General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

PROGRESS REPORT FOR NASA GRANT NGR 44-007-028, TO DECEMBER 31, 1968

by John E. Walsh, Principal Investigator

SUMMARY

Work on this grant, during 1968, produced 12 documents. In addition, a graduate student, Charles Anderson, has completed most of a doctoral thesis. Of the 12 documents, four have been accepted for publication. Also, three have either been presented, or will be presented, at national or international meetings. The other document have been submitted for publication in various journals of a statistical nature. The 12 documents are listed in the References, along with statements about where they are submitted, have been accepted, etc.

Three technical meetings were attended with some support from this grant. They were the Annual Meetings of: The Institute of Mathematical Statistics, The Society of Actuaries, and The Casualty Actuarial Society of America. One trip was made to NASA Headquarters.

Many useful results were obtained for order statistics of sets of univariate observations. Both sample and nonsample cases were considered. Some results were for large sets (asymptotic case) but many are also applicable for small sets. An extension of asymptotic distributions for extremes of samples was considered (by Charles Anderson). Also, conditions were developed for approximate independence of some order statistics. The generality of the possible distributions for order statistics, and relations between sample and nonsample cases, were investigated. Also, some results were obtained for a generalized form of median regression that uses order statistics.

A technical outline of the work, and its extension of previous

results, is given next. This starts with a discussion of the usefulness of extreme-value theory. Finally, a few conclusions related to this research are stated.

TECHNICAL DISCUSSION

Usefulness of Extreme-Value Theory

Consider a set of univariate observations (perhaps with a large set size). An extreme order statistic is an identified one of the largest, or the smallest, of these observations. For example, the largest and next to largest, the smallest and next to smallest, etc., of the observations are extremes of this nature. Here, the set of observations is often considered to be a random sample, but this is not necessarily the case.

To provide an intuitive understanding of the practical utility of extremes, consider their use with regard to breaking strength of materials (as an example). One theory hypothesizes that a piece of material contains a number, usually very large, of flaws and that a breaking strength (for the piece) is associated with each flaw. Thus, the breaking strength of the piece is the smallest of the breaking strengths for the flaws and can be investigated on the basis of smallest extremes. Another theory asserts that each flaw has a size and that the larger the size of the flaw the more likely that the piece will break because of this flaw. Combined with some additional theory, this leads to an analysis based on largest extremes. As another example, largest gusts on the wing of a flying aircraft can have a strong influence on structural failure of the wing. Here, analysis is based on the largest values of gusts of wind.

As still another example, the failure of the rotor blade system for a helicopter can be analyzed as a problem in fatigue failure that involves extreme values. Breaking of materials and fatigue failures for space vehicles, exposure of space vehicles to large meteors, etc., are other examples of situations where analysis by use of extreme-value theory can be helpful. In addition, extreme-value theory has been found useful for investigation of extreme temperatures, barometric pressures, vapor pressures, floods, etc.

Asymptotic Distributions (Sample Cases)

Suppose that the set size n is very large and that the observations are a random sample. Various types of asymptotic distributions ($n \to \infty$) can often be developed for an extreme. The best known but most crude types are the three "asymptotes" (for example, see Gumbel, 1958). The first two types of asymptotes have infinite tails, with the tail for the second type being very heavy. The third type does not have an infinite tail (is truncated in the direction of the extreme). The first type involves two parameters, the second type two or three parameters (one of the three parameters may be known), and the third type three parameters. Transformations exist such that the transformation of an extreme that originally had an asymptotic distribution of the second or third asymptote type has a first asymptote type. Thus, in some respects, it is sufficient to consider the first asymptote.

In general, two of the parameters can depend on n (although this is undesirable). In (Walsh, 1965), the desirable case, where each parameter is constant for sufficiently large n, or a constant plus a completely specified function of n, is called Situation (I). The other case is called Situation (II). For use of an asymptote, the normal and lognormal distributions belong to Situation (II). Much more of the data is usable for estimation and investigation when Situation (I) holds for the asymptote used.

Extension of Situation (I) to a much wider class of distributions sampled, including the normal and lognormal, seemed worthwhile. Also, it would appear desirable to have one functional form that is usable for both the first and second asymptotes. This has been accomplished by use of a form of asymptotic distribution that has four or five parameters (one of the five parameters may have a known value). These extended asymptotic distributions have great curve-fitting flexibility. For the largest extreme the cumulative distribution function (cdf) is of the form

$$(1) \qquad \exp\left\{-e^{-ax^2 + bx + c + d\log_e(x-h) + \log_e n}\right\} \quad (x \geq x_L),$$

where $-ax^2 + bx + c + d\log_e(x-h)$ is a monotonically decreasing function of x for $x \geq x_L$ and x_L has a determined value. This extended distribution is directly applicable for anything between light and heavy tails.

The cdf (1) can be applied with still heavier tails by considering that it is the distribution of $\log_e(X[n;n] - x_L)$, where $X[1;n] \leq \cdots \leq X[n;n]$ are the order statistics of the n observations. It can be applied to truncated situations similar to the third asymptote of X[n;n] by considering that (1) is the cdf of $-\log_e(U - X[n,n])$, where the parameter U is the truncation point. Then, there are five or six parameters. Also, the cdf (1) can be applied with much lighter tails by considering that it is the distribution of $\exp(X[n;n] - x_L)$. Of course, more abrupt truncation can be accomplished by considering that $\log_e\{-\log_e(\delta - X[n;n])\}$ has a cdf of the form (1).

It is anticipated that, for most cases of practical interest, the extended asymptotic distributions involving five parameters will belong to Situation (I). That is, for X[n;n], all of a, b, c, d, h do not depend on n, and similarly for other extremes. Then, virtually all of the data in the upper tail can be used for estimations and investigation purposes. This form of asymptotic distribution should exist and be usable for nearly all cases of practical interest if a, b, c, d, and h are allowed to depend on n; that is, when Situation (II) holds, with its disadvantages in use of past data for estimation or investigation.

The extended asymptotic distributions exist for a much wider class of distributions sampled than do the first and second asymptotes. Also, the class of distributions sampled that result in Situation (I) is much more extensive. Identification of these two classes of distributions sampled, and properties of the extended asymptotic distributions, have been partially examined by the graduate student (Anderson, 1968).

For an upper extreme X[n+1-t;n], the asymptotic cdf is a monotonic function of $-ax^2 + bx + c + d\log_e(x-h)$ that increases as x increases $(x \ge x_L)$. For a lower extreme X[t;n], the asymptotic cdf is a monotonic function of an expression of the form $-ax^2 + bx + c + d\log_e(h-x)$, where $-ax^2 + bx + c + d\log_e(h-x)$ decreases as x decreases $(x \le x_U)$.

Incidentally, it should be noted that, for given n, a stated extreme can have any possible distribution. In fact it was shown that this is true for any specified order statistic of a sample of size n (Walsh, 1968a).

<u>Nonsample Cases</u>

As shown in (Walsh, 1964, 1965) and extended in (Walsh, 1968b), the results for samples are approximately usable for extremes in many nonsample cases. That is, consider a specified extreme of n observations that have an arbitrary joint distribution. There exists a distribution such that this extreme is a sample of size n from this population has the same cdf as that of this extreme in the n observations (Walsh, 1968b). If n is large and the observations are independent (or have a form of mild independence), the distribution considered to be sampled approximately equals the arithmetic average of the distributions for the individual observations (Walsh, 1964, 1965). In fact, for independence, the distribution of any specified order statistic (not necessarily an extreme) can often be approximated by considering that the observations are a sample from this average distribution (Walsh, 1959).

More refined approximations have been developed for the upper part of the cdf of $X[n;n]$ for the case of independence (Walsh, 1968c and the related material Walsh, 1968d) and can be directly extended to the general nonsample case by use of the results in (Walsh and Kelleher, 1968a). Similar results are developed for the lower part of the cdf of $X[1;n]$. Also, an approximation has been developed for joint probabilities involving $X[1;n]$ and $X[n;n]$ in the case of independence (Walsh, 1968e). These results are applicable for $n \geq 1$.

In all of the independence cases mentioned, the distribution of an upper extreme can be approximately expressed in terms of $n[1 - \bar{\bar{F}}(x;n)]$, and the distribution of a lower extreme in terms of $n\bar{\bar{F}}(x;n)$. Here,

$\bar{F}(x;n)$ is the arithmetic average of the cdfs for the individual observations. Use of an extended asymptotic distribution for an upper extreme is accomplished by letting $-ax^2 + bx + c + d\log_e(x-h)$ represent $\log_e[1 - \bar{F}(x;n)]$ for $x \geq x_L$. Likewise, for a lower extreme, the form $-ax^2 + bx + c + d\log_e(h-x)$ is used to represent $\log_e\bar{F}(x;n)$ for $x \leq x_U$. This converts the expressions stated in terms of $n[1 - \bar{F}(x;n)]$, and in terms of $n\bar{F}(x;n)$, into their asymptotic forms.

In many respects, the cdf $\bar{F}(x;n)$ for the nonsample case and independence has been found to play the same role as the population cdf for a random sample. This is especially the case when n is large. To use past data for estimation or investigation, the average of the cdfs should, of course, equal $\bar{F}(x;n)$ for the n new observations or bear a completely determined relation to $\bar{F}(x;n)$.

Properties for Sample and Nonsample Cases

It is, of course, allowable to obtain the joint distribution of $X[1;n]$ and $X[n;n]$ as the product of their separate distributions when these extremes are very nearly independent. This simplifies the estimation problem for the joint distribution to separate estimation of the distribution for each extreme.

Asymptotic independence of $X[1;n]$ and $X[n;n]$ has been examined for the case of independent observations (Walsh, 1968f). Asymptotic independence does not always occur but should happen in most situations of practical interest.

Asymptotic independence of $X[1;n]$ and $X[n;n]$ does always occur for the random sample case. Minimum values of n to assure at least a stated level of independence have been developed (Walsh, 1968g). Also, minimum n for assuring a stated level of independence between X[1;n] and the sample median, or X[n;n] and the sample median, have been developed (Walsh, 1968h).

The approach developed in (Walsh, 1968b) is helpful in establishing regression functions whose probability properties are approximately determined under very general conditions (Walsh and Kelleher, 1968b). Approximately, these results represent an extension of median regression, and can be used to develop a very general form of paired comparisons (Walsh, 1968i).

A general discussion of recent results for extreme values (oriented toward extreme ages), that includes many of the results for this NASA project, is given in (Walsh, 1968j).

CONCLUSIONS

It seems that asymptotic distributions that are much more useful than those now considered can be developed. A first step has been made in this direction. Also, a number of results have been developed that should be useful in applications involving extreme values and other order statistics.

REFERENCES

Anderson, Charles, (1968), Research on Asymptotic Distributions of Extreme Values (Preliminary Rept.), Statistics Dept., Southern Methodist Univ.

Gumbel, Emil J., (1958), Statistics of Extremes, Columbia Univ. Press.

Walsh, John E., (1959), "Definition and use of generalized percentage points," Sankhyā, Vol. 21, pp. 281-288.

_____________, (1964), "Approximate distribution of extremes for nonsample cases," Jour. Amer. Stat. Assoc., Vol. 59, pp. 429-436.

_____________, (1965), Handbook of Nonparametric Statistics, II, Van Nostrand, pp. 573-629.

List of Reports Written Under NASA Grant
NGR 44-007-028 to December 31, 1968
Statistics Department, Southern Methodist University

(a) Walsh, John E., Exact Existence of Every Possible Distribution for Any Sample Order Statistic, 3 pages. Submitted to Australian Journal of Statistics.

(b) _____________, Sample-Like Distribution of An Order Statistic Under General Nonsample Conditions and Some Asymptotic Implications, 7 pages. Submitted to South Africa Statistics Journal.

(c) _____________, Asymptotic Distributions for Largest and for Smallest of a Set of Independent Observations, 12 pages. Submitted to South Africa Statistics Journal.

(d) _____________, Chance of All Successes for Independent Binomial Events with Large Success Probabilities, 5 pages. Presented at National AAAS meeting, Dallas, Texas, December 26, 1968.

(e) _____________, Approximate Joint Probabilities for Largest and Smallest of a Set of Independent Observations, 9 pages. Submitted to Sankhyā.

(f) _____________, Asymptotic Independence Between Largest and Smallest of a Set of Independent Observations, 4 pages. Submitted to Annals of the Institute of Statistical Mathematics.

(g) ______________, Sample Sizes for Approximate Independence of Largest and Smallest Order Statistics, 5 pages. Accepted for publication in Journal of the American Statistical Association.

(h) ______________, Sample Sizes for Approximate Independence Between Sample Median and Largest (or Smallest) Order Statistic, 7 pages. Accepted for publication in Australian Journal of Statistics.

(i) ______________, Development and Analysis of Modified Paired Comparisons by Use of Linearized Nonlinear Regression, 12 pages. Accepted for publication in Biometrische Zeitschrift.

(j) ______________, Recent Probability Results for Extreme Ages, 8 pages. Submitted to Journal of the Institute of Actuaries.

(a) Walsh, John E., and Kelleher, G. J., Approximation to Large Probabilities of All Successes for General Case and Some Operations Research Implications, 5 pages. To be presented at 3rd Israel Operations Research Conference, July 2-4, 1969.

(b) ______________, and Kelleher, G. J., Extended Use of Linearized Nonlinear Regression for Random Nature Simulations, 12 pages. To be presented at Fifth International Conference of International Federation of Operational Societies, June 23-27, 1969, and published in the Proceedings.

1968 LIST OF PAPERS FOR NGR 44-007-028

(By John E. Walsh)

1. "Exact existence of every possible distribution for any sample order statistic," submitted to Australian Journal of Statistics.
2. "Recent probability results for extreme ages," submitted to Journal of the Institute of Actuaries.
3. "Joint probabilities for largest and smallest of a set of independent observations," submitted to Sankhyā.
4. "Sample sizes for approximate independence between sample median and largest (or smallest) order statistic," accepted for publication in Australian Journal of Statistics.
5. "Asymptotic independence between smallest and largest of a set of independent observations," submitted to Annals of the Institute of Statistical Mathematics.
6. "Approximate distributions for largest and for smallest of a set of independent observations," submitted to South Africa Statistics Journal.
7. "Extended uses of linearized nonlinear regression for random-nature simulations," (with G. J. Kelleher), Invited paper to be presented at the Fifth International Conference of the International Federation of Operational Research Societies (Italy, June 23-27, 1969), and included in the published proceedings for the conference.
8. "Sample sizes for approximate independence of largest and smallest order statistics," submitted to Journal of the American Statistical Association.
9. "Sample-like distribution of an order statistic under general nonsample conditions and some asymptotic implications," submitted to Journal of the American Statistical Association.
10. "Chance of all successes for independent binomial events with large success probabilities." Invited paper presented at National AAAS Meeting, Dallas, Texas on December 26, 1968.
11. "Development and analysis of modified paired comparisons by use of linearized nonlinear regression," accepted for publication in Biometrische Zeitschrift.
12. "Approximation to large probabilities of all successes for general case and some operations research implications," (with G. J. Kelleher). Invited paper to be presented at Israel International Operations Research Meeting, July 2-4, 1969.