AD 747064

NPS55LW71122A

United States Naval Postgraduate School

SIMULTANEOUS ESTIMATION OF LARGE NUMBERS OF EXTREME QUANTILES IN SIMULATION EXPERIMENTS

by

A. S. Goodman
IBM Research Lab

P. A. W. Lewis
Naval Postgraduate School

H. E. Robbins
Columbia University

December 1971

Approved for public release; distribution unlimited.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Simultaneous Estimation of Large Numbers of Extreme Quantiles in Simulation Experiments

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Peter A. W. Lewis

6. REPORT DATE December 1971	7a. TOTAL NO. OF PAGES 37	7b. NO. OF REFS 13
8a. CONTRACT OR GRANT NO. b. PROJECT NO c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The large random access memory and high internal speeds of present day computers can be used to increase the efficiency of large-scale simulation experiments by estimating simultaneously several quantiles of each of several statistics. In order to do this without inordinately increasing programming complexity, quantile estimation schemes are required which are simple and do not depend on special features of the distributions of the statistics considered. We discuss limitations, when the probability level is very high or very low, of two basic methods of estimating quantiles. One method is the direct use of order statistics; the other is based on the use of stochastic approximation.

Several modifications of these two estimation schemes are considered. In particular a simple and computationally efficient transformation of the simulation data is proposed and the properties (i.e. bias and variance) of quantile estimates based on this scheme are discussed.

J

DD FORM 1 NOV 65 1473 (PAGE 1)

N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Computers						
Simulation						
Estimation						
Quantiles						
Stochastic approximation						
Order statistics						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

UNCLASSIFIED
Security Classification

A-31409

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral A. S. Goodfellow, USN
Superintendent

M. U. Clauser
Provost

ABSTRACT

The large random access memory and high internal speeds of present day computers can be used to increase the efficiency of large-scale simulation experiments by estimating simultaneously several quantiles of each of several statistics. In order to do this without inordinately increasing programming complexity, quantile estimation schemes are required which are simple and do not depend on special features of the distributions of the statistics considered. We discuss limitations, when the probability level α is very high or very low, of two basic methods of estimating quantiles. One method is the direct use of order statistics; the other is based on the use of stochastic approximation.

Several modifications of these two estimation schemes are considered. In particular a simple and computationally efficient transformation of the simulation data is proposed and the properties (i.e. bias and variance) of quantile estimates based on this scheme are discussed.

Prepared by:

P. A. W. Lewis

P. A. W. Lewis
Department of Operations Research
and Administrative Sciences

Approved by:

J. R. Borsting, Chairman
Department of Operations Research
and Administrative Sciences

Released by:

C. E. Menneken
Dean of Research Administration

P. A. W. Lewis is on leave of absence from the IBM Research Center.

NPS55LW71122A

INTRODUCTION

Consider a situation in which we have a collection of random variables $X_1, X_2, \ldots, X_n$ with joint distribution $F(x_1, \ldots, x_n)$, and several statistics (functions of these n random variables), say $S(n) = g_1(X_1, \ldots, X_n)$, $T(n) = g_2(X_1, \ldots, X_n)$, etc. It is required to estimate the distributions $F_S(s)$, $F_T(t)$, ... of these statistics (or some characteristics of the distributions) by obtaining m samples $x_{1,i}, x_{2,i}, \ldots, x_{n,i}$, $i = 1, \ldots, m$, from $F(x_1, \ldots, x_n)$ and hence m values for each of the statistics $S(n)$, $T(n)$, ... Two examples of this type of situation are as follows:

i) In testing for independence in a time series $\{X_i\}$, many test statistics have been proposed. These are functions of finite sets of the $\{X_i\}$, namely $X_1, \ldots, X_n$, and the hypothesis is that $F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$. Typical statistics are the sample serial correlation coefficients with various delays (lags), i.e.

$$\rho_\ell(n) = \frac{\frac{1}{n-\ell} \sum_{i=1}^{n-\ell} (x_i - \bar{x})(x_{i+\ell} - \bar{x})}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad \ell = 1, 2, \ldots, n-1,$$

where $\bar{x} = (x_1 + \ldots + x_n)/n$, statistics based on the finite Fourier transform of the x_i's which test essentially that the spectrum of $\{X_i\}$ is flat, and several non-parametric test statistics such as those based on runs. The distributions of most of these statistics are known for independent, normally distributed X_i's, but not when the assumption of a

normal distribution is removed. In testing for a renewal hypothesis in series of events (Cox and Lewis, 1966, p. 164) an exponential distribution for the X_i's may be reasonable. The null distributions of the test statistics are then unknown, as are the rate of convergence to the limiting $(n \to \infty)$ distributions (some of which are known).

In examining and tabulating the finite sample distributions, it may be required to estimate the distributions of several of these test statistics for many different values of n.

ii) There is much interest in analyzing very complex queueing and congestion systems, particularly those that arise in computing and communication contexts. Here one might be interested in estimating by simulation the distributions of the waiting times at several points in the system at several different times during its evolution.

In looking at the problem of estimating these distributions from m replications of the statistics, several general problems arise which need to be considered.

First, it is neither practical nor desirable to save all of the information generated on a statistic by the simulation in the form of the empirical distribution function or, equivalently, in the form of the complete set of m order statistics. Some compact characterization of the distribution is required. In situations such as that of the first example given above, out of which this present study in fact arose, the characteristics of the distribution function chosen were the first four moments

and sixteen quantiles of the distribution. (A quantile x_α of a distribution $F(x)$ is defined as the solution of the equation

$$\alpha = F(x_\alpha), \qquad 0.0 < \alpha < 1.0,$$

and we shall be assuming throughout this paper that x_α is unique). The probability levels chosen for the quantiles were $\alpha = 0.001$, 0.002, 0.005, 0.010, 0.020, 0.025, 0.050, 0.100, and $\alpha = 0.900$, 0.950, 0.975, 0.980, 0.990, 0.995, 0.998, 0.999. The choice of some of these α's is based on the levels customarily used in testing statistical hypotheses; the more extreme values have been chosen rather arbitrarily to characterize the extreme tails of the distributions. In many queueing situations it is these extreme values, rather than moments, which are of interest.

Another possible characteristic is the probability of the statistic being less than a given value. These percentiles are clearly important in studies of the power of test statistics against non-null hypotheses. Their estimation is fairly straightforward and will not be considered here.

A second point concerns the measurement of <u>statistical efficiency</u> by either the variance or the mean square error of the estimator. The mean square error is the variance of the estimator plus its bias squared. In large scale applications of simulations, as treated in this study, it is important to obtain internal assessments of the variability of the estimation procedures by, for instance, estimating a quantile as the average of r estimates from samples of size m_r, where $r\,m_r = m$. The sample

standard deviation calculated from the r estimates, divided by $(r)^{1/2}$, then estimates the sampling standard deviation of the quantile estimator (see Mosteller and Tukey, 1968, for more details). In order to assess the internal variability of the estimation in this way the bias of the estimator must be small compared to the standard deviation of the estimator. Otherwise we obtain, from one point of view, a significant bias component in the mean square error, or from another point of view, an estimate with very small sampling variance of the wrong quantity, i.e., true quantile plus bias.

Bias thus becomes a very important factor in assessing the quantile estimates discussed in this paper.

A third consideration is that in some cases one can find particular properties of a statistic whose distribution is to be estimated that allow one to obtain estimates that are more "efficient" than those obtained by straight synthetic sampling. By "efficiency" here we mean <u>statistical efficiency</u>, or the relative variances or mean square errors of different estimating procedures. However there are other less tangible costs involved in simulation. One is the time involved in deriving a particular procedure for a given statistic, another the time involved in programming and debugging such a procedure and the delay in obtaining results. Still another cost is the <u>actual computing time</u> involved though this is seldom mentioned in the statistical literature. (This latter point will become clearer later in the paper.) These less tangible costs are an important

component of overall computational efficiency. The point of view taken in this paper is that the rapidly accelerating availability of large memory, high internal speed computers makes it usually more "efficient", in the general computational sense, to forego using special techniques for each particular statistic and to use very simple straightforward simulation techniques. Thus a criterion for the quantile estimation techniques discussed here is programming and computing simplicity.

We do not mean to imply by this that all sophisticated statistical and Monte Carlo techniques are not useful or applicable. Global techniques such as jackknifed estimates (Quenouille, 1956) or the use of variance reduction techniques such as control variables (Gaver, 1969) can be used with the quantile estimation methods discussed in this paper. The jackknife technique will be discussed in the next section and the use of the quantile estimation techniques in the context of sophisticated Monte Carlo will be discussed elsewhere.

Finally, it is perhaps worthwhile to give some idea of the numbers envisioned in connection with the estimation procedures. Clearly no scheme involving only raw simulation will work satisfactorily in estimating a 0.001 quantile ($x_{0.001}$) unless the number m of replications involved is substantially greater than 1000. Typically in the COMPSTAT program (Goodman and Lewis, 1972), for which these techniques were developed, replications of 100,000 or more are common. These are not excessive on the IBM 360/91 on which the runs were made. In addition,

the large core memory of this machine enabled us to run sampling experiments on up to 31 statistics simultaneously.

The following discussion uses as examples, for the most part, the extreme 0.001 and 0.999 quantiles. The techniques are still useful, though not as much so, for the inner quantiles. Some discussion of the simultaneous estimation of, for example, all sixteen quantiles listed above is given. The dependence of the utility and efficiency of the various quantile estimation schemes on the particular quantile or set of quantiles plus, for example, variations in the complexity of computation of the series $X_1, \ldots, X_n$ and the statistics $S(n)$, $T(n), \ldots$ in different problems make it difficult to be dogmatic about the relative utility of various quantile estimation schemes. In addition, most of the results required are finite sample results. These are difficult to obtain analytically and expensive, as yet, to obtain computationally.

II. QUANTILE ESTIMATION

a) Overall Considerations

Two general methods of quantile estimation are considered, one based on the order statistics of the sample, the other based on stochastic approximation (Robbins-Monro) techniques (see, e.g., Robbins and Monro [1951], Hodges and Lehmann [1956], Cochran and Davis [1965]).

For simplicity we drop the notation which indicates dependence on n and write $S(n)$ as S and write its distribution function simply as $F(s)$. A collection of m independent random

variables with distribution $F(s)$ will be denoted by $S_1, \ldots, S_i, \ldots, S_m$ and the corresponding order statistics by $S_{(1)} < S_{(2)} < \ldots < S_{(i)} < \ldots < S_{(m)}$.

The order statistic estimator of the α-quantile s_α, where $\alpha = F(s_\alpha)$, is

$$\bar{s}_\alpha = S_{([\alpha m])}, \tag{2.1}$$

where $[\alpha m]$ denotes the integral part of αm. Thus for $\alpha = 0.999$ and $m = 10,000$, $\bar{s}_{0.999} = S_{(9990)}$.

The stochastic approximation estimate $\hat{s}_\alpha(m)$ is defined to be the m th value in the sequence defined by

$$\hat{s}_\alpha(i) = \hat{s}_\alpha(i-1) - \frac{C}{i}\left[\frac{1 - \operatorname{sgn}\{S_i - \hat{s}_\alpha(i-1)\}}{2} - \alpha\right] \quad (i = 1, 2, \ldots, m), \tag{2.2}$$

where $\operatorname{sgn} x = 1$ if $x > 0$ and -1 if $x \leq 0$, and $\hat{s}_\alpha(0)$ is an arbitrary initial value. If the constant C is chosen to be $1/f(s_\alpha)$, where $f(s_\alpha)$ is the density associated with $F(s)$ evaluated at the quantile s_α, then the asymptotic variance $(m \to \infty)$ of $\hat{s}_\alpha(m)$ is minimized. In fact,

$$E\{\hat{s}_\alpha(m)\} \sim s_\alpha \tag{2.3}$$

and

$$\operatorname{var}\{\hat{s}_\alpha(m)\} \sim \frac{\alpha(1-\alpha)}{m f^2(s_\alpha)} . \tag{2.4}$$

Remarkably, the estimate $\bar{s}_\alpha$ has the same asymptotically normal distribution as does $\hat{s}_\alpha(m)$. Results on rates of convergence are known for $\bar{s}_\alpha$ but not for $\hat{s}_\alpha(m)$. This will be discussed later, but significant

comparisons can be made on the basis of the asymptotic results and on known computational results.

Order Statistics, $\bar{s}_\alpha$

- computation time (to order the m realizations of S) is proportional to m ln(m);
- computer memory required for the sorting process is proportional to m. (Actually, to α m if $\alpha < \frac{1}{2}$, $(1-\alpha)$m if $\alpha > \frac{1}{2}$);
- no initial values or knowledge of F(s) is required;
- the rate of convergence of the estimate is known.

Stochastic Approximation, $\hat{s}_\alpha = \hat{s}_\alpha(m)$

- computation time (binary comparison) is proportional to m;
- computer memory required is proportional to 2;
- initial values $\hat{s}_\alpha(0)$ and values for $f(s_\alpha)$ are needed, presumably previous estimates or guesses based on prior knowledge;
- no reliable results known on the rate of convergence of $\hat{s}_\alpha$, or even of $E(\hat{s}_\alpha)$;
- it is not necessary to know S_i exactly, only that it is greater than or less than $\hat{s}_\alpha(i)$. This can be very advantageous if computation of S is time-consuming.

In summary, $\hat{s}_\alpha$ has very definite advantages over $\bar{s}_\alpha$ in terms of computational considerations. One might say that the asymptotic relative efficiency of $\bar{s}_\alpha$ compared to $\hat{s}_\alpha$, in terms of real time and not sample size m, is zero. However, initial values are needed for $\hat{s}_\alpha$, so

that the asymptotic results really beg the question. Further differences appear in terms of finite sample properties of the estimators, and these are discussed next.

III. QUANTILE ESTIMATION - Finite Sample Considerations for $\bar{s}_\alpha$

It is well known (David, 1971) that for the ordered sample $S_{(1)} < S_{(2)} < \ldots < S_{(i)} < \ldots < S_{(m)}$

$$F_{S_{(i)}}(s) = \text{prob}\,\{S_{(i)} \le s\} = \sum_{k=i}^{m} \binom{m}{k} [F(s)]^k\,[1-F(s)]^{m-k}, \tag{3.1}$$

$$f_{S_{(i)}}(s) = F'_{S_{(i)}}(s) = \binom{m}{i}\, i\,[F(s)]^{i-1}\,[1-F(s)]^{m-i}\, f(s), \tag{3.2}$$

$$E\{\bar{s}_\alpha\} = E\{S_{([\alpha m])}\} = s_\alpha + O\left(\frac{1}{m}\right), \tag{3.3}$$

$$\text{var}\,\{\bar{s}_\alpha\} = \text{var}\,\{S_{([\alpha m])}\} = \frac{\alpha(1-\alpha)}{m f^2(s_\alpha)} + O\left(\frac{1}{m^2}\right). \tag{3.4}$$

Asymptotic expansions for $E\{s_\alpha\}$ are known (David and Johnson [1954] and Clark and Williams [1958]), the first terms in the expansion being

$$E\{s_\alpha\} = s_\alpha - \frac{f'(s_\alpha)}{2 f^3(s_\alpha)} \cdot \frac{\alpha(1-\alpha)}{m+2} + \frac{2[(f'(s_\alpha))^2 - f(s_\alpha) f''(s_\alpha)]}{6 f^5(s_\alpha)} \cdot \frac{2\alpha(1-\alpha)(1-2\alpha)}{(m+2)(m+3)} + \ldots, \tag{3.5}$$

where derivatives are denoted by primes and powers by arabic numerical exponents. No precise conditions for these asymptotic expansions seem to be known, and they must be used with care. For example, if S is uniformly distributed between 0 and 1, $s_\alpha = \alpha$ and, using (3.2), $E\{S_{(i)}\} = i/(m+1) = (i/m) - i/[(m)(m+1)]$.

Thus if αm is the integer i, we have

$$E\{\bar{s}_\alpha\} = \alpha - \frac{\alpha}{(m+1)},$$

but the second term of the expansion (3.5) is zero. Similarly, in the extreme but computationally useful case $S = \frac{U}{1-U}$, where U is uniform, we have

$$F(s) = \frac{s}{1+s}, \tag{3.6}$$

a distribution with an infinite mean. Surprisingly, however, $\bar{s}_\alpha$ can be shown, using (3.2) to be an unbiased estimator of s_α. In this case the expansion (3.5) does not even converge.

An important consequence of (3.3) is that the jackknife technique (Quenouille [1956], Mosteller and Tukey [1966]) for eliminating an $O(\frac{1}{m})$ term in the bias can be applied to the order statistic estimate $\bar{s}_\alpha$. By way of illustration consider the technique being applied with just a simple splitting of the data. Thus, assume m is even and $\bar{s}_\alpha(1)$ is the order statistic estimator from the first m/2 S_i's, $\bar{s}_\alpha(2)$ the estimator from the second m/2 S_i's. The "typical values" are defined to be

$$\bar{\bar{s}}_\alpha(1) = 2\bar{s}_\alpha - \bar{s}_\alpha(1), \qquad \bar{\bar{s}}_\alpha(2) = 2\bar{s}_\alpha - \bar{s}_\alpha(2), \tag{3.7}$$

and the jackknifed estimate is

$$\bar{\bar{s}}_\alpha = \frac{\bar{\bar{s}}_\alpha(1) + \bar{\bar{s}}_\alpha(2)}{2} = 2\bar{s}_\alpha - \frac{\bar{s}_\alpha(1)}{2} - \frac{\bar{s}_\alpha(2)}{2}. \tag{3.8}$$

The $O(\frac{1}{m})$ term in $E[\bar{\bar{s}}_\alpha]$ is zero. There is no appreciable computing cost involved in obtaining the jackknifed estimate. The two halves of the sample can be sorted in place to obtain $\bar{s}_\alpha(1)$ and $\bar{s}_\alpha(2)$, and then these sorted halves are merged to obtain the complete, sorted sample $S_{(1)} < \ldots < S_{(m)}$. This in fact is just the usual binary sorting procedure which results in $m \ln(m)$ sorting time.

The main drawback to the jackknife procedure is that it may increase the variance of the estimator, and this variance is difficult to obtain analytically. We have

$$\text{var}\{\bar{\bar{s}}_\alpha\} = 5\,\text{var}\{\bar{s}_\alpha\} - 2\,\text{cov}\{\bar{s}_\alpha\ \bar{s}_\alpha(1)\} \leq 5\,\text{var}\{\bar{s}_\alpha\}. \tag{3.9}$$

The covariance term involves the covariance between the $[\alpha(m/2)]$ order statistic in half the sample, and the $[\alpha m]$ order statistic in the whole sample. If in fact $\bar{\bar{s}}_\alpha(1)$ and $\bar{\bar{s}}_\alpha(2)$ are approximately uncorrelated, then there is no increase in variance. Even if the variance is inflated enough to make the mean square error of the estimates approximately equal, there is a gain in that the smaller bias of the jackknifed estimator allows for sectioning the complete sample of size m into r smaller sections of size m_r. This gives a more precise empirical variance estimate and smaller computation time, the latter following since shorter sections are sorted.

Unfortunately, there is some evidence (Miller, 1964) that jackknifing is a poor technique to use with estimators involving extreme order

statistics. We give now an illustrative example.

Example - The exponential distribution

Consider the estimation of the 0.999 quantile $s_{0.999}$ for a unit exponential distribution. For simplicity we assume that the sample size m is such that 0.999m is an integer k. We have

$$y = F(s) = 1 - e^{-s},$$

$$s = F^{-1}(y) = -\ln(1-y), \tag{3.10}$$

$$s_\alpha = -\ln(1-\alpha).$$

By direct methods (Cox and Lewis, 1966) one gets

$$E\{\bar{s}_\alpha\} = E\{S_{(k)}\} = -\ln(1-\alpha) - \frac{1}{2m}\left(\frac{\alpha}{1-\alpha}\right) + \frac{1}{12m^2}\left[\frac{(1-\alpha)^2 - 1}{(1-\alpha)^2}\right] + O\left(\frac{1}{m^3}\right), \tag{3.11}$$

$$\text{var}\{\bar{s}_\alpha\} = \frac{\alpha}{m(1-\alpha)} - \frac{1}{2m^2}\left[\frac{1}{(1-\alpha)^2} - 1\right] + O\left(\frac{1}{m^3}\right). \tag{3.12}$$

These results are used to give Table 1. The ratio (column 8) of the standard deviation (σ, column 7) of the estimate to the bias of the estimate (column 3) indicates roughly how feasible it is to use averages of estimates $\bar{s}_{0.999}$ from samples of size m to estimate $\bar{s}_{0.999}$ more precisely, along with an estimate of the sampling variance of that estimate. Thus 36 samples of size m = 10,000 produces an estimate with a standard deviation approximately equal to the absolute value of the bias, clearly an undesirable situation. Moreover the bias is -0.051, so that this estimate gives us accuracy of at best two decimal places. This would not be acceptable in many cases.

Table 1

Unit Exponential - Estimation of $s_{0.999} = 6.908$ by $\bar{s}_{0.999} = s_{[0.999\,m]}$

m	k=.999m	Bias	$\frac{1}{m}$ term	$\frac{1}{m^2}$ term	var$\{\bar{s}_{0.999}\}$	σ	$\lvert\sigma/\text{Bias}\rvert$	m. s. e.
1,000	999	-0.583	-0.500	-0.083	0.499	0.706	1.21	0.839
10,000	9990	-0.051	-0.050	-0.001	0.095	0.308	6.07	0.098
100,000	99900	-0.005	-0.005	-0.830×10^{-5}	0.010	0.100	19.93	0.010

Table 2

Unit Exponential - Estimation of $s_{0.999}$ by jackknifed estimator $\bar{\bar{s}}_{0.999}$

m	Bias	$\lvert\sigma/\text{Bias}\rvert$	Variance factor for m. s. e.	var $\{\bar{\bar{s}}_{0.999}\}$	m. s. e.
1,000	-0.167	4.2	1.53	1.151 (0.0351)	1.179
10,000	-0.002	154.0	1.03	0.136 (0.00381)	0.136
100,000	-0.166×10^{-4}	6,000.0	1.00	0.0111 (0.000269)	0.0111

The leading term of the bias of the jackknifed estimate (3.8) is shown in Table 2 (column 2) as a function of m. Column 3 shows the ratio of the standard deviation of the unjackknifed estimator to this bias. Clearly, sectioning and averaging is much more feasible. However, as indiated in column 4 an increase by only a factor of 1.03 in the variance of $\bar{\bar{s}}_{0.999}$ makes the mean square error of this estimate as large as that of $s_{0.999}$ at $m = 10,000$. Clearly, jackknifing is of greater utility in the range of m from about 1,000 to 5,000, if the variance does not blow up and moreover, it is desirable to use sections as short as this if possible, to reduce computation time.

Estimates of the variance of $\bar{\bar{s}}_{0.999}$ obtained by synthetic sampling are given in column 5 of Table 2. The quantities in parentheses are estimates (63 degrees of freedom) of the standard deviation of these variance estimates. There is enough increase in the variance over the unjackknifed situation to characterize the gain from using the jackknife as being marginal rather than categorical with this type of quantile estimate. For the exponential case, however, detailed calculations on computation times, bias, etc., show that in large simulations there would be an advantage in using $\bar{\bar{s}}_{0.999}$ with sections of length approximately $m = 5,000$.

An alternative scheme is to use as a "typical value" an order statistic estimate, s^*_α, from a small <u>independent</u> sample of size m/ℓ. Then the estimate $(s^*_\alpha - \ell\,\bar{s}_\alpha)/(1-\ell)$ has a mean value free of the $1/m$ term and a variance decreasing to var $\{\bar{s}_\alpha\}$ as ℓ increases. For $\ell = 10$ the variance is 1.36 var $\{\bar{s}_\alpha\}$; for $\ell = 20$ it is 1.13 var $\{\bar{s}_\alpha\}$, an increase smaller than that found for $\bar{\bar{s}}_{0.999}$. However, this technique is difficult to use for an m much less than 10,000 in estimating a 0.001 quantile.

IV. QUANTILE ESTIMATION - Finite Sample Considerations for $\hat{s}_\alpha$

Experience shows that for extreme quantiles, convergence of $\hat{s}_\alpha(m)$ is so slow as to be unacceptable. Though problems might be anticipated in the tails of extremely skew distributions, e.g., the distribution $F(s) = s/(1+s)$ discussed above, they occur elsewhere too, as can be seen roughly in the following example.

Example - The unit exponential distribution

Consider again the unit exponential and the $\alpha = 0.001$ quantile. This is

$$s_{0.001} = -\ln(1 - 0.001) \approx 0.001 .$$

We have

$$f(s_{0.001}) = e^{-0.001} \approx 1 .$$

Assume the initial estimate is $\hat{s}_{0.001}(0) = 0.001$, and that S_1 is less than 0.001. This has probability 0.001. Then

$$\hat{s}_{0.001}(1) \approx -1 \times 0.999 \approx -1 .$$

Clearly successive S_i's are greater than the estimates until the estimates get back to zero. After ℓ steps the estimate will have moved $0.001 \times 1 \times \sum_{i=2}^{\ell} 1/i$ in the positive direction and since

$$\sum_{i=1}^{\ell} \frac{1}{i} \sim 0.5772 + \ln(\ell) + \frac{1}{2\ell} ,$$

the return to the origin takes about $\ell = 10^{400}$ steps. By following through this example it can be seen that the estimate is almost sure to become negative and take an enormously long time to return to zero.

The jackknife technique is not suitable as a means of overcoming this difficulty, partly because the order of the leading term in the

asymptotic expansion for the bias is not known, but also because if the correct jackknifing technique was applied (possibly, for a $m^{-1/2}$ term) the example makes it clear that the estimator would still be unsuitable.

Two other techniques suggest themselves. Kesten [1958] seems to be one of the few authors to have noted the problem of long runs occurring in the stochastic approximation scheme. He suggested higher-order memory schemes, for example, not increasing the divisor of C in (2.2) by one if all ρ previous differences of sample values and estimators were of the same sign. Although this would clearly help in the example, it violates the need for simplicity. In fact, the divisors in the estimators of different quantiles for different statistics become different, and this creates very complex and time consuming programming procedures.

Another technique is to bound the estimator $\hat{s}_\alpha(i)$. Thus requiring $\hat{s}_\alpha(i) \geq 0$ in the example would have obviated the problem, but uses specific information about the statistics. Empirically derived bounds can be obtained at the same time that initial values $\hat{s}_\alpha(0)$ and $f(s_\alpha)$ are obtained. For example a small pilot run using order statistics will give these initial estimates and bounds for outer quantiles (.001 and .999) from the known properties of the order statistics. Inner quantiles are bounded by outer quantiles. Thus it has been found empirically that if $\Delta = |\text{bound} - \bar{s}_\alpha(0)|$, then when $C/\Delta > 100$, the stochastic convergence scheme works reasonably well. It is, however, ponderous when compared to the quantile estimation scheme discussed in the following section.

V. THE MAXIMUM (OR MINIMUM) TRANSFORMATION

It is natural to look for a computationally simple transformation of the data to overcome some of the problems of estimating extreme quantiles, and the following transformation appears to solve most of the computational and statistical problems.

Assume, e.g., that $\alpha > 1/2$, and consider the first v S's, i.e., $S_1, \ldots, S_v$. Then the distribution $\overline{F}(s)$ of the maximum of the v S's is

$$\overline{F}(s) = \text{prob}\,\{(\max_{1 \le i \le v} S_i) \le s\} = \{F(s)\}^v . \tag{5.1}$$

Note that

$$\overline{F}(s_\alpha) = F^v(s_\alpha) = \alpha^v = \alpha' \tag{5.2}$$

and

$$s_\alpha = s_{\alpha'} , \tag{5.3}$$

so that the α-quantile of $F(s)$ is the same as the $\alpha' = \alpha^v$ quantiles of $\overline{F}(s)$. If we assume that v divides m, and $m' = m/v$, taking maxima of successive groups of v S's gives a reduced sample of size m', i.e.,

$$S'_1, S'_2, \ldots, S'_{m'} .$$

Thus we have a reduced sample and have transformed the problem from estimating an extreme quantile (for level α) to estimating a more reasonable quantile (for level α^v). For example, one might take v large enough to reduce the problem to estimating a median. Since $\alpha^v = 1/2$ in this case, $v = (\ln 1/2)/(\ln \alpha)$. The consequent values of v are shown for eight α's in Table 3.

Table 3

Size of v for $\alpha^v = 1/2$

α	0.900	0.950	0.975	0.980	0.990	0.995	0.998	0.999
v	6.6	13.5	27.4	34.3	69.0	138.3	346.2	692.8

Some general points about the use of the transformation in quantile estimation follow.

(1) The transformation of the original sample $\{S_i\}$ into the reduced sample $\{S_i'\}$ is very efficient computationally since only v binary comparisons and one memory cell are needed to obtain S_1' from $S_1, \ldots, S_v$.

(2) The transformation uses no special information about the properties of S, e.g., $S > 0$.

(3) Although transforming to the median is not necessarily optimal, it is known that stochastic approximation estimates of the median work well (Cochran and Davis, 1965).

(4) If $\alpha < 1/2$ the minimum S_1'' of the v S's is used instead.

VI. THE MAXIMUM TRANSFORMATION: ASYMPTOTIC VARIANCE

Using (2.4) we now compare the asymptotic variance of a quantile estimate (order statistic or stochastic approximation) $\tilde{s}_{\alpha'}$ from the reduced sample with that of the quantile estimate $\tilde{s}_\alpha$ from the original sample.

We have, from (5.1),

$$\bar{f}(s) = v\, f(s)\, \{F(s)\}^{v-1}$$

and, therefore, using (5.3)

$$\bar{f}(s_{\alpha'}) = v\, f(s_{\alpha'})\, \{F(s_{\alpha'})\}^{v-1}$$

$$= v\, f(s_{\alpha})\, \{F(s_{\alpha})\}^{v-1}$$

$$\bar{f}(s_{\alpha'}) = v\, f(s_{\alpha})\alpha^{v-1} = v\, f(s_{\alpha})\frac{\alpha'}{\alpha}\,. \tag{6.1}$$

Using these expressions in (2.4) we have

$$\operatorname{var}\{\tilde{s}_{\alpha'}\} \approx \frac{\alpha^{v}(1-\alpha^{v})}{\frac{m}{v}\,[\bar{f}(s_{\alpha'})]^{2}} \qquad (0.5 < \alpha < 1)$$

$$= \frac{\alpha(1-\alpha)}{m f^{2}(s_{\alpha})} \cdot \frac{(1-\alpha^{v})}{v(1-\alpha)\alpha^{v-1}} \tag{6.2}$$

$$= \operatorname{var}\{\tilde{s}_{\alpha}\} \cdot g(\alpha;v)\,. \tag{6.3}$$

It can be shown that $g(\alpha;v)$ increases from 1 to infinity as v increases from 1 to infinity. Moreover, for the median transformation, $v \approx (\ln 1/2)/(\ln \alpha)$, we have, as α approaches 1

$$g(\alpha;v) \approx \frac{1}{0.6932} \approx 1.443\,.$$

The function $g(\alpha;v)$ at the value of v generating the median transformation varies little with α, as is shown in Table 4.

Table 4

Variation of $g(\alpha;v)$ with α at median transformation

α	0.9000	0.9990	0.9999
v	7	693	6931
$g(\alpha;v)$	1.4020	1.4426	1.4426

Thus the statistical efficiency of the estimate is decreased by approximately 1/(1.443), but the speed of the maximum transformation, compared to the computations involved in, for example, the stochastic approximation estimate would probably make their computational efficiency about equal. By computational efficiency we mean the relative computing times required to achieve the same variance. Both estimates (reduced and non-reduced samples) are asymptotically unbiased.

The variation of $g(\alpha;v)$ with v for $\alpha = 0.999$ is shown in Table 5.

Table 5

Variation of $g(\alpha;v)$ with v at $\alpha = 0.999$

v	10	50	100	200	500	700	1000
α'	0.990	0.951	0.905	0.819	0.606	0.496	0.368
$g(0.999;v)$	1.005	1.025	1.051	1.107	1.297	1.448	1.718

The choice of v in particular situations depends on computational considerations, particularly the meshing of estimates for various α's for a given statistic, and although no global results can be given, we discuss these considerations in the next two sections.

VII. THE MAXIMUM TRANSFORMATION: COMPUTATIONAL CONSIDERATIONS

(i) Order statistics

The use of the maximum transformation with order statistic quantile estimates gives very little gain. Bias is not an extreme consideration here, and computations for several examples with the asymptotic ex-

pansion (3.5) show that there is very little change in small sample bias from one estimate to the other. Memory size is not much affected. For $\alpha = 0.999$ we require 0.001 m memory cells with the original sample; using $v = 693$ to give $\alpha' \approx 1/2$ and a reduced sample, we need $\frac{1}{2} m' = \frac{1}{2} \cdot \frac{m}{693}$ memory cells. The slight advantage is lost when it is noticed (6.2) that a larger sample size, m, is required to achieve equivalent variance.

If the eight quantiles with α's given in Table 3 are required, the unreduced sample ordering (done at one time for all eight quantiles) requires 0.1 m memory cells. The eight quantiles require eight separate reduced samples with memory requirements $\frac{1}{2}\frac{m}{693}, \frac{1}{2}\frac{m}{347}, \ldots, \frac{1}{2}\frac{m}{7}$, or a total roughly equivalent to 0.1 m.

The time gain from sorting smaller samples is again marginal.

(ii) Stochastic approximation

The greatest gain in using the maximum (or minimum) scheme comes from the reduction of bias with the stochastic approximation estimator (2.2). There are other gains; the maximum operation is faster than the the computation (2.2), and changing v and α' is very simple. One could do the eight quantiles of Table 3 from eight reduced samples obtained from v's of 7, 14, 28, 35, 70, 140, 350, 700. These values are close to the v's given in Table 3 and all divide 700. Thus it is easy to compute the eight reduced samples simultaneously in nested loops.

Again only a fixed number of memory locations is required.

A possible scheme for selecting initial values for the stochastic approximation estimator is the following. Let $S'_1, \ldots, S'_{m'}$ be the reduced sample. Take the first three values, S'_1, S'_2, S'_3 and order them as $S'_{(1)}, S'_{(2)}, S'_{(3)}$. If v is such as to make $\alpha' \approx 1/2$, use $S'_{(2)}$ as $\tilde{s}_{\alpha'}(0)$, the initial value, and since $S'_{(1)}$ and $S'_{(3)}$ estimate the $\frac{1}{4}$ and $\frac{3}{4}$ quantiles of $\bar{F}(s)$, estimate $f(s_{\alpha'})$ as:

$$\frac{1}{2}\left[\frac{1}{4\left(S'_{(2)}-S'_{(1)}\right)}+\frac{1}{4\left(S'_{(3)}-S'_{(2)}\right)}\right] . \tag{7.1}$$

(This is the simplest of many density estimates.)

The stochastic approximation estimate

$$\tilde{s}_{\alpha'}(i) = \tilde{s}_{\alpha'}(i-1) - \frac{C}{i}\left[\frac{1-\operatorname{sgn}\{S'_i - \tilde{s}_{\alpha'}(i-1)\}}{2} - \alpha'\right], \qquad i = 1,2,\ldots,m, \tag{7.2}$$

uses S'_4 as the S'_i, i=1, in the equation.

The value of C is not critical and this estimate should be well-behaved. The jackknife technique (3.8) can be applied, using alternate values $S'_{(i)}$ for the sub-estimates, although it is not known if the leading term in the bias is $O(1/m)$.

Analytical results for $\tilde{s}_{\alpha'}$ and the jackknifed estimate $\tilde{\tilde{s}}_{\alpha'}$, cannot be obtained and sampling results are given in Tables 6, 7 and 8 for S having a unit exponential distribution and S having the distribution $F(s) = s/(1+s)$.

(iii) Sampling results

Results are given in Tables 6, 7 and 8 from an extensive simulation to investigate the properties of quantile estimates based on (7. 1) and (7. 2). The columns in these tables show successively m, the sample size; m', the reduced sample size; the expected values and standard deviations of the jackknifed estimator $\tilde{\tilde{s}}_{\alpha'}$ based on (7. 1) and (7. 2) and the expected values and standard deviations of the straightforward estimator $\tilde{s}_{\alpha'}$ based on (7. 1) and (7 2). The last column in the Tables gives the asymptotic standard deviations from (6. 2). Note that $v = 700$. To indicate the precision obtained in the sampling experiment the quantities in parentheses in the last row of each table are estimates (7 degrees of freedom) of the standard deviations of the estimates.

In Table 6, the minimum transformation gives estimates of $s_{0.001} = 0.001001$, the 0.001 quantile of the unit exponential distribution. The jackknifed estimator $\tilde{\tilde{s}}_{\alpha'}$, where $\alpha' = 1 - [1-0.001]^{700}$, has converged by the time m reaches roughly 30,000. There is still a small bias in the unjackknifed estimator $\tilde{s}_{\alpha'}$, approximately one tenth of the standard deviation of the estimator at $m = 28,000$. There is a penalty paid in terms of a larger standard deviation for the jackknifed estimator ($\approx 20\%$) and the standard deviation for $\tilde{s}_{\alpha'}$ is larger than predicted by the asymptotic formula (6. 2). This is due to the added variability introduced by the density estimate (7. 1) and the finite sample size m.

TABLE 6

Estimation of 0.001 quantile with minimum transformation and stochastic approximation, with ($\tilde{\tilde{s}}_{\alpha'}$) and without ($\tilde{s}_{\alpha'}$) jackknife. $\alpha' = 1-[1-0.001]^v$, $v = 700$. Unit exponential distribution.

$$F(s) = 1 - e^{-s}; \qquad \bar{F}(s) = 1 - [1 - F(s)]^v; \qquad s_{0.001} = 0.001001$$

m	m'	$E(\tilde{\tilde{s}}_{\alpha'})$	st. dev. ($\tilde{\tilde{s}}_{\alpha'}$)	$E(\tilde{s}_{\alpha'})$	st. dev. ($\tilde{s}_{\alpha'}$)	Asymptotic st. dev.
4200	6	0.00107	0.00095	0.00111	0.00065	0.000587
5600	8	0.00106	0.00081	0.00109	0.00058	0.000509
7000	10	0.00104	0.00073	0.00108	0.00053	0.000455
8400	12	0.00103	0.00067	0.00107	0.00050	0.000415
9800	14	0.00103	0.00062	0.00106	0.00047	0.000385
11200	16	0.00102	0.00058	0.00106	0.00045	0.000360
12600	18	0.00102	0.00055	0.00105	0.00043	0.000339
14000	20	0.00102	0.00053	0.00105	0.00042	0.000322
16800	24	0.00101	0.00049	0.00104	0.00040	0.000294
19600	28	0.001008	0.000456	0.001034	0.000378	0.000272
28000	40	0.001003	0.000400	0.001025	0.000344	0.000228
42000	60	0.000999 (0.000001)	0.000346 (0.000001)	0.001016 (0.000001)	0.000310 (0.000002)	0.000186

TABLE 7

Estimation of 0.999 quantile with maximum transformation and stochastic approximation, with ($\tilde{\tilde{s}}_{\alpha'}$) and without ($\tilde{s}_{\alpha'}$) jackknife. $\alpha' = 0.999^{v}$, $v = 700$. Unit exponential distribution.

$F(s) = 1 - e^{-s}$; $\bar{F}(s) = [F(s)]^{v}$; $s_{0.999} = 6.908$

m	m'	$E(\tilde{\tilde{s}}_{\alpha'})$	st. dev. ($\tilde{\tilde{s}}_{\alpha'}$)	$E(\tilde{s}_{\alpha'})$	st. dev. ($\tilde{s}_{\alpha'}$)	Asymptotic st. dev.
4200	6	6.946	0.970	6.962	0.640	0.587
5600	8	6.942	0.825	6.957	0.569	0.508
7000	10	6.931	0.736	6.949	0.522	0.455
8400	12	6.929	0.670	6.946	0.488	0.415
9800	14	6.926	0.619	6.942	0.461	0.384
11200	16	6.923	0.581	6.940	0.440	0.359
12600	18	6.920	0.548	6.936	0.423	0.339
14000	20	6.919	0.520	6.934	0.406	0.321
16800	24	6.919	0.481	6.933	0.386	0.293
19600	28	6.916	0.448	6.929	0.365	0.272
28000	40	6.912	0.384	6.923	0.326	0.227
42000	60	6.912 (0.001)	0.332 (0.001)	6.920 (0.001)	0.294 (0.001)	0.186

TABLE 8

Estimation of 0.999 quantile with maximum transformation and stochastic approximation, with $(\tilde{\tilde{s}}_{\alpha'})$ and without $(\tilde{s}_{\alpha'})$ jackknife. $\alpha' = 0.999^{v}$, $v = 700$.

$F(s) = s/(1+s)$; $\bar{F}(s) = [s/(1+s)]^{v}$; $s_{0.999} = 999$

m	m'	$E(\tilde{\tilde{s}}_{\alpha'})$	st. dev. $(\tilde{\tilde{s}}_{\alpha'})$	$E(\tilde{s}_{\alpha'})$	st. dev. $(\tilde{s}_{\alpha'})$	Asymptotic st. dev.
4200	6	1048	2902	1233	1375	587
5600	8	1196	1729	1217	1233	508
7000	10	1076	2416	1197	1195	455
8400	12	1126	1412	1180	1042	415
9800	14	1063	1422	1163	872	384
11200	16	1076	1078	1148	802	359
12600	18	1052	1082	1137	738	339
14000	20	1060	1123	1127	753	321
16800	24	1048	872	1112	685	293
19600	28	1042	828	1103	698	272
28000	40	1025	747	1077	624	227
42000	60	1016 (1)	551 (23)	1055 (1)	494 (28)	186

In Table 7 the maximum transformation gives estimates of $s_{0.999} = 6.908$, the 0.999 quantile of the unit exponential distribution. The bias is smaller for the jackknifed estimator, and extrapolation of figures in column 3 gives convergence in the mean for most purposes by $m = 50,000$. Note, however, that the standard deviation/bias ratio is large (~ 100) down the column. Again, there is an inflation in the standard deviation of the jackknifed estimator.

The bias is examined more closely in Figure 1 where the absolute value of the (estimated) bias is plotted on a log scale against m for both estimators. The curvature indicates that higher order terms than $1/m$ are still important for these sample sizes. Except for the $m = 42,000$ point, the bias in the jackknifed estimator appears to be falling off more rapidly than the bias for $\tilde{s}_{\alpha'}$ (here $\alpha' = 0.999^{700}$). No formal regression analyses of these sampling results have been done.

Note that for both estimators the absolute value of the bias (except for the last point) is less than the absolute value of the bias in the order statistic estimator. This is given by (3.1) and plotted as a sequence of crosses in Figure 1.

Bias is more serious for the extreme case of the distribution $F(s) = s/(1+s)$, as shown in Table 8. The jackknifed estimator is advantageous here where the standard deviation/bias ratio is smaller than for the exponential distribution and convergence is a little slower.

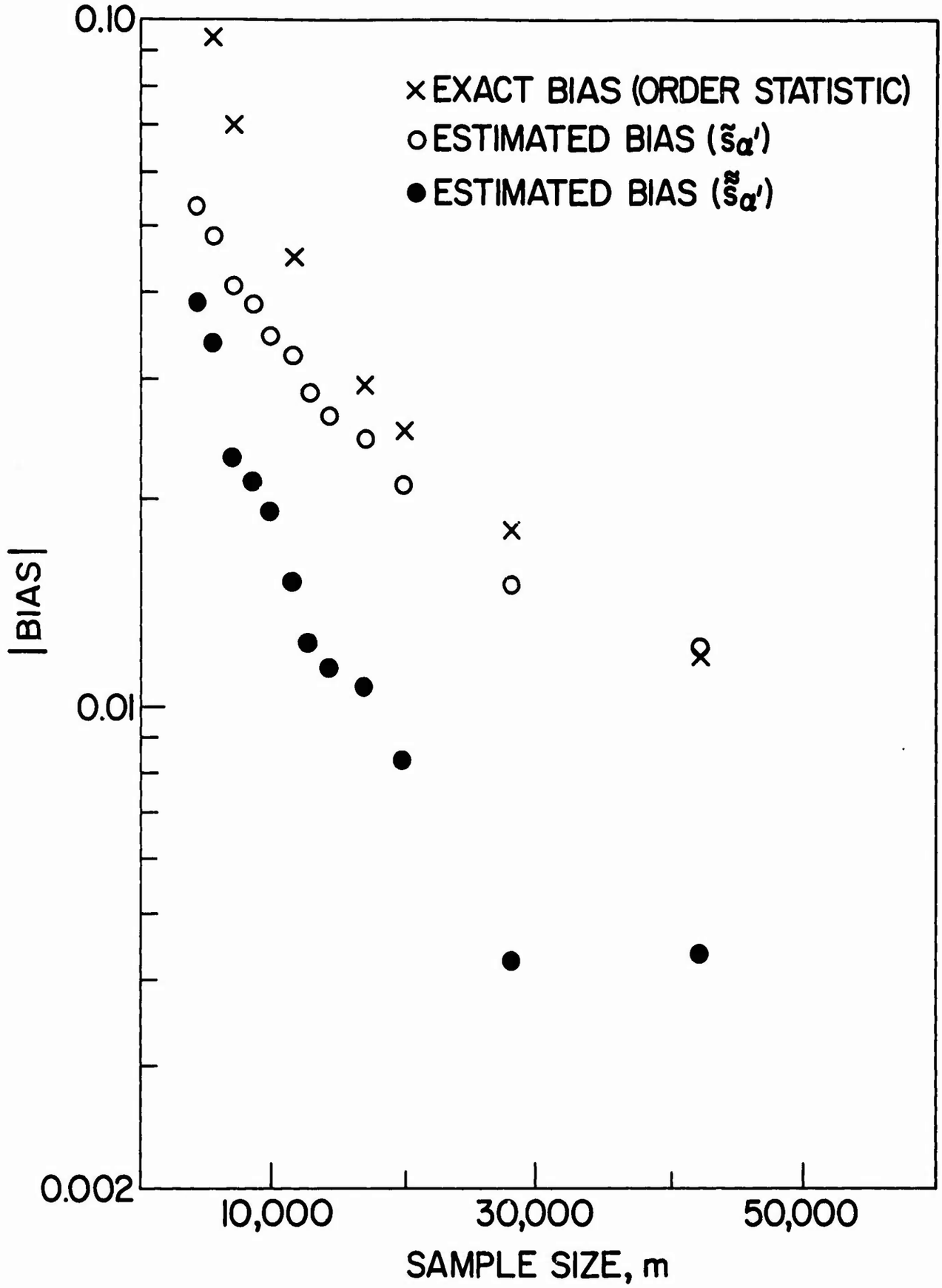

Figure 1. Exact bias for the order statistic estimator $\bar{s}_{\alpha'}$ of the 0.999 quantile of the unit exponential distribution, and estimated bias for estimators $\tilde{s}_{\alpha'}$ and $\tilde{\tilde{s}}_{\alpha'}$ using the maximum transformation ($\alpha' = 0.999^{v}$, $v = 700$) and stochastic approximation.

VIII. CONCLUSIONS AND FURTHER WORK

The conclusion of this investigation is that the maximum transformation and the stochastic approximation scheme (7. 2) yields a quantile estimator for extreme quantiles (e. g., $x_{0.999}$) which is fast and linear in sample size m; requires a small, fixed amount of storage, and without using external information provides a virtually bias free estimate even in extreme cases (such as the $s/(1+s)$ distribution) for sample sizes of $m = 50,000$. In most cases smaller samples will be satisfactory.

There are possibilities, based on the properties of extreme value distributions, of improving the quantile estimators performance even more.

Note too that the estimator can be used to advantage for smaller α than 0. 999, and it is completely suited for use with global variance reduction techniques such as the use of control variables (Gaver, 1969).

REFERENCES

Clark, C. E. and Williams, G. T. (1958), "Distributions of the members of an ordered sample," Annals of Math. Stat., 29, 862-870.

Cochran, W. G. and Davis, M. (1965), "The Robbins-Monro method for estimating the median lethal dose," J. Roy. Statist. Soc. B, 27, 28-44.

Cox, D. R. and Lewis, P. A. W. (1966), The statistical analysis of series of events. Methuen: London and Barnes and Noble: New York.

David, F. N. and Johnson, N. L. (1954), "Statistical treatment of censored data, I. Fundamental formulae," Biometrika, 41, 228-240.

David, H. A. (1971), Order statistics. Wiley: New York.

Gaver, D. P. (1969), "Statistical methods for improving simulation efficiency," Proceedings of Third Annual Conference on Applications of Simulation: Los Angeles, December 1969.

Goodman, A. S. and Lewis, P. A. W. (1972), "The null distributions of test statistics based on sample serial correlations and the periodogram for renewal processes with exponential and 1/2 Weibull intervals," IBM Research Report, to appear.

Hodges, J. L. and Lehman, E. L. (1956), "Two approximations to the Robbins-Monro process," Proceedings Third Berkeley Symposium on Mathematical Statistics and Probability, 1, 95-104.

Kesten, H. (1958), "Accelerated stochastic approximation," Annals of Math. Stat., 29, 41-49.

Mosteller, F. and Tukey, J. W. (1968), "Data analysis, including statistics," Handbook of Social Psychology, Vol. II, second edition (edited by G. Lindzey and E. Aronson), Addision-Wesley: New York, 80-203.

Miller, R. G. (1964), "A trusty jackknife," Annals of Math. Stat., 35, 1594-1605.

Quenouille, M. H. (1956), "Notes on bias estimation," Biometrika, 43, 353-360.

Robbins, H. E. and Monro, S. (1951), "A stochastic approximation method," Annals of Math. Stat., 22, 400-407.