THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

RECENT PROBABILITY RESULTS FOR EXTREME AGES

by

John E. Walsh

Technical Report No. 11
Department of Statistics THEMIS Contract

September 20, 1968

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

DEPARTMENT OF STATISTICS
Southern Methodist University

RECENT PROBABILITY RESULTS FOR EXTREME AGES

John E. Walsh

Southern Methodist University*

ABSTRACT

Consider a very large number of persons, and probability distributions for the age at death of the last survivor, next to last survivor, etc. First, suppose that the persons are statistically independent with the same probability distribution for age at death (random sample case). Then, some approximations to distributions of extremes are often usable. These approximations are completely specified except for at most three parameters. This simplifies distribution estimation to estimation of the parameters. Moreover, previous large samples (possibly different sizes) from the same population of persons, and much of their data on extremes, can be used for estimation. Also, the sample results often remain applicable for the more general case of independence (or mild dependence) but possibly different distributions for the ages at death. Here, the average of these distributions is "sampled." Very recent results show that distributions of extremes are of the sample type for any joint distribution of the ages at death. However, the distribution "sampled" can be greatly different for the last survivor, next to last survivor, etc. This effectively limits estimation of parameters to previous groups having very nearly the same joint distribution and use of one observed extreme per group.

*Research partially supported under NASA Grant NGR 44-007-028.
Also associated with ONR Contract N00014-68-A-0515.

INTRODUCTION

This paper is of a descriptive nature, outlining various probability results for extremes but not stating any of them in technical detail. However, references are given that contain the material in detail. An exception in references is the material stated for the general nonsample case. This is so recent that it is not yet published.

Probability distributions of age at death for the last survivors of a very large group of persons are interesting in at least two respects. They can be of some interest in calculating premiums (especially for annuities). They are also of interest as an indication of the maximum longevity for humans.

The somewhat unrealistic situations where the ages at death constitute a random sample is discussed first. Here, the recent results consist of improved methods of estimation. The nonsample case with independence or mild dependence is considered next. This case seems to be the most appropriate for extreme ages. Finally, the general nonsample case is discussed.

SAMPLE CASE

Consider the idealized situation where, for all members of the group, the ages at death are statistically independent and have the same probability distribution. Then, the observed values for the ages constitute a random sample and the interest is in the probability distribution of the largest sample value, the next to largest value, etc. Large sample approximations to these distributions have been developed for cases where

the population sampled is continuous (virtually always the case for ages at death). Three types of approximating distributions, which satisfy a special stability condition, have received the predominant consideration. They are often referred to as the first, second, and third asymptotes for largest extremes (Gumbel, 1958).

The first two asymptotes, which do not place any upper bound on possible values, are of principal interest for the case of extreme ages. They are completely specified except for two parameters (sometimes an additional parameter occurs for the second asymptote). The second asymptote is characterized by having a heavy tail (to the right). The first asymptote also has a tail extending toward increasingly large values, but this tail is not of a heavy nature. Although a first or second asymptote does not necessarily occur for all distributions (with tails) that could be sampled, this is ordinarily the case. The first asymptote would be used if the upper tail of the distribution is believed to not be heavy (not much heavier than for the exponential distribution). The second asymptote is used if the tail is believed to be quite heavy.

Use of a distribution (for an extreme) that is known except for the values of a few parameters is convenient. The problem of estimating the distribution is reduced to that of estimating these parameters. In general, this estimation has been based on previous independent samples having the same size and from (approximately) the same distribution as for the new sample to be obtained. Consider the extreme from each previous sample that corresponds to the extreme whose distribution is

to be estimated. These observed extremes constitute a random sample (of this type of extreme) and can be directly used to estimate the parameters of their asymptote (Gumbel, 1958).

The ability to use only past samples having the same size as the new sample, and only one extreme of each sample, severely limits the data available for estimation. Fortunately, if the distribution sampled is of a reasonable nature, the past samples need only be large and nearly all of their extremes can be used in the estimation. Here "reasonable nature" implies that asymptotes for increased sample sizes can be obtained by using those for smaller (but still very large) sample sizes in the usual ways. For example, consider the joint distribution of the largest value for each of six independent samples of size n (from the same distribution). The joint distribution of these six extremes is (approximately) the product of their individual asymptotes, and the distribution of the largest of their values can be determined. This should be (approximately) the same as the asymptote of the largest value of a sample of size 6n if the distribution sampled is of a reasonable nature. ("Reasonable nature" is the same as Situation I in [Walsh, 1965].)

Some recent results (Walsh, 1965) provide many ways for estimating the parameters of an asymptote when the population sampled is of a reasonable nature. Included is a least-squares procedure which makes use of all of a specified class of upper extremes for every previous sample. The past samples can have different sizes (all very large) for each of these estimation methods. The basis for these estimation methods is the fact that the same parameters occur in the asymptote for all of the large extremes.

INDEPENDENT OR MILDLY DEPENDENT

The idealized situation of the ages at death constituting a random sample should seldom occur to any reasonable approximation in practice. However, it seems that statistical independence should occur to a reasonable approximation in many cases. More generally, an m-dependence situation, with m small and mild dependence, should often be closely approximated. That is, the age at death for a person is independent of the ages at death except for at most m other persons, with m small (say, at most 10). Also, the dependence is so mild that corresponding conditional and unconditional probabilities are of about the same absolute magnitude but not necessarily of nearly the same relative magnitude. For example, the unconditional probability of living after age x might be .005 while the largest conditional probability is .05. These differ by a factor of 10 but are both of small absolute magnitude.

Recent results (Walsh, 1963, 1965) show that the asymptotic distributions of extremes for this m-dependence (m = 0 for independence) situation are of the same nature as those for the sample case in most instances. In fact, the observations behave as if they were a random sample (size equal to the number of observations) from a distribution that equals the arithmetic average of the unconditional distributions for the individual observations. Thus, the asymptotes that were developed for the sample case remain applicable. Also, the estimation methods for the sample case can be used. However, the average of the distributions for each of the previous sets of observations must be approximately the same as that for the new set of observations.

GENERAL NONSAMPLE CASE

The m-dependence situation seems capable of approximating most cases that yield extreme ages. There is some possibility, however, that situations can occur where the dependence is greater than that allowed for m-dependence. For example, a fire in a home for elderly people could result in a very strong dependence among the ages at death of some persons who survived to extreme ages.

Very recent results (Walsh, 1968) that are not yet published show that an extreme has a sample-like distribution even when the sample condition is grossly violated. In fact, this is the case for any joint probability distribution of the ages at death. That is, consider any number n of observations and a stated extreme. There exists a distribution such that this extreme in a sample of size n has exactly the same distribution as the extreme in the set of observations with the joint probability distribution. However, the distribution considered to be "sampled" for one extreme need not be even roughly the same as that "sampled" for another extreme.

The estimation of parameters is complicated by the fact that different parameters occur in the asymptotes for different extremes. Also, the previous data for estimation must yield approximately the same distribution for the extreme as that occurring for this extreme in the new set of observations. Ordinarily, this implies that the previous sets of observations must be of the same size as the new set and also from approximately the same source. Moreover, the only observation usable for a previous set is this extreme of that set. Then, if the

previous groups are statistically independent, these observed extremes are an approximate random sample from the asymptote to be estimated. The parameters of this asymptote can be estimated by the procedures that have been developed for using data of this kind in the sample case (Gumbel, 1958, etc.).

REFERENCES

Gumbel, Emil J., Statistics of extremes, Columbia Univ. Press, 1958.

Walsh, John E., "Approximate distribution of extremes for nonsample cases," Jour. Amer. Stat. Assoc., Vol. 59, 1964, pp. 429-436.

Walsh, John E., Handbook of nonparametric statistics, II, Van Nostrand, 1965.

Walsh, John E., "Sample-like distribution of an order statistic under general nonsample conditions and some asymptotic implications," NASA Project NGR 44-007-028, Southern Methodist University, Statistics Department, July, 1968.