THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

GENERALIZED ASYMPTOTES FOR EXTREME VALUE DISTRIBUTIONS

by

Charles L. Anderson

Technical Report No. 34
Department of Statistics THEMIS Contract

May 14, 1969

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

GENERALIZED ASYMPTOTES FOR EXTREME VALUE DISTRIBUTIONS

A Dissertation Presented to the Faculty of the Graduate School

of

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistics

by

Charles Lloyd Anderson
(M.S., Statistics, Southern Methodist University, 1967)

April 19, 1969

Anderson, Charles Lloyd B.A., Southern Methodist University, 1963
M.S., Southern Methodist University, 1967

<u>Generalized Asymptotes for Extreme Value Distributions</u>

Adviser: Professor John E. Walsh

Doctor of Philosophy Degree conferred May 25, 1969

Dissertation completed April 19, 1969

Some new asymptotic forms for extreme value distributions are given, the most important being

$$\exp\ (-\ \exp\ (-\ a^2 (X - b)^2 \operatorname{sgn}(X - b) + c^2))$$

where X is the identity function on the extended real line and a, b, and c are real parameters. The family of distributions of this form is called the quadratic type of dne.

It is shown that the sequence of extreme value distributions from a normal distribution is asymptotically attracted to the quadratic type of dne in a stronger sense, related to Walsh's "Situation I," than the sense in which it is attracted to the linear type of dne, or first asymptotic type of extreme value distributions, which consists of all distributions of the form exp(- exp(-aX + b)) with a positive.

It is also shown that the distribution of the largest value in a sample from a normal population can be approximated rather closely by a distribution in the quadratic type of dne even when the sample size is fairly small.

Various senses of asymptotic attraction and asymptotic equivalence for sequences of distribution functions are discussed and compared *re* normal and Poisson populations.

ACKNOWLEDGEMENTS

The research embodied in this dissertation was supported in part by NASA Grant NGR 44-007-028.

I should like to thank my adviser, Professor John E. Walsh, for constant guidance and encouragement. I should also like to thank Professor Paul D. Minton and Professor D. B. Owen for their help and encouragement throughout the course of my graduate work.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The theory of asymptotic types of extreme value distributions may be said to have been worked out completely, in the case of sampling from a fixed univariate population, in the papers of Frechét [1], Fisher and Tippett [2], and Gnedenko [3]. Gnedenko emphasized analogies with the theory of sums of independent random variables and made it a point to answer in one paper all of the questions concerning extreme values whose analogues in the theory of sums had been the central problems of probability theory. More recent work has been centered on non-sample or multivariate generalizations, or else on problems of statistical inference. For discussions of this work, see Walsh [4] and Gumbel [5], each of which have extensive bibliographies. Two earlier papers, those of de Finetti [6] and von Mises [7], have remained somewhat out the mainstream of work on extreme values inasmuch as they do not start with the classical functional equation of Fisher and Tippett.

Although Gnedenko's paper evidently wrapped up the theory of asymptotic types of extreme value distributions, one is nevertheless led to seek for generalizations because of the well-known fact that, for samples from a normal population, the asymptotic form for the distribution of the largest value does not furnish a decent approximation to the true distribution unless the sample size is extremely large. The discrepancy may be seen in the numerical results of Dronkers [8], who also gives more exact expressions for the corresponding frequency function.

In this paper we consider a simple asymptotic form for extreme value distributions which seems to be useable for normal populations even when the sample size is fairly small. The asymptotic form which we consider, which we call the quadratic type of the double negative exponential distribution, proves to be of some interest in itself, and we show that it represents the asymptotic form of the distribution of the largest value in a sample from a normal population in a stronger sense than does the usual form, which is the (linear) type of the double negative exponential distribution. What we show, in fact, is that the quadratic type yields a weakened form of what Walsh [9] calls "Situation I."

Much of what we shall have to say about the senses in which extreme value distributions are or are not attracted to the linear and/or quadratic types of the double negative exponential (dne) distribution could be applied more generally to give other asymptotic forms of extreme value distributions. We are giving a method for extending the Gnedenko theory of extreme values, but we are presenting this method mainly by means of the one important example which motivated our research in the first place. However, we have phrased our remarks somewhat more generally when this could be done without getting away from the central issue, and, as a second and somewhat more pathological example, we have included a short discussion of extreme value distributions from a Poisson distribution.

To conclude this introductory chapter, we describe our notation and summarize the basic facts about extreme value distributions.

Notation. Since we will be dealing with sequences of distribution functions and related functions, it will be convenient to denote by X the identity function on the extended real line. Except where otherwise noted, we also use n to denote the identity function on the positive integers.

For construction of test statistics and for related purposes, we adopt the convention that if Q is a proposition formula, then [Q] denotes unity when Q is true and zero when Q is false.

Distribution functions and probability integral transformations are imbedded in a larger class of functions which we shall call distribution transformations. By a distribution transformation (dt), we mean a monotonic non-decreasing function $F = F(X)$ from the extended real line into the extended real line which is continuous on the right at each finite point and continuous at the ideal limit points $\pm\infty$.

If F is a dt, then the unique dt F^- satisfying $F(F^-(F)) = F$ will be called the pseudoinverse of F. Of course, $F^{--} = F$, and $F^-(x)$ is the infimum of extended real numbers y such that $F(y) \geq x$ or else $y = \infty$.

<u>The exact distribution of extreme values and the Poisson approximation.</u> If m and n are positive integers with m no greater than n, let $V(X;\ m,\ n)$ be the distribution function for the m-th largest value in a sample of size n from a population which is distributed uniformly on the interval between zero and one. The same statistic, taken from a population with arbitrary distribution function F would have $V(F;\ m,\ n)$ as its distribution function. In such an experiment, the number of sample values exceeding a given value has a binomial distribution, and thus one obtains the formula

$$V(F;\ m,\ n) = \sum_{k=0}^{m-1} \binom{n}{k} F^{n-k}(1-F)^k \quad . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . (1)$$

An equally simple argument, considering the probability that the m-th largest value will fall in a fixed small interval, yields the following equation, which could also be derived analytically from (1):

$$V(F;\ m,\ n) = \int_0^F p^{n-m}(1-p)^{m-1}\, dp/B(n-m+1,\ m) = IB(F;\ n-m+1,\ m), \ . \ . \ . \ . \ (2)$$

where B and IB are the Beta function and the incomplete Beta function respectively. One may consider $V(F;\ m,\ n)$ to be defined by (2) even when m and n are not integers.

Equation (1) may be rewritten as follows:

$$V(F;\ m,\ n) = F^n \sum_{k=0}^{m-1} \frac{(-\log F^n)^k}{k!} \cdot \frac{n^{[k]}(1-F)^k}{n^k(-\log F)^k F^k} \quad \ldots\ldots\ldots \quad (3)$$

As $n \to \infty$ with m fixed, $V^-(p;\ m,\ n) \to 1$, for each fixed p in the open interval (0, 1), and hence the difference between $V(F;\ m,\ n)$ and $W(F^n;\ m)$ approaches zero uniformly, where

$$W(p;\ m) = p \sum_{k=0}^{m-1} (-\log p)^k/k! \quad \ldots\ldots\ldots\ldots\ldots \quad (4)$$

Note that if N is a Poisson random variable and if p is the probability that N = 0, then $W(p;\ m)$ is the probability that N is less than m. Note also that $F^n = V(F;\ 1,\ n)$ is the distribution function for the maximum in the sample. The uniform approximation of $V(F;\ m,\ n)$ by $W(F^n;\ m)$, or alternatively by $W(\exp(-n/(1-F));\ m)$, which is more common in the literature, will be referred to as the Poisson approximation. It allows us to concentrate our attention on the distribution of the largest value, assuming the large sample size, and we need not consider $V(F;\ m,\ n)$ especially for fixed values of m other than 1.

It may also be pointed out that the m-th smallest order statistic in a sample of size n from F has distribution function equal to $1-V(1-F(X-0);\ m,\ n)$, and hence our results on the largest order statistics apply immediately to the smallest order statistics.

Finally, we note that $W(F;\ m)$ is given by the incomplete Gamma function $I\Gamma$. The well-know equation gives

$$W(F;\ m) = 1 - I\Gamma(-\log F;\ m) = \int_{-\log F}^{\infty} x^{m-1}\, e^{-x}\, dx/\Gamma(m) \quad \ldots\ldots \quad (5)$$

This equation may be taken as a definition when m is not an integer.

For large n, it is often convenient to replace $-\log(F^n)$ by $n(1-F)$. The justification is that if n approaches infinity while x varies with n so $F^n(x)$ remains more or less constant, we then have $F(x)$ approaching unity and so $1-F(x)$ is asymptotic to $-\log F(x)$. This is the same argument which would be used in demonstrating the approximation of $V(F; m, n)$ by $W(F^n; m)$.

We consider the distribution of the largest value by transforming to the double negative exponential distribution function $dne = dne(X) = \exp(-\exp(-X))$. More generally, we may transform to any distribution of this same linear type. (Two distribution functions F and G are of the same linear type if there is a linear dt $A = aX + b$, with $a > 0$, such that $F(A) = G$.) All distribution functions of the type of dne are of the form

$$dne(X;\ a,\ b) = (1/a)^{(1/b)^X} = dne(X \log b - \log \log a)\ ,$$

where a and b are constants greater than unity. The distributions functions all have the property that exponentiation reduces to translation. Thus $V(dne(X;\ a,\ b);\ 1,\ n) = (dne(X;\ a,\ b))^n = dne(X-\log_b n;\ a,\ b)$. Among the distributions of the dne type, perhaps more interesting for practical purposes is $dne(X;\ 2,\ 10)$, which has the property that the median largest value in a sample of size n is $\log_{10} n$. We consider the dt $T = T_F = dne^-(F;\ 2,\ 10) =$ $= -\log_{10}(-\log_2 F)$, which, if F is continuous and strictly increasing, is the unique dt which, when applied to a random variable having distribution function F, gives a random variable having distribution function $dne(X;\ 2,\ 10)$.

A graph of T contains much information. When the ordinate is $\log_{10} n$, the abscissa is the median largest value. When the ordinate is the common logarithm ($\log_{10}$) of n times $(\log_2)/\log(1/p)$, the abscissa is the p-quantile of the largest value. If q_p is the p-quantile of F^n, then $T(q_p) - T(q_{1/2}) =$ $= \log_{10} \log_{10} p - \log_{10} \log_{10} (1/2)$ is a constant not depending on n. Thus a vertical shift of the graph of T amounts to replacing the median by a different quantile.

CHAPTER II

ASYMPTOTIC FORMS FOR EXTREME VALUE DISTRIBUTIONS

A dt of the form aX + b will be called a linear dt only if it is strictly increasing, i.e. only if the constant a is strictly positive. A dt of the form

$$a^2 (X - b)^2 \text{ sgn } (X - b) - c^2$$

will be called a quadratic dt provided that the constant a is strictly positive. We will be interested usually in the form of a dt only where its values are fairly large. Hence it is mainly for convenience that we define a quadratic dt in this fashion, i.e. symmetrically about the point $(b, - c^2)$.

The linear type of a distribution function F is the set of distributions F(A), where A ranges over the linear dt's. The set of distributions F(A), where A ranges over the quadratic dt's, will be called the quadratic type of F.

The linear types form a partition of the set of all distribution functions into equivalence classes. We point out that this is not true of the quadratic types. Nevertheless, the notion of quadratic type may prove useful in studying the useful asymptotic forms of extreme value distributions, and this for two reasons. In the first place, a quadratic type appears naturally, as we shall see, as the asymptotic form for the distribution of the largest value in a sample from a normal population.

In the second place, quadratic types are among the most simply structured classes of distributions, next to linear types, and hence they form a good testing ground for a generalized theory of asymptotic form.

Generally speaking, the goal of our research into generalized asymptotic forms of extreme value distributions is to find fairly simple parametric families K of distribution functions such that, for distributions F belonging to a fairly large non-parametric family, the sequence F^n tends to be attracted asymptotically in some sense to the family K. The sense of this attraction should be that when sample size is large, one may assume without significant error that the distribution of the maximum belongs to K. We shall assume that attraction of the sequence F^n to K may be defined to be equivalent to the existence of a sequence G_n from K, possibly satisfying certain stability conditions related to the behavior of a sequence of extreme value distributions in general, which is asymptotically equivalent to the sequence F^n in some meaningful sense.

Among the many possible families K, other than the linear types, we choose to work with one of the simplest, which is the quadratic type of dne. Also we try to consider asymptotic equivalence in as simple a sense as possible.

If F_n and G_n are sequences of distribution functions, we shall say that F_n and G_n are asymptotically equivalent, written $F_n \doteq G_n$, in Sense 0 if $F_n - G_n$ converges uniformly to zero as n tends to infinity. If K is a class of distribution functions and if F_n is a sequence of distribution functions, we shall say that F_n is attracted to K in Sense 0 if there is a sequence G_n from K such that $F_n \doteq G_n$ in Sense 0. We shall say that a sequence G_n of distribution functions satisfies Condition 0 if there is a distribution function F such that $G_n \doteq F^n$ in Sense 0.

Theorem 1. Let G_n be a sequence of distribution functions. If G_n is attracted to the linear type of dne in Sense 0, then G_n is attracted to the quadratic type of dne in Sense 0. Conversely, if G_n is attracted to the quadratic type of dne in Sense 0, and if G_n satisfies Condition 0, then G_n is attracted to the linear type of dne.

Proofs of this and other theorems in this chapter are collected in an appendix.

The following lemma is worth noting since it relates our discussion to the treatment of asymptotic linear types of extreme value distributions given by Gnedenko.

Lemma 1. Let F be a distribution function. Then F^n is attracted to the linear type of dne if and only if there is a sequence A_n of linear dt's such that $F^n(A_n)$ converges to dne pointwise, i.e. in the weak-star topology for distributions (defined by the convergence of the expected values of each fixed bounded continuous statistic.)

Gnedenko gave necessary and sufficient conditions for F^n to be attracted to the linear type of dne in Sense 0, or rather in the equivalent sense given in the lemma. According to Theorem 1, exactly the same conditions are necessary and sufficient for F^n to be attracted to the quadratic type of dne in Sense 0. Nevertheless, it may be true that, for moderate values of n, the approximation of F^n by a distribution in the quadratic type of dne may be much better than any approximation from the linear type.

Normal distributions afford an appropriate example. Letting Φ = dne(T) be the standardized normal distribution function, we have the familiar asymptotic formula

$$e^{-T(x)} \sim 1 - \Phi(x) \sim \Phi'(x)/x$$

as x approaches infinity. It follows that the difference between T(x) and $\frac{1}{2}(x^2 + \log\ (2\ x))$ approaches zero as x approaches infinity, and thence it follows that

$$\Phi^n \doteq \mathrm{dne}(\frac{1}{2}\, X^2\ \mathrm{sgn}(X)\ -\ (\log n - 2\sqrt{\log n}))$$

in Sense 0. To understand this formula, it is convenient to study the conditions under which $\mathrm{dne}(A_n) \doteq \mathrm{dne}(B_n)$ in Sense 0, where A_n and B_n are sequences of dt's which fix the ideal points $\pm\infty$. When $\mathrm{dne}(A_n)$ and $\mathrm{dne}(B_n)$ are asymptotically equivalent in Sense 0, we shall say that $A_n \doteq B_n$ in Sense 0'.

<u>Lemma 2</u>. Let A_n and B_n be sequences of dt's. For $A_n \doteq B_n$ in Sense 0' it is necessary and sufficient that the following condition hold for all real z, for all positive h, and for all sufficiently large integers m: For any real x, if $A_m(x) \leq z$, then $B_m(x) < z + h$, and if $A_m(x) \geq z$, then $B_m(x) > z - h$.

<u>Corollary</u>. If A_n and B_n are sequences of dt's and if each term of A_n is continuous and strictly increasing, then $A_n \doteq B_n$ in Sense 0' if and only if $B_n(A_n^-) \to X$ pointwise.

For the most part, it will be the Corollary which will be used in this paper, since linear and quadratic dt's are continuous and strictly increasing.

It is worth remarking that the dt $T = \mathrm{dne}^-(F)$ is a very convenient way of specifying the distribution F for the population from which extreme values are taken, charting as it does the logarithm of sample size versus a certain quantile of the largest value.

In Figure 1, we give a graph of the common logarithm of sample size versus the median largest value, assuming that the underlying distribution function is

$$\Phi = \mathrm{dne}(T) = \left(\frac{1}{2}\right)^{(1/10)^{S}} .$$

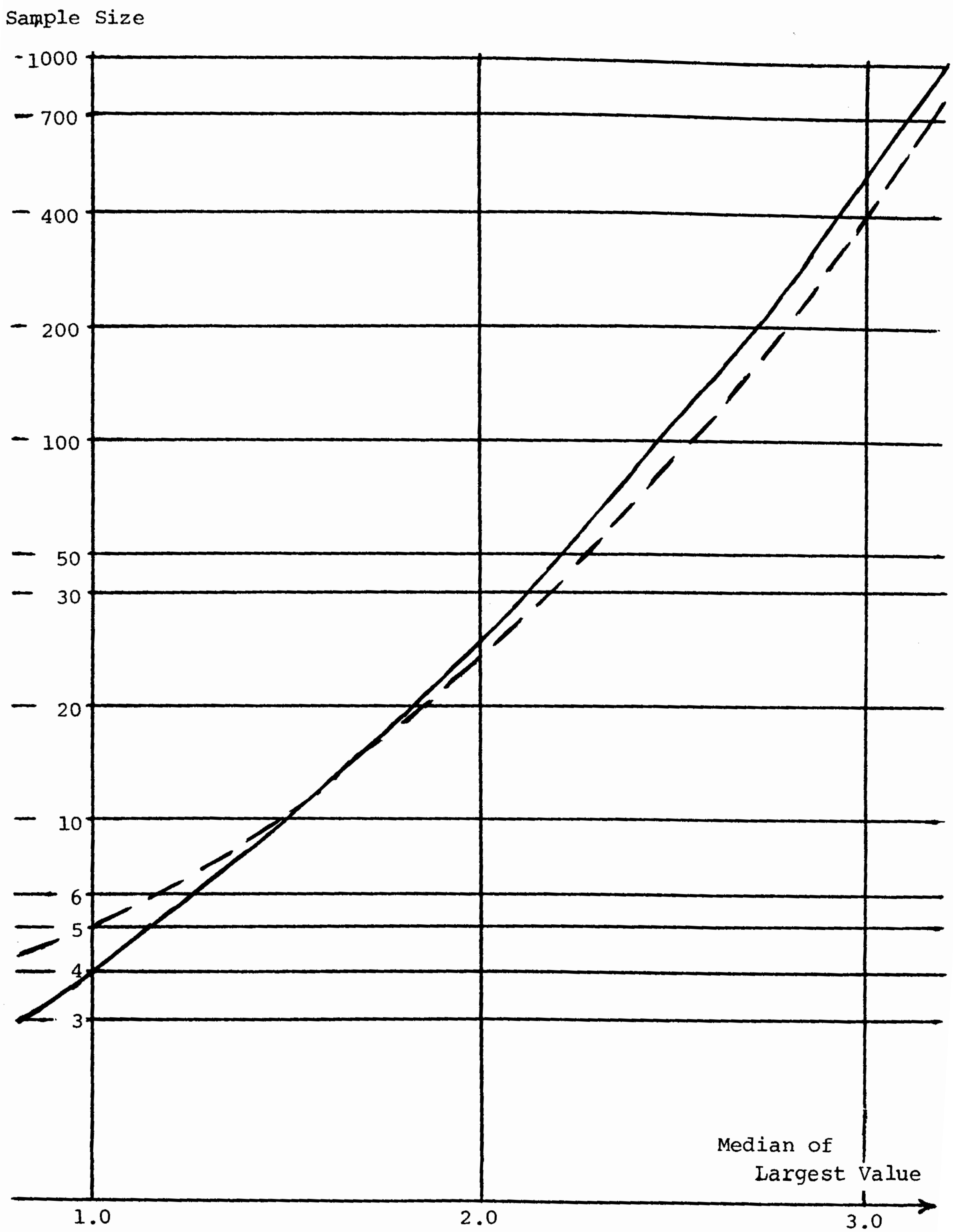
Notice that S is T $\log_{10}$ e plus a constant. Hence in Figure 1 we draw an approximation to S of the form $(\frac{1}{2} \log_{10} e)\, X^2 \,\mathrm{sgn}(X) - c$. The approximation is dotted, and it is clearly a better approximation to the solid graph S than could be obtained by a straight line, even for these relatively small sample sizes. The graph of S is based on a table given by de Finetti [6].

It is also clear that we could write a different quadratic approximation to S which would be virtually indistinguishable from S in the graph, which contains most of the important information about extreme values for moderate sample sizes. Hence there would be some real utility in studying the quadratic type of dne as a parametric family of distributions for which statistical procedures ought to be developed.

We shall do nothing at this time with the problems of estimating or testing the parameters of the quadratic type of dne, on the basis of the various kinds of data which may be available in work with extreme order statistics. We mention only that the quadratic dt is essentially linear in its three parameters in the region of interest and that standard procedures for fitting a quadratic polynomial to a set of points may be used to give estimation procedures, for instance, whose properties would be well worth studying. We note that many of the obvious difficulties with such procedures have already been encountered in work

FIGURE 1

GRAPH OF dne^{-}(Φ; 2, 10)



with the linear type of dne. See Gumbel, for instance, for references to this work, and, in particular, see Gumbel's discussion of "plotting position," but note that Gumbel is using the usual least-squares technique to fit a straight line to T^-, which is rather different from our problem of fitting a quadratic to T by least squares or other curve-fitting methods.

We return to more theoretical questions by introducing the subject of stability conditions, by which we mean conditions which it might be proper to impose on a sequence G_n which is to be used to approximate the sequence F^n of extreme value distributions for some distribution function F. We note that Condition 0 is a rather minimal stability condition for G_n. Of considerable historical importance is the stability condition for G_n which was imposed by Fisher and Tippett namely that each term of the sequence be of the same linear type. This condition, together with Condition 0, or the apparently but not actually weaker condition that there be a distribution function F, not of the unitary type, and a sequence A_n of linear dt's such that $G_n(A_n) = G_1$, and $F_n(A_n) \to G_1$ in weak-star, implies that the linear type of the distribution G_n is the linear type of (exactly) one of the following: (1) dne; (2) $[X > 0]\ \mathrm{dne}(\log_B X)$ for some $B > 1$; (3) $[X < 0]\ \mathrm{dne}(-\log_B(-X)) + [X \geq 0]$ for some $B > 1$; and (4) $[X \geq 0]$, representing the unitary type. These types were given by Fisher and Tippett. A complete proof of their uniqueness was given by Gnedenko.

The need for stability conditions was made apparent by Walsh [4] when he made the distinction between "Situation I" and "Situation II." Basically, Walsh pointed out that an approximation of Φ^m by a distribution dne(A) in the linear type of dne cannot be used to approximate Φ^k when k and m are very different, however large they may be. On the other hand, a distribution such as the exponential $F = (1 - e^{-X})[X > 0]$ has the property that $F^n \doteq (\text{dne})^n = \text{dne}(X - \log n)$, and hence the sequence of approximations from the linear type of dne possesses, or rather can possess, a strong stability allowing one to pass from one sample size to another under the sole condition that each sample size be sufficiently large. The importance of being able to make such a passage is one of the reasons for our study of the quadratic type of dne as a source of approximations to the extreme value distributions from normal populations.

If K is a set of distributions and if F is a distribution function, we shall say that the sequence F_n is attracted to K in Sense 1 if there is a distribution function G such that G^n is a sequence in K, a trivial condition when K is the linear or quadratic type of some distribution, and such that $F^n \doteq G^n$ in Sense 0, which means simply that $1 - F(x) \sim 1 - G(x)$ as $F(x)$ approaches unity.

<u>Theorem 2</u>. Let $F = \text{dne}(T)$ be a distribution function. For F^n to be attracted to the linear (quadratic) type of dne in Sense 1, it is necessary and sufficient that there be a linear (quadratic) dt A such that $T(x) - A(x)$ approaches zero as $F(x)$ approaches unity. Hence F^n cannot be attracted to both the linear and the quadratic types of dne in Sense 1. If F^n is attracted to the linear or quadratic type of dne in Sense 1, then so is $F^n(A)$ where A is an arbitrary linear dt.

<u>Corollary</u>. Let F be a distribution function. If F is of the linear type of $(1 - e^{-X})$ $[X > 0]$, then F^n is attracted to the linear type of dne in Sense 1. If F is of the linear type of Φ, then F^n is not attracted to either the linear or the quadratic type of dne in Sense 1. If F is of the linear type of Φ, then F^n is attracted in Sense 1 to the set of distributions which can be written in the form

dne(T) where

$$T = a^2(X - b)^2 \operatorname{sgn}(X - b) - c^2 + d^2 \log X$$

For practical purposes, the disadvantages of introducing the logarithm term and the fourth parameter d^2 would seem to outweigh the slight advantages, which are not of the same order as the advantages gained by introducing a quadratic term. Hence, we introduce a weaker but still significant stability condition and a corresponding sense of attraction. We shall say that the sequence F^n is attracted to a set K of distributions in Sense 2, assuming that K contains with each member G the entire sequence G^n, if there is a distribution G in K and a sequence c(n) of constants such that $F^n \doteq G^{c(n)}$ in Sense 0.

Theorem 3. Let F be a distribution function. Then F^n cannot be attracted in Sense 2 to both the linear and the quadratic types of dne. If F is normal, then F^n is attracted in Sense 2 to the quadratic type of dne and hence not to the linear type of dne.

To emphasize the distinction between attraction to the linear and the quadratic types of dne, we consider the law of large numbers, which was a point of departure in the pioneering work of von Mises [7] and de Finetti. It was studied by Gnedenko also.

A sequence Y_n of random variables is said to satisfy the law of large numbers if there is a sequence c_n of constants such that $Y_n - c_n$ converges to zero in probability.

Theorem 4. Let X_n be a sequence of independent random variables each having the distribution function F. Let Y_n be the maximum of $X_1, \ldots, X_n$. If F^n is attracted to the quadratic type of dne in Sense 2, then the sequence Y_n satisfies the law of large numbers, but if F^n is attracted to the linear type of dne in Sense 2, then Y_n does not satisfy the law of large numbers. Generally, if $F = \text{dne}(T)$ then Y_n satisfies the law of large numbers if and only if the difference between $T(x + h)$ and $T(x)$ diverges, as x increases without bound, for every fixed h.

For purposes of comparison with Theorem 4, we state a theorem of Gnedenko's in the notation of our Theorem 4: For there to exist a sequence c(n) of constants such that $Y_n/c(n) - 1$ converges to zero in probability, in which case Gnedenko says that Y_n is relatively stable, it is sufficient that one or both of the following conditions hold: Either there is a finite point x with $F(x) = 1$, or F^n is attracted to the linear type of dne in Sense 0. On the other hand, Gnedenko's Theorem 2 and Gnedenko's Theorem 4 may be compared to show that, if F^n is attracted to the linear type of $[X > 0]\text{dne}(\log_B X)$ for some $B > 1$, then Y_n is not relatively stable.

The bulk of the present paper has thus far been aimed at demonstrating properties, both practical and theoretical, possessed by the quadratic type of dne which seem to justify it as an object for further study, particularly with regard to the very complex problems of inference. In concluding this argument, we wish to make some remarks about the general idea of using as the asymptotic form of extreme value distributions a distribution containing parameters other than location and scale parameters, i.e. using approximations from a parametric family other than a linear type.

In using parametric families other than linear types, we are making a small break with the elegant tradition of such works as that of Gnedenko. Our main reason for doing so is the well-known fact that, for a normal population, enormous sample sizes are required before the distribution of the largest value can be satisfactorily approximated by a distribution of the linear type of dne. The graph in Figure 1 illustrates this fact clearly enough, but for a more exact description of the discrepancy, one can refer to Dronkers [8], who also gives a variety of asymptotic expressions for the densities of the extreme value distributions from normal and other populations with smooth density functions. It is hoped that this present paper may complement Dronker's paper by giving asymptotic forms of the distribution function rather than the density function, thus avoiding most of Dronker's regularity conditions, and by concentrating on theoretical limit theorems which are similar to those of Gnedenko et al. for linear types.

Thus the theory of linear types gives poor approximations in the case of the normal distribution. More generally, the theory is somewhat restrictive because of the paucity of linear types which attract any sequence G^n of extreme value distributions.

Von Mises was apparently the first to give an example of a distribution function F = dne(T) for which F^n is not attracted to any of the linear types, even in Sense 0. His example can be specified by letting $T = X + \frac{1}{2} \sin X$. One verifies by a glance at the graph of T, with its regular oscillation about the equiangular line, that T - log n is asymptotically equivalent in Sense 0' to no sequence of linear dt's or, still less, to a sequence of dt's of one of the logarithmic forms $[X > 0] \log_B X$ or $[X < 0] (-\log_B (-X))$.

If von Mises' example seems far-fetched, consider the case of a Poisson distribution

$$F = e^{-1}\sum_{k=0}^{\infty} (1/k!)\ [k \leq X] = \text{dne}(T),$$

which fails to be attracted in Sense 0 to the linear (or quadratic) type of dne for exactly the same reason that von Mises' example fails to be so attracted, namely because of a "wave" in T which is in this case solely due to the fact that T increases only at regularly spaced points, namely at the integers.

Calculations given in the appendix show that, as x approaches infinity through the integers only, $T(x) = S(x) + o(1)$, where

$$S = 1 + \log \sqrt{2\pi X} + X(\log X - 1).$$

Theorem 5. Let F be the Poisson distribution with mean 1 as above, and let S be the dt defined by the above equation. Then F^n is not attracted even in Sense 0 to either the linear or the quadratic type of dne, but the difference between F^n and $(\text{dne}(S))^n$ approaches zero uniformly on the positive integers. Moreover, $(\text{dne}(S))^n$ is attracted in Sense 0 to the linear (or quadratic) type of dne, but it is attracted to neither of these in Sense 1 or even in Sense 2. However, $\text{dne}(S - \log n) = (\text{dne}(S))^n$ is attracted in Sense 2 to the family of all positive powers of the distribution function (dne(X)) exp X. Even if the Poisson distribution F is made continuous by convolving it with the uniform distribution on the interval (0, 1), the sequence of extreme value distributions (powers) of the resulting distribution function is still not attracted in Sense 0 to the linear type of dne.

One further point needs to be made in favor of the study of families other than linear types. Gnedenko and others, in their theory of linear types, have underscored analogies between the asymptotic theory of extreme values and the asymptotic theory of sums of random variables. Recently, however, Walsh [9] pointed out an essential difference: In the theory of extreme values, every distribution is, in effect, infinitely divisible. In fact, if m and k are positive integers with $k \leq m$ and if G is a distribution function then there exists a distribution function F such that G is the distribution of the k-th largest value in a sample of size m from F. This extra freedom is reflected in the variety of limit theorems which can be obtained.

We consider now weakened forms of the stability conditions underlying attraction in Sense 1 and attraction in Sense 2. We recall that in the definitions of these senses of attraction we referred to the existence of an approximating sequence G_n in which the entire sequence was already determined by the first term, in the case of attraction in Sense 1, and determined except for one simple parameter, representing the appropriate vertical shift of $\text{dne}^-(G_n)$, in the case of attraction in Sense 2. In practice, however, the sequence of approximations would usually be based on data whose information about the parameters increased with sample size. One might wish to impose upon the actual sequence of approximating distribution functions a stability condition which asymptotically represents the conditions underlying attraction in one of the two stronger senses which have been considered. The conditions that we want should guarantee that one can pass from one sample size to another provided that both are large but without the restriction, which would be needed in the general situation, that their ratio be close to unity.

Let $G_n = \text{dne}(A_n)$ be a sequence of distribution functions. We are interested in two cases, the terms of A_n being linear in the first case and quadratic in the second case. However, we shall at present assume only that the terms of A_n are continuous and strictly increasing and that they fix the ideal points $\pm\infty$.

We shall say that the sequence G_n satisfies Condition 1 if $G_k^{m/k} - G_m$ converges uniformly to zero as m and k tend to infinity independently except that k must remain larger than m. Thus in view of Lemma 1, in which generalized sequences could be used instead of sequences, Condition 1 says that $A_k(A_m^-) - \log k + \log m \to X$ pointwise as $m, k \to \infty$ with $k > m$. In particular, we may let $x_n = A_n^-(0)$ and observe that, for any $h > 0$, if m is sufficiently large, then $U(x_m) < \log m + h$ and also $L(x_m) > \log m - h$, where, for any extended real number x, U(x) and L(x) are the limits superior and inferior, respectively, of $A_k(x) + \log k$ as $k \to \infty$. It is clear that L is a non-decreasing function and hence that there is a distribution function F, which could be taken to agree with dne(L) at every point of continuity, such that $G_n \doteq F^n$ in Sense 0. Thus G_n satisfies Condition 0.

Now let us assume that each term of A_n is a linear dt. We continue to assume that G_n satisfies Condition 1. We want to show that there is a linear dt A such that $A_n \doteq A - \log n$ in Sense 0 and hence G_n is attracted in Sense 1 to the linear type of dne. Let $B_n = A_n + \log n =$ $= b_n X + c_n$. One can find x and y such that $-\infty < L(x) \leq U(x) < L(y) \leq U(y) < \infty$ and this implies that the sequences b_n and c_n eventually lie in certain finite intervals and hence possess finite limit points $b \neq 0$ and c respectively, which are easily seen to be unique. Thus $b_n \to b$ and $c_n \to c$, whence $B_n \to A = bX + c$ and $A_n \doteq A - \log n$ in Sense 0'.

Now we consider the case that the terms of A_n are quadratic dt's.

Let $B_n = A_n - \log n = a_n^2(X - b_n)^2 \operatorname{sgn}(X - b_n) - c_n^2$. Now there are two cases to consider. First suppose that the sequence b_n is bounded. As in the linear case, this will force the sequences a_n^2 and c_n^2 to be bounded also, and, as in the linear case, the sequences b_n, a_n^2, c_n^2 will converge to finite numbers b, $a^2 \neq 0$, c^2 respectively. Then G_n will be attracted to the quadratic type of dne in Sense 1.

On the other hand, let us assume that the sequence b_n is not bounded. We shall show that G_n is attracted in Sense 1 to the linear type of dne. First suppose that there is a subsequence N of the positive integers such that b_N diverges monotonically to $-\infty$. We shall show that for each real number u, there is a linear dt C_u such that, for $0 < x < u$, $L(x) \leq C_u(x) \leq U(x)$. Thus for a properly chosen sequence u(n) of real numbers, $C_{u(n)} - \log n \doteq A_n$ in Sense 0 and $C_{u(n)}$ satisfies Condition 1, whence G_n is attracted in Sense 1 to the linear type of dne by what was just shown. Let C_u be the linear dt which agrees with L at 0 and at u. For large m, the second derivative of B_m is positive for arguments between 0 and u. Hence $C_u(x) > B_m(x) - o(m)$ when $0 < x < u$ and $m \to \infty$. This proves that $C_u(x) \geq L(x)$ for $0 < x < u$. On the other hand, suppose that there is a value x such that $0 < x < u$ and such that $C_u(x) \geq U(x)$. This puts a lower bound on the amount by which B_m "bends" in the interval (0, u), for large m, and thus puts a lower bound on b_m, which is contrary to hypothesis.

Secondly, suppose that there is a subsequence N of the positive integers such that b_N diverges monotonically to $+\infty$. This case is merely a reflection of the first case and may be treated similarly. This concludes our discussion of Condition 1 and its relation to attraction in Sense 1.

One can define a weaker stability condition which is related to attraction in Sense 2 just as Condition 1 is related to attraction in Sense 1. We say that the sequence G_n satisfies Condition 2 if it satisfies Condition 1 and if there is a sequence $c(n)$ of constants such that $G_k^{c(k)/c(m)} \pm G_m$ in Sense 0 (i.e. the difference converges uniformly to zero) as $m, k \to \infty$ with $m < k$. The analysis of Condition 2 parallels that of Condition 1, and we may summarize these results in a final theorem.

Theorem 6. If G_n is a sequence of distributions from the linear type of dne which satisfies Condition 1 (respectively Condition 2), then G_n is attracted to the linear type of dne in Sense 1 (respectively Sense 2). If G_n is a sequence of distributions from the quadratic type of dne which satisfies Condition 1 (respectively Condition 2), then G_n is attracted to the quadratic type of dne in Sense 1 (respectively Sense 2) if the sequence b_n is bounded, b_n being the point of inflection of $\text{dne}^-(G_n)$, and G_n is attracted to the linear type of dne in Sense 1 (respectively Sense 2) otherwise.

APPENDIX

PROOFS OF THEOREMS

Proof of Lemma 1. To prove Lemma 1, let F be a distribution function for which F^n is attracted to the linear type of dne. We shall show that $F_n(A_n(x)) \to \text{dne}(x)$, where A_n is a sequence of linear dt's for which $F^n \doteq \text{dne}(A_n^-)$ in Sense 0. Let $h > 0$ and let x be a real number. Let m be sufficiently large that $|F^n - \text{dne}(A_n^-)| < h$. Then, in particular, $|F^n(A_n(x)) - \text{dne}\ (A_n^-(A_n(x)))| < h$. Since $A_n^-(A_n(x)) = x$, we have completed the proof.

Conversely, suppose that A_n is a sequence of linear dt's such that $G_n = F^n(A_n)$ converges to dne pointwise. Let $h > 0$. Find a finite increasing sequence $x_0, \ldots, x_{m+1}$, with $x_{m+1} = -x_0 = \infty$, such that $\text{dne}(x_{i+1}) - \text{dne}(x_i) < h$ for $i = 0, \ldots, m$. Let x be any real number. Take i so that $x_i \leq x \leq x_{i+1}$. Then $\text{dne}(x) \leq \text{dne}(x_{i+1}) < \text{dne}(x_i) + h <$ $< G_k(x_i) + 2h \leq G_k(x) + 2h$. Similarly $\text{dne}(x) < G_k(x) - 2h$. Thus $|G_k - \text{dne}| < 2h$, and so $|F^k - \text{dne}(A_k^-)| = |G_k(A_k^-) - \text{dne}(A_k^-)| < 2h$. Thus $\text{dne}(A_n^-) \doteq F^n$ in Sense 0.

Proof of Lemma 2. First suppose that the condition holds for each real number z, for each $h > 0$, and for m sufficiently large. Let $k > 0$. Choose a finite, strictly increasing sequence $z_0, \ldots, z_{M+1}$ such that $z_{M+1} = -z_0 = \infty$ and such that $\text{dne}(Z_{i+1}) - \text{dne}(z_i) < k$ for $i = 0, \ldots, M$. Take m sufficiently large that the condition holds with $z = z_1, \ldots, z_M$ and $h = \frac{1}{2}\min\{z_i - z_{i-1}, z_{i+1} - z_i\}$, where $i = 1, \ldots, M$. For any real x, one can find i among $0, \ldots, M$

to Gnedenko's lemma stating that the law of large numbers holds for Y_n if and only if $(1 - F(x + h))/(1 - F(x)) \to 0$ as $x \to \infty$, for every fixed real number h.

Proof of Theorem 5. If $F = \text{dne}(T)$ is the Poisson distribution with mean 1, then

$$1 - F(n - 1) = e^{-1} \sum_{k=n}^{\infty} (1/k!)$$

$$= \frac{1}{en!} \left(1 + \sum_{k=1}^{\infty} (1/(n + k)^{[k]})\right)$$

$$= \frac{1}{en!} (1 + o(1)).$$

Thus $T(n - 1) = 1 + \log(n!) + o(1) = T(n) + o(1)$. By Sterling's formula, $T(n) = 1 + \log \sqrt{2\pi} + (n + \frac{1}{2}) \log n - n + o(1) = S(n) + o(1)$. Since $S(n) - S(n - 0) \to \infty$, there can be no attraction of F^n to the linear type of dne, even in Sense 0.

Looking now at $S = X \log X + A(X)$, we see that $A(X)$ rises much more slowly than $X \log X$ when S is large and hence that

$$S - \log n \doteq X \log X - \log n + \log A(x_n)$$

in Sense 0', where $x_n \log x_n = \log n$. Thus $(\text{dne}(S))^n \doteq (\text{dne}(X \log X))^{c(n)}$ in Sense 0, where $c(n) = n/A(x_n)$.

Note on Theorem 6. The proof of this theorem precedes its statement in the text, beginning on page 19.

LIST OF REFERENCES

1. Frechet, M., Sur la loi de probabilite de l'ecart maximum, Annales de la Societe Polonaise de Mathematiques, VI, 93, Cracovie, (1927).

2. Fisher, R. A. and Tippett, L. H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proceedings of the Cambridge Philosophical Society, 24, Pt. 2, 180-190 (April, 1928).

3. Gnedenko, B., Sur la distribution limite du terme maximum d'une serie aleatoire, Annals of Mathematics, 44, 423-453 (1943).

4. Walsh, John E., Handbook of Nonparametric Statistics, II, New York, Van Nostrand, (1965).

5. Gumbel, E. J., Statistical Theory of Extreme Values and Some Practical Applications, Washington, D. C., National Bureau of Standards, Applied Mathematics Series, 33 (1954).

6. de Finetti, B., Sulla legge di probabilita degli estremi, Metron, 9, 127-138 (1932).

7. von Mises, R., La distribution de la plus grande de n valeurs, Revue Mathematique de l'Union Interbalkanique, 1, 1-20 (1936).

8. Dronkers, J. J., Approximate formulae for the statistical distributions of extreme values, Biometrika, 45, 447-470 (1958).

9. Walsh, John E., Sample-like distributions of an order statistic under general nonsample conditions and some asymptotic implications, Southern Methodist University, Statistics Department (July, 1968).