

SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 627

March 1984

NONLINEAR SCALING OF SAMPLE MAXIMA

by

Eric V. Slud[†]

[†]Author is on leave from the Mathematics Department, University of Maryland, College Park.

Summary

It is shown that if F is a distribution function and $\{d_n\} \in \mathcal{D}$ a sequence of invertible continuous functions from \mathbb{R} to itself, where \mathcal{D} is a space of such functions (typically a finite-dimensional parametric family) satisfying a certain natural technical condition, then $F^n \circ d_n$ can have a nondegenerate (in a slightly modified sense) weak limit G only if there exists a homeomorphism h between some interval (a,b) and \mathbb{R} for which $n[1 - F(d_n \circ h^{-1}(-\log t))] \rightarrow t$ as $n \rightarrow \infty$ for all $t > 0$, and in this case $[a,b] = \text{support}(G)$ and $G(x) = \exp[-\exp(-h(x))]$. In this context, the Fisher-Tippett-Gnedenko theory of extreme-value distributions amounts to a characterization of the F for which this can happen when $d_n(x)$ must have the form $a_n x + b_n$. More general examples of families \mathcal{D} are given with corresponding enlarged domains of attraction to weak limits for sample maxima. These results are also related via theorems of Darling (1952) to the problem of nonlinear scaling of sample sums.

Introduction. In standard discussions of extreme-value theory--especially in works describing applications (eg. Barlow and Proschan, 1981; Gumbel, 1958)--one is left with an impression of inevitability of the "three types of extreme-value distribution" obtainable as weak limits of $F^n(a_n x + b_n)$ as $n \rightarrow \infty$, where F is some distribution function (d.f.) and $a_n > 0$ and b_n are constants. This impression is slightly misleading for a few reasons: first, because the most general weak-convergence theorem for sample maxima does not require linear (or any parametric) rescaling; second, because the family of linear rescalings is not uniquely distinguished among families of parametric rescalings of maxima with respect to weak convergence; and third, because there is not enough variety within the standard "three types" to accommodate efforts at modelling lifetime distributions in Reliability and Actuarial Science. We now discuss these three points in greater detail.

If $\{X_i\}_{i=1}^n$ is an independent identically distributed (i.i.d.) sequence of random variables (r.v.'s) with continuous d.f. F , then it is well known that $P\{\max_{1 \leq i \leq n} X_i \leq F^{-1}(1 - t/n)\} \rightarrow e^{-t}$ as $n \rightarrow \infty$. In other words, $n(1 - F(\max_{1 \leq i \leq n} X_i))$ has asymptotically a unit exponential distribution; a simple fact which lies at the heart of all asymptotic theory about sample maxima, which makes no appeal to linear rescalings $a_n^{-1}(\max_{1 \leq i \leq n} X_i - b_n)$ of the maximum, and which allows the greatest possible generality for domains of attraction to weak limits.

Our second remark above concerns the main subject of this paper. We are interested in the question: for which families $\{d_n(\cdot)\}$ of increasing functions from the real line R to itself and which d.f.'s F can $F^n \circ d_n$ have a non-degenerate weak limit, and what limits are possible? If

no restriction is placed upon allowable scaling sequences $(d_n(\cdot))$, then this question is rather trivial: every F for which

$$\delta_F \equiv \lim_{x \uparrow x_F} \sup (1 - F(x-)) / (1 - F(x)) = 1 \text{ where } x_F = \sup\{x: F(x) < 1\} \text{ (and only such } F)$$

allows every d.f. G as limit of $F^n \circ d_n$ for the choice $d_n(s) \equiv F^{-1}(1 + (\log G(s))/n)$. If one restricts $d_n(x)$ to have the linear form $a_n x + b_n$, one is led back to the usual Fisher-Tippett-Gnedenko theory. In Section 1 of this paper, we restrict $\{d_n(\cdot)\}$ to lie in a space \mathcal{D} of invertible continuous functions from \mathbb{R} to \mathbb{R} and derive some simple consequences of a general technical assumption on \mathcal{D} which will often be satisfied by finite-dimensional parametric families $\{f(\cdot, \theta): \theta \in \mathbb{R}^k\}$. After presenting examples in Section 2 to illustrate the assumption on \mathcal{D} , we prove our main theorem in Section 3 showing how the tail behavior of F and the structure of \mathcal{D} must interact to allow nondegenerate weak limits of $F^n \circ d_n$ for $\{d_n\} \in \mathcal{D}$. The Theorem, together with some further examples in Section 4 of domains of attraction, justifies our contention that the family of linear scalings for maxima is not distinguished in any abstract way, although it deserves pride of place as the parametric family easiest to write down which already includes in the domains of attraction to its possible weak limits (for $F^n \circ d_n$) nearly all the common d.f.'s F of practical interest. A similar situation exists with respect to nonlinear scalings for sums of i.i.d. variables, and we discuss in Section 5 the nonlinear scaling of sums for which it is known that no linear scaling can give nondegenerate limits.

Finally, our third remark about the standard extreme-value distributions is that the motivation to choose life distributions from among them or their counterparts for sample minima is usually rather weak. One

observes that the standard extreme-value distributions all have monotone hazard rate functions on their intervals of support. If one allows the slightly greater generality of distributions of maxima (or minima) of finitely many independent r.v.'s X_1, \dots, X_k with different extreme-value distributions, then one includes important life distributions with U-shaped hazards, such as the Makeham-Gompertz distributions of actuarial usage (Jordan, 1967). This is essentially the idea of Brillinger (1961).

Acknowledgements. The author is grateful to George O'Brien for introducing him to the subject of maxima and for many valuable conversations, and to Marjorie Hahn for making him aware of Darling's (1952) paper.

1. Formulation of problems. Let \mathcal{D} be a family of continuous strictly increasing functions $d: \mathbb{R} \rightarrow \mathbb{R}$ such that $d(-\infty) = -\infty$, $d(+\infty) = +\infty$. We assume \mathcal{D} satisfies

(D) if for sequences $\{e_n\}, \{d_n\} \in \mathcal{D}$ there exist real numbers $a < b$, $\alpha < \beta$, such that $\bigcup_{n \geq 1} d_n^{-1} \circ e_n([a, b])$ and $\bigcup_{n \geq 1} e_n^{-1} \circ d_n([\alpha, \beta])$ are bounded sets, then there exists a subsequence $\{n_k\}$ of integers such that $d_{n_k}^{-1} \circ e_{n_k}$ converges pointwise to a continuous strictly increasing limit Δ .

We recall in passing that pointwise convergence of nondecreasing functions to a continuous function is equivalent to uniform convergence on compact subsets of \mathbb{R} . Let \mathcal{D}^* denote the set of such limit functions Δ .

Definition. We call a d.f. G on \mathbb{R} \mathcal{D} -degenerate if there exists $\Delta \in \mathcal{D}^*$ not the identity function such that $G \equiv G \circ \Delta$ (or if G is degenerate).

Remark. Each Δ is continuously invertible on its range, which might a priori be smaller than \mathbb{R} . If G is such that $G \circ \Delta \equiv G$, then $\Delta(x) = x$ for every point of increase of G . Thus G \mathcal{D} -degenerate implies there exists $\Delta \in \mathcal{D}^*$, not the identity, which fixes the support of G . Typically, if \mathcal{D} has k -dimensional smooth parameterization, the \mathcal{D} -degenerate d.f.'s will concentrate at $k - 1$ or fewer atoms.

Lemma 1. (generalizing a Theorem of Khinchin, Theorem 1.2.3 in Leadbetter et al., 1983). Let $\{F_n\}$ be a sequence of d.f.'s and G a non \mathcal{D} -degenerate d.f., and suppose $d_n \in \mathcal{D}$ are such that

$$F_n(d_n(\cdot)) \xrightarrow{W} G(\cdot) \text{ as } n \rightarrow \infty.$$

Then for some nondegenerate d.f. G_* and $\{\bar{d}_n\} \subset \mathcal{D}$,

$$F_n(\bar{d}_n(\cdot)) \xrightarrow{w} G_*(\cdot)$$

iff $d_n^{-1} \circ \bar{d}_n$ converges uniformly on compact sets to some $\Delta \in \mathcal{D}^*$ and $G \circ \Delta \equiv G_*$.

Proof. (\Rightarrow) Since G_* and G are nondegenerate, there exist $a < b$ and $\alpha < \beta$ such that $0 < G_*(a) \leq G_*(b) < 1$, $0 < G(\alpha) \leq G(\beta) < 1$. Now

$\bigcup_n d_n^{-1} \circ \bar{d}_n([a, b])$ and $\bigcup_n \bar{d}_n^{-1} \circ d_n([\alpha, \beta])$ must be bounded, for otherwise, e.g. if $d_{n_k}^{-1} \circ \bar{d}_{n_k}(b) \rightarrow \infty$, then the weak convergence of $F_n(d_n(\cdot))$ to a proper d.f. $G(\cdot)$ implies $F_{n_k}(\bar{d}_{n_k}(b)) = F_{n_k} \circ d_{n_k}(d_{n_k}^{-1} \circ \bar{d}_{n_k}(b)) \rightarrow 1$ as $k \rightarrow \infty$: a contradiction. Then property (D) of \mathcal{D} implies every subsequence of $\{d_n^{-1} \circ \bar{d}_n\}$ possesses a further subsequence $d_{n_k}^{-1} \circ \bar{d}_{n_k}(\cdot)$ converging uniformly on compact sets to some $\Delta \in \mathcal{D}^*$ (which a priori may depend on the subsequence.) Therefore, $F_{n_k} \circ d_{n_k}(d_{n_k}^{-1} \circ \bar{d}_{n_k}(x))$ lies, for each continuity point of $G \circ \Delta$, each $\epsilon > 0$, and all sufficiently large k , between $G \circ \Delta(x - \epsilon) - \epsilon$ and $G \circ \Delta(x + \epsilon) + \epsilon$. Thus $F_{n_k} \circ \bar{d}_{n_k} \xrightarrow{w} G \circ \Delta$, so that $G \circ \Delta = G_*$. By non \mathcal{D} -degeneracy of G , this relation determines Δ uniquely, and we conclude $d_n^{-1} \circ \bar{d}_n \rightarrow \Delta$ pointwise as $n \rightarrow \infty$.

(\Leftarrow) Easy, by steps like those just used to show weak convergence of $F_{n_k} \circ \bar{d}_{n_k}$ to $G \circ \Delta$. \square

Now we can ask our main question:

(1) For what d.f.'s $F(\cdot)$ with corresponding i.i.d. sequences $\{X_i\}_{i=1}^{\infty}$ do there exist sequences $\{d_n\} \in \mathcal{D}$ such that $d_n^{-1}(\max_{1 \leq i \leq n} X_i)$ (with d.f. $F^n \circ d_n$) converges in distribution to some non (\mathcal{D}) -degenerate r.v. with d.f. G ?

What is of interest here is primarily the interaction between tail behavior of $F(\cdot)$ and necessary structure of \mathcal{D} , not the class of possible limits, since for any fixed homeomorphism ϕ of \mathbb{R} with itself, if $\{\phi(ax + b): a > 0, b \in \mathbb{R}\} \in \mathcal{D}$ then one may obtain all $H \circ \phi$ as limits G where H is one of the standard "extremal types" of Fisher-Tippett (1928) and Gnedenko (1943).

A related (but much harder) question, dealt with only briefly in this paper, is:

(2) When can $d_n^{-1}(\sum_{i=1}^n X_i)$ converge in distribution to a r.v. with non-degenerate d.f. G ?

2. Examples of nonlinear scaling families.

(i) The example of \mathcal{D} on which virtually all previous work has been done is $\mathcal{D}_L \equiv \{ax + b: a > 0, b \in \mathbb{R}\}$. If $d_n(x) \equiv a_n x + b_n$ for $n \geq 1$ and $\bigcup_{n \geq 1} d_n([\xi, \eta])$ is bounded, where $\xi < \eta$, then obviously both $\{a_n\}$ and $\{b_n\}$ are bounded and have convergent subsequences: however a_n might converge to 0. If also $\bigcup_{n \geq 1} d_n^{-1}([\alpha, \beta])$ is bounded for some $\alpha < \beta$, then $\{1/a_n\}$ is also bounded. Since \mathcal{D} is a group under composition, it satisfies (D). Also, in this case \mathcal{D} -degeneracy and degeneracy are the same.

(ii) For every strictly monotone continuous function ϕ sending \mathbb{R} onto itself, the family $\mathcal{D}_\phi \equiv \{\phi(a \cdot \phi^{-1}(\cdot) + b): a > 0, b \in \mathbb{R}\}$ is conjugate-equivalent to \mathcal{D}_L and provides another example of \mathcal{D} satisfying (D).

(iii) Examples based on finite-order polynomials can take the form

$\mathcal{D}_{P,M} \equiv \left\{ \sum_{i=0}^M (a_i x + b_i)^{2i+1} : a_i \geq 0, \sum a_i > 0, b_i \in \mathbb{R} \right\}$. Here again the condition (D) holds, although it is more laborious to verify. The main idea

is the following: suppose $d_n(x) \equiv \sum_{i=0}^M (a_i^{(n)} x + b_i^{(n)})^{2i+1}$ and

$\bar{d}_n(x) \equiv \sum_{i=0}^M (\bar{a}_i^{(n)} x + \bar{b}_i^{(n)})^{2i+1}$ with $\bigcup_n d_n^{-1} \circ \bar{d}_n([a, b])$ and

$\bigcup_n \bar{d}_n^{-1} \circ d_n([\alpha, \beta])$ bounded and not all coefficients $a_i^{(n)}, b_i^{(n)}, \bar{a}_i^{(n)}, \bar{b}_i^{(n)}$ bounded as $n \rightarrow \infty$. Then one can find dominant values of i, j such that,

along subsequences of integers $n = n_k$, for x in bounded intervals

$(a_i^{(n)} x + b_i^{(n)})^{2i+1}$ and $(\bar{a}_j^{(n)} x + \bar{b}_j^{(n)})^{2j+1}$ are of the same order of

magnitude as $d_n(x)$ and $\bar{d}_n(x)$ respectively. There may be more than one

choice for each of i, j ; but in case they are unique, then (at least along

subsequences of n_k) one can check that $d_n^{-1} \circ \bar{d}_n(x)$ is asymptotic to a

convergent sequence of functions $[(\bar{a}_j^{(n)} x + \bar{b}_j^{(n)})^{\frac{2j+1}{2i+1}} - b_i^{(n)}] / a_i^{(n)}$ which

can have limits only of the form $(\xi x + \eta)^{\frac{2j+1}{2i+1}}$. The most general limits in $\mathcal{D}_{P,M}^*$ will always be contained in $\{d^{-1} \circ \bar{d} : d \in \mathcal{D}_{P,M_1}, \bar{d} \in \mathcal{D}_{P,M_2}, 0 \leq M_i \leq M\}$.

(iv) A particularly fruitful example for our later considerations is $\mathcal{D}_E \equiv \{(ax + b)^c : a, c > 0, b \in \mathbb{R}\}$, where for $z < 0$ we understand z^c to mean $(\operatorname{sgn} z)|z|^c$. If $d_n(x) \equiv (a_n x + b_n)^{c_n}$ and $\bar{d}_n(x) \equiv (\bar{a}_n x + \bar{b}_n)^{\bar{c}_n}$, then

$$d_n^{-1} \circ \bar{d}_n(x) = \frac{(\bar{a}_n x + \bar{b}_n)^{\bar{c}_n/c_n} - b_n}{a_n}$$

and an examination of cases (to which we will return later) shows that condition (D) holds.

(v) If \mathcal{D} is the collection of all homeomorphisms $d: \mathbb{R} \xrightarrow{\text{onto}} \mathbb{R}$, then (D) does not hold, since the subcollection of all homeomorphisms of \mathbb{R} which fix $[a,b]$ is not relatively compact with respect to the topology of uniform convergence on $[a,b]$. In this example, G is \mathcal{D} -degenerate iff $\operatorname{support}(G) \neq \mathbb{R}$.

3. Reductions and main theorem. Consider question (1) re-stated as: when can $F^n(d_n(\cdot)) \xrightarrow{w} G(\cdot)$ for non- \mathcal{D} -degenerate G with $\{d_n\} \in \mathcal{D}$? It makes sense to restrict to non- \mathcal{D} -degenerate G , especially when dealing with classes \mathcal{D} of parametric functions with finite-dimensional parameter, since \mathcal{D} -degenerate d.f.'s are typically concentrated at a finite number of atoms (related to the dimension of parameter-space) and can occur as weak-limiting d.f.'s for $d_n^{-1}(\max_{1 \leq i \leq n} X_i)$ with $d_n \in \mathcal{D}$.

We further restrict attention to d.f.'s for which

$$\delta_F \equiv \limsup_{x \uparrow x_F} (1 - F(x-)) / (1 - F(x)) = 1,$$

where $x_F \equiv \sup\{t: F(t) < 1\}$, since O'Brien (1974) shows that there can be no nondegenerate limits $G(\cdot)$ (indeed, no limits of $F^n(d_n(x))$ strictly between 0 and 1) if $\delta_F > 1$.

Our method of deriving functional equations for $G(\cdot)$ in this setting follows closely the ideas of de Haan (1970) as expounded in Leadbetter, Lindgren and Rootzén (1983, Chapter 1). Suppose now that $F^n(d_n(\cdot)) \xrightarrow{w} G(\cdot)$ as $n \rightarrow \infty$, where δ_F is 1, $d_n \in \mathcal{D}$ satisfying (D), and $G(\cdot)$ is non- \mathcal{D} -degenerate. For each fixed $s > 0$, we have $F^{[ns]}(d_{[ns]}(\cdot)) \xrightarrow{w} G(\cdot)$ as $n \rightarrow \infty$, where $[\]$ here denotes greatest-integer function. Thus $F^n(d_{[ns]}(\cdot)) \xrightarrow{w} G^{1/s}(\cdot)$ and our Lemma implies $d_n^{-1} \circ d_{[ns]} \rightarrow \Delta_s \in \mathcal{D}^*$ uniformly on compact sets, while $G \circ \Delta_s \equiv G^{1/s}$.

Remark. With not much greater difficulty, one finds the analogous functional equation for (2) to be:

$$\begin{aligned} &\text{for each integer } k \geq 1, \text{ there exists } \Delta_{(k)} \in \mathcal{D}^* \\ &\text{such that } G^{*k} \circ \Delta_{(k)} = G \end{aligned}$$

where G^{*k} denotes k -fold convolution-product. However, since characteristic functions do not necessarily behave well under composition with \mathcal{D}^* functions, the author has made no further progress on (2), which remains open.

We require one further definition for the statement of our result for (1).

Definition: for a given d.f. $F(\cdot)$ with $\delta_F = 1$, we call $\{c_n(\cdot)\}$ (a sequence of right-continuous functions from $[0, \infty]$, to \mathbb{R}) a scaled quantile sequence if for each $s > 0$,

$$n\bar{F}(c_n(s)) \rightarrow s \quad \text{as } n \rightarrow \infty.$$

Here and from now on, we use the notation $\bar{F}(x)$ to denote $1 - F(x)$.

To motivate our Theorem, suppose there exists some increasing homeomorphism $h(\cdot)$ of a subinterval of \mathbb{R} onto \mathbb{R} itself and some scaled quantile sequence $\{c_n(\cdot)\}$ for F such that $c_n(e^{-h(\cdot)}) \in \mathcal{D}$ for each $n \geq 1$. Then putting $d_n(\cdot) \equiv c_n(e^{-h(\cdot)})$, we have by definition for each $t \in \mathbb{R}$, $\lim_{n \rightarrow \infty} n\bar{F} \circ d_n(t) = e^{-h(t)}$, and

$$F^n(d_n(t)) = (1 - \bar{F} \circ d_n(t))^n \rightarrow \exp(-e^{-h(t)})$$

as $n \rightarrow \infty$.

Theorem. Assume F has $\delta_F = 1$ and \mathcal{D} satisfies (D). The d.f.'s $F^n \circ d_n$ for $\{d_n\} \subset \mathcal{D}$ have a non- \mathcal{D} -degenerate weak limit G if and only if there exists a continuous strictly increasing function $h(\cdot)$ of a subinterval (ξ, η) of \mathbb{R} such that $h(\xi) = -\infty$, $h(\eta) = +\infty$, for which $\{d_n \circ h^{-1}(-\log t)\}_n$ is a scaled quantile sequence for F .

Remark. There would be nothing more to prove if we knew G had points of increase (ξ, η) and support $[\xi, \eta]$, for then we could reverse the steps of our "motivation" above. As it is, we must use our functional equation for G to prove these properties and extract the function $h(\cdot)$.

Proof. In deriving a functional equation, we have found a measurable curve $s \rightarrow \Delta_s(\cdot)$ from $[0, \infty)$ to \mathcal{D}^* . Letting $t = \log(s)$, $\sigma_t(\cdot) \equiv \Delta_s(\cdot)$, and $\phi(\cdot) \equiv -\log(-\log G(\cdot))$ (as an extended-real-valued function) we can rewrite our functional equation as

$$(3.1) \quad \phi \circ \sigma_t - t \equiv \phi \quad \text{for each } t \in \mathbb{R}$$

The relation $G \circ \Delta_2 \equiv G^{1/2}$ already implies that G cannot have a jump at either extreme of its support, so we can regard ϕ as a right-continuous monotone function on a closed interval $[-\infty, \infty]$ to itself, with $\phi(-\infty) = -\infty$, $\phi(+\infty) = +\infty$, which is continuous at the endpoints. Let $\phi(x) \equiv \phi^{-1}(x) \equiv \inf\{t: \phi(t) > x\}$ define the right-continuous inverse of ϕ on $[-\infty, \infty]$. We want to re-write the last equation in terms of inverse-functions, but the \mathcal{D}^* functions $\sigma_t(\cdot)$ are so far known only to be continuously invertible on their ranges $R_t \subset \mathbb{R}$.

Now if $\phi(b) = \infty$ for $b < \infty$, then (3.1) implies for all t , $\phi(\sigma_t(b)) = \infty$. Putting $b = x_G \equiv \sup\{x: G(x) < 1\}$, we find $\sigma_t(x_G) \geq x_G$, while for each $\epsilon > 0$, $\phi(x_G - \epsilon) < \infty$ implies $\sigma_t(x_G - \epsilon) < x_G$. By continuity of $\sigma_t(\cdot)$, we conclude x_G and similarly $w_G \equiv \inf\{x: G(x) > 0\}$ must be fixed points of all $\{\sigma_t(\cdot)\}_{t \in \mathbb{R}}$, and $\text{range}(\sigma_t) \supset (w_G, x_G)$. Thus we can regard $\phi(\cdot)$ and $\sigma_t(\cdot)$ as functions from the (possibly infinite) closed interval $[w_G, x_G]$ respectively to $[-\infty, \infty]$ and $[w_G, x_G]$, and each

σ_t is continuous and continuously invertible. Taking inverse functions in (3.1), we can now write

$$(3.2) \quad \phi(y) = \sigma_t^{-1}(\phi(t + y))$$

or $\sigma_t(\phi(y)) = \phi(t + y)$ for all $t \in \mathbb{R}$. Therefore, if $z_0 \in (-\infty, \infty)$ is not a point of (left or right) increase for $\phi(\cdot)$, then neither is $t + z_0$. (Recall that σ_t is strictly increasing). So $\phi: [-\infty, \infty] \rightarrow [w_G, x_G]$ must in fact be strictly increasing. Similarly, if $\phi(\cdot)$ has a jump at y , then it must also have a jump at $t + y$, so that $\phi(\cdot)$ is also continuous. Hence for each $x \in [w_G, x_G]$, $\sigma_t(x)$ is strictly increasing and continuous in $t \in \mathbb{R}$. From (3.2) we deduce

$$(3.3) \quad \phi(t) = \sigma_t(\phi(0)), \quad \sigma_{y+t}(\phi(z)) = \sigma_y(\sigma_t(\phi(z))).$$

That is, $\{\sigma_t(\cdot)\}_{t \in \mathbb{R}}$ is a family of transformations on $[w_G, x_G]$ which forms a group under composition. Putting $z = 0$ in (3.3), and recalling $\phi^{-1}(\cdot) \equiv \phi(\cdot)$, we have (writing $\xi = \phi(t) = \sigma_t(\phi(0))$)

$$(3.4) \quad \begin{aligned} \sigma_y(\xi) &= \sigma_y(\sigma_t(\phi(0))) = \sigma_{t+y}(\phi(0)) = \sigma_{\phi(\xi)+y}(\phi(0)) \\ &= \phi(\phi(\xi) + y) = \phi^{-1}(y + \phi(\xi)). \end{aligned}$$

We return now to make use of the primary assumption of our Theorem, namely that $F^n(d_n(\cdot)) \xrightarrow{W} G(\cdot)$. On the one hand, we know from the definition of $\phi(\cdot)$ that $G(\cdot) = \exp(-e^{-\phi(\cdot)})$, so that $G(\cdot)$ is continuous and

strictly increasing on $[w_G, x_G]$. Therefore our weak convergence is point-wise convergence, and we have for every $t \in R$,

$$(1 - (1 - F(d_n(t))))^n \xrightarrow{n \rightarrow \infty} \exp(-e^{-\phi(t)})$$

so that

$$n(1 - F(d_n(t))) \xrightarrow{n \rightarrow \infty} e^{-\phi(t)}$$

which implies $\{d_n \circ \phi^{-1}(-\log t)\}_n$ is a scaled quantile sequence for F .

Thus the function $h(\cdot)$ in the statement of the Theorem is $\phi(\cdot)$. \square

4. Domains of attraction. The Theorem of Section 3 describes the interplay between the tail behavior of a d.f. F and the structure of scaling family \mathcal{D} allowing weak limits for $F^n \circ d_n$ where $d_n \in \mathcal{D}$. In the present Section, for the various families \mathcal{D} given in examples (i)-(iv) of Section 2, we summarize which classes of F allow weak limits $F^n \circ d_n \rightarrow G$ for $d_n \in \mathcal{D}$.

(i) The large body of work (taken farthest by Gnedenko, 1943; Marcus and Pinsky, 1969; and de Haan, 1970) on domains of attraction for non-degenerate weak limits of maxima of i.i.d. random samples, has focussed exclusively on the question re-expressed by our Theorem as:

(4.1) for which d.f.'s F (with $\delta_F = 1$) does there exist a homeomorphism h from a subinterval I_h of \mathbb{R} onto \mathbb{R} and a scaled quantile sequence $\{c_n(\cdot)\}$ for F such that

$$c_n(e^{-h(\cdot)}) \in \mathcal{D}_L \quad \text{for each } n > 1, \text{ i.e.,}$$

$$\text{for some } a_n > 0, b_n \in \mathbb{R}, \quad c_n(\exp(-h(t))) \equiv a_n t + b_n?$$

Gnedenko (1943) already showed that h could have only three possible forms: (I) $I_h = \mathbb{R}$, $h(x) \equiv ax + b$ with $a > 0$, $b \in \mathbb{R}$; (II) $I_h = (-b/a, \infty)$ $h(x) \equiv \alpha \log(at + b)$, with $\alpha, a > 0$, $b \in \mathbb{R}$; (III) $I_h = (-\infty, -b/a)$, $h(x) \equiv -\alpha \log(-at-b)$, with $\alpha, a > 0$, $b \in \mathbb{R}$. Gnedenko showed also that the d.f.'s F corresponding in (4.1) to these three forms were respectively (I) those d.f.'s for which there exists positive measurable $g(\cdot)$ such that $\lim_{t \uparrow x_F} \bar{F}(t + xg(t))/\bar{F}(t) = e^{-x}$; (II) those d.f.'s with $x_F = \infty$ such that $\lim_{t \uparrow \infty} \bar{F}(tx)/\bar{F}(t) = x^{-\alpha}$ for $\alpha > 0$; and (III) those d.f.'s with

$x_F < \infty$ such that $\lim_{h \rightarrow 0} \bar{F}(x_F - xh)/\bar{F}(x_F - h) = x^\alpha$ for some $\alpha > 0$. In what follows, we refer to $G(\cdot) \equiv \exp(-\exp(-h(\cdot)))$ in (I)-(III) as the three standard forms of extreme-value distributions, and the allowable F as the three standard domains of attraction.

(ii) For each homeomorphism ϕ of R onto itself, there exist $\{a_n, b_n\}$ such that $F^n(\phi(a_n \phi^{-1}(\cdot) + b_n)) \xrightarrow{W} H(\cdot)$ (nondegenerate) as $n \rightarrow \infty$ if and only if $F \circ \phi$ is in one of the three standard domains of attraction, and the possible limits $H(\cdot)$ are of the form $G \circ \phi^{-1}$ with G a standard extreme-value d.f. Thus all H with support a half-line or R itself are possible limits of $F^n \circ d_n$ with $\{d_n\} \subset \mathcal{D}_\phi$ for some ϕ .

(iii) Suppose $\{d_n\} \subset \mathcal{D}_{p, M}$ is such that as $n \rightarrow \infty$, $F^n \circ d_n \xrightarrow{W} G$ nondegenerate. Possibly after passing to a subsequence, we can find monomials

$\bar{d}_n(x) \equiv (a_i^{(n)} x + b_i^{(n)})^{2i+1}$ such that $d_n^{-1} \circ \bar{d}_n$ converges pointwise and uniformly on compact sets, say to Δ (see example (iii) of Section 2.)

Then $F^n \circ \bar{d}_n \xrightarrow{W} G \circ \Delta$, and upon letting $\hat{F}(y) \equiv F(y^{2i+1})$, we find that \hat{F} must be in one of the standard three domains of attraction. If \hat{F} is of type (II), then obviously so is F (with α replaced by $\alpha/(2i+1)$). If \hat{F} is of type (I) with given positive $\hat{g}(\cdot)$, then F is also with $g(\cdot)$ given by $g(s) \equiv (2i+1)s^{2i/(2i+1)} \hat{g}(s^{1/(2i+1)})$ for $s \in R$. Finally, if \hat{F} is of type (III), we assume $x_F \neq 0$ (since for $x_F = 0$, F is obviously of type (III)). Then for fixed $x > 0$, as $h \rightarrow 0$, $\bar{F}((x_F^{1/(2i+1)} - xh)^{2i+1})/\bar{F}((x_F^{1/(2i+1)} - h)^{2i+1}) \rightarrow x^\alpha$. Putting $x_F - K \equiv (x_F^{1/(2i+1)} - h)^{2i+1}$, and observing for each $\epsilon > 0$ that $(x_F^{1/(2i+1)} - (x+\epsilon)K)^{2i+1} \leq x_F - xK \leq (x_F^{1/(2i+1)} - (x-\epsilon)K)^{2i+1}$, we conclude that $\lim_{K \rightarrow 0} \bar{F}(x_F - xK)/\bar{F}(x_F - K) = x^\alpha$ and F is of type (III). Thus we have proved (via Gnedenko's 1943 results)

Proposition. The set of d.f.'s F such that for some $M \geq 0$ there exist $\{d_n\} \in \mathcal{D}_{P,M}$ with $F^n \circ d_n \xrightarrow{w} H$ for some non-degenerate d.f. H , is precisely the union of the three standard domains of attraction (I)-(III).

The possible limits H are the standard extreme-value distributions.

(iv) The family $\mathcal{D}_E \equiv \{(ax + b)^c : a, c > 0, b \in \mathbb{R}\}$ is an excellent example to show how domains of attraction can be expanded by scaling maxima using transformations $d_n(\cdot)$ outside \mathcal{D}_L . As we have just seen, choosing $d_n(x) = (a_n x + b_n)^c$ for fixed $c > 0$ cannot possibly give weakly-convergent $F^n \circ d_n$ unless F were already in one of the standard domains of attraction. However, if $d_n(x) \equiv \alpha_n(x + \beta)^c$ with fixed $\beta \in \mathbb{R}$, then the possible cases of non-degenerate limits for $F^n \circ d_n$ are derived by simple transformations from the standard three (in all cases a and α are positive parameters): (I') if $x_F > 0$, and there exists $g(\cdot) > 0$ such that $\bar{F}(s \exp[xg(s)])/\bar{F}(s) \rightarrow e^{-x}$ as $s \uparrow x_F$, then there is a weak limit of the form $\exp(-a(t + \beta)^{-\alpha})$ for $t > -\beta$; (I'') if $x_F \leq 0$ and there exists $g(\cdot) > 0$ with $\bar{F}(s \exp[-xg(s)])/\bar{F}(s) \rightarrow e^{-x}$ as $s \uparrow x_F$, then a weak limit has the form $\exp(-a|t + \beta|^\alpha)$ for $t \leq -\beta$; (II') if $x_F = +\infty$ and $\bar{F}(s^x)/\bar{F}(s) \rightarrow x^{-\alpha}$ as $s \rightarrow \infty$, then a weak limit has the form $\exp[-a(\log(t + \beta) + c)^{-\alpha}]$ for $t > e^{-c} - \beta$; (III') if $0 < x_F < \infty$ and $\bar{F}(x_F \lambda^x)/\bar{F}(x_F \lambda) \rightarrow x^\alpha$ as $\lambda \uparrow 1$, then the limit is of the form $\exp[-\alpha|\log(t + \beta) + c|^\alpha]$ for $-\beta < t \leq e^{-c} - \beta$; (II'') if $x_F = 0$ and $\bar{F}(-|s|^x)/\bar{F}(s) \rightarrow x^{-\alpha}$ as $s \uparrow 0$, then a limit has the form $\exp[-a(c - \log|t + \beta|)^{-\alpha}]$ for $-e^c - \beta < t < -\beta$; and (III'') if $x_F < 0$ and $\bar{F}(x_F \lambda^x)/\bar{F}(x_F \lambda) \rightarrow x^\alpha$ as $\lambda \uparrow 1$, then the weak limit has the form $\exp[-a(\log|t + \beta| - c)^\alpha]$ for $t < -\beta - e^c$.

It is easy to see that domains (II'), (III'), (II''), and (III'') (the last two of which are relevant to Reliability and allow limiting distributions with hazards not monotone on their supports, e.g. U-shaped hazards) are disjoint from the standard domains of attraction for extremal distributions and include, for example, d.f.'s with slowly varying tails such as $F(x) \sim a/(\log x)^\alpha$ as $x \rightarrow \infty$ which will play a role in our subsequent discussion. The domains (I') and (I''), which respectively include domain (II) and those parts of domain (III) with $x_F < 0$, actually contain more. For example, the d.f. $F(x) \equiv 1 - \exp(-(\log x)^2)$ for $x \geq 1$ belongs to domain (I') with $g(s) \equiv (2 \log s)^{-1}$ and $x_F = \infty$, but evidently does not belong to (II).

The other scaling sequences $\{d_n\} \in \mathcal{D}_E$ which deserve special mention are $d_n(x) \equiv (a_n x + b_n)^{c_n}$ such that $|b_n| \rightarrow \infty$ with $c_n(a_n/|b_n|)^2$ converging to 0. Such $d_n(x)$ are asymptotically of the form $\pm a'_n \exp(b'_n x)$ and one can again by simple transformations of the standard domains and limits read off the domains of attraction and limits for $F^n \circ d_n$ with such d_n .

5. Complements and discussion. As we have argued in the Introduction, with justification from our Theorem, results about limiting distributions for maxima are most naturally presented in terms of limiting behavior of $P(\max(X_1, \dots, X_n) \leq c_n(s))$ for scaled quantile sequences $c_n(\cdot)$. Indeed, the most far-reaching weak-convergence results related to maxima have been stated in this way, and these carry over without difficulty to statements about $\{d_n^{-1}(\max(X_1, \dots, X_k)): 1 \leq k \leq n\}$ whenever our Theorem applies. The generalizations of the Gnedenko (1943) theory of asymptotic, distributions for sample maxima are of two main types: first, functional limit theorems for partial maxima and for record-value processes (implying, for example, asymptotic joint distributions of the k largest order-statistics from i.i.d. samples) based on Resnick's (1975) point-process approach; the second direction of generalization (in the work of Loynes, 1965; O'Brien, 1974 and 1984; and Leadbetter, 1983) is from theorems about i.i.d. sample maxima to maxima of stationary sequences satisfying conditions on mixing and local dependence. The theorems of O'Brien (1974) especially have shown quite generally that up to an additional scaling parameter the behavior of i.i.d. sample maxima and stationary-sequence maxima are the same. Thus for marginal d.f.'s F and scaling families \mathcal{D} to which our Theorem applies, O'Brien's (1974) Theorem 3 implies the painless generalization to maxima for uniformly mixing stationary sequences (satisfying additional conditions on local dependence). Leadbetter's (1983) and O'Brien's (1984) results imply slightly broader generalizations.

Our question (2) in Section 1 about weak convergence of nonlinearly scaled sums of i.i.d. random variables was discussed more than thirty years ago by Darling (1952). Of course, as he noted, for any d.f. F for which

the tails $F(-x) + 1 - F(x)$ vary regularly at ∞ with positive exponent or are of smaller order than $x^{-\alpha}$ for every $\alpha > 0$ as $x \rightarrow \infty$, it is well known that a linear scaling $F^{*n}(a_n x + b_n)$ has a nondegenerate weak limit. In addition, Darling (1952) proved the following

Proposition. Suppose $\lim_{t \rightarrow \infty} \bar{F}(tx)/\bar{F}(t) = 1$ for each $x > 0$ and $\lim_{t \rightarrow \infty} F(-t)/(F(-t) + \bar{F}(t)) = p < 1$ exists, where p may equal 0. Let $q = 1 - p$, $\{X_i\}_{i=1}^n$ be i.i.d. F -distributed, $S_n \equiv \sum_{i=1}^n X_i$, and let X_n^* be the term X_i with largest absolute value, $1 \leq i \leq n$.

Put

$$H(x) \equiv \begin{cases} \bar{F}(x) & \text{if } x \geq 0 \\ F(x) & \text{if } x < 0. \end{cases}$$

Then as $n \rightarrow \infty$,

$$P\{nH(S_n) \geq y\} \rightarrow p \exp(-y/p) + q \exp(-y/q)$$

and

$$E|S_n/X_n^* - 1| \rightarrow 0.$$

(Quite generally, the condition $\delta_F = 1$ with $F(x_F-) = 1$ is equivalent to the property used by Darling that

$$P\{\max_{1 \leq i \leq n} X_i = X_j = X_k \text{ for some } j \neq k, 1 \leq j, k \leq n\} \rightarrow 0.$$

In particular if F has slowly varying tail, i.e. $\bar{F}(tx)/\bar{F}(t) \rightarrow 1$ as $t \rightarrow \infty$, $\delta_F = 1$ and $x_F = \infty$.)

The foregoing Proposition implies, for the significant class of d.f.'s F satisfying its hypothesis, that while no linear scaling for S_n leads to a non-degenerate limiting distribution, any nonlinear scaling $\{d_n\}$ with $d_n^{-1}(X_n^*)$ convergent in distribution will also give $d_n^{-1}(S_n)$ a non-degenerate asymptotic distribution. For example, if $\bar{F}(e^x)$ is regularly varying at ∞ with exponent $\alpha > 0$ and, say, $F(-t)/\bar{F}(t) \rightarrow 0$ as $t \rightarrow +\infty$, then F is in the domain of attraction (II') for \mathcal{D}_E in Section 4 (iv), and by the Proposition we conclude (for $\{X_i\}_{i=1}^\infty$ i.i.d. with d.f. F)

$$(X_1 + \dots + X_n)^{n^{-1/\alpha}} \rightarrow Y \text{ in distribution as } n \rightarrow \infty$$

with $P\{Y \geq y\} = \exp(-y^{-\alpha})$ for $y > 0$. This example (for $\alpha = 1$) was given by Darling. It is clear that other F for which $\bar{F}(x)$ is regularly varying at ∞ as a function of a slowly varying function $L(x)$ (other than $\log x$) can by other scaling-functions d_n be brought within domains of attraction for weak convergence of maxima and sums.

References

- Barlow, R. and Proschan, F. (1981), Statistical Theory of Reliability and Life Testing. Silver Spring: To Begin With.
- Brillinger, D. (1961), Justification of some common laws of mortality. Trans. Soc. of Actuaries 13, 116-119.
- Darling, D. (1952), The influence of the maximum term in the addition of independent random variables. Trans. Amer. Math. Soc. 72, 95-107.
- De Haan, L. (1970), On regular variation and its application to weak convergence of sample extremes. Math. Centre Tracts 32, Amsterdam.
- Fisher, R.A. and Tippett, L. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Camb. Philos. Soc. V. XXIV: 180-190.
- Gnedenko, B. (1943), Sur la distribution du terme maximum d'une série aléatoire. Ann. Math. 44, 423-453.
- Gumbel, E.J. (1958), Statistics of Extremes. New York: Columbia Univ. Press.
- Jordan, C.W. (1967), Life Contingencies, 2nd ed. Chicago: Soc. of Actuaries.
- Leadbetter, M.R. (1983), Extremes and local dependence in stationary sequences. Z. Wahrscheinlichkeitstheorie verw. Geb. 65, 291-306.
- Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983), Extremes and Related Properties of Random Sequences and Processes. New York: Springer-Verlag.
- Loynes, R.M. (1965), Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Statist. 36, 993-999.
- Marcus, M. and Pinsky, M. (1969), On the domain of attraction of $e^{-e^{-x}}$. J. Math. Anal. Appl. 28, 440-449.
- O'Brien, G. (1974), Limit theorems for the maximum term of a stationary process. Ann. Prob. 2, 540-545.
- O'Brien, G. (1984), Extreme values in stationary sequences and functions of Markov sequences. Preprint.
- Resnick, S. (1975), Weak convergence to extremal processes. Ann. Prob. 3, 951-960.