup\{|P(A \cap B) - P(A)P(B)|: A \in \sigma(\xi_1, \dots, \xi_n),\ B \in \sigma(\xi_{n+l}, \xi_{n+l+1}, \dots),\ n = 1, 2, \dots\} \to 0$ as $l \to \infty$, where $\sigma(\ )$ denotes the σ-field generated by the indicated r.v.'s]. Weaker (distributional) conditions will suffice and will be used here. The difference is not too important for our present purposes since the main ideas of proof are essentially the same. The main condition to be used [termed $D(u_n)$] is defined with reference to a sequence $\{u_n\}$ of constants in terms of the finite-dimensional d.f.'s $F_{i_1, \dots, i_n}(x_1, \dots, x_n) = P\{\xi_{i_1} \le x_1, \dots, \xi_{i_n} \le x_n\}$ of the stationary sequence $\{\xi_n\}$. Writing $F_{i_1, \dots, i_n}(u) = F_{i_1, \dots, i_n}(u, u, \dots, u)$, define

$$\alpha_{n,l} = \max\Big\{\big|F_{i_1, \dots, i_p, j_1, \dots, j_{p'}}(u_n) - F_{i_1, \dots, i_p}(u_n) F_{j_1, \dots, j_{p'}}(u_n)\big|:$$
$$1 \le i_1 < \cdots < i_p < j_1 < \cdots < j_{p'} \le n,\ j_1 - i_p \ge l\Big\}.$$

Then $D(u_n)$ is said to hold if $\alpha_{n, l_n} \to 0$ for some sequence $l_n = o(n)$.

It is, incidentally, obviously possible to weaken the condition $D(u_n)$ very slightly to involve "intervals" of consecutive integers. (See O'Brien [78] for the details of such a procedure and for some advantages in application to periodic Markov chains.)

The following result ([64], Lemma 2.1) is basic for the discussion of M_n and shows the form in which $D(u_n)$ entails approximate independence.

LEMMA 2.1.1. *Let $\{u_n\}$ be a sequence of constants and let $D(u_n)$ be satisfied by the stationary sequence $\{\xi_n\}$. Let $\{k_n \geq 1\}$ be constants such that $k_n = o(n)$ and [in the notation used before for $D(u_n)$] $k_n l_n = o(n)$, $k_n \alpha_{n, l_n} \to 0$. Then*

$$P\{M_n \leq u_n\} - P^{k_n}\{M_{r_n} \leq u_n\} \to 0, \quad \text{as } n \to \infty,$$

where $r_n = [n/k_n]$.

The proof of this result is perhaps the key method in dependent extremal theory. The type of argument was used first in this context by Loynes [72] but was used earlier in dependent central limit theory (cf. [22]). The basic idea is to divide the integers $1, 2, \ldots, n$ into k_n "intervals" of length r_n and clip out small (but expanding) intervals of length l_n from the right-hand end of each. Then M_n is approximated by the largest submaximum over each remaining interval, the submaxima having a degree of independence from $D(u_n)$, which allows the conclusion to be obtained.

The extremal types theorem now follows simply from this result by showing max stability of the limit G ([66], Theorem 3.3.3).

THEOREM 2.1.2 (Extremal types theorem for stationary sequences). *Let $\{\xi_n\}$ be a stationary sequence such that $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ has a nondegenerate limiting distribution G as in* (1.2.1). *Suppose that $D(u_n)$ holds for each u_n of the form $u_n = x/a_n + b_n$, for x with $0 < G(x) < 1$. Then G is one of the three classical extremal types.*

2.2. *The extremal index.* Whereas the introduction of dependence into a sequence can significantly affect various extremal properties, it does not, within broad limits, affect the distributional type for the maximum. The purpose of this section is to make that rough statement precise and to explore the explicit changes brought by a dependence structure. This depends essentially on a single parameter sometimes called the "extremal index" of the (stationary) sequence $\{\xi_n\}$.

Following Loynes [72], it will be convenient, for a given stationary sequence $\{\xi_n\}$, to define the *associated independent sequence* $\{\hat{\xi}_n\}$ to be i.i.d. with the same d.f. F as ξ_n and to write $\hat{M}_n = \max(\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_n)$, with $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ as before. As noted originally for strongly mixing sequences in [72], if $u_n = u_n(\tau)$ satisfies (1.2.3) for each τ, then any limit (function) for $P(M_n \leq u_n(\tau))$ must be of the form $e^{-\theta\tau}$ with fixed $\theta \in [0, 1]$ rather than just the function $e^{-\tau}$ given by (1.2.2) in the i.i.d. case.

If $P\{M_n \le u_n(\tau)\} \to e^{-\theta\tau}$ for each $\tau > 0$, with $u_n(\tau)$ satisfying (1.2.3), we say that the stationary sequence $\{\xi_n\}$ *has extremal index* θ (≥ 0). This definition does not involve any dependence restriction on the sequence $\{\xi_n\}$. If, however, $\{\xi_n\}$ is a stationary sequence with $D\{u_n(\tau)\}$ holding for each $\tau > 0$ [$u_n(\tau)$ satisfying (1.2.3)], it may be shown by Lemma 2.1.1 that there exist constants θ, θ', $0 \le \theta \le \theta' \le 1$, such that $\limsup_{n\to\infty} P\{M_n \le u_n(\tau)\} = e^{-\theta\tau}$, $\liminf_{n\to\infty} P\{M_n \le u_n(\tau)\} = e^{-\theta'\tau}$ for each τ, so that if $P\{M_n \le u_n(\tau)\}$ converges for some $\tau > 0$, then $\theta' = \theta$ and $P\{M_n \le u_n(\tau)\} \to e^{-\theta\tau}$ for all $\tau > 0$ and $\{\xi_n\}$ has extremal index θ, $0 \le \theta \le 1$. (See [64] for details.)

If $P\{M_n \le u_n(\tau)\} \to e^{-\theta\tau}$ it is clear that $\theta \ge 0$. One might suspect that also $\theta \le 1$ on the grounds that one feels intuitively that the maximum $\hat{M}_n$ of i.i.d. r.v.'s should be stochastically at least as large as M_n, which would imply $e^{-\tau} = \lim P\{\hat{M}_n \le u_n(\tau)\} \le \lim P\{M_n \le u_n(\tau)\} = e^{-\theta\tau}$. In fact, it follows simply that $\theta \le 1$ since

$$\begin{aligned} P\{M_n \le u_n(\tau)\} &= 1 - P\left\{\bigcup_1^n (\xi_i > u_n(\tau))\right\} \ge 1 - nP\{\xi_1 > u_n(\tau)\} \\ &= 1 - n\left[1 - F(u_n(\tau))\right] \\ &\to 1 - \tau, \quad \text{as } n \to \infty. \end{aligned}$$

Since the left-hand side tends to $e^{-\theta\tau}$, it follows that $e^{-\theta\tau} \ge 1 - \tau$, which is only possible for *all* $\tau > 0$ if $\theta \le 1$.

Clearly, any i.i.d. sequence for which $u_n(\tau)$ may be chosen satisfying (1.2.3) has extremal index $\theta = 1$. A stationary sequence $\{\xi_n\}$ satisfying $D(u_n(\tau))$ for each $\tau > 0$ also has extremal index $\theta = 1$ if

$$\limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P\{\xi_1 > u_n, \xi_j > u_n\} \to 0, \quad \text{as } k \to \infty. \tag{2.2.1}$$

For proof see [66], Theorem 3.4.1, where (2.2.1) is referred to as $D'(u_n)$.

Many stationary sequences satisfy (2.2.1), including normal sequences with covariance sequence $\{r_n\}$ satisfying the "Berman condition" $r_n \log n \to 0$. Sufficient conditions for values of $\theta < 1$ are given in [64], and an example with $\theta = 1/2$ appears later in this section. Examples can be found where the extremal index is zero, or does not even exist. This obviously has some theoretical interest but appears to occur in somewhat pathological cases and will not be pursued in the present discussion.

It may be shown by obvious arguments ([64]) that if a stationary sequence $\{\xi_n\}$ has extremal index $\theta > 0$, $\{v_n\}$ is any sequence of constants and ρ any constant with $0 \le \rho \le 1$, then $P\{\hat{M}_n \le v_n\} \to \rho$ if and only if $P\{M_n \le v_n\} \to \rho^\theta$. (This result makes no assumption about dependence.) By taking $v_n = x/a_n + b_n$ one then obtains the following important result.

THEOREM 2.2.1. *Let the stationary sequence $\{\xi_n\}$ have extremal index $\theta > 0$. If $P\{a_n(\hat{M}_n - b_n) \le x\} \to G(x)$, then $P\{a_n(M_n - b_n) \le x\} \to G^\theta(x)$ and conversely. That is $\hat{M}_n$ has an asymptotic distribution if and only if M_n does, with the power relation between the limits and the same normalizing constants.*

By way of comment, note that G^θ is of the same type as G if one of them is of extreme value type [e.g., $[\exp(-e^{-x})]^\theta = \exp[-e^{-(x-\log\theta)}]$, and similarly for types II and III]. If $\theta = 1$ the limits for M_n and $\hat{M}_n$ are precisely the same. Indeed, for $0 < \theta < 1$ the limits may also be taken to be the same by a simple change of normalizing constants.

The practical implication of this result is that often dependence in data does not invalidate application of classical extreme value theory. Indeed, one may not have to worry about the precise value of the extremal index since this only alters parameters of the distribution, which usually must be estimated in any case. Furthermore, if $\theta > 0$, the fact that the distributional type under dependence is the same as under independence means that the classical domain of attraction criteria may be applied to the marginal d.f. of the terms to determine which type applies.

The following simple example provides a case where $\theta < 1$ and will also be useful later when the effects of the value of θ on the clustering of exceedances will be discussed.

EXAMPLE 2.2.2. Let $\eta_1, \eta_2 \dots$ 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 an easy calculation shows that if $u_n(\tau)$ satisfies (1.2.3), then $n[1 - H(u_n(\tau))] \to \tau/2$ and

$$P\{M_n \le u_n(\tau)\} = P\{\max(\eta_1, \dots, \eta_n) \le u_n(\tau)\}P\{\eta_{n+1} \le u_n(\tau)\} \to e^{-\tau/2},$$

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

Criteria for determining the extremal index are discussed in [64]. Finally, we note that an interesting and informative approach to the relating of dependent and i.i.d. cases has been given recently by O'Brien [78] (cf. also [91]). This is based on the general result

$$P\{M_n < u_n\} - F(u_n)^{nP\{\max(\xi_2, \xi_3, \dots, \xi_{p_n}) \le u_n | \xi_1 > u_n\}} \to 0,$$

which is shown in [78] to hold under weak dependence conditions, for a wide variety of sequences $\{u_n\}$ and integers $p_n \to \infty$ with $p_n = o(n)$.

2.3. *Relevant point process concepts.* In dealing with dependent cases, it will be necessary to be somewhat more formal than previously in the use of point process methods. Here we establish the notation and framework (substantially following Kallenberg [63]) and review a few key concepts that will be needed.

In general, a point process is often defined on a locally compact second countable (hence complete separable metric) space S, though here S will invariably be a subset of the line or plane. Write $\mathscr{S}$ for the class of Borel sets on S and $\mathscr{B} = \mathscr{B}(S)$ for the bounded (i.e., relatively compact) sets in $\mathscr{S}$. A point process ξ on S is a random element in $M = M(S)$, the space of locally finite [i.e., finite on $\mathscr{B}(S)$] integer-valued measures on $\mathscr{S}$, where M has the vague topology and Borel σ-field $\mathscr{M} = \mathscr{M}(S)$.

Write $\mathscr{F} = \mathscr{F}(S)$ for the class of nonnegative $\mathscr{S}$-measurable functions, $\mu f = \int f\, d\mu$ for $\mu \in M$, $f \in \mathscr{F}(S)$. The distribution $P\xi^{-1}$ of a point process ξ is uniquely determined by the distributions of $(\xi(I_1),\dots,\xi(I_k))$, $k=1,2,\dots$, $I_j\in\mathscr{T}$, if $\mathscr{T}$ is any semiring whose generated ring is $\mathscr{B}$. The distribution of ξ is also uniquely determined by the Laplace transform $L_\xi(f)=Ee^{-\xi f}$, $f\in\mathscr{F}$.

A (general) Poisson process with intensity measure λ has the Laplace transform $L_\xi(f)=\exp\{-\lambda(1-e^{-f})\}$, whereas a compound Poisson process has Laplace transform

$$L_\xi(f)=\exp\{-\lambda(1-L_\beta\circ f)\}, \tag{2.3.1}$$

where β is a positive integer-valued random variable with Laplace transform $L_\beta(t)=Ee^{-\beta t}$. (No confusion should arise with this dual use of L.) This consists of multiple events of (independent) sizes β located at the points of a Poisson process having intensity measure λ.

Convergence in distribution of a sequence $\{\xi_n\}$ of point processes to a point process ξ is, of course, simply weak convergence of $P\xi_n^{-1}$ to $P\xi^{-1}$. It may be shown (cf. [63]) that $\xi_n\to_d\xi$ if and only if $L_{\xi_n}(f)\to L_\xi(f)$ for every $f\in\mathscr{F}_c$, the subclass of $\mathscr{F}$ consisting of the nonnegative continuous functions with compact support. Point process convergence is also equivalent to convergence of finite-dimensional distributions. Even more simply $\xi_n\to_d\xi$ if and only if $(\xi_n(I_1),\dots,\xi_n(I_k))\to_d(\xi(I_1),\dots,\xi(I_k))$, $k=1,2,\dots$, $I_j\in\mathscr{T}$, where $\mathscr{T}\subset\mathscr{B}$ is a semiring such that $\xi(\partial B)=0$ a.s. for each $B\in\mathscr{T}$, and such that for any $B\in\mathscr{B}$, $\varepsilon>0$, B may be covered by finitely many sets of $\mathscr{T}$ with diameter less than ε (cf. [63], Theorem 4.2). The results of Section 1.3 use the facts that semiclosed intervals and rectangles form such classes.

2.4. *Convergence of point processes associated with extremes.* We return now to the stationary sequence $\{\xi_n\}$ and consider point process convergence results along the same lines as for the i.i.d. case in Section 1.3. The notation of that and other previous sections will be used. In particular, N_n will denote the point process of exceedances on $(0,1]$ as defined in Section 1.3, viz., $N_n(E)=\#\{i/n\in E\colon \xi_i>u_n, 1\le i\le n\}$, for a given sequence of constants u_n.

When $\{\xi_n\}$ has extremal index $\theta=1$, the Poisson convergence result of Section 1.3 for exceedances may be proved provided $D(u_n)$ holds. This leads again to the classical form (1.2.5) for the asymptotic distributions of extreme order statistics. Similarly, r-level convergence results hold under an r-level version $D_r(u_n)$ of $D(u_n)$ (cf. [66], page 107), leading to the classical forms for the asymptotic joint distributions of extreme order statistics when $\theta=1$ [cf. (1.3.1)]. The "complete convergence" result Theorem 1.3.1 also holds giving again a Poisson limit in the plane when $\theta=1$ provided the multilevel conditions $D_r(u_n)$ all hold. These results are described in [66]; here we indicate the new features that arise when $0<\theta<1$.

As noted in Section 2.2, cases when $\theta<1$ occur when there is "high local dependence" in the sequence so that one exceedance is likely to be followed by others (see Example 2.2.2 as an illustration of this). The result is a clustering of exceedances, leading to a compounding of events in the limiting point process.

To include cases where such clustering occurs (i.e., $0<\theta<1$), we require a modest strengthening of the $D(u_n)$ condition (cf. [60]). Let $\mathscr{B}_i^j(u_n)$ be the σ-field generated by the events $\{\xi_s \le u_n\}$, $i \le s \le j$. For $1 \le l \le n-1$ write

$$\beta_{n,l} = \max\{|P(A \cap B) - P(A)P(B)|: A \in \mathscr{B}_1^k(u_n),\ B \in \mathscr{B}_{k+l}^n(u_n),\ 1 \le k \le n-l\}. \tag{2.4.1}$$

Then the condition $\Delta(u_n)$ is said to hold if $\beta_{n,l_n} \to 0$ for some sequence l_n with $l_n = o(n)$. $\{\beta_{n,l}\}$ will be called the *mixing coefficients* for Δ. The condition Δ is, of course, stronger than D but still significantly weaker than strong mixing.

The condition Δ will be applied through the following lemma, which is a special case of [97], Equation I′.

LEMMA 2.4.1. *For each n and $1 \le l \le n-1$ write $\gamma_{n,l} = |\sup E\eta\zeta - E\eta E\zeta|$, where the supremum is taken over all η, ζ measurable with respect to $\mathscr{B}_1^j(u_n)$, $\mathscr{B}_{j+l}^n(u_n)$, respectively, $0 \le \eta$, $\zeta \le 1$, $1 \le j \le n-l$. Then $\beta_{n,l} \le \gamma_{n,l} \le 4\beta_{n,l}$, where $\beta_{n,l}$ is the mixing coefficient for Δ given by* (2.4.1). *In particular, $\{\xi_n\}$ satisfies $\Delta(u_n)$ if and only if $\gamma_{n,l_n} \to 0$ for some $l_n = o(n)$.*

It will be convenient to have the following simple notion of clusters. Divide the $\{\xi_i\}$ into successive groups $(\xi_1, \dots, \xi_{r_n}), (\xi_{r_n+1}, \dots, \xi_{2r_n}), \dots$ of r_n consecutive terms, where r_n $[= o(n)]$ is appropriately chosen. Then all exceedances of u_n within a group are regarded as forming a cluster. Note that since $r_n = o(n)$ the positions of the members of a single cluster will coalesce after the time normalization, giving nearly multiple points in the point process N_n on $(0,1]$. The following lemma (proved similarly to Lemma 2.1.1 but using Lemma 2.4.1—cf. [60]) shows that the clusters are asymptotically independent.

LEMMA 2.4.2. *Let $\tau > 0$ be constant and let $\Delta(u_n)$ hold with $u_n = u_n(\tau)$ satisfying* (2.2.1). *Suppose $\{k_n\}, \{l_n\}$ are sequences of integers such that $k_n l_n/n \to 0$ and $k_n\beta_{n,l_n} \to 0$, where $\beta_{n,l}$ is the mixing coefficient of $\Delta(u_n)$. Then, for each nonnegative continuous f on $(0,1]$,*

$$E\exp\left(-\sum_{j=1}^{n} f(j/n)\chi_{n,j}\right) - \prod_{i=1}^{k_n} E\exp\left(-\sum_{j=(i-1)r_n+1}^{ir_n} f(j/n)\chi_{n,j}\right) \to 0, \tag{2.4.2}$$

where $\chi_{n,j}$ is the indicator $1_{\{\xi_j > u_n\}}$ and $r_n = [n/k_n]$.

The number of exceedances in the ith cluster is $N_n(((i-1)r_n/n, ir_n/n]) = \sum_{j=(i-1)r_n+1}^{ir_n}\chi_{n,j}$ and the cluster-size distribution is therefore conveniently defined to be given by

$$\pi_n(i) = P\left\{\sum_{j=1}^{r_n}\chi_{n,j} = i \,\middle|\, \sum_{j=1}^{r_n}\chi_{n,j} > 0\right\}, \qquad i = 1, 2, \dots. \tag{2.4.3}$$

The following result of [60], giving sufficient conditions for N_n to have a compound Poisson limit, is proved by using Lemma 2.4.2 (cf. [60]).

THEOREM 2.4.3. *Let the stationary sequence* $\{\xi_n\}$ *have extremal index* $\theta > 0$ *and suppose that the conditions of Lemma* 2.4.2 *hold. If* $\pi_n(i)$ [*defined by* (2.4.3)] *has a limit* $\pi(i)$ *for each* $i = 1, 2, \dots,$ *then* π *is a probability distribution on* $1, 2, \dots$ *and the exceedance point process* N_n *converges in distribution to a compound Poisson process* N *with Laplace transform*

$$L_N(f) = \exp\left\{-\theta\tau\int_0^1\left(1 - \sum_{i=1}^{\infty} e^{-f(t)i}\pi(i)\right)dt\right\}. \tag{2.4.4}$$

The Laplace transform (2.4.4) is of the form (2.3.1) with the integer-valued r.v. β satisfying $P\{\beta = i\} = \pi(i)$ and intensity measure simply $\theta\tau m$, where m is Lebesgue measure. That is, N consists of multiple events of size whose distribution is $\pi(i)$, located at the points of a Poisson process having intensity $\theta\tau$.

The following result, showing that the compound Poisson process is the only possible limit for N_n under the conditions Δ is proved along similar lines to Theorem 2.4.3. (Full details may be found in [60].)

THEOREM 2.4.4. *Suppose the condition* $\Delta(u_n)$ *holds for* $u_n = u_n(\tau)$ *satisfying* (1.2.3) *for a* $\tau > 0$, *for the stationary sequence* $\{\xi_j\}$. *If* N_n *converges in distribution to some point process* N, *then the limit must be a compound Poisson process with Laplace transform* (2.4.4), *where* π *is some probability measure on* $\{1, 2, \dots\}$ *and* $\theta = -\tau^{-1}\log\lim_{n\to\infty}P\{M_n \le u_n(\tau)\} \in [0, 1]$. *If* $\theta \ne 0$, *then* $\pi(i) = \lim \pi_n\{i\}$, *where* π_n *is defined by* (2.4.3) *for* $r_n = [n/k_n]$, $k_n(\to \infty)$ *being any sequence chosen as in Lemma* 2.4.2.

EXAMPLE 2.2.2 (continued). It is evident that the exceedances of u_n by the process $\xi_j = \max(\eta_j, \eta_{j+1})$ in Example 2.2.2 occur in (at least) pairs, since if $\xi_{j-1} \le u_n$ but $\xi_j > u_n$, then $\eta_{j+1} > u_n$ and hence $\xi_{j+1} > u_n$. It is readily seen by direct evaluation that $\pi_n(2) \to 1$ and hence $\pi(i) = 1$ or 0 according as $i = 2$ or $i \ne 2$. Thus the limiting point process N consists entirely of double events and (2.4.4) gives $L_N(f) = \exp\{-(\tau/2)\int_0^1(1 - e^{-2f(t)})\,dt\}$.

The most important application of the compound Poisson limit is to give the asymptotic distribution of the kth largest value $M_n^{(k)}$ of $\xi_1, \dots, \xi_n$, when $\theta < 1$, using the equivalence $\{M_n^{(k)} \le u_n(\tau)\} = \{N_n((0, 1]) < k\}$.

THEOREM 2.4.5. *Suppose that for each* $\tau > 0$, $\Delta(u_n)$ *holds with* $u_n = u_n(\tau)$ *satisfying* (1.2.3) *and that* $N_n (= N_n^{(\tau)})$ *converges in distribution to some nontrivial point process* $N (= N^{(\tau)})$ (*which will occur, e.g., if the conditions of Theorem* 2.4.3 *hold*). *Assume that the maximum* M_n *has the nondegenerate asymptotic distribution* G *as given in* (1.2.1). *Then for each* $k = 1, 2, \dots,$

$$P\left\{a_n\left(M_n^{(k)} - b_n\right) \le x\right\} \to G(x)\left[1 + \sum_{j=1}^{k-1}\sum_{i=1}^{k-1}\left((-\log G(x))^j/j!\right)\pi^{*j}(i)\right] \tag{2.4.5}$$

[*zero if* $G(x) = 0$], *where* π^{*j} *is the j-fold convolution of the probability* $\pi = \lim \pi_n$, π_n *being given as in Theorem* 2.4.4.

Note that the form (2.4.5) differs from the (classical) case $\theta = 1$ [i.e., (1.2.5)], by inclusion of the convolution terms. These arise since, e.g., the second largest may be the second largest in the cluster where the maximum occurs or the largest in some other cluster. This contrasts with the case $k = 1$ for the maximum itself involving only the relatively minor change (Theorem 2.2.1) of replacing the classical limit by its θth power.

Finally, in this section we note that the "complete" convergence result, Theorem 1.3.1, still holds giving a Poisson limit under appropriate conditions, when $\{\xi_n\}$ has extremal index 1. However, as for the exceedance point process, the limit may undergo "compounding" when $\theta < 1$.

The possible limiting forms for N_n' (defined as in Theorem 1.3.1) were discussed first by Mori [75] under strong mixing conditions. More recently, a transparent derivation has been given by Hsing [57] under weaker conditions, of $\Delta(u_n)$ type but involving multiple levels $u_n(\tau_i)$. A derivation similar to that for the exceedance process shows that any limit in distribution of N_n', N' say, must have independent increments, be infinitely divisible and have certain stationarity properties. These properties restrict the Laplace transform of N' to a form that can be readily determined (though requiring further notation). It is also possible to give an illuminating "cluster representation," which exhibits N' as a Poisson process in the plane together with a countable family of points with integer-valued masses on vertical lines above and emanating from each Poisson point (cf. [58]).

As noted in Section 1.3, results of this type summarize the relevant information concerning asymptotic joint distributions of extreme order statistics, in contrast to the individual marginal distributions obtained in Theorem 2.4.5.

2.5. *Normal sequences*: *The comparison method*. For stationary normal sequences with covariances $\{r_n\}$, the condition $D(u_n)$ holds—as also does the sufficient condition (2.2.1) for the extremal index to be 1 provided the "Berman condition" holds, viz.,

$$r_n \log n \to 0, \quad \text{as } n \to \infty. \tag{2.5.1}$$

These results are simply proved by means of a widely used comparison method, which, in particular, bounds the difference between two (standardized) normal d.f.'s by a convenient function of their covariances. This result—here given in a general form ([66], page 81)—was essentially developed by Berman [10] and Slepian [94].

THEOREM 2.5.1 (Normal comparison lemma). *Suppose that* $\xi_1, \dots, \xi_n$ *are standard normal random variables with covariance matrix* $\Lambda^1 = (\Lambda^1_{ij})$ *and* $\eta_1, \dots, \eta_n$ *similarly with covariance matrix* $\Lambda^0 = (\Lambda^0_{ij})$, *and let* $\rho_{ij} =$

$\max(|\Lambda^1_{ij}|, |\Lambda^0_{ij}|)$. *Then, for any real numbers* $u_1, u_2, \dots, u_n$,

$$P\{\xi_j \le u_j,\ j = 1,2,\dots,n\} - P\{\eta_j \le u_j,\ j = 1,2,\dots,n\}$$

$$\le (2\pi)^{-1} \sum_{1 \le i < j \le n} \left(\Lambda^1_{ij} - \Lambda^0_{ij}\right)^+ \left(1 - \rho^2_{ij}\right)^{-1/2} \tag{2.5.2}$$

$$\times \exp\left[-\left(u_i^2 + u_j^2\right)/\left(2(1 + \rho_{ij})\right)\right],$$

where $x^+ = \max(x, 0)$. *Furthermore, replacing* $(\Lambda^1_{ij} - \Lambda^0_{ij})^+$ *by the absolute value on the right-hand side of* (2.5.2) *yields an upper bound for the absolute value of the difference on the left-hand side.*

By taking $\xi_1, \xi_2 \dots$ to be a stationary sequence of standardized normal r.v.'s with covariance sequence $\{r_n\}$ and $\eta_1, \eta_2 \dots$ to be i.i.d. standard normal r.v.'s it follows simply from the theorem that if $\sup_n |r_n| < 1$, then for any real sequence $\{u_n\}$,

$$\left|F_{i_1,\dots,i_s}(u_n) - \Phi^s(u_n)\right| \le Kn \sum_{j=1}^{n} |r_j| \exp\left(-u_n^2/(1 + |r_j|)\right), \tag{2.5.3}$$

where $F_{i_1,\dots,i_s}$ is the joint (normal) distribution of $\xi_{i_1}, \dots, \xi_{i_s}$ and Φ is the standard normal d.f., $i_1, \dots, i_s$ being any choice of distinct integers from $1, 2, \dots, n$.

Now if $n(1 - \Phi(u_n))$ is bounded and (2.5.1) holds, it can be shown (by some routine calculation) that the right-hand side of (2.5.3) tends to zero as $n \to \infty$, showing that $P\{\xi_{i_1} \le u_n, \dots, \xi_{i_s} \le u_n\}$ is approximately the same as it would be if the r.v.'s were i.i.d. instead of being correlated.

One can clearly (by identifying $i_1, \dots, i_s$ with $1, \dots, n$) then show directly that $P\{M_n \le u_n\}$ is approximately the same as for the i.i.d. standard normal sequence, leading to the following result of Berman [10]. This result may also be proved from Theorem 2.2.1 by verifying the condition $D(u_n)$ and (2.2.1).

THEOREM 2.5.2. *Let* $\{\xi_n\}$ *be a (standardized) stationary normal sequence with covariances* $\{r_n\}$ *such that* $r_n \log n \to 0$ *as* $n \to \infty$. *Then*

$$P\{a_n(M_n - b_n) \le x\} \to \exp(-e^{-x}),$$

where a_n, b_n *are given by* (1.2.4).

Thus if $r_n \log n \to 0$, the maximum M_n from the stationary normal sequence has precisely the same asymptotic distribution as an i.i.d. normal sequence. The same is true of the distributions of all extreme order statistics. Although a slight weakening of (2.5.1) is possible, this condition is close to being necessary for Theorem 2.5.2. Indeed, if $r_n \log n \to \gamma > 0$ and $u_n = x/a_n + b_n$ [with a_n, b_n given by (1.2.4)], then the time normalized point processes of exceedances converge in distribution to a certain doubly stochastic Poisson process. This leads to the asymptotic distribution of the maximum given by the convolution of a normal and type I extreme value distribution. (See [66], Section 6.5, for details.) Furthermore, Mittal and Ylvisaker [74] have shown that if $r_n \downarrow 0$ but $r_n \log n \to \infty$, then M_n has an asymptotic normal distribution. Thus in these "highly dependent" cases where $D(u_n)$ fails, the classical theory no longer applies.

As noted previously, stationarity has been assumed in many of the results to avoid the complications of notation and calculation, which a nonstationary framework entails. For normal sequences, however, the sufficient correlation conditions still remain quite simple in nonstationary cases. For example, the following result holds.

THEOREM 2.5.3. *Suppose that $\{\xi_n\}$ is a standardized normal sequence with correlations r_{ij} satisfying $|r_{ij}| \le \rho_{|i-j|}$ for $i \ne j$, where $\rho_n < 1$ for all n and $\rho_n \log n \to 0$ as $n \to \infty$. Let u_{ni}, $1 \le i \le n$, $n = 1, 2, \ldots$, be constants such that $\lambda_n = \min u_{ni} > c(\log n)^{1/2}$ for some $c > 0$. If for some $\tau \ge 0, \Sigma_1^n(1 - \Phi(u_{ni})) \to \tau$, then $P\{\bigcap_{i=1}^n(\xi_i \le u_{ni})\} \to e^{-\tau}$ as $n \to \infty$.*

Theorem 2.5.3 has a very useful corollary in the case where a sequence $\{\eta_n\}$ is obtained from a stationary normal sequence $\{\xi_n\}$ by adding a varying mean—such as a seasonal component or trend. Calculations then show that the double exponential limit for the maximum still holds, but the normalizing constant b_n can require an appropriate modification. Specifically, suppose that $\eta_i = \xi_i + m_i$, where $\{\xi_i\}$ is a standard (zero mean, unit variance) normal sequence (not necessarily covariance stationary) and m_i are added deterministic components with the property that

$$\beta_n = \max_{1 \le i \le n} |m_i| = o(\log n)^{1/2}, \quad \text{as } n \to \infty. \tag{2.5.4}$$

Under this condition it may be shown that a sequence of constants $\{m_n^*\}$ may be found such that

$$\frac{1}{n}\sum_{i=1}^{n} \exp\left(a_n^*(m_i - m_n^*) - \frac{1}{2}(m_i - m_n^*)^2\right) \to 1, \quad \text{as } n \to \infty, \tag{2.5.5}$$

in which $a_n^* = a_n - \log\log n/(2a_n)$, with a_n as in (1.2.4). With this notation, the following result holds.

THEOREM 2.5.4. *Let $\eta_i = \xi_i + m_i$ as before, where $\{\xi_n\}$ is a standard normal sequence with correlations r_{ij} satisfying $|r_{ij}| < \rho_{|j-i|}$ for $i \ne j$ with $\rho_n < 1$ and $\rho_n \log n \to 0$. Suppose that* (2.5.4) *holds and m_n^* satisfies* (2.5.5). *Then $M_n = \max(\eta_1, \eta_2, \ldots, \eta_n)$ satisfies*

$$P\{a_n(M_n - b_n - m_n^*) \le x\} \to \exp(-e^{-x}),$$

with a_n and b_n given by (1.2.4).

Thus the nonstationarity in the correlation structure has no effect on the limit law, and that introduced by the added deterministic component is adjusted for by the change of b_n to $(b_n + m_n^*)$. For details see [66], Chapter 6.

Normal processes provide a widely used source of models for describing physical phenomena, and it is gratifying that extremal theory applies so simply to them. Another convenient source of models is, of course, Markov chains, whose extremal behavior we discuss next.

2.6. *Regenerative and Markov sequences.* Most limit results for Markov chains are intimately tied to the theory of regenerative processes. For extreme values, this has been used in [2] and [10], some further references on extremes of Markov chains being [12], [78] and [91]. The "classical" case, exemplified by the GI/G/1 queue, is when a recurrent atom exists. However, recently regeneration techniques have been extended, in [6], [7] and [77], to show that any Harris recurrent chain $\{\eta_n\}$ on a general state space is regenerative or 1-dependent regenerative (concepts to be defined later), and to give effective criteria for regeneration. Furthermore, clearly a function $\xi_n = f(\zeta_n)$ of a (1-dependent) regenerative sequence is (1-dependent) regenerative. An example where this added generality is useful is given by ARMA(p, q)-processes. They are naturally considered as functions of a Markov chain in $\mathbb{R}^d$, for $d = \max(p, 1) + q$ and can be shown to be 1-dependent regenerative under weak conditions but usually not to be regenerative (cf. [91]).

Regenerative and 1-dependent regenerative sequences are strongly mixing, and hence the theory from Sections 2.1–2.4 applies; in particular, the extremal types theorem and the compound Poisson limit for exceedances hold. However, this can also be obtained directly, and the direct approach gives some added insight also into the results for general stationary sequences. In the present section this will be briefly outlined, along with some results directly tailored to Markov chains.

A sequence $\{\xi_t\colon t = 1, 2, \dots\}$ is regenerative if there exist integer-valued random variables $0 < S_0 < S_1 < \cdots$, which divide the sequence into "cycles"

$$c_0 = \{\xi_n\colon 0 \le n < S_0\}, \qquad c_1 = \{\xi_n\colon S_0 \le n < S_1\},$$

$$c_2 = \{\xi_n\colon S_1 \le n < S_2\}, \dots,$$

which are independent and such that, in addition, $c_1, c_2, \dots$ follow the same probability law. Then $\{S_k\}$ is a renewal process, i.e., $T_0 = S_0$, $T_1 = S_1 - S_0$, $T_2 = S_2 - S_1, \dots$ are independent and $T_1, T_2, \dots$ have the same distribution. We shall here assume that $m = ET_1 < \infty$ and that the distribution of T_1 is aperiodic, i.e., that the only integer for which $P(T_1 \in \{d, 2d, \dots\}) = 1$ is $d = 1$. The sequence $\{\xi_n\}$ is 1-dependent regenerative if there exists a renewal process $\{S_k\}$ as before, which makes $c_0, c_1, \dots$ 1-dependent (i.e., cycles separated by at least one cycle are independent) and $c_1, c_2, \dots$ stationary.

Suppose now that $\{\xi_n\colon n = 0, 1, \dots\}$ is a stationary regenerative sequence, let $\zeta_0 = \max\{\xi_i\colon\ 0 \le i < S_0\}$, $\zeta_1 = \max\{\xi_i\colon\ S_0 \le i < S_1\}$, $\zeta_2 = \max\{\xi_i\colon\ S_1 \le i < S_2\}, \dots$ be the cycle maxima and define $\nu_t = \inf\{k \ge 1\colon\ S_k > t\}$. By the law of large numbers $\nu_t/t \to 1/m$ a.s. and $M_n = \max\{\xi_1, \dots, \xi_n\}$ is easily approximated by $\max\{\zeta_1, \dots, \zeta_{\nu_n}\}$, which then in turn can be approximated by

$\max\{\zeta_1, \ldots, \zeta_{[n/m]}\}$. Since $\zeta_1, \zeta_2, \ldots$ are i.i.d., this can be shown to lead to

$$\sup_x |P(M_n \le x) - G^n(x)| \to 0, \quad \text{as } n \to \infty, \tag{2.6.1}$$

with $G(x) = P(\zeta_1 \le x)^{1/m}$, see, e.g., [9] and [91]. Since G is a d.f. it follows at once that the extremal types theorem holds for $\{\xi_n\}$, and criteria for domains of attraction are obtained by applying the criteria for i.i.d. variables to $G(x)$.

In particular, it follows from (2.6.1) and Lemma 1.2.2 that

$$P(M_n \le u_n) \to e^{-\eta}, \quad \text{as } n \to \infty, \tag{2.6.2}$$

if and only if $n(1 - G(u_n)) \to \eta$. As in Section 2.2 let $\hat{\xi}_1, \hat{\xi}_2, \ldots$ be the associated independent sequence, which has the same marginal d.f. F as $\xi_1, \xi_2, \ldots$ and write $\hat{M}_n = \max\{\hat{\xi}_1, \ldots, \hat{\xi}_n\}$. If, in addition, $n(1 - F(u_n)) = nP(\xi_1 > u_n) \to \tau > 0$, then $P(\hat{M}_n \le u_n) \to e^{-\tau}$ and $\{\xi_n\}$ hence has extremal index $\theta = \eta/\tau$. Since $1 - G(u_n) \sim P(\zeta_1 > u_n)/m$, this can be formulated as follows. If there exists a sequence $\{u_n\}$ such that $n(1 - F(u_n)) \to \tau > 0$ and

$$\frac{P(\zeta_1 > u_n)/m}{P(\xi_1 > u_n)} \to \theta > 0, \tag{2.6.3}$$

then $\{\xi_t\}$ has extremal index θ. In the same way it can be seen that conversely if $\{\xi_t\}$ has extremal index $\theta > 0$, then for any $\tau > 0$ there exists a sequence $\{u_n\}$, which satisfies $n(1 - F(u_n)) \to \tau$ and (2.6.3). However, it should be noted that there are examples of regenerative sequences $\{\xi_t\}$ that satisfy (2.6.2), even for $u_n = u_n(x) = x/a_n + b_n$ for all x, but for which $(P(\zeta_1 > u_n)/m)/P(\xi_1 > u_n)$ does not converge, and hence the extremal index does not exist, even if this is not expected to occur in cases of practical interest.

A counterpart to the compound Poisson limit theorem 2.4.3 for the exceedance point process N_n, given by $N_n(E) = \#\{i/n \in E: \xi_i > u_n\}$, is also easy to obtain for stationary regenerative sequences. Let N_n' be the point process on $(0,1]$, which has points of multiplicity $\gamma_i = \#\{t: \xi_t > u_n, S_{i-1} \le t < S_i\}$ at i/n for each i for which $\gamma_i > 0$, i.e., N_n' is defined by $N_n'(E) = \sum_{i/n \in E} \gamma_i$. Then $\{\gamma_i\}_{i=1}^{\infty}$ is an i.i.d. sequence, and if (2.6.2) holds so that $nP(\gamma_1 > 0) = nP(\zeta_1 > u_n) \to \eta m$ and if

$$\pi_n(i) = P(\gamma_1 = i | \gamma_1 > 0) \to \pi(i), \quad \text{as } n \to \infty, \tag{2.6.4}$$

for all i for some $\{\pi(i);\ i = 1, 2, \ldots\}$, then it follows at once that N_n' converges in distribution to a compound Poisson process N' with Laplace transform $L_{N'}(f) = \exp\{-\eta m \int_0^1 (1 - \sum_{i=1}^{\infty} e^{-f(t)i} \pi(i))\, dt\}$. By definition, a nonzero γ_i corresponds to a cluster of γ_i exceedances of u_n by ξ_t for $S_{i-1} \le t < S_i$, and since $S_i/i \to m$ as $i \to \infty$, there is hence a cluster of γ_i points located approximately at mi/n in N_n. Hence for an interval E, $N_n(E)$ is approximated by $N_n'(m^{-1}E)$ (for $m^{-1}E = \{x: mx \in E\}$) and asymptotically $N_n(E)$ should have the same distribution as $N'(m^{-1}E)$. This argument can easily be extended and made stringent, to give the following result of [91] (cf. also [92]).

THEOREM 2.6.1. *Let $\{\xi_n: n = 0, 1, \dots\}$ be a stationary aperiodic regenerative sequence with $m < \infty$ and let $\{u_n\}$ be constants such that*

$$nP(\gamma_1 > 0)/m = nP(\zeta_1 > u_n)/m \to \eta$$

and (2.6.4) *holds. Then N_n converges in distribution to a compound Poisson process N with Laplace transform* (2.4.4), *i.e.,*

$$L_N(f) = \exp\left\{-\eta \int_0^1 \left(1 - \sum_{i=1}^{\infty} e^{-f(t)i}\pi(i)\right) dt\right\}.$$

These results may also be extended to 1-dependent regenerative sequences, however with some extra complexity. Here we mention that the criterion (2.6.3) for the extremal index to be θ then is replaced (cf. [91]) by

$$\frac{P(\zeta_1 \le u_n, \zeta_2 > u_n)/m}{P(\zeta_1 > u_n)} \to \theta. \tag{2.6.5}$$

In [91], (2.6.5) is further used to find conditions for $\theta = 1$ for a function $\xi_t = f(\eta_t)$ of a Markov chain on a general state space. This result is expressed directly in terms of the transition probabilities $P_n(x) = P(f(\eta_1) > u_n | \eta_0 = x) = P(\xi_1 > u_n | \eta_0 = x)$ as follows.

THEOREM 2.6.2. *Let $\{\eta_n\}$ be a stationary regenerative Markov chain with the cycle length T_1 aperiodic and satisfying $ET_1^\alpha < \infty$ for some $\alpha > 1$. If $u_n = u_n(\tau)$ satisfies* (1.2.3) *for some $\tau > 0$ and*

$$E\big(P_n(\eta_0)^s\big)n^{1+s/\alpha} \to 0, \quad \textit{as } n \to \infty,$$

for some $s > 1$ with $1/\alpha + 1/s < 1$, then $\{\xi_n\}$ has extremal index $\theta = 1$.

We also refer to [91], Theorem 4.1, and [78], Theorem 2.1, for additional results on the extremal index and compound Poisson convergence, for general distributionally mixing sequences, in a form that is particularly convenient for applications to Markov chains. Finally, the restriction that the Markov chain (or regenerative sequence) is started with the stationary initial distribution is not essential. All the results hold for arbitrary initial distributions, provided only that

$$P(\zeta_0 > \max\{\zeta_1, \dots, \zeta_k\}) \to 0, \quad \text{as } k \to \infty.$$

2.7. *Moving averages.* Here, a stationary sequence $\{\xi_t\}$ is a moving average if it can be written in the form

$$\xi_t = \sum_{i=-\infty}^{\infty} c_i \zeta_{t-i}, \qquad t = 0, \pm 1, \dots, \tag{2.7.1}$$

where $\{\zeta_t\}$ is an i.i.d. sequence (the "noise sequence") and $\{c_i\}$ is a sequence of constants (the "weights") and where the sums are assumed to converge with probability 1. If a stationary normal sequence has a spectral density—this holds, e.g., if $\Sigma r_t^2 < \infty$, it can be represented in a nonunique way, as a moving average with normally distributed ζ's. Furthermore, (2.7.1) includes the ARMA-processes (which satisfy a finite linear difference equation in the ζ's and hence also are multidimensional Markov chains), which are extensively used in time-series analysis. Thus, in particular, some of the themes from Sections 2.5 and 2.6 will be taken up again here, but from a slightly different point of view.

The extremal behavior of $\{\xi_t\}$ depends on both the weights and the two tails of the marginal d.f. of the noise variables in an intricate and interesting way. To reduce the amount of detail, we shall only describe the asymptotic distribution of the maxima, for the case of *nonnegative* c_i's. The general case involves some extra complexity, since then an extreme negative noise variable, which is multiplied by a negative c_i, may contribute to a large ξ_t-value. In addition to this, the references cited later prove point process convergence and give rather detailed information on the sample path behavior near extremes, including the clustering that occurs when the extremal index is less than one. Here we will only exhibit the limiting form of the sample paths near extreme values without going into technicalities, referring to [86], [89] and [36] for further details.

In cases when (1.2.1) holds, i.e., when

$$P\{a_n(M_n - b_n) \le x\} \to_d G(x), \quad \text{as } n \to \infty, \tag{2.7.2}$$

the asymptotic behavior of the maximum is specified by the constants $a_n > 0$, b_n and the d.f. G. However, this involves an arbitrary choice, since if a_n, b_n are replaced by a'_n, b'_n, where $a_n/a'_n \to a > 0$ and $a_n(b'_n - b_n) \to b$, then (2.7.2) still holds, but with $G(x)$ replaced by $G(ax + b)$. In the sequel we will keep the G's fixed, as the standard d.f.'s displayed in Theorem 1.2.1 and hence describe extremal behavior by a_n, b_n and the type of G.

Extremal behavior of the moving average $\{\xi_t\}$ can be put into perspective by comparing with extremes of the noise sequence and of the associated i.i.d. sequence $\{\hat{\xi}_t\}$ with the same marginal d.f. $\{\xi_t\}$. Specifically, for $\tilde{M}_n = \max\{\zeta_1, \dots, \zeta_n\}$ and $\hat{M}_n = \{\hat{\xi}_1, \dots, \hat{\xi}_n\}$ there are norming constants $\tilde{a}_n$, $\hat{a}_n > 0$ and $\tilde{b}_n$, $\hat{b}_n$ such that for the cases we consider here,

$$P\{\tilde{a}_n(\tilde{M}_n - \tilde{b}_n) \le x\} \to G(x) \tag{2.7.3}$$

and

$$P\{\hat{a}_n(\hat{M}_n - \hat{b}_n) \le x\} \to G(x), \tag{2.7.4}$$

with the same G as in (2.7.2), and we shall indicate the relations between the different norming constants.

The articles by Rootzén [86] and Davis and Resnick [36] are concerned with noise variables that are in the domain of attraction of the type II extreme value distribution, or, equivalently, when the noise variables have a regularly varying tail,

$$P(\zeta_0 > x) = x^{-\alpha}L(x), \tag{2.7.5}$$

with $\alpha > 0$ and L slowly varying at infinity. Hence, using the prescription for norming constants given after Lemma 1.2.2, if γ_n satisfies $P(\zeta_0 < \gamma_n) \le 1 - 1/n \le P(\zeta_0 \le \gamma_n)$, so that γ_n is roughly of the order $n^{1/\alpha}$, then (2.7.3) holds, with

$$\tilde{a}_n = \gamma_n^{-1}, \qquad \tilde{b}_n = 0,$$
$$G(x) = \exp(-x^{-\alpha}), \qquad x > 0.$$

Let $c_+ = \max\{c_i;\ i = 0, \pm 1, \dots\}$. Then also (2.7.2) is satisfied, with

$$a_n = c_+^{-1}\tilde{a}_n, \qquad b_n = 0, \tag{2.7.6}$$
$$G(x) = \exp(-x^{-\alpha}).$$

This is elegantly proved in [36], by first noting that complete Poisson convergence of extremes of the ζ-sequence is immediate (cf. Section 1.3) and then obtaining the corresponding result for the ξ's by a "continuous mapping" and approximation argument. [36] uses some summability assumptions on the c_i's, and for convenience that $c_i = 0$ for $i = -1, -2, \dots$. However, it seems clear that the results hold without any restrictions beyond the assumption that the sums in (2.7.1) converge (cf. [86]).

An intuitive explanation of (2.7.6) is that when the tails of the noise variables decrease slowly, as in (2.7.5), then the extreme noise values are very much larger than the typical ones, and that hence the maximal ξ_t-value asymptotically is achieved when the largest ζ_t-value is multiplied by the largest weight c_+. This, of course, agrees with (2.7.6), since the norming constants there are the same as those that apply to $\max\{c_+\zeta_1, \dots, c_+\zeta_n\}$. These heuristic arguments also easily lead to the following form of the normalized sample path $\xi_{t+\tau}/\xi_\tau$ near an extreme value at, say, the time point τ; asymptotically, this ratio has the same distribution as the function y_t given by

$$y_t = Uc_{-t}, \tag{2.7.7}$$

where U is a certain random variable which takes values in the set $\{\dots 1/c_{-1}, 1/c_0, 1/c_1, \dots\}$. Thus, except for a random height, sample paths near extremes are asymptotically deterministic.

The special case of (2.7.5) when the noise variables are stable (or "sum-stable," as opposed to max-stable) was studied first in [86]. It has the appealing feature that then also the moving average, and indeed all linear functions of the noise variables, are jointly stable. For such variables, it is easily seen that (2.7.4) holds, with

$$\hat{a}_n = \left(\sum c_i^\alpha\right)^{1/\alpha}\tilde{a}_n, \qquad \hat{b}_n = 0,$$
$$G(x) = \exp\{-x^{-\alpha}\},$$

and hence also that the extremal index is $c_+^\alpha/\Sigma c_i^\alpha$, for the case of nonnegative c's discussed here. Although not considered in [36], this can be shown to hold also for the general case (2.7.5), provided the sums involved are convergent.

The other class of moving averages, which has been studied in [89], is specified by

$$P(\zeta_0 > x) \sim Kx^\alpha e^{-x^p}, \quad \text{as } x \to \infty, \tag{2.7.8}$$

where K, $p > 0$ and α are constants. Again it follows, using Lemma 1.2.2, that (2.7.3) holds, with

$$\tilde{a}_n = p(\log n)^{1-1/p},$$
$$\tilde{b}_n = (\log n)^{1/p} + p^{-1}((\alpha/p)\log\log n + \log K)(\log n)^{1/p-1},$$
$$G(x) = \exp\{-e^{-x}\}.$$

Thus the center of the distribution of $\tilde{M}_n$ tends to infinity roughly as $(\log n)^{1/p}$, and the "scale parameter" a_n^{-1} is of the order $(\log n)^{1/p-1}$, which shows that for $p > 1$ the distribution of $\tilde{M}_n$ becomes more and more concentrated as $n \to \infty$, and that it becomes increasingly spread out for $0 < p < 1$, whereas the order of the scale does not change for $p = 1$. As we shall see, the same holds for $\hat{M}_n$ and M_n.

The case when (2.7.8) holds with $p = 1$ leads to intermediate behavior, and we will only discuss the remaining cases. For $0 < p < 1$ again a large ξ-value is caused by just one large noise variable, in a similar way to the behavior when (2.7.5) holds. However, the nonzero $\tilde{b}_n$-terms cause some extra complications. Thus (2.7.2) holds with

$$a_n = c_+^{-1}\tilde{a}_n, \qquad b_n = c_+\tilde{b}_n,$$
$$G(x) = \exp\{-e^{-x}\},$$

in analogy with (2.7.6), but, writing k for the number of i's with $c_i = c_+$, the appropriate version of (2.7.4) involves

$$\hat{a}_n = c_+^{-1}\tilde{a}_n, \qquad \hat{b}_n = c_+\left(\tilde{b}_n + (\log k)/\tilde{a}_n\right),$$
$$G(x) = \exp\{-e^{-x}\}.$$

Also the asymptotic form of the sample path $\xi_{t+\tau}/\xi_\tau$ near an extreme value at τ is similar. For $k = 1$ it is given by the deterministic function

$$y_t = c_{-t}/c_+,$$

whereas in the general case it is a random translate of this.

The case when (2.7.8) holds with $p > 1$ is more intricate, since then an extreme ξ-value is caused by many moderately large noise variables in conjunction, and since extremal behavior is determined by the constant $\|c\|_q = (\Sigma|c_i|^q)^{1/q}$ and the function

$$y_t = \sum_i c_{i-t}c_i^{q/p}/\|c\|_q^q, \tag{2.7.9}$$

with $q = (1 - 1/p)^{-1}$. In fact, the normalized sample path $\xi_{t+\tau}/\xi_\tau$ near an extreme at τ asymptotically has the deterministic form (2.7.9), and (2.7.2) and (2.7.4) hold, with

$$\begin{aligned} a_n = \hat{a}_n = \|c\|_q^{-1}\tilde{a}_n, \qquad b_n = \hat{b}_n,\\ G(x) = \exp\{-e^{-x}\}. \end{aligned} \tag{2.7.10}$$

Here $b_n = \hat{b}_n$ is not determined by (2.7.8) alone; except for finite moving averages, it is also influenced by the center of the distribution of the ζ's. However, it is roughly of the order $\|c\|_q\tilde{b}_n$, but still $a_n|(b_n - \|c\|_q\tilde{b}_n)|$ may, in general, tend to infinity. It, of course, follows at once from (2.7.10) that the extremal index is one for $p > 1$.

For $p = q = 2$, which includes the normal case, (2.7.9) is the correlation function and $\|c\|_q$ is proportional to the standard deviation, in agreement with Section 2.5. However, it is interesting to note that for $p \neq 2$ covariances seem to have little bearing on extremes.

The results for the case (2.7.8) use the assumption that $|c_i| = 0(|i|^{-\eta})$, for some $\eta > \max(1, 2/q)$, and for $p > 1$ in addition a number of smoothness restrictions on the distribution of the noise variables. These are mainly used in the derivations of the behavior of the tail of $\xi_0 = \Sigma c_i \zeta_{-i}$, which for $p > 1$ is the main difficulty (cf. [90]). It is fairly easy to see that $D(u_n)$ holds for all the moving averages considered here, and the results above for $p > 1$ are obtained along the lines set out in Section 2.2 by verifying (2.2.1). For $0 \le p \le 1$, i.e., in the cases when θ may be less than one, the proofs use ad hoc methods, closely related to the heuristic arguments given previously.

Finally, it should be mentioned that Finster [45] obtains some related results using autoregressive representations of the processes and Chernick [25] provides an example with qualitatively different behavior.

2.8. *Rates of convergence.* Rates of convergence for the distribution of the maximum have mainly been studied for i.i.d. variables. In the present section we briefly review this work, discussing in turn pointwise rates, uniform convergence of d.f.'s, so-called "penultimate" approximations, uniform convergence over the class of all sets and "large deviation" type results. Although generalizations seem straightforward, the only dependent sequences that have been considered are the normal ones. The quite precise results available for this case are discussed at the end of the section. A useful general observation, which applies to i.i.d. and dependent cases with extremal index $\theta = 1$, is that once rates of convergence of the maximum have been found, then it is typically quite easy to find similar rates for kth order statistics.

For i.i.d. random variables and a given u_n, the error $P(M_n \le u_n) - e^{-\tau}$ in the approximation (1.2.2) is easy to compute directly, since then $P(M_n \le u_n) = F^n(u_n)$, where F is the common d.f. of the variables. Furthermore, if F is continuous one can always make the difference zero for any $n, \tau > 0$ [by taking $u_n = F^{-1}(e^{-\tau/n})$]. However, often u_n is determined from other considerations, e.g., in (1.2.1) it is chosen as $u_n = u_n(x) = x/a_n + b_n$ and correspondingly $\tau = \tau(x) = -\log G(x)$. Then the behavior of the approximation error

$$\Delta_n(x) = P(M_n \le u_n(x)) - e^{-\tau(x)},$$

perhaps over a range of x-values, and, in particular, of

$$d_n(a_n, b_n) = \sup_x |\Delta_n(x)| = \sup_x |P(a_n(M_n - b_n) \le x) - G(x)|$$

is less immediate. If (1.2.1) is used as an approximation or, more importantly, if it motivates statistical procedures when a_n, b_n have to be estimated, interest centers on which rate of decrease is attainable when the "best" a_n, b_n are used, i.e., on

$$d_n = \inf_{a>0,\, b} d_n(a, b) = \inf_{a>0,\, b} \sup_x |P\{a(M_n - b) \le x\} - G(x)|.$$

It is easy to give examples of distributions F for which d_n tends to zero arbitrarily slowly, and to any exponential rate there is an F that achieves this rate. However, faster than exponential decrease of d_n implies that F is max-stable, and then $d_n = 0$ for all n ([8] and [88]). Also different standard distributions give quite different rates, e.g., for the normal distribution d_n is of the order $1/\log n$, whereas for the uniform and exponential distributions the order is $1/n$.

Let $\tau_n = \tau_n(x) = n(1 - F(u_n(x)))$. In the sequel we will usually, for brevity, delete the explicit dependence on x. An obvious approach to analyzing Δ_n $[= \Delta_n(x)]$ in the i.i.d. case is to introduce

$$\Delta'_n = (1 - \tau_n/n)^n - e^{-\tau_n}, \qquad \Delta''_n = e^{-\tau_n} - e^{-\tau},$$

so that

$$|\Delta_n| = |F(u_n)^n - e^{-\tau}| = |(1 - \tau_n/n)^n - e^{-\tau}| \le |\Delta'_n| + |\Delta''_n|. \tag{2.8.1}$$

Here $0 \le \tau_n \le n$, and for such values the satisfying uniform bound

$$|\Delta'_n| \le n^{-1}(a + n^{-1})e^{-2} \tag{2.8.2}$$

is derived by Hall and Wellner [55]. Furthermore, for fixed τ, by Taylor's formula

$$|\Delta''_n| \sim e^{-\tau}|\tau_n - \tau|, \tag{2.8.3}$$

as $\tau_n \to \tau$. However, (2.8.3) is only uniform for $\tau_n - \tau = \tau_n(x) - \tau(x)$ in intervals that are bounded from below, and to bound d_n a further argument has to be added. Often this runs as follows; (2.8.2) and (2.8.3) give sharp estimates of $\sup_{x > a}|\Delta_n(x)|$ for any $a > x_0$, where x_0 is the left-hand endpoint of the d.f. G, and then also for $\sup_{x > x_n}|\Delta_n(x)|$ if x_n is taken to converge to x_0 suitably slowly. Combining this with

$$\sup_{x \le x_n} |\Delta_n(x)| \le \max\{F^n(x_n/a_n + b_n), G(x_n)\} \tag{2.8.4}$$

leads to a bound for $d_n(a_n, b_n)$, and then, by varying a_n, b_n, to bounds for d_n. This approach is used, with some variations, by Hall and Wellner [55], Davis [33], Cohen [26] and [27] and Leadbetter, Lindgren and Rootzén [66]. Here the bounds corresponding to (2.8.2) and (2.8.3) are asymptotically sharp, but there is a possibility that Δ'_n and Δ''_n can at least partially cancel. However, this happens only if $\tau_n = \tau - \tau^2/(2n) + o(1/n)$, and hence in fairly special cases, as is readily seen (cf. Davis [33]).

A number of papers, some of the later references being Cohen [26] and [27], Smith [95] and Resnick [84], have introduced conditions that permit more explicit bounds than (2.8.1)–(2.8.4) to be calculated. Their approach is to take some set of conditions for attraction to an extreme value distribution, typically involving convergence of some quantity related to the tail of F, and show that if this holds at a specific rate then $d_n(a_n, b_n)$ converges at a corresponding rate. In this a set of simple sufficient conditions due to von Mises [98] (cf. [66], page 16) have been particularly useful. There are many possible versions of such conditions, and hence many partially overlapping results have been obtained. As a typical example, we cite the following result of Resnick [84].

Suppose F is differentiable and that there exists a continuous function g that tends monotonically to zero and satisfies

$$\left| \frac{xF'(x)}{F(x)(-\log F(x))} - \alpha \right| \le g(x), \qquad x > 0, \tag{2.8.5}$$

for some $\alpha > 0$. Then, if a_n is chosen to satisfy $-\log F(a_n^{-1}) = n^{-1}$,

$$\sup_{x \ge 1} |F^n(x/a_n) - \exp\{-x^{-\alpha}\}| \le 0.2701 g(a_n^{-1})/(\alpha - g(a_n^{-1})),$$

for n such that $g(a_n^{-1}) < \alpha$. Here (2.8.5) is a slight variation of von Mises' condition for attraction to the type II extreme value distribution, and the proof is somewhat different from the method outlined previously, the main ingredient being an estimate of $-\log(-\log F(x))$. Resnick also obtains a somewhat more complicated bound for the supremum $d_n(a_n, 0)$ over all x.

For i.i.d. variables bounds on the rate of convergence of the maximum automatically lead to bounds for the rate of convergence also of kth largest values. This follows as in (1.2.5), by using any of the known bounds for the difference between the binomial and Poisson distributions, since S_n is binomial with parameters n, τ_n/n (see, e.g., [66], Section 2.4).

The normal case, briefly mentioned above, of course, has attracted special attention. Straightforward calculations show that for a_n, b_n given by (1.2.4)

$$\Delta_n(x) \sim \left[\exp(-e^{-x}) e^{-x} (\log\log n)^2\right] / (16 \log n), \quad \text{as } n \to \infty,$$

and in Hall [53] it is shown that for i.i.d. normal variables there are constants $0 < c_1 < c_2 < 3$ such that $c_1/\log n \le d_n \le c_2/\log n$ for $n \ge 3$, i.e., the best rate of convergence is of the disconcertingly slow order $1/\log n$. However, this is partially offset by the fact that d_n is, nevertheless, fairly small for small n, e.g., for $n \le 10{,}000$ it compares well with the error in the normal approximation to the binomial distribution.

In their pioneering paper [46], Fisher and Tippett had already noticed the slow convergence rate for the normal case, and suggested improved "penultimate" approximations. The idea is that since the type I extreme value d.f. can be approximated arbitrarily well by type II (or type III) d.f.'s, if a d.f. can be approximated by a type I d.f., the same error can (in the limit) be achieved by a type II (or III) d.f., and there is always a possibility they can do better. This has been further developed by Cohen [26] and [27], who, in particular, shows that a penultimate approximation of the maximum of normal random variables by a type II extreme value d.f. improves the rate of convergence to $1/(\log n)^2$. The disadvantage with this approach is that the exponent α in the approximating d.f. then has to be chosen differently for different values of n. A related approach is to consider a function $|M_n|^a$ instead of M_n itself. This is pursued in Hall [54] and Haldane and Jayakar [52] and gives the rate of convergence $1/(\log n)^2$ for $a = 2$, whereas other values of a lead to the same order $1/\log n$ as for M_n itself. Numerical computations show that these approximations also do better for small and moderate values of n, as could be expected.

A further statistically relevant question is to find rates of uniform convergence, i.e., to bound

$$d'_n = \inf_{a>0,\, b} \sup_{B\in\mathscr{B}} |P(a(M_n - b) \in B) - G(B)|,$$

where $\mathscr{B}$ denotes the Borel sets in R, and $G(B)$ is the probability that a random variable with d.f. G belongs to B. The obvious approach is to bound the difference between the density (which is assumed to exist) of $a(M_n - b)$ and G'. Let $G'(x) = G(x)\gamma(x)$, so that $\gamma(x) = e^{-x}$, $\alpha x^{-\alpha-1}$ and $\alpha(-x)^{\alpha-1}$ for the type I, II and III extreme value distributions, respectively. Since (for i.i.d. variables),

$$\begin{aligned} d/dx\, P\{a_n(M_n - b_n) \le x\} &= d/dx\, F^n(x/a_n + b_n) \\ &= F(x/a_n + b_n)^{n-1} n a_n^{-1} F'(x/a_n + b_n), \end{aligned}$$

where the first factor tends to G at a rate given by the references cited above, the main problem is to bound the difference $na^{-1}(F'(x/a + b) - \gamma(x)$. The recent thesis by Falk [44] contains a survey of results in this direction, some further recent work being that of de Haan and Resnick [41] and Weissman [100].

Another problem, which has attracted some attention partly because of reliability applications, is the uniformity of the convergence of

$$P\{a_n(M_n - b_n) > x\}/(1 - G(x))$$

for large x; see Anderson [2] and de Haan and Hordijk [39].

For a stationary dependent sequence with extremal index $\theta = 1$, a further source of error is the approximation by the associated independent sequence, i.e., the difference

$$\Delta_n'''(x) = P\{a_n(M_n - b_n) \le x\} - F^n(x/a_n + b_n),$$

where F is the marginal d.f. of the sequence. Cohen [26] shows, under weak covariance conditions, that for a stationary normal sequence Δ_n''' is $o(1/\log n)$, and hence that the rate of convergence in (1.2.1) is determined by the difference $F^n(x/a_n + b_n) - G(x)$, and hence is the same as in the i.i.d. case. Let ρ be the maximal correlation in the stationary normal sequence. Rootzén [87] gives a first order approximation and bounds for Δ_n''' that are roughly of the order $1/n^{(1-\rho)/(1+\rho)}$ for $\rho \ge 0$.

By using an embedding technique, these rates are extended also to $M_n^{(k)}$ and to point processes of exceedances. This embedding can be used more generally, and hence also in dependent cases rates for the maximum often easily lead to similar rates for kth largest values.

2.9. *Multivariate extremes.* We shall discuss here only one multivariate problem, the extremal types theorem for i.i.d. random vectors, and its extension to dependent sequences. As shown by de Haan and Resnick [40] and Pickands [82], the problem of characterizing the possible limit laws of the vector of coordinatewise maxima splits into two independent problems, to find the marginal d.f.'s that may occur—by the one-dimensional result this is just the class of extreme value d.f.'s—and to characterize the limiting dependence between components. Following Deheuvels [42] and Hsing [59], we will use the concept of dependence functions to discuss this.

Let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)$ be a d-dimensional random vector with d.f. $\mathbf{G}$ and marginal d.f.'s G_j, $1 \le j \le d$. The *dependence function* $\mathbf{D}$ of $\boldsymbol{\xi}$ (or of $\mathbf{G}$) is defined by

$$\mathbf{D}(x_1, \dots, x_d) = P\{G_1(\xi_1) \le x_1, \dots, G_d(\xi_d) \le x_d\}.$$

$\mathbf{D}$ is the d.f. of a distribution on $[0,1]^d$, and it has uniform marginal distributions if the G_j's are continuous. The marginal distributions together with the dependence function determine $\mathbf{G}$, since

$$\mathbf{G}(x_1, \dots, x_d) = \mathbf{D}(G_1(x_1), \dots, G_d(x_d)), \qquad x_1, \dots, x_d \in R. \tag{2.9.1}$$

This is a consequence of the relation

$$\begin{aligned} \{G_j(\xi_j) \le G_j(x_j);\, 1 \le j \le d\} &\setminus \bigcup_{j=1}^{d} \{G_j(\xi_j) \le G_j(x_j),\, \xi_j > x_j\} \\ &\subset \{\xi_j \le x_j;\, 1 \le j \le d\} \subset \{G_j(\xi_j) \le G_j(x_j),\, 1 \le j \le d\}, \end{aligned}$$

since it is readily seen that $P\{G_j(\xi_j) \le G_j(x_j),\ \xi_j > x_j\} = 0$ for each j.

A further useful property is that convergence of d-dimensional distributions is equivalent to convergence of the dependence function and the marginal distributions, provided the limit has continuous marginal d.f.'s. This can be proved rather easily using (2.9.1). Similarly to the one-dimensional case, a d-dimensional d.f. $\mathbf{G}$ is said to be max-stable if there exist constants $a_{n,i} > 0$, $b_{n,i}$, $i = 1, \dots, d$, such that

$$\mathbf{G}^n(a_{n,1}x_1 + b_{n,1}, \dots, a_{n,d}x_d + b_{n,d}) = \mathbf{G}(x_1, \dots, x_d), \qquad x_1, \dots, x_d \in R, \tag{2.9.2}$$

for each $n = 1, 2, \dots$. Furthermore, a dependence function $\mathbf{D}$ is max-stable if

$$\mathbf{D}^n(x_1^{1/n}, \dots, x_d^{1/n}) = \mathbf{G}(x_1, \dots, x_d), \qquad x_1, \dots, x_d \in R, \tag{2.9.3}$$

for $n = 1, 2, \dots$.

THEOREM 2.9.1. *A d-dimensional $(d \ge 2)$ d.f. with nondegenerate marginal distributions is max-stable if and only if its marginal d.f.'s and its dependence function are nondegenerate max-stable.*

PROOF. If $G_1, \dots, G_d$ are max-stable, or if $\mathbf{G}$ is max-stable, then there are constants $a_{n,i} > 0$, $b_{n,i}$ with $G_i^n(a_{n,i}x + b_{n,i}) = G_i(x)$, for $i = 1, \dots, d$. Hence, in either case,

$$\begin{aligned} &\mathbf{G}^n(a_{n,1}x_1 + b_{n,1}, \dots, a_{n,d}x_d + b_{n,d}) \\ &\quad = \mathbf{D}^n(G_1(a_{n,1}x_1 + b_{n,1}), \dots, G_d(a_{n,d}x_d + b_{n,d})) \\ &\quad = \mathbf{D}^n\big(G_1(x_1)^{1/n}, \dots, G_d(x_d)^{1/n}\big). \end{aligned} \tag{2.9.4}$$

Thus (2.9.2) follows at once if $\mathbf{D}$ is max-stable, by (2.9.1). The converse, i.e., that

$\mathbf{G}^n(y_1^{1/n},\ldots,y_d^{1/n}) = \mathbf{D}(y_1,\ldots,y_d)$ for $y_i \in (0,1)$, $i = 1,\ldots,d$, if $\mathbf{G}$ is max-stable also follows from (2.9.4), by taking $x_i = G_i^{-1}(y_i)$ there (note that each G_i is nondegenerate max-stable and hence continuous and strictly increasing on its support). □

Let $\{\boldsymbol{\xi}_n\} = \{(\xi_{n,1},\ldots,\xi_{n,d})\}_{n=1}^{\infty}$ be a sequence of i.i.d. random vectors, write $M_{n,i} = \max\{\xi_{1,i},\ldots,\xi_{n,i}\}$ and suppose there are constants $a_{n,i} > 0$, $b_{n,i}$ such that

$$P\{a_{n,i}(M_{n,i} - b_{n,i}) \le x_i, 1 \le i \le d\} \to_d \mathbf{G}(x_1,\ldots,x_d), \tag{2.9.5}$$

where we may assume without loss of generality that the marginal distributions of $\mathbf{G}$ are nondegenerate. It then follows exactly as in the one-dimensional case that the possible limits $\mathbf{G}$ in (2.9.5) are precisely the max-stable d.f.'s. Thus by Theorem 2.9.1 each marginal d.f. is max-stable and hence one of the three extreme value types, and the dependence function is max-stable. Furthermore, the distribution of $a_{n,i}(M_{n,i} - b_{n,i})$ tends to G_i, for $i = 1,\ldots,d$ and the dependence function of $\{M_{n,i}\colon i = 1,\ldots,d\}$ converges to the dependence function of $\mathbf{G}$. To complete the characterization of the limits, it only remains to describe the max-stable dependence functions. Again, this is a purely analytical problem, to solve the functional equation (2.9.3), and we thus only cite the result, which is obtained in somewhat varying forms in [40], [82], [42] and [59].

THEOREM 2.9.2. *A function* $\mathbf{D}$ *on* $[0,1]^d$ *is a max-stable dependence function if and only if it has the representation*

$$\mathbf{D}(y_1,\ldots,y_d) = \exp\left\{\int_S \min_{1\le i\le d}\{x_i \log y_i\}\, d\mu\right\},$$

where S is the simplex $\{(x_1,\ldots,x_d)\colon x_i \ge 0,\ i = 1,\ldots,d,\ \Sigma_i^d x_i = 1\}$, *for some finite measure* μ *on S that satisfies* $\int_S x_i\, d\mu = 1$ *for* $i = 1,\ldots,d$.

Hsing [59] also makes the observation that whereas the characterization of the limiting marginal d.f.'s is crucially tied to linear normalizations, this is not so for the dependence function. Specifically, if $\{u_{n,i}(x)\}$ are levels that are continuous and strictly increasing in x, and if

$$P\{M_{n,i} \le u_{n,i}(x_i),\ i = 1,\ldots,d\} \to_d \mathbf{G}(x_1,\ldots,x_d),$$

where $\mathbf{G}$ has continuous marginal distributions, then the dependence function of $\mathbf{G}$ is max-stable. The basic reason for this is the obvious fact that if $T_1,\ldots,T_d$ are continuous and strictly increasing, then $(\xi_1,\ldots,\xi_d)$ and $(T_1(\xi_1),\ldots,T_d(\xi_d))$ have the same dependence function.

Hsing also extends these results to stationary dependent sequences $\{\boldsymbol{\xi}_n\}$, along rather similar lines as for the one-dimensional case, as treated in Sections 2.1 and 2.2. Specifically, for given constants $\{u_{n,j}\colon j = 1,\ldots,d,\ n \ge 1\}$, the condition $D(u_{n,1},\ldots,u_{n,d})$ is defined to hold if there is a sequence $l_n = o(n)$ such that

$\alpha_{n,l_n} \to 0$ as $n \to \infty$ for

$$\alpha_{n,l} = \max\Big\{\Big|P\big(\xi_{i,j} \le u_{n,j}\colon j = 1,\dots,d,\ i \in A \cup B\big) \\ -P\big(\xi_{i,j}u_{n,j}\colon j = 1,\dots,d,\ i \in A\big) \\ P\big(\xi_{i,j} \le u_{n,j}\colon j = 1,\dots,d,\ i \in B\big)\Big|\Big\},$$

where the maximum is taken over all sets A, B such that $A \subset \{1,\dots,k\}$, $B \subset (k + l,\dots,n\}$ for some k. If $D(u_{n,1},\dots,u_{n,d})$ holds the only possible limits in (2.9.5) again are the max-stable d.f.'s. Furthermore, if, in addition,

$$\limsup_{n\to\infty} n \sum_{i_1=1}^{d} \sum_{i_2=1}^{d} \sum_{j=2}^{[n/k]} P\big\{\xi_{1,i_1} > u_{n,i_1}, \xi_{j,i_2} > u_{n,i_2}\big\} \to 0, \tag{2.9.6}$$

as $k \to \infty$,

then $P(M_{n,i} < u_{n,i},\ i = 1,\dots,d) \to \rho > 0$ if and only if $P^n(\xi_{1,i} \le u_{n,i}\colon i = 1,\dots,d) \to \rho$, i.e., the asymptotic distribution of maxima is the same as if the vectors were independent. [(2.9.6), of course, reduces to (2.2.1) for $d = 1$.]

A further question considered by Hsing is independence of the marginals in the limiting distribution. In particular, he shows that if

$$\limsup n \sum_{\substack{i_1, i_2=1 \\ i_1 \ne i_2}}^{d} \sum_{j=1}^{[n/k]} P\big\{\xi_{1,i_1} > u_{n,i_1}, \xi_{j,i_2} > u_{n,i_2}\big\} \to 0, \tag{2.9.7}$$

as $k \to \infty$, and if $D(u_{n,1},\dots,u_{n,d})$ is satisfied, then (2.9.5) holds if and only if $P\{a_{ni}(M_{n,i} - b_{n,i}) \le x\} \to G_i(x)$ as $n \to \infty$ for $i = 1,\dots,d$, and $\mathbf{G}$ then is of the form $\mathbf{G}(x_1,\dots,x_d) = G_1(x_1)G_2(x_2)\cdots G_d(x_d)$.

Now, let $\{\xi_n\}$ be normally distributed with $E\xi_{1,i} = 0$, $V(\xi_{1,i}) = 1$ and let $r_{ij}(n)$ be the covariance between $\xi_{1,i}$ and $\xi_{1+n,j}$. If $r_{ij}(k) < 1$ for $1 \le i \ne j \le d$ for all k and $r_{ij}(n)\log n \to 0$ as $n \to \infty$ for $i, j = 1,\dots,d$ and $u_{n,i} = x_i/a_n + b_n$ with a_n, b_n as in (1.2.4), then $\mathbf{D}(u_{n,1},\dots,u_{n,d})$, (2.9.6) and (2.9.7) are satisfied, so that the asymptotic distributions of maxima are the same as for a sequence of independent normal vectors with independent components (see [59] and [1]).

2.10. *Convergence of sums to nonnormal stable distributions.* The central limit problem of convergence of sums to nonnormal stable distributions hinges on the convergence of extreme order statistics, and perhaps the most natural approach to it, and its extensions to dependent settings, is via extreme value theory. In Theorem 2.10.1, which is new, this is made precise. The theorem, which builds on ideas of Durrett and Resnick [43] and Resnick [85], contains a functional central limit theorem, and the corresponding extreme value result is the "complete" convergence of upper and of lower extremes, which is discussed in Sections 1.3 and 2.4. A similar one-dimensional approach via the joint distribution of extreme order statistics is used in [67] and [35] and will be briefly discussed at the end of this section.

The results depend essentially on the Itô–Lévy representation of the stable process, and we shall now list the relevant properties, referring to Itô [61], Section 1.12, for proofs and further information. Let $\{\eta(t): 0 \le t \le 1\}$ be a nonnormal stable stationary independent increments process [briefly, $\{\eta(t)\}$ will be referred to as a stable process]. $\{\eta(t)\}$ can—and will throughout—be assumed to have sample paths in $D[0,1]$ the space of real functions on $[0,1]$, which are right-continuous and have left limits at each point. Let $S = [0,1] \times \overline{R}$, with $\overline{R} = [-\infty, \infty] \setminus \{0\}$, and define the Itô process N of jumps of $\{\eta(t)\}$ by

$$N(A) = \#\{t: (t, \eta(t) - \eta(t-)) \in A\}, \tag{2.10.1}$$

for Borel sets $A \subset S$, where $\eta(t) - \eta(t-)$ is the jump of $\eta(\cdot)$ at time t. Then $N(A)$ is (measurable and) finite a.s. for each rectangle A such that $A \subset [0,1] \times [-\infty, -\varepsilon] \cup [\varepsilon, \infty]$ for some $\varepsilon > 0$. Hence N is a point process, and, in fact, it is a Poisson process with intensity measure ν, which is the product of Lebesgue measure and the measure ν' on $\overline{R}$ with density $\gamma_+ y^{-\alpha-1}$ for $y > 0$ and $\gamma_- |y|^{-\alpha-1}$ for $y < 0$ for some constants $\gamma_+, \gamma_- \ge 0$, which are not both zero [i.e., in shorthand notation, $d\nu = dt \times d\nu' = dt \times (\gamma_\pm |y|^{-\alpha-1}\, dy)$].

Let $m(\varepsilon) = 0$ for $0 < \alpha < 1$, let $m(\varepsilon) = \int_{\varepsilon < |y|} y(1+y^2)^{-1}\, d\nu'(y)$ for $\alpha = 1$ and let $m(\varepsilon) = \int_{\varepsilon < |y|} y\, d\nu'(y)$ for $1 < \alpha < 2$, and define

$$\eta^{(\varepsilon)}_{(t)} = \iint_{\substack{0 \le s \le t \\ \varepsilon < |y|}} y\, dN - t m(\varepsilon). \tag{2.10.2}$$

Here the integral is just a finite sum: If N has the points $\{(t_j, y_j): j > 1\}$, then $|y_j| > \varepsilon$ and $0 \le t_j \le 1$ only for finitely many j's, and

$$\iint_{\substack{0 \le s \le t \\ \varepsilon < |y_j|}} y\, dN = \sum_{\substack{j:\, t_j \le t \text{ and} \\ \varepsilon < |y_j|}} y_j.$$

With this notation

$$P\Big(\sup_{0 \le t \le 1} |\eta(t) - \eta^{(\varepsilon)}(t)| > \delta\Big) \to 0, \quad \text{as } \varepsilon \to 0, \tag{2.10.3}$$

for any $\delta > 0$.

Let $\{\xi_n\}_{n=1}^\infty$ be arbitrary random variables, let $\{a_n > 0, b_n\}_{n=1}^\infty$ be norming constants, define stochastic processes: $\{\eta_n(t): 0 < t < 1\}_{n=1}^\infty$ in $D[0,1]$ by

$$\eta_n(t) = \sum_{j=1}^{[nt]} a_n(\xi_j - b_n), \tag{2.10.4}$$

and in analogy with (2.10.1) let N_n be the point process of jumps of η_n, defined as

$$\begin{aligned} N_n(A) &= \#\{t: (t, \eta_n(t) - \eta_n(t-)) \in A\} \\ &= \#\{j: (j/n, a_n(\xi_j - b_n)) \in A\}, \end{aligned} \tag{2.10.5}$$

for Borel sets $A \subset S = [0,1] \times \overline{R}$. The following theorem specifies the connection between convergence in distribution of η_n to η and of N_n to N. In this convergence is in $D[0,1]$ given the Skorokhod topology, see, e.g., [23], Section 16.

THEOREM 2.10.1. *Let $\{\eta_n(t)\colon 0 \le t \le 1\}$ and N_n be given by* (2.10.4) *and* (2.10.5) *and let $\{\eta(t)\colon 0 \le t \le 1\}$ be a nondegenerate nonnormal stable process with Itô process N defined by* (2.10.1). *Then $\eta_n \to_d \eta$ as $n \to \infty$ in $D[0,1]$ if and only if the following two conditions hold*:

$$N_n \to_d N, \quad \text{as } n \to \infty \text{ on } S, \tag{2.10.6}$$

and, writing $I_{n,j} = 1$ if $|a_n(\xi_j - b_n)| > \varepsilon$ and $I_{n,j} = 0$ otherwise,

$$\limsup_{n\to\infty} P\left\{ \sup_{0\le t\le 1} \left| \sum_{j=1}^{[nt]} a_n(\xi_j - b_n)(1 - I_{n,j}) + tm(\varepsilon) \right| > \delta \right\} \to 0, \tag{2.10.7}$$

as $\varepsilon \to 0$ for each $\delta > 0$.

PROOF. Let $N^{(\varepsilon)}$ and $N_n^{(\varepsilon)}$ be the restrictions of N and N_n to $[0,1] \times [-\infty, -\varepsilon] \cup [\varepsilon, \infty]$ for $\varepsilon > 0$. Let $\eta^{(\varepsilon)}$ be given by (2.10.2) and set

$$\eta_n^{(\varepsilon)} = \int_{\substack{0\le s\le t\\ \varepsilon<|y|}} y\, dN_n - tm(\varepsilon) = \sum_{j=1}^{[nt]} a_n(\xi_j - b_n) I_{n,j} - tm(\varepsilon).$$

First, suppose that $\eta_n \to_d \eta$. The function that maps η into $N^{(\varepsilon)}$ and η_n into $N_n^{(\varepsilon)}$ is a.s. continuous with respect to the distribution of η (see [85]) and hence $N_n^{(\varepsilon)} \to_d N^{(\varepsilon)}$ for each $\varepsilon > 0$. This implies that $N_n \to_d N$, i.e., (2.10.6) holds. Similarly, $|\eta_n(\cdot) - \eta_n^{(\varepsilon)}(\cdot)| \to_d |\eta(\cdot) - \eta^{(\varepsilon)}(\cdot)|$ in $D[0,1]$, and hence

$$P\left\{ \sup_{0\le t\le 1} |\eta_n(t) - \eta_n^{(\varepsilon)}(t)| > \delta \right\} \to P\left\{ \sup_{0\le t\le 1} |\eta(t) - \eta^{(\varepsilon)}(t)| > \delta \right\}, \quad \text{as } n \to \infty,$$

since

$$P\left(\sup_{0\le t\le 1} |\eta(t) - \eta^{(\varepsilon)}(t)| = \delta \right) = 0, \quad \text{for } \delta > 0.$$

Now,

$$\eta_n(t) - \eta_n^{(\varepsilon)}(t) = \sum_{j=1}^{[nt]} a_n(\xi_j - b_n)(1 - I_{n,j}) + tm(\varepsilon), \tag{2.10.8}$$

and (2.10.7) thus follows immediately from (2.10.3).

Conversely, suppose (2.10.6) and (2.10.7) hold. The map that takes N_n into $\eta_n^{(\varepsilon)}$ is a.s. N-continuous, and hence $\eta_n^{(\varepsilon)} \to_d \eta^{(\varepsilon)}$ as $n \to \infty$ in $D[0,1]$, and together with (2.10.7) and (2.10.8) this implies that $\eta_n \to_d \eta$ by [23], Theorem 4.2. □

The main condition, $N_n \to N$, of "complete" convergence of extremes, requires much weaker asymptotic mixing conditions than those needed for convergence of sums to the normal distribution (cf. the end of Section 2.4). However, the local dependence restrictions, such as (2.2.1), may instead be rather restrictive and are not even in general satisfied for 1-dependent processes (cf. Example 2.2.2).

The conditions, of course, become particularly simple when $\xi_1, \xi_2, \dots$ are i.i.d. Then $N_n \to N$ is equivalent to $nP(a_n(\xi_1 - b_n) \in A) \to \nu'(A)$ for each Borel set $A \subset [-\infty, -\varepsilon] \cup [\varepsilon, \infty]$ for some $\varepsilon > 0$, which in turn is the same as

$$\begin{aligned} &nP\{a_n(\xi_1 - b_n) > x\} \to \gamma_+ \int_x^\infty y^{-\alpha-1}\,dy, && \text{for } x > 0, \\ &nP\{a_n(\xi_1 - b_n) \le x\} \to \gamma_- \int_{-\infty}^x |y|^{-\alpha-1}\,dy, && \text{for } x < 0, \end{aligned} \tag{2.10.9}$$

as $n \to \infty$. Another way of expressing (2.10.9) is to say that the marginal d.f. F of the ξ's should belong to the domain of attraction of the type II distribution for both maxima (if $\gamma_+ > 0$) and minima (if $\gamma_- > 0$), with the same norming constants $\{a_n > 0, b_n\}$. Furthermore, Resnick [85] shows that (2.10.9) actually implies also (2.10.7) for i.i.d. sequences. Thus in this case $\eta_n \to_d \eta$ in $D[0,1]$ is equivalent to (2.10.9). It may also be noted that b_n can be taken to be zero here.

If one is not interested in full convergence in $D[0,1]$, but only in "marginal" convergence of $\eta_n(1) = \sum_{j=1}^n a_n(\xi_j - b_n)$ to a nonnormal stable distribution, sufficient conditions are easily found by "projecting onto the y-axis." Let N' be the point process of jump heights of η, given by

$$N'(A) = \#\{t \in [0,1]\colon \eta(t) - \eta(t-) \in A] = N([0,1] \times A),$$

for Borel sets $A \subset \overline{R}$, so that N' is a Poisson process with intensity ν' and similarly let

$$N_n'(A) = \#\{j \in [1,n]\colon a_n(\xi_j - b_n) \in A\} = N_n([0,1] \times A).$$

By the same considerations as in the last part of the proof of Theorem 2.10.1, if

$$N_n' \to_d N', \quad \text{as } n \to \infty \text{ in } \overline{R}, \tag{2.10.10}$$

and if, as before with $I_{n,j} = 1$ if $|a_n(\xi_j - b_n)| > \varepsilon$ and $I_{n,j} = 0$ otherwise,

$$\limsup_{n\to\infty} \mathrm{P}\left(\left|\sum_{j=1}^n a_n(\xi_j - b_n)(1 - I_{n,j}) + m(\varepsilon)\right| > \delta\right) \to 0, \quad \text{as } \varepsilon \to 0,$$

for each $\delta > 0$, then $\eta_n(1) \to_d \eta(1)$ in R. Moreover, it can be seen that (2.10.10) holds if and only if the joint distribution of the k largest and k smallest order statistics of $\xi_1, \dots, \xi_n$ tends to the distribution of the k largest and k smallest jumps of $\{\eta(t)\colon 0 \le t \le 1\}$ for each k, again emphasizing the connection with extreme value theory. This approach to convergence of $\sum a_n(\xi_j - b_n)$ to nonnormal stable distributions is, with some variations, pursued in detail for i.i.d. ξ's by LePage, Woodroofe and Zinn [67] and for stationary sequences satisfying distributional mixing conditions by Davis [35].

Finally, the results of this section easily carry over to nonstationary situations with $[nt]$ replaced by an arbitrary time scale, to convergence of row-sums in a doubly indexed array $\{\xi_{n,j}\}$ to a Lévy (independent increments) process without continuous component, to multidimensional ξ's and also to convergence of so-called self-normalized sums.

2.11. *Miscellanea.*

(a) *Minima and maxima.* Since the minimum $m_n = \min\{\xi_1, \dots, \xi_n\}$ can be obtained as $m_n = -\max(-\xi_1, \dots, -\xi_n\}$, results for maxima carry directly over to minima. In particular, it follows from the extremal types theorem that, under distributional mixing assumptions, limiting d.f.'s of linearly normalized minima must be of the form $1 - G(-x)$, where G is an extreme value d.f. Furthermore, it is trivial to see that for i.i.d. variables minima and maxima are asymptotically independent (cf. [66], page 28).

In a series of papers ([30], [31] and [34]), Davis studies the joint distribution of m_n and M_n for stationary sequences $\{\xi_n\}$ under a number of different dependence restrictions. Here we only note that some of his results alternatively may be obtained as corollaries of the multivariate theory discussed in Section 2.9 by making the identification $\xi_{i,1} = \xi_i$, $\xi_{i,2} = -\xi_i$, so that $M_{n,1} = M_n$, $M_{n,2} = -m_n$. For example, writing $u_{n,1} = u_n$, $u_{n,2} = -v_n$ for $v_n \le u_n$, the mixing condition $D(u_{n,1}, u_{n,2})$ then translates to $\alpha_{n,l_n} \to 0$ for some sequence $l_n = o(n)$, with

$$\alpha_{n,l} = \max\{|P(\xi_1 \le u_n, \xi_i \ge v_n : i \in A \cup B) \\ -P(\xi_i \le u_n, \xi_i \ge v_n : i \in A)P(\xi_i \le u_n, \xi_i \ge v_n : i \in B)|\},$$

where the maximum is taken over all sets $A \subset \{1, \dots, k\}$, $B \subset \{k + l, \dots, n\}$ for $k = 1, \dots, n - l$. Thus if this holds for $u_n = x/a_n + b_n$ and $v_n = y/c_n + d_n$ for all x and y, it follows that any limiting d.f. of $(a_n(M_n - b_n), c_n(m_n - d_n))$ must be of the form $\mathbf{G}(x, \infty) - \mathbf{G}(x, -y)$ where $\mathbf{G}$ is a bivariate extreme value d.f. Furthermore, the criterion (2.9.7) for independence of componentwise maxima, i.e., here for asymptotic independence of M_n and m_n translates to

$$\limsup_{n \to \infty} n \sum_{j=2}^{[n/k]} \{P(\xi_1 > u_n, \xi_j < v_n) + P(\xi_1 < v_n, \xi_j > u_n)\} \to 0, \quad \text{as } k \to \infty.$$

(b) *Poisson limit theorems.* Although somewhat less generally formulated, the Poisson and compound Poisson limits discussed in Section 2.4 amount to convergence of point processes N_n defined from a triangular array $\{\varepsilon_{n,i}: i = 1, \dots, n, n \ge 1\}$ of zero–one variables, with stationary rows $\varepsilon_{n,1}, \dots, \varepsilon_{n,n}$, by

$$N_n(E) = \sum_{i:\, i/n \in E} \varepsilon_{n,i},$$

for Borel subsets E of $(0, 1]$. Thus the proof of the Poisson limit for $\theta = 1$ (see [66], Section 2.5) is easily seen to show that if $D(u_n)$ and (2.2.1) hold with $\xi_i \le u_n$ and $\xi_i > u_n$ replaced by $\varepsilon_{n,i} = 0$ and $\varepsilon_{n,i} = 1$, respectively, then N_n converges to a Poisson process with intensity τ if and only if $nP(\varepsilon_{n,i} = 1) \to \tau$.

Conversely, the literature contains many sufficient conditions for convergence, which may be applied to extremes by setting $\varepsilon_{n,i}$ equal to zero or one according to whether $\xi_i \le u_n$ or $\xi_i > u_n$. Two further sets of such conditions seem particularly useful here. For the first, let $\mathscr{B}_{n,i}$ be the σ-algebra generated by $\varepsilon_{n,1}, \dots, \varepsilon_{n,i}$.

Then the relation

$$\sum_{i=0}^{[nt]} E\{\varepsilon_{n,i+1}|\mathscr{B}_{n,i}\} \to t\tau, \quad \text{as } n \to \infty, \tag{2.11.1}$$

in probability for each $t \in (0,1]$ is sufficient for convergence of N_n to a Poisson process with intensity τ ([48] and [43]). For the second one, which is due to Berman [12] and [17], we assume that each row has been extended to a doubly infinite sequence $\ldots, \varepsilon_{n,-1}, \varepsilon_{n,0}, \varepsilon_{n,1}, \ldots$ and write $\overline{\mathscr{B}}_{n,i}$ for the σ-algebra generated by $\ldots, \varepsilon_{n,i-1}, \varepsilon_{n,i}$. Berman's result is that if $nP(\varepsilon_{n,1} = 1) \to \tau$ and if there exists a sequence γ_n of integers with $\gamma_n = o(n)$, such that

$$n \sum_{i=2}^{\gamma_n} P(\varepsilon_{n,1} = 1, \varepsilon_{n,i} = 1) \to 0, \quad \text{as } n \to \infty,$$

$$nP(\varepsilon_{n,1} = 1|\overline{\mathscr{B}}_{n,-\gamma_n}) \to \tau, \qquad n \to \infty,$$

in probability, then N_n again converges to a Poisson process with intensity τ.

Neither one of these three sets of conditions imply any of the others; in particular, they are not necessary, and each of them might be the most convenient one in some situation. However, e.g., for normal sequences with $r_n \log n \to 0$ they all seem to lead to about the same amount of work. One useful feature of (2.11.1) is that it also directly gives rate of convergence results (cf. [87]).

3. Extremes of continuous parameter processes.

3.1. *The extremal types theorem for stationary processes.* Let $\{\xi(t): t \geq 0\}$ be a strictly stationary process having a.s. continuous sample functions and continuous one-dimensional distributions. Then, assuming that the underlying probability space is complete, $M(I) = \sup\{\xi(t): t \in I\}$ is a r.v. for any finite interval I and, in particular, so is $M(T) = M([0,T])$. The extremal types theorem may be proved even in this continuous context, showing that, under general dependence restrictions, the only nondegenerate limits G in

$$P\{a_T(M(T) - b_T) \leq x\} \to G(x), \quad \text{as } T \to \infty, \tag{3.1.1}$$

are the three classical types.

Though the general result requires considerable details of proof, the method involves the very simple observation that for (any convenient) $h > 0$,

$$M(nh) = \max(\zeta_1, \zeta_2, \ldots, \zeta_n), \tag{3.1.2}$$

where $\zeta_i = \max\{\xi(t): (i-1)h \leq t \leq ih\}$. Thus if (3.1.1) holds and the (stationary) sequence $\zeta_1, \zeta_2, \ldots$ satisfies $D(u_n)$ for each $u_n = x/a_{nh} + b_{nh}$, then it follows from the discrete parameter extremal types result (Theorem 2.1.2) that G must be one of the extreme value types. Hence the extremal types theorem certainly holds for strongly mixing stationary processes since then the sequence $\{\zeta_n\}$ is also strongly mixing and thus trivially satisfies $D(u_n)$. However, a more general form of the theorem results from showing that the $D(u_n)$ condition holds for the ζ's when the ξ's satisfy certain conditions—in particular, a continuous version $C(u_T)$ of $D(u_n)$. In fact, the condition $C(u_T)$ may be defined in terms of the process properties only at "time-sampled" points jq_T for a suitable sampling interval $q_T \to 0$.

In the following definition $F_{t_1,\dots,t_n}(u)$ will be written for $F_{t_1,\dots,t_n}(u,\dots,u)$, where $F_{t_1,\dots,t_n}(x_1,\dots,x_n) = P\{\xi(t_1) \le x_1,\dots,\xi(t_n) \le x_n\}$.

The condition $C(u_T)$ will be said to hold for the process $\xi(t)$ and the family of constants $\{u_T\colon T > 0\}$, with respect to the constants $q_T \to 0$ if for any points $s_1 < s_2 < \cdots < s_p < t_1 < \cdots < t_{p'}$ belonging to $\{kq_T\colon 0 < kq_T \le T\}$ and satisfying $t_1 - s_p \ge \gamma$, we have

$$\left|F_{s_1,\dots,s_p,t_1,\dots,t_{p'}}(u_T) - F_{s_1,\dots,s_p}(u_T)F_{t_1,\dots,t_{p'}}(u_T)\right| \le \alpha_{T,\gamma_T},$$

where $\alpha_{T,\gamma_T} \to 0$ for some family $\gamma_T = o(T)$ as $T \to \infty$.

The ζ_i are, of course, maxima of $\xi(t)$ in fixed intervals of length h [e.g., $\zeta_1 = M(h)$] and the sampling interval q must be taken small enough so that these are well approximated by the maxima at the sample points jq. This is conveniently done by assuming that for each $a > 0$ there is a family $\{q\} = \{q_a(u)\}$ tending to zero as $u \to \infty$ such that

$$\text{(3.1.3)} \quad \limsup_{u\to\infty} \frac{P\{M(h) > u,\ \xi(jq) \le u, 0 \le jq \le h\}}{\psi(u)} \to 0, \quad \text{as } a \to 0.$$

Here $\psi(u)$ is a function, which will later be taken to represent the tail of the distribution of $M(h)$ but which for the present need only dominate $P(\xi(0) > u\}$, i.e.,

$$\text{(3.1.4)} \qquad P\{\xi(0) > u\} = o(\psi(u)).$$

The following result ([66], Theorem 13.1.5) then holds.

THEOREM 3.1.1 (Extremal types theorem for stationary processes). *With the notation given previously suppose that* (3.1.1) *holds for the stationary process* $\{\xi(t)\}$ *and some constants* a_T, b_T *and a nondegenerate* G. *Suppose also that* $\psi(u)$ *is a function such that* (3.1.4) *holds and* $T\psi(u_T)$ *is bounded for* $u_T = x/a_T + b_T$ *for each* x. *If* $C(u_T)$ *holds for the families of constants* $\{q_a(u)\}$ *satisfying* (3.1.3), *then* G *must be one of the three classical extreme value types.*

3.2. *Domains of attraction.* In the classical theory of extremes of i.i.d. sequences, the type of limiting distribution for the maximum was determined by the asymptotic form of the tail of the distribution of ξ_1. This remained true for dependent stationary cases with nonzero extremal index since the limiting type was that of the associated independent sequence. For continuous parameter processes, however, it is clearly the tail of the distribution of ζ_1 [in view of (3.1.2)] rather than that of ξ_t, which determines the limiting type. More specifically, if $\hat{\zeta}_1, \hat{\zeta}_2 \dots$ are i.i.d. random variables with the same distribution as $\zeta_1 = M(h)$, then $\{\hat{\zeta}_n\}$ is called the *independent sequence associated with* $\{\xi_t\}$. If the ζ_n-sequence has extremal index $\theta > 0$, then any asymptotic distribution for $M(T)$ is of the same type as that for $\hat{M}_n = \max\{\hat{\zeta}_1,\dots,\hat{\zeta}_n\}$. Again the case $\theta = 1$

is of special interest and sufficient conditions may be given. In particular, the following condition [analogous to (2.2.1) for sequences] is useful:

The condition $C'(u_T)$ will be said to hold for the process $\{\xi(t)\}$ and the family of constants $\{u_T\colon T > 0\}$ with respect to the constants $\{q = q(u_T) \to 0\}$ if

$$\limsup_{T\to\infty} (T/q) \sum_{h<jq<\varepsilon T} P\{\xi(0) > u_T, \xi(jq) > u_T\} \to 0, \quad \text{as } \varepsilon \to 0.$$

We assume also as needed that for some function ψ,

(3.2.1) $P\{M(h) > u\} \sim h\psi(u)$, as $u \to \infty$ for $0 < h < \delta$ some $\delta > 0$.

The following result may also be shown (see [66], Section 13.2).

THEOREM 3.2.1. *Suppose that* (3.2.1) *holds for some function* ψ *and let* $\{u_T\}$ *be a family of constants such that for each* $a > 0$, $C(u_T), C'(u_T)$ *hold with respect to the family* $\{q_a(u)\}$ *of constants satisfying* (3.1.3) *with* h *in* $C'(u_T)$ *not exceeding* $\delta/2$, *where* δ *is from* (3.2.1). *Then as* $T \to \infty$,

$$T\psi(u_T) \to \tau > 0, \tag{3.2.2}$$

if and only if

$$P\{M(T) \le u_T\} \to e^{-\tau}. \tag{3.2.3}$$

Hence the function ψ may be conveniently used in the domain of attraction criteria and also plays the role of $1 - F$ in the continuous parameter analog of Lemma 1.2.2. In particular, if $M(T)$ has a limiting distribution as in (3.1.1) the constants a_T, b_T must satisfy $T\psi(u_T) \to \tau$ with $u_T = x/a_T + b_T$, $(\tau = \tau(x)) = -\log G(x))$ from which a_T, b_T may sometimes be conveniently obtained. In some cases the important function ψ is readily obtained (as in Section 3.4), but in others (cf. Section 3.3), its calculation can be quite intricate.

3.3. *Extremes of stationary normal processes.* Let $\xi(t)$ be a stationary normal process (assumed standardized to have zero mean, unit variances) and covariance function $r(t)$ satisfying

$$r(t) = 1 - C|t|^\alpha + o(|t|^\alpha), \quad \text{as } t \to 0, \tag{3.3.1}$$

for some $C > 0$, $0 < \alpha \le 2$. This includes all the mean-square differentiable cases ($\alpha = 2$) and a wide variety of cases with less regular sample functions ($0 < \alpha < 2$), such as the Ornstein–Uhlenbeck process ($\alpha = 1$). It may be shown that for such a process a function $\psi(u)$ satisfying (3.2.1) is given by

$$\psi(u) = C^{1/\alpha} H_\alpha u^{(2-\alpha)/\alpha} (2\pi)^{-1/2} \exp(-u^2/2), \tag{3.3.2}$$

but the proof involves quite intricate computations when $\alpha < 2$. The H_α are constants whose numerical values are known only in the cases $\alpha = 1, 2$ ($H_1 = 1$, $H_2 = \pi^{-1/2}$). The "regular" case $\alpha = 2$ is simpler and $\psi(u)$ may then be alternatively obtained as in the next section.

It can be shown (cf. [66], Theorem 2.5.1), using the normal comparison lemma that the (standard) stationary normal process $\xi(t)$ satisfying (3.3.1) satisfies the required dependence conditions for the general theory provided, that

$$r(t)\log t \to 0, \quad \text{as } t \to \infty. \tag{3.3.3}$$

The function $\psi(u)$ given by (3.3.2) satisfies the domain of attraction criteria for the type I extreme value distribution. Indeed, some calculation shows that $T\psi(u_T) \to \tau$ with $\tau = e^{-x}$, $u_T = x/a_T + b_T$, for

$$\begin{aligned} a_T &= (2\log T)^{1/2}, \\ b_T &= a_T + \Big\{((2-\alpha)/2\alpha)\log\log T \\ &\qquad + \log\Big(C^{1/\alpha}H_\alpha(2\pi)^{-1/2}2^{(2-\alpha)/2\alpha}\Big)\Big\}/a_T. \end{aligned} \tag{3.3.4}$$

Hence (using the last remark of the previous section) (3.1.1) holds with a_T, b_T given by (3.3.4) and $G(x) = \exp(-e^{-x})$.

This result was obtained by Cramér [28] for the case $\alpha = 2$ and a somewhat more restrictive condition on the rate of decay of $r(t)$ as $t \to \infty$. The result in its present generality was obtained by Pickands ([79], [80]), though the proof was not quite complete and was subsequently corrected by Qualls and Watanabe [83]. In particular, considerable generality is afforded by the family of covariances satisfying (3.3.1), and the requirement $r(t)\log t \to 0$ imposes only a very mild assumption on the rate of convergence of $r(t)$ to zero as $t \to \infty$.

3.4. *Finite upcrossing intensities and point processes of upcrossings.* In the continuous parameter case exceedances of a level typically occur on intervals and do not form a point process. However, a natural analog is provided by the upcrossings (i.e., points where excursions above a level begin), which can form a useful point process for discussing extremal properties. Furthermore, in many cases the intensity of this point process provides the function $\psi(u)$ needed for the determination of extremal type. Before proceeding it is of interest to note that an alternative to discussing upcrossings is to consider the amount of time that the process spends above a level. This approach, used by Berman, is briefly indicated in Section 3.7.

Let then (as before) $\{\xi(t)\colon t \geq 0\}$ be stationary with a.s. continuous sample functions, and continuous one-dimensional d.f. If u is a constant, $\xi(t)$ is said to have an *upcrossing of u at t_0* > 0 if for some $\varepsilon > 0$, $\xi(t) \leq u$ in $(t_0 - \varepsilon, t_0)$ and $\xi(t) \geq u$ in $(t_0, t_0 + \varepsilon)$.

Under the given assumptions, the number $N_u(I)$ of upcrossings of u by $\xi(t)$ in an interval I is a (possibly infinite valued) r.v. If $\mu(u) = EN_u((0,1)) < \infty$, then $N_u(I) < \infty$ a.s. for bounded I, and the upcrossings form a stationary point process N_u with intensity parameter $\mu = \mu(u)$.

For stationary normal processes satisfying (3.3.1), μ is finite when $\alpha = 2$ and is then given by *Rice's formula*, $\mu(u) = (C/2)^{1/2}\pi^{-1}\exp(-u^2/2)$ and for general processes μ may be calculated under weak conditions as

$$\mu(u) = \int_0^\infty z p(u, z)\, dz, \tag{3.4.1}$$

where $p(u, z)$ is the joint density of $\xi(t)$ and its (q.m.) derivative $\xi'(t)$. In fact, these relations can be shown simply since $\mu(u) = \lim_{q \downarrow 0} J_q(u)$, where

$$J_q(u) = q^{-1}P\{\xi(0) \le u < \xi(q)\}, \qquad q > 0, \tag{3.4.2}$$

and hence depends only on the bivariate distribution of $\xi(0)$ and $\xi(q)$. Under general conditions, it is also the case when $u \to \infty$ as $q \to 0$ in a suitably coordinated way that $J_q(u) \sim \mu(u)$. We shall use a variant of this property, assuming that for each $a > 0$ there are constants $q_a(u) \to 0$ as $u \to \infty$ with

$$\liminf_{u \to \infty} \frac{J_{q_a}(u)}{\mu(u)} \to 1, \quad \text{as } a \to \infty, \tag{3.4.3}$$

and that

$$P(M(q) > u) = o(\mu(u)). \tag{3.4.4}$$

It may then be readily shown that (3.1.3) holds if $\psi(u) = \mu(u)$. Also (3.2.1) is often satisfied in regular cases. Under such conditions it thus follows that $\psi(u)$ may be replaced by $\mu(u)$ in previous results. (For details see [66], Section 13.5.)

Thus the intensity $\mu(u)$ can provide a convenient means for determining the type of limiting distribution for $M(T)$. However, the point process of upcrossings has further interesting properties analogous to those for exceedances in discrete parameter cases. Specifically, let $u = u_T$ and T tend to infinity in such a way that $T\mu(u_T) \to \tau > 0$. Define a normalized point process N_T^* of upcrossings having points at t/T when ξ has an upcrossing of u at t, i.e., $N_T^*(I) = \#\{\text{upcrossings of } u_T \text{ by } \xi(t) \text{ for } t/T \in I\}$. Then the following result holds.

THEOREM 3.4.1. *Suppose that the conditions of Theorem* 3.2.1 *hold, with* $\psi(u) = \mu(u)$ *and with* (3.1.3) *replaced by* (3.4.3) *and* (3.4.4). *Then* N_T^* *converges in distribution to a Poisson process with intensity* τ *as* $T \to \infty$. *This, in particular, holds for the stationary normal processes satisfying* (3.3.1) *with* $\alpha = 2$ *and* (3.3.3).

Similar results may be obtained under appropriate conditions for the point process of local maxima of height at least u as $u \to \infty$, leading, in particular, to the asymptotic distribution of $M^{(k)}(T)$, the kth largest local maximum in $[0, T]$. Indeed, "complete Poisson convergence" results, analogous to those indicated for sequences in Sections 1.3 and 2.4, may be obtained for the point process in the plane consisting of the locations and heights of the local maxima (cf. [66], Sections 9.5 and 13.6, for details).

Finally, it is also possible to obtain Poisson limits in cases with irregular sample paths when $\mu(u) = \infty$ (e.g., normal with $0 < \alpha < 2$) by the simple device of using the "ε-upcrossings" of Pickands [80] in lieu of ordinary upcrossings. Specifically, for given $\varepsilon > 0$, $\xi(t)$ has an ε-upcrossing of the level u at t_0 if $\xi(t) \le u$ for $t \in (t_0 - \varepsilon, t_0)$ and $\xi(t) > u$ for some $t \in (t_0, t_0 + \eta)$ for each $\eta > 0$, so that clearly the number of ε-upcrossings in a finite interval I is finite [indeed bounded by $(m(I)/\varepsilon) + 1$ where $m(I)$ is the length of I]. This device was used in [79] to give one of the first proofs of Theorem 3.3.1. (See also Section 3.7 for a different approach.)

3.5. *χ^2-processes.* The proofs for normal processes in Section 3.3 and also for the sequence case (Section 2.5) use the normal comparison lemma (Theorem 2.5.1) in an essential way. It will also be the basis for the present section on functions $\{\chi(t)\}$ of stationary d-dimensional ($d \geq 2$) normal processes $\xi(t) = (\xi_1(t), \ldots, \xi_d(t))$ defined as

$$\chi(t) = \sum_{i=1}^{d} \xi_i^2(t). \tag{3.5.1}$$

We shall assume that the components are standardized to have mean 0 and the same variance 1—here this is a real restriction and not just a question of normalization—and also that the components are independent. Then $\chi(t)$ has a χ^2-distribution and the process $\{\chi(t);\ t \geq 0\}$ is called a χ^2-process (with d degrees of freedom). Extremal properties of χ^2-processes, and of some related functions of $\xi(t)$, have been studied in detail by Sharpe [93], Arònowich and Adler [4] and [5] and Lindgren [68]–[70]. Here we will follow the "geometrical" approach of [69], and use the fact that $\chi(t)$ is the radial part of $\xi(t)$ to find the asymptotic double exponential distribution of maxima of $\chi(t)$, referring the reader to [5] for results on minima. However, we will indicate how the results can be obtained quite smoothly from the general theory of Section 3.4, rather than by using Lindgren's direct calculations.

Now, suppose further that the component processes $\{\xi_i(t)\}$, $i = 1, \ldots, d$, are continuously differentiable a.s. and have the same covariance function $r(t)$. We shall presently show that $\mu(u)$, the mean number of u-upcrossings by $\chi(t)$, $0 \leq t \leq 1$, is easily found from (3.4.1), and then apply Theorem 3.4.1. For $i = 1, \ldots, d$, $\xi_i(0)$ and $\xi_i'(0)$ are jointly normal, and hence independent, since

$$\operatorname{cov}(\xi_i'(0), \xi_i(0)) = \lim_{h \to 0} E\{h^{-1}(\xi_i(h) - \xi_i(0))\xi_i(0)\} = r'(0) = 0,$$

where the last equality holds because $r(t)$ is symmetric around zero. Similarly, if $\lambda = -r''(0)$ is the second spectral moment, $\xi_i'(t)$ has variance λ. Thus the conditional distribution of $\chi'(0) = \Sigma_{i=1}^d 2\xi_i(0)\xi_i'(0)$ given $\chi(0) = \Sigma_{i=1}^d \xi_i^2(0) = u > 0$ is normal with mean 0 and variance $\Sigma_{i=1}^d 4\lambda\xi_i^2(0) = 4\lambda u$. Let $p(z|u)$ be the density of this conditional distribution and let $p(u)$ be the density of $\chi(0)$, i.e.,

$$p(u) = 2^{-d/2}\Gamma(d/2)^{-1}u^{d/2-1}e^{-u/2}. \tag{3.5.2}$$

Then, using (3.4.1), it follows that

$$\begin{aligned} \mu(u) &= p(u)\int_0^\infty zp(z|u)\,dz \\ &= 2^{-(d-1)/2}\Gamma(d/2)^{-1}(\lambda/\pi)^{1/2}u^{(d-1)/2}e^{-u/2}, \end{aligned} \tag{3.5.3}$$

for $u > 0$. For u fixed, $J_q(u) = P(\chi(0) \leq u < \chi(q))/q \to \mu(u)$ as $q \to 0$, and similarly (3.4.3) holds for $u^{1/2}q = u^{1/2}q_a(u) \to a > 0$ (cf. [69], Lemma 2.5).

THEOREM 3.5.1. *Let $\xi(t) = (\xi_1(t), \dots, \xi_d(t))$ be a continuously differentiable d-dimensional standardized normal process with independent components and the same covariance function $r(t)$, as before. Suppose further that $r(t)\log t \to 0$ as $t \to \infty$ and that*

$$T\mu(u_T) \to \tau, \quad \text{as } T \to \infty, \tag{3.5.4}$$

and let N_T^* be the point process of upcrossings of u_T by $\{\chi(tT): t \in [0,1]\}$. Then N_T^* converges in distribution to a Poisson process with intensity τ, and, in particular,

$$P\Big\{\max_{0 \le t \le T} \chi(t) \le u_T\Big\} \to e^{-\tau}, \quad \text{as } T \to \infty. \tag{3.5.5}$$

PROOF. We shall briefly indicate how the conditions of Theorem 3.4.1 can be checked. We assume that $d = 2$, the extension to $d > 2$ being straightforward. The main idea in [69] is to introduce the normal random field $\{\chi_\theta(t): 0 \le \theta < 2\pi, t \ge 0\}$, where

$$\chi_\theta(t) = \xi_1(t)\cos\theta + \xi_2(t)\sin\theta$$

is the component of $\xi(t)$ in the direction $(\cos\theta, \sin\theta)$, and to note that then

$$\chi(t) = \sup_{0 \le \theta < 2\pi} \chi_\theta(t)^2. \tag{3.5.6}$$

Thus $\sup_{0 \le t \le h}\chi(t) = \max_{0 \le t \le h,\, 0 \le \theta < 2\pi}\chi_\theta(t)^2$, and it follows at once from the extremal theory for normal random fields that (3.2.1) holds for $\psi(u) = \mu(u)$ and any $h > 0$, (see [69], Lemma 2.2). As noted before, for fixed $a > 0$ (3.4.3) holds for $q = a/u^{1/2}$, and (3.4.4) is an easy consequence of (3.2.1). Thus it only remains to establish $C(u_T)$ and $C'(u_T)$ for an arbitrary h, say $h = 1$, and with $q = q_a(u)$ for each $a > 0$. For this we introduce a further sampling, in the θ-direction, given by a parameter $r = b/u^{1/2}$ with $b > 0$. Let $\tilde{\chi}_r(t) = \max\{\chi_{ir}(t);\ i = 0, \dots, [2\pi/r]\}$. Then, by (3.5.6) and an easy geometrical argument,

$$\chi(t)\cos^2 r \le \tilde{\chi}_r(t)^2 \le \chi(t), \tag{3.5.7}$$

for $0 < r < \pi/2$. To show that $C'(u_T)$ holds let $u_T' = (u_T)^{1/2}\cos r$, so that by (3.5.7) and stationarity,

$$\begin{aligned} &\frac{T}{q}\sum_{1 \le jq \le \varepsilon T} P(\chi(0) > u_T, \chi(jq) > u_T) \\ &\quad \le \frac{T}{q}\sum_{1 \le jq \le \varepsilon T} P(\tilde{\chi}_r(0) > u_T', \tilde{\chi}_r(jq) > u_T') \\ &\quad \le \frac{T}{q}\sum_{1 \le jq \le \varepsilon T} \big|P(\tilde{\chi}_r(0) > u_T', \tilde{\chi}_r(jq) > u_T') \\ &\qquad\qquad - P(\tilde{\chi}_r(0) > u_T')P(\tilde{\chi}_r(jq) > u_T')\big| \\ &\quad + \varepsilon(T/q)^2 P(\tilde{\chi}_r(0) > u_T')^2. \end{aligned} \tag{3.5.8}$$

It is readily seen that $T\mu((u'_T)^2) \to \tau' = \tau\exp\{b^2/2\}$ and that $\chi_\theta(t)$ has mean 0 and variance 1, and that $|\mathrm{cov}(\chi_\theta(0), \chi'_\theta(t))| \le |r(t)|$ for any θ, θ'. The normal comparison lemma can then be applied in a straightforward way to show that the sum on the right-hand side of (3.5.8) tends to zero. Furthermore, it follows from (3.5.7) and (3.5.2)–(3.5.4) that

$$\varepsilon(T/q)^2 P(\tilde{\chi}_r(0) > u'_T)^2 \le \varepsilon(T/q)^2\Big(P\big(\chi(0) > (u'_T)^2\big)\Big)^2 \to \varepsilon\big(\tau e^{b^2/2}\big)^2 2\pi/(\lambda a^2),$$

and thus $C'(u_T)$ is satisfied.

Next, with the notation of $C(u_T)$,

$$\begin{aligned}
&\left|F_{s_1,\dots,s_p,t_1,\dots,t_{p'}}(u_T) - F_{s_1,\dots,s_p}(u_T)F_{t_1,\dots,t_{p'}}(u_T)\right| \\
&\quad\le \Big|P\big(\tilde{\chi}_r(t) \le u_T^{1/2}\colon t \in \{s_1,\dots,s_p,t_1,\dots,t_{p'}\}\big) \\
&\qquad - P\big(\tilde{\chi}_r(t) \le u_T^{1/2}\colon t \in \{s_1,\dots,s_p\}\big) \\
&\qquad\quad \times P\big(\tilde{\chi}_r(t) \le \sqrt{u_T}\colon t \in \{t_1,\dots,t_{p'}\}\big)\Big| \\
&\qquad + \sum_{1 \le jq \le T} P\big(u_T \le \chi(jq) \le u_T/\cos^2 r\big).
\end{aligned} \tag{3.5.9}$$

Here the normal comparison lemma may be applied, similarly as for $C'(u_T)$, to show that the first expression on the right tends to zero as $T \to \infty$ if $t_1 - s_p \ge \gamma_T$ for suitable $\gamma_T = o(T)$. Furthermore, the last sum in (3.5.9) is bounded by

$$\begin{aligned}
\frac{T}{q}P\big(u_T \le \chi(0) \le u_T/\cos^2 r\big) &= \frac{T}{q}\left\{e^{-2^{-1}u_T} - e^{-2^{-1}u_T/\cos^2 r}\right\} \\
&\to (2\pi/\lambda)^{1/2}\tau\big(1 - e^{-b^2/2}\big)/a, \quad \text{as } T \to \infty,
\end{aligned}$$

by straightforward computations. Since this limit tends to zero as $b \to 0$ for a fixed, this may be seen to prove $C(u_T)$.

It is easy to "solve" (3.5.3), to show that (3.5.4) implies that

$$P\left\{a_T\left(\max_{0 \le t \le T}\chi(t) - b_T\right) \le x\right\} \to \exp(-e^{-x}), \quad \text{as } T \to \infty,$$

for

$$a_T = 1/2, \quad b_T = 2\log T + (d-1)\log\log T - \log\big(\Gamma(d/2)^2\pi/\lambda\big).$$

It might also be noted that this proof of $C(u_T)$ and $C'(u_T)$ applies, with obvious changes, also when the components of $\xi(t)$ are dependent and have different covariance functions.

3.6. *Diffusion processes.* Diffusion processes have many useful special properties, and, correspondingly, several different approaches to their extremal behavior are possible. E.g., Darling and Siegert [29], Newell [76] and Mandl [73] apply transform techniques and the Kolmogorov differential equations (cf. also the survey [24]), Berman [11] exploits the regenerative nature of stationary diffusions, similarly to Section 2.6, and Davis [32] and Berman [14] use a representation of the diffusion in terms of an Ornstein–Uhlenbeck process. Here we shall discuss some aspects of Davis' methods, and, in particular, state his main result [relation (3.6.6)].

A diffusion process $\{\xi(t);\ t \geq 0\}$ can be specified as the solution of a stochastic differential equation

$$d\xi(t) = \mu(\xi(t))\,dt + \sigma(\xi(t))\,dB(t), \tag{3.6.1}$$

where $\{B(t):\ t \geq 0\}$ is a standard Brownian motion. We refer to [62] for the precise definition and for the properties of $\{\xi(t)\}$ used in the following discussion. For simplicity, we will consider a somewhat more restrictive situation than in [62], and, in particular, we assume that $\{\xi(t)\}$ is defined on some open, possibly infinite, interval $I = (r_1, r_2)$ and that μ and σ are continuous with $\sigma > 0$ on I.

Let $\{s(x);\ x \in I\}$ be a solution of the ordinary differential equation

$$\sigma^2(x)s''(x) + 2\mu(x)s'(x) = 0, \tag{3.6.2}$$

i.e., let it have the form $s(x) = c_1 + c_2\int_{x_0}^{x}\exp\{-\int_{x_0}^{y}(2\mu(z)/\sigma^2(z))\,dz\}\,dy$, with $c_2 > 0$, c_1 real constants, for some point $x_0 \in I$. Then s is strictly increasing and by Itô's formula $\eta_t = s(\xi_t)$ satisfies $d\eta_t = f(\eta_t)\,dB_t$ for $f(x) = s'(s^{-1}(x))\sigma(s^{-1}(x))$, i.e., s is a *scale function* and $\{\eta_t:\ t \geq 0\}$ is the diffusion on natural scale. The *speed measure* m, corresponding to this scale function, then has density $1/f(x)$, i.e., $m(dx) = (1/f(x))\,dx$. We further assume that the speed measure is finite, $|m| = \int_I m(dx) = \int_I(1/f(x))\,dx < \infty$, and that $s(x) \to \infty$ as $x \to r_2$, $s(x) \to -\infty$ as $x \to r_1$. It then follows that the boundaries r_1, r_2 are inaccessible, that the diffusion is recurrent and that there exists a stationary distribution so that $\{\xi(t)\}$ becomes a stationary process if $\xi(0)$ is given this distribution.

The Ornstein–Uhlenbeck process, which will be denoted by $\{\tilde{\xi}(t)\}$ here, is the stationary diffusion process (3.6.1) specified by $I = R$, $\mu(x) = x/2$, $\sigma(x) = 1$, $x \in I$. For the present purposes, a convenient choice of scale function for $\{\tilde{\xi}(t)\}$ is $\tilde{s}(x) = (2\pi)^{1/2}\int_0^x e^{y^2/2}\,dy$, and the corresponding speed measure is $\tilde{m}(dx) = (2\pi)^{-1/2}e^{-x^2/2}\,dx$. Furthermore, it can be seen that $\{\tilde{\xi}(t)\}$ is a standardized stationary normal process with covariance function $r(t) = e^{-|t|}$ and that $\tilde{s}(x) \sim (2\pi)^{1/2}x^{-1}e^{x^2/2} = (x\phi(x))^{-1}$ as $x \to \infty$. Hence Theorem 3.3.1 may be applied with $C = \alpha = 1$ and its conclusion can, e.g., by a simple "subsequence argument," be written as

$$\sup_{u > \tilde{u}_0}\left|P(\tilde{M}(T) \leq u) - e^{-T/\tilde{s}(u)}\right| \to 0, \quad \text{as } T \to 0, \tag{3.6.3}$$

for any $\tilde{u}_0 > 0$ and with $\tilde{M}(t) = \sup\{\tilde{\xi}(t):\ 0 \leq t \leq T\}$.

The main additional fact needed is that the Ornstein–Uhlenbeck process on natural scale can, by a change of time, be made to have the same distribution as $\{\eta(t)\}$. More precisely ([32], Theorems 2.1 and 2.2), there exists a strictly increasing random function $\{\tau(t):\ t \geq 0\}$ such that the processes $\{s(\xi(t)):\ t \geq 0\}$

and $\{\tilde{s}(\tilde{\xi}(\tau(t))): t \geq 0\}$ have the same distribution, and which satisfies

$$T^{-1}\tau(T) \to 1/|m|, \quad \text{as } T \to \infty, \tag{3.6.4}$$

almost surely.

As in Section 2.6 it follows easily from (3.6.3) and (3.6.4) that

$$\sup_{u > \tilde{u}_0} \left| P\big(\tilde{M}(\tau(T)) \leq u\big) - e^{-T/(\tilde{s}(u)|m|)} \right| \to 0, \quad \text{as } T \to \infty. \tag{3.6.5}$$

Since for $M(T) = \sup\{\xi(t): 0 \leq t \leq T\}$,

$$\begin{aligned} P(M(T) \leq u) &= P\big(\sup\{s(\xi(t)): 0 \leq t \leq T\} \leq s(u)\big) \\ &= P\big(\sup\{\tilde{s}(\tilde{\xi}(\tau(t))): 0 \leq t \leq T\} \leq s(u)\big) \\ &= P\{\tilde{M}(\tau(T)) \leq \tilde{s}^{-1}(s(u))\}, \end{aligned}$$

(3.6.5) is readily seen to imply the main result of [34], that

$$\sup_{u < u_0} \left| P(M(T) \leq u) - e^{-T/(s(u)|m|)} \right| \to 0, \quad \text{as } T \to \infty, \tag{3.6.6}$$

for any $u_0 \in I$ with $s(u_0) > 0$. This is a quite explicit description of $M(T)$, "as the maximum of T i.i.d. random variables with d.f. $G(u) = \exp\{-1/(s(u)|m|)\}$," and, in particular, domains of attraction for $M(T)$ are found by applying the classical criteria to $\exp\{-1/(s(u)|m|)\}$. Finally, as for Markov chains, the hypothesis of stationarity is not essential, (3.6.6) holds for any initial distribution, as can be seen, e.g., by a simple "coupling argument."

3.7. *Miscellanea.*

(a) *Moving averages of stable processes.* These are continuous time processes of the form $\xi(t) = \int c(t-x)\, d\zeta(x)$, with $\{\zeta(x)\}$ a nonnormal stable independent increments process. Their extremal behavior, which is similar to that of the corresponding discrete parameter moving average (cf. Section 2.7), is studied in detail in [86].

(b) *Sample path properties.* As mentioned in Section 2.7, the asymptotic distribution of sample paths near extremes is studied in [86], [89] and [36]. A different approach to this problem, via so-called Slepian model processes, has been pursued by Lindgren in a series of papers (cf. the survey [71] and the references therein).

(c) *Extremal properties and sojourn times.* In an important series of papers, Berman studies "the sojourn of $\xi(t)$ above u," defined as $L_T(u) = \int_0^T 1\{\xi(t) > u\}\, dt$, where $1\{\cdot\}$ is the indicator function. For a wide variety of cases, including many normal processes, χ^2-processes, Markov processes and random Fourier sums, he finds the asymptotic form of the distribution of $L_T(u)$ as $u \to \infty$ for fixed T, and as $u, T \to \infty$ in a coordinated way. Furthermore, he uses the equivalence of the events $\{M(T) > u\}$ and $\{L_T(u) > 0\}$ to study the maximum of $\{\xi(t)\}$. The earlier work on these topics is reviewed in the present journal ([12]) by Berman himself. For later work see [13], [15], [16] and [18]–[21].

(d) *Exceedance random measure.* The sojourn time of $\xi(t)$ above u can, of course, be defined for processes whose sample functions are continuous but not sufficiently regular to define upcrossings (though "ε-upcrossings" may be defined and useful). However, a unifying viewpoint may be obtained by considering the "exceedance random measure" $\zeta_T(B) = \int_{t \in T \cdot B} 1\{\xi(t) > u\}\, dt$ for Borel subsets of $[0, 1]$, which extends the notion of $L_T(u)$ in an obvious way. Similar limiting theorems hold for $\zeta_T(B)$ as for the exceedance point process N_n of Section 2.4. In particular, compound Poisson limits typically occur (with multiplicities that are not necessarily now integer-valued). In cases where upcrossings are defined, this limit has the pleasant interpretation that the positions of the events represent upcrossing points, and the associated multiplicities represent the immediately following exceedance time above the level u. For details see [65].

Acknowledgments. It is a pleasure to thank Patrick Albin, Simeon Berman, Richard Davis and Sidney Resnick for comments on a number of sections.

Note added in proof. The following new book has recently appeared and is an important addition to the literature on extreme value theory: RESNICK, S. I. (1987), *Extreme Values, Regular Variation and Point Processes*, Springer, New York. This book provides an elegant and scholarly treatment of many areas for independent random variables, especially surrounding regular variation, records and extremal processes, and multivariate extremes.

REFERENCES

Because of space limitations we have not attempted to give a complete set of references. For further references we refer to the bibliographies compiled by Harter [56] and Tiago de Oliveira and co-workers [96] (cf. also [49] and [66]).

[1] AMRAM, F. (1985). Multivariate extreme value distributions for stationary Gaussian sequences. *J. Multivariate Anal.* **16** 237–240.

[2] ANDERSON, C. W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* **7** 99–113.

[3] ANDERSON, C. W. (1978). Super-slowly varying functions in extreme value theory. *J. Roy. Statist. Soc. Ser. B* **40** 197–202.

[4] ARONOWICH, M. and ADLER, R. J. (1985). Behaviour of χ^2 processes at extrema. *Adv. in Appl. Probab.* **17** 280–297.

[5] ARONOWICH, M. and ADLER, R. J. (1986). Extrema and level crossings of χ^2 processes. *Adv. in Appl. Probab.* **18** 901–920.

[6] ASMUSSEN, S. (1987). *Applied Probability and Queues.* Wiley, New York.

[7] ATHREYA, K. B. and NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

[8] BALKEMA, A. A., DE HAAN, L. and RESNICK, S. I. (1984). Uniform rates of convergence to extreme value distributions. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 341-346. Reidel, Dordrecht.

[9] BERMAN, S. M. (1962). Limiting distribution of the maximum term in sequences of dependent random variables. *Ann. Math. Statist.* **33** 894–908.

[10] BERMAN, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35** 502–516.

[11] BERMAN, S. M. (1964). Limiting distribution of the maximum of a diffusion process. *Ann. Math. Statist.* **35** 319–329.

[12] BERMAN, S. M. (1982). Sojourns and extremes of stationary processes. *Ann. Probab.* **10** 1–46.

[13] BERMAN, S. M. (1982). Sojourns and extremes of a diffusion process on a fixed interval. *Adv. in Appl. Probab.* **14** 811–832.

[14] BERMAN, S. M. (1983). High level sojourns of a diffusion process on a long interval. *Z. Wahrsch. verw. Gebiete* **62** 185–199.

[15] BERMAN, S. M. (1983). Sojourns of stationary processes in rare sets. *Ann. Probab.* **11** 847–866.

[16] BERMAN, S. M. (1983). Sojourns and extremes of random Fourier sums and series with random coefficients. *Stochastic Process. Appl.* **15** 213–238.

[17] BERMAN, S. M. (1984). Limiting distribution of sums of nonnegative stationary random variables. *Ann. Inst. Statist. Math.* **36** 301–321.

[18] BERMAN, S. M. (1986). The supremum of a process with stationary independent and symmetric increments. *Stochastic Process. Appl.* **23** 281–290.

[19] BERMAN, S. M. (1987). Extreme sojourns for random walks and birth-and-death processes. *Stochastic Models.* To appear.

[20] BERMAN, S. M. (1987). Poisson and extreme value limit theorems for Markov random fields. *Adv. in Appl. Probab.* **19** 106–122.

[21] BERMAN, S. M. (1988). Extreme sojourns of diffusion processes. *Ann. Probab.* **16** 361–374.

[22] BERNSTEIN, S. (1927). Sur l'extension du theoreme du calcul des probabilites aux sommes de quantites dependantes. *Math. Ann.* **97** 1–59.

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

[24] BLAKE, I. F. and LINDSEY, W. C. (1973). Level-crossing problems for random processes. *IEEE Trans. Inform. Theory* **19** 295–315.

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

[26] COHEN, J. P. (1982). The penultimate form of approximation to normal extremes. *Adv. in Appl. Probab.* **14** 324–339.

[27] COHEN, J. P. (1982). Convergence rates for the ultimate and penultimate approximations in extreme value theory. *Adv. in Appl. Probab.* **14** 833–854.

[28] CRAMÉR, H. (1965). A limit theorem for the maximum values of certain stochastic processes. *Theory Probab. Appl.* **10** 126–128.

[29] DARLING, D. A. and SIEGERT, A. J. F. (1953). The first passage problem for a continuous Markov process. *Ann. Math. Statist.* **24** 624–639.

[30] DAVIS, R. (1979). Maxima and minima of stationary sequences. *Ann. Probab.* **7** 453–460.

[31] DAVIS, R. (1982). Limit laws for the maximum and minimum of stationary sequences. *Z. Wahrsch. verw. Gebiete* **61** 31–42.

[32] DAVIS, R. (1982). Maximum and minimum of one-dimensional diffusions. *Stochastic Process. Appl.* **13** 1–9.

[33] DAVIS, R. (1982). The rate of convergence in distribution of the maxima. *Statist. Neerlandica* **36** 31–35.

[34] DAVIS, R. (1983). Limit laws for upper and lower extremes from stationary mixing sequences. *J. Multivariate Anal.* **13** 273–286.

[35] DAVIS, R. (1983). Stable limits for partial sums of dependent random variables. *Ann. Probab.* **11** 262–269.

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

[37] DE HAAN, L. (1976). Sample extremes: An elementary introduction. *Statist. Neerlandica* **30** 161–172.

[38] DE HAAN, L. (1984). Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 31–48. Reidel, Dordrecht.

[39] DE HAAN, L. and HORDIJK, A. (1972). The rate of growth of sample maxima. *Ann. Math. Statist.* **43** 1185–1196.

[40] DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. verw. Gebiete* **40** 317–337.

[41] DE HAAN, L. and RESNICK, S. I. (1982). Local limit theorems for sample extremes. *Ann. Probab.* **10** 396–413.

[42] DEHEUVELS, P. (1983). Point processes and multivariate extreme values. *J. Multivariate Anal.* **13** 257–272.
[43] DURRETT, R. and RESNICK, S. I. (1978). Functional limit theorems for dependent variables. *Ann. Probab.* **6** 829–846.
[44] FALK, M. (1984). Uniform rates of convergence of extreme order statistics. Thesis, Dept. Mathematics, Univ. Siegen.
[45] FINSTER, M. (1982). The maximum term and first passage time for autoregressions. *Ann. Probab.* **10** 737–744.
[46] FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24** 180–190.
[47] FRÉCHET, M. (1927). Sur la loi de probabilite de l'ecart maximum. *Ann. Soc. Math. Polon.* **6** 93–116.
[48] FREEDMAN, D. (1974). The Poisson approximation for dependent events. *Ann. Probab.* **2** 256–269.
[49] GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York.
[50] GNEDENKO, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)* **44** 423–453.
[51] GUMBEL, E. J. (1958). *Statistics of Extremes.* Columbia Univ. Press, New York.
[52] HALDANE, J. B. S. and JAYAKAR, S. D. (1963). The distribution of extremal and nearly extremal values in samples from a normal population. *Biometrika* **50** 89–94.
[53] HALL, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–439.
[54] HALL, P. (1980). Estimating probabilities for normal extremes. *Adv. in Appl. Probab.* **12** 491–500.
[55] HALL, W. J. and WELLNER, J. A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica* **33** 151–154.
[56] HARTER, H. L. (1978). A bibliography of extreme-value theory. *Internat. Statist. Rev.* **46** 279–306.
[57] HSING, T. (1984). Point processes associated with extreme value theory. Dissertation, Technical Report 83, Center for Stochastic Processes, Univ. North Carolina.
[58] HSING, T. (1987). On the characterization of certain point processes. *Stochastic Process. Appl.* To appear.
[59] HSING, T. (1987). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.* To appear.
[60] HSING, T., HUSLER, J. and LEADBETTER, M. R. (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields.* To appear.
[61] ITÔ, K. (1969). *Stochastic Processes.* Lecture Notes Series 16, Dept. Mathematics, Aarhus Univ.
[62] ITÔ, K. and MCKEAN, H. P. (1965). *Diffusion Processes and Their Sample Paths.* Springer, Berlin.
[63] KALLENBERG, O. (1976). *Random Measures.* Academic, New York.
[64] LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. verw. Gebiete* **65** 291–306.
[65] LEADBETTER, M. R. (1987). The exceedance random measure for a stationary process. Technical Report, Center for Stochastic Processes, Univ. North Carolina.
[66] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.
[67] LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to stable distributions via order statistics. *Ann. Probab.* **9** 624–632.
[68] LINDGREN, G. (1980). Extreme values and crossings for the χ^2-process and other functions of multidimensional Gaussian processes, with reliability applications. *Adv. in Appl. Probab.* **12** 746–774.
[69] LINDGREN, G. (1980). Point processes of exits by bivariate Gaussian processes and extremal theory for the χ^2-process and its concomitants. *J. Multivariate Anal.* **10** 181–206.

[70] LINDGREN, G. (1984). Extremal ranks and transformation of variables for extremes of functions of multivariate Gaussian processes. *Stochastic Process. Appl.* **17** 285–312.

[71] LINDGREN, G. (1984). Use and structure of Slepian model processes for prediction and detection in crossing and extreme value theory. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 261–284. Reidel, Dordrecht.

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

[73] MANDL, P. (1968). *Analytic Treatment of One-Dimensional Markov Processes*. Springer, New York.

[74] MITTAL, Y. and YLVISAKER, D. (1975). Limit distributions for the maxima of stationary Gaussian processes. *Stochastic Process. Appl.* **3** 1–18.

[75] MORI, T. (1977). Limit distributions of two-dimensional point processes generated by strong mixing sequences. *Yokohama Math. J.* **25** 155–168.

[76] NEWELL, G. F. (1962). Asymptotic extreme value distributions for one-dimensional diffusion processes. *J. Math. Mech.* **11** 481–496.

[77] NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. verw. Gebiete* **43** 309–318.

[78] O'BRIEN, G. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.* **15** 281–291.

[79] PICKANDS, J., III (1969). Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.* **145** 75–86.

[80] PICKANDS, J., III (1969). Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145** 51–73.

[81] PICKANDS, J., III (1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Probab.* **8** 745–756.

[82] PICKANDS, J., III (1980). Multivariate negative exponential and extreme value distributions. Manuscript, Wharton School, Univ. Pennsylvania.

[83] QUALLS, C. and WATANABE, H. (1972). Asymptotic properties of Gaussian processes. *Ann. Math. Statist.* **43** 580–596.

[84] RESNICK, S. I. (1985). Uniform rates of convergence to extreme value distributions. Technical Report, Dept. Statistics, Colorado State Univ.

[85] RESNICK, S. I. (1986). Point processes, regular variation and weak convergence. *Adv. in Appl. Probab.* **18** 66–138.

[86] ROOTZÉN, H. (1978). Extremes of moving averages of stable processes. *Ann. Probab.* **6** 847–869.

[87] ROOTZÉN, H. (1983). The rate of convergence of extremes of stationary normal sequences. *Adv. in Appl. Probab.* **15** 54–80.

[88] ROOTZÉN, H. (1984). Attainable rates of convergence for maxima. *Statist. Probab. Lett.* **2** 219–221.

[89] ROOTZÉN, H. (1986). Extreme value theory for moving average processes. *Ann. Probab.* **14** 612–652.

[90] ROOTZÉN, H. (1987). A ratio limit theorem for the tails of weighted sums. *Ann. Probab.* **15** 728–747.

[91] ROOTZÉN, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. in Appl. Probab.* To appear.

[92] SERFOZO, R. (1980). High-level exceedances of regenerative and semi-stationary processes. *J. Appl. Probab.* **17** 423–431.

[93] SHARPE, K. (1978). Some properties of the crossings process generated by a stationary χ^2-process. *Adv. in Appl. Probab.* **10** 373–391.

[94] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell. System Tech. J.* **41** 463–501.

[95] SMITH, R. L. (1982). Uniform rates of convergence in extreme-value theory. *Adv. in Appl. Probab.* **14** 600–622.

[96] TIAGO DE OLIVEIRA, J. (1987). *Bibliography of Extreme-Value Theory.* To appear.

[97] VOLKONSKII, V. A. and ROZANOV, YU. A. (1959). Some limit theorems for random functions. I. *Theory Probab. Appl.* **4** 178–197.

[98] VON MISES, R. (1936). La distribution de la plus grande de n valeurs. Reprinted in *Selected Papers II* (1954) 271–294. Amer. Math. Soc., Providence, R.I.

[99] WATTS, V., ROOTZÉN, H. and LEADBETTER, M. R. (1982). Convergence in distribution of intermediate order statistics from stationary sequences. *Ann. Probab.* **10** 653–662.

[100] WEISSMAN, I. (1984). On the asymptotic joint distribution of an unbounded number of sample extremes. *Probab. Theory Related Fields*. To appear.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN
5, UNIVERSITETSPARKEN
DK-2100 COPENHAGEN Ø
DENMARK