NON-PARAMETRIC QUANTILE ESTIMATION THROUGH STOCHASTIC APPROXIMATION

David Walter Robinson

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

Non-parametric Quantile Estimation
Through
Stochastic Approximation

by

David Walter Robinson

June 1975

Thesis Advisor: Peter A.W. Lewis

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Non-parametric Quantile Estimation Through Stochastic Approximation

5. TYPE OF REPORT & PERIOD COVERED
Ph.D. Thesis; June 1975

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
David Walter Robinson

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

12. REPORT DATE
June 1975

13. NUMBER OF PAGES
231

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
Naval Postgraduate School
Monterey, California 93940

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Statistical Estimators; Stochastic Approximation; Quantiles; Robbins-Monro Process; Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The extreme values which a random variable X may take on are usually best characterized by the quantiles of the random variable. Known non-parametric methods for the statistical estimation of extreme quantiles all suffer from serious shortcomings, however. In this thesis a robust and efficient method for quantile estimation is described; both

Block 20. (Continued)

the asymptotic and finite sample properties of the estimator are determined and computer implementations are given. Possible applications for the technique include the analysis of computer simulations and data analysis in large data bases or real time computer systems.

Non-parametric Quantile Estimation Through Stochastic Approximation

by

David Walter Robinson
Lieutenant, United States Navy
B.S., United States Naval Academy, 1965
M.S., Naval Postgraduate School, 1973

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

from the
NAVAL POSTGRADUATE SCHOOL
June 1975

ABSTRACT

The extreme values which a random variable X may take on are usually best characterized by the quantiles of the random variable. Known non-parametric methods for the statistical estimation of extreme quantiles all suffer from serious shortcomings, however. In this thesis a robust and efficient method for quantile estimation is described; both the asymptotic and finite sample properties of the estimator are determined and computer implementations are given. Possible applications for the technique include the analysis of computer simulations and data analysis in large data bases or real time computer systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Peter Lewis for his guidance and assistance during the course of this research. My wife Kathleen is also due a share of the credit for this thesis; without her understanding and support the work would never have been completed. Finally I would like to thank the Lord for so many answered prayers.

Chapter I. INTRODUCTION

A. Description of the Problem

The problem addressed in this thesis is the non-parametric estimation of population _quantiles_. Given a random variable X with continuous distribution function $F(\cdot)$, we define the a-quantile s_a as the solution to the equation

$$(1) \qquad F(s_a) = a$$

for some given value of a between 0 and 1. We shall assume in what follows that s_a is unique, i.e. that we are dealing with continuous or partly continous distributions. Completely discrete distributions with relatively small numbers of atoms present a much simpler estimation problem. Quantiles find application, for example, in testing statistical hypotheses and in characterizing the extreme values of the distribution of X when a is near 0 or 1.

At the outset we note that there is a related problem, namely, given a value s, to estimate the quantity p_s given by

$$(2) \qquad F(s) = p_s .$$

The value p_s found in this way will be called a _percentile_. Percentiles may be used, for example, to find the power of a statistical test under a non-null hypothesis. By way of

contrast we note that a is the known value in (1) while s is the known in (2).

The non-parametric estimation of percentiles is relatively straightforward; the number of values of the random variable less than s in a random sample $X_1, X_2, \ldots, X_n$ is clearly a binomial random variable with parameters n and p_s so that this number divided by n is an unbiased estimator of p_s.

If the distribution function $F(\bullet)$ in (1) is completely known, finding s_a becomes a problem of numerical approximation, i.e. one must evaluate

(3) $$s_a = F^{-1}(a).$$

Note that if the random variable X has an infinite support the slope of $F(\bullet)$ will be very small in one or both tails of the distribution (i.e. as the quantile level a approaches 0 or 1); this means that in evaluating (3) for extreme quantiles one is likely to encounter serious numerical instabilities. If the distribution function $F(\bullet; \theta)$ is known except for a finite vector θ of unknown parameters we may still proceed as in (3) provided we have some estimate $\tilde{\theta}$ of the parameters. The resulting <u>parametric</u> estimate of s_a is given by

(4) $$\tilde{s}_a = F^{-1}(a; \tilde{\theta}).$$

The properties of $\tilde{s}_a$ will depend on both the underlying

distribution $F(\bullet)$ and the nature of the estimate θ; the sampling variation of θ, however, is likely to increase the numerical difficulties with extreme quantiles.

If nothing is known about $F(\bullet)$, one must resort to non-parametric or distribution-free methods for estimating s_a. Non-parametric quantile estimation is considerably more complex than non-parametric percentile estimation. Two solutions have been proposed for this problem (Goodman, Lewis and Robbins [14]): the order statistic estimator, $\hat{s}_a$, and a class of stochastic approximation estimators, $\bar{s}_a$.

The order statistic estimator is obtained by sorting the random sample $X_1, X_2, \ldots, X_n$ into order, thus determining the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Then the estimator is

(5) $$\hat{s}_a = X_{([a(n+1)])},$$

where $[z]$ denotes the integer part of z. It is known (David [5]) that $\hat{s}_a$ has an asymptotically normal distribution with

(6) $$E[\hat{s}_a] = s_a + O(1/n)$$

and

$$(7) \qquad \mathrm{Var}[\hat{s}_a] = \frac{a(1-a)}{nf^2(s_a)} + O(n^{-2}) ,$$

where $f(x) \equiv F'(x)$ is the density function of the random variable X. Unfortunately, the time required to order a complete sample of size n is proportional to n ln n; thus the computational effort for this estimator increases faster than the sample size. Furthermore, considerations of finite computer memory size limit order statistic estimators to samples of perhaps 10,000 observations (less if several distributions must be investigated at once as might be the case in a systems simulation study). We discuss some other considerations relating to order statistic estimators in Chapter III; because partial sorting can be done in time proportional to n some improvement is possible, but these estimators still suffer from serious shortcomings.

To overcome these drawbacks, we consider a sequential estimation scheme. This may be defined by a sequence of functions $\{h_n\}$; our estimates are given recursively by

$$(8) \qquad \bar{s}_a(j+1) = h_j(\bar{s}_a(j), X_{j+1}), \qquad j=1,\ldots,n-1,$$

where $\bar{s}_a(j)$ is the estimator at step j of the procedure. In the sequel, we denote this j-th sequential estimator by $\bar{s}_j$, suppressing the dependence on a when this will cause no confusion.

B. Stochastic Approximation Estimators

The most important class of functions to be used in sequential quantile estimation schemes are stochastic approximation estimators. There is an extensive literature on so-called stochastic approximation methods; these methods are intended to find the root $x = \theta$ of the regression function

$$(9) \qquad E[Y(x)] = M(x) = a,$$

where the only information available consists of independent observations on the random variable $Y(x)$. We note that this is a more general problem than the quantile estimation problem considered here. Most work on stochastic approximation has been concerned with specifying conditions under which the sequence of estimators converges probabilistically to the correct value. Many of these conditions are trivially satisfied in the quantile estimation case; for example, the regression function will always be bounded since it is a distribution function, $F(x)$.

The simplest type of stochastic approximation quantile estimators are based on the work of Robbins and Monro [30]. They are defined by the relationship

$$(10) \qquad \bar{s}_{n+1} = \bar{s}_n - a_n Y(\bar{s}_n), \qquad n=1,2,\ldots \quad .$$

In this formulation $\{a_n\}$ is a sequence of positive constants of the form

$$(11) \qquad a_n = \frac{1}{nA}, \qquad A > 0,$$

and $Y_n(\bar{s}_n)$ is a random variable which depends only on X_n and $\bar{s}_n$ and which is defined by

$$(12) \qquad Y_n(\bar{s}_n) = \begin{cases} -a & \text{if } X_n > \bar{s}_n \\ 1-a & \text{if } X_n \leq \bar{s}_n \end{cases}.$$

The initial estimate $\bar{s}_1$ and the parameter A may be chosen arbitrarily or at random.

The procedure given by (10) is called a Robbins-Monro (RM) process; under suitable conditions (which are satisfied by (10)-(12) as long as $Var[\bar{s}_1] < \infty$), Blum [2] and Dvoretzky [7] have shown

$$(13) \qquad \bar{s}_n \rightarrow s_a \quad \text{almost surely (a.s.)},$$

$$(14) \qquad \lim_{n\to\infty} E[(\bar{s}_n - s_a)^2] = 0.$$

Furthermore, Sacks [33] has shown that if F(x) has a continuous derivative f(x) at s_a then

$$(15) \qquad \bar{s}_n \xrightarrow{L} N\left(s_a, \frac{a(1-a)}{nA(2f(s_a) - A)}\right),$$

as long as $0 < A < 2f(s_a)$. The asymptotic variance is minimized by taking $A = f(s_a)$; this results in the same

asymptotic normal distribution for $\bar{s}_n$ as for the order statistic estimator, $\hat{s}_a$.

C. Improving the RM Estimators

An intuitive discussion of the operation of the RM process (10) will serve to point out ways in which the resulting quantile estimators can be improved. First, we note that the sequence $\{\bar{s}_n\}$ is a Markov process, although a non-homogeneous one. Moreover, as long as A is fixed, $\bar{s}_n$ may take on one of only 2^n distinct values at stage n. This is because Y_n is a discrete random variable: it increases the estimate value ("step up") when the latest observation is larger than the current estimate and decreases the value ("step down") when the observation is smaller.

The actual magnitude of the step is governed by the gain sequence $\{a_n\}$. The factor 1/n in (11) is necessary so that successive steps become smaller, thus allowing the estimator to converge; however, since $\sum_{n=1}^{\infty} (1/n) = \infty$ the sequence of estimators can reach any quantile value s_a starting from an arbitrary initial value $\bar{s}_1$. Note however that if $\bar{s}_n$ is still far from s_a for even moderately large n,

a prohibitive number of steps may be needed to obtain a reasonable estimate.

The first improvement to the basic RM process was suggested by Kesten [18]. To cut down the number of steps required to converge to the true value after the difference $\bar{s}_n - s_a$ becomes large, the divisor n in (11) is modified so that it is increased only when the current step direction differs from the step taken at the previous stage. This suggests that we have "straddled" the true quantile value. Although the stochastic approximation estimator obtained in this way has the same asymptotic distribution as the RM estimator (Davis [6]), its convergence properties in small samples seem to be superior (Cochran and Davis [4]; Davis [6]). The Kesten procedure does have the disadvantage, however, that it often fails to reduce the step size even when $\bar{s}_n$ is close to s_a. The optimum procedure is probably to keep the step size constant until $\bar{s}_n$ is "close" to s_a and then to carry out the usual RM procedure. Such a "delayed" process has been studied by Cochran and Davis [4] and Davis [6].

A related difficulty with the basic RM process is that it does not work well at all for the estimation of even moderately extreme quantiles ($a < 0.25$ or $a > 0.75$). This problem was first noted by Wetherill [36]; he traced the difficulty to the slow rate of increase of the harmonic series $\sum_{n=k} (1/n)$ when $k >> 1$.

A solution to this problem was developed by Goodman, Lewis and Robbins [14]. Instead of carrying out the operation (10) for every sample value X_n we use only the maximum (or minimum for $a < 0.5$) of some number of observations, say v, where v is chosen so that

$$(16) \qquad a^v \equiv a' \doteq 0.5 \quad .$$

The RM process can then be applied to estimate the a'-quantile of the maxima (or minima); this has the same value as the a-quantile of X. The basic idea is to use a data transformation to shift the problem to the estimation of a population median, for which RM is known to be well-behaved. It is unnecessary to go all the way to the median; good results are obtained for $0.3 < a' < 0.7$. Convergence rates are apparently much improved by this procedure; the cost, as Goodman, Lewis and Robbins [14] show, is an inflation of the asymptotic variance

$$(17) \qquad \mathrm{Var}[\bar{s}'_n] = \mathrm{Var}[\bar{s}_n] \frac{a(1-a')}{va'(1-a)} \quad .$$

In most cases the inflation is less than 40 %.

A natural extension of this so-called maximum transformation process is to consider a next-to-maximum transformation, i.e. applying the RM process (10) to the second largest (or smallest) in a sample of size w where

$$(18) \qquad wa^{w-1} - (w-1)a^w \equiv a'' \doteq 0.5 \quad .$$

The appeal of this procedure in dealing with highly skewed real world data is that it may give a more robust estimation procedure. Once again, there is an inflation of the asymptotic variance

$$\text{(19)} \qquad \mathrm{Var}[\bar{s}''_n] = \mathrm{Var}[\bar{s}_n] \, \frac{a''(1 - a'')}{w(w-1)^2 \, a^{2w-3} \, (1-a)^3} \, .$$

The inflation is somewhat greater in this case than for the maximum transform but it may still be limited to less than 50 % by the proper choice of w.

In the remainder of this thesis, a single prime (as in a' or $\bar{s}'_n$) will denote an estimate or parameter which is based on the maximum transform while the double prime (e.g., $\bar{s}''_n$) will denote a next-to-maximum transformed value. Except for equations (17) and (19), a subscript n appended to a primed value will indicate the number of <u>steps</u> taken by the corresponding stochastic approximation process and not the X sample size, which will be larger. In fact, we will need at least n•v X observations to obtain $\bar{s}'_n$; more will be needed if the initial estimate $\bar{s}_1$ is chosen at random.

For efficient estimation of a set of several quantiles we prefer to use v (or w) values for higher quantiles which are integral multiples of the values for lower quantiles; this greatly simplifies determination of sample maxima and minima. In this research, a set of 19 quantiles has been arbitrarily selected; these include the 16 quantiles of Goodman, Lewis and Robbins [14] together with the median (a = 0.5) and the quartiles (a = 0.25, 0.75). The values of v and w for each of the transformation schemes together with the respective variance inflation factors are shown in Table I.

Having dealt with the effects of the 1/n term in the gain sequence $\{a_n\}$ we now consider the parameter A. The $O(n^{-1})$ variance implied by (15) will result when A is not too large, i.e. when the initial step size is not too small. It is known (Major and Revesz [26]) that the order of

a	v	a'	V'	w	a"	V"
.001	672	.4895	1.425	1536	.4542	1.476
.002	336	.4897	1.425	768	.4543	1.476
.005	112	.4296	1.338	384	.5726	1.608
.010	56	.4304	1.336	192	.5732	1.608
.020	28	.4320	1.331	96	.5745	1.606
.025	28	.5078	1.437	48	.3383	1.423
.050	14	.5123	1.426	24	.3392	1.420
.100	7	.5217	1.402	12	.3410	1.414
.250	1	.2500	1.000	6	.4661	1.414
.500	1	.5000	1.000	3	.5000	1.333
.750	1	.7500	1.000	6	.5339	1.414
.900	7	.4783	1.402	12	.6590	1.414
.950	14	.4877	1.426	24	.6608	1.420
.975	28	.4922	1.437	48	.6617	1.423
.980	28	.5680	1.331	96	.4255	1.606
.990	56	.5696	1.336	192	.4268	1.608
.995	112	.5704	1.338	384	.4274	1.608
.998	336	.5103	1.425	768	.5457	1.476
.999	672	.5105	1.425	1536	.5458	1.476

Table I. Sample sizes, transformed levels and variance inflation factors for maximum transformation (v, a' and V') and next-to-maximum transformation (w, a" and V") stochastic approximation quantile estimation designs.

convergence may be substantially worse when $A \geq 2f(s_a)$. When the optimum value $A = f(s_a)$ is chosen, the RM process acts like steepest descent approximation with small steps; the steps are the same as those for a linear approximation to the distribution function through the point (s_a, a) (Fabian [10]).

Evidently the initial choice of A has an important influence on the efficiency of the basic RM process, but in general the magnitude of the effect cannot be determined since $f(s_a)$ is unknown. In fact, the asymptotic normality of $\bar{s}_n$ stated by (15) cannot even be asserted since it will not be known whether $A < 2f(s_a)$. _For this reason, we consider procedures which simultaneously estimate s_a and $f(s_a)$ and are thus more generally applicable._

Practical application of stochastic approximation quantile estimation then requires that we have both a starting value $\bar{s}_1$ and an estimate of $f(s_a)$. Although there is an improvement over order statistic estimators in both speed and memory, the additional values required in the stochastic approximation case introduce a degree of complexity. In fact the selection of these two values is critical to the feasibility of stochastic approximation quantile estimation and is one of the main problems addressed and solved in this thesis.

D. Venter's Method and Confidence Intervals

The first method for simultaneously estimating s_a and $f(s_a)$ is due to Venter [37]. Note that although this solves the problem of finding a suitable A value we must still select an initial estimate $\bar{s}_1$; this is not nearly as crucial or as difficult as the choice of A. In Venter's method we observe two X values at each stage of the procedure and determine

$$(20) \qquad Y'_n = \begin{cases} -a & \text{if } X_{2n-1} > \bar{s}_n + c_n \\ 1-a & \text{if } X_{2n-1} \leq \bar{s}_n + c_n \end{cases}$$

and

$$(21) \qquad Y''_n = \begin{cases} -a & \text{if } X_{2n} > \bar{s}_n - c_n \\ 1-a & \text{if } X_{2n} \leq \bar{s}_n - c_n \,. \end{cases}$$

The sequence $\{c_n\}$ is a sequence of positive constants called the finite difference sequence; it must satisfy

$$(22) \qquad c_n n^r \text{ --> } c, \qquad c > 0, \quad 0.25 < r < 0.50 \,.$$

A sequential estimator of $f(s_a)$ is then given by

$$(23) \qquad A_n = \frac{1}{n} \sum_{j=1}^{n} \frac{Y'_j - Y''_j}{2c_j} \,.$$

Finally to estimate s_a we apply the basic RM recursion relation (10) with

$$(24) \qquad Y_n = (Y'_n + Y''_n) / 2 \quad .$$

The latest estimate A_n of $f(s_a)$ is used in the gain sequence in the place of the arbitrary value A, i.e. we use the random value $1/(nA_n)$ for a_n in (10). In a practical application of the method to quantile estimation, we accumulate only the sum in (23) thus obtaining nA_n; this quantity is used directly as the denominator of the gain sequence (11).

The chief practical difficulty encountered in using the estimator (23) is that A_n may become negative, in which case the RM process will take steps in the wrong direction, or else A_n may get too large in which case the $O(n^{-1})$ variance will be lost. For this reason, Venter uses as an estimate of $f(s_a)$ in the gain sequence the value A^*_n, where

$$(25) \qquad A^*_n = \begin{cases} a^* & \text{if } A_n < a^* \\ A_n & \text{if } a^* \leq A_n \leq b^* \\ b^* & \text{if } A_n > b^* \end{cases}$$

and where it is known a priori that $a^* < f(s_a) < b^*$. As long as b^* is not too large, we have (Venter [37]):

(26) $\bar{s}_n \to s_a$ a.s.,

(27) $A_n \to f(s_a)$ a.s.,

and

(28) $\bar{s}_n \xrightarrow{L} N\left(s_a, \frac{a(1-a)}{2nf^2(s_a)}\right)$.

Thus, the Venter estimator has the same asymptotically normal distribution as the other stochastic approximation estimators we have considered. (Recall that $\bar{s}_n$ is based on a total X sample of size 2n in this case.)

The advantage of the Venter procedure is that we no longer need an independent initial estimate of $f(s_a)$ since the procedure converges for any initial value of $f(s_a)$ in the interval (a^*, b^*). We also obtain (asymptotically) the minimum possible variance and we have the additional estimate A_n which may be used to determine a confidence interval on s_a. Sielken ([34] and [35]) has investigated the application of the Venter process to the estimation of confidence intervals and stopping times.

The problem of finding the interval (a^*, b^*) was solved by Fabian [9]; he suggested the use of

$$(29)\qquad a^* = C_1 n^{-L}, \quad 0 < L < 1/2,$$

$$b^* = C_2 \log(n+1),$$

$$0 < C_1 < C_2 .$$

From a practical point of view, we may establish the lower bound by setting nA_n to some small positive constant whenever the accumulated sum becomes negative. Venter's results also indicate that the upper bound b^* may be arbitrarily large when the density function is analytic in some neighborhood of s_a, so that this does not represent a restriction in many applications.

E. A New Method

A modification of the basic RM stochastic approximation process along the lines of Venter's work is the major contribution of this thesis. The new process is asymptotically equivalent to the other processes discussed in this Chapter but its finite sample properties seem to be much better. Just as in the case of the Venter process, we obtain an estimate of $f(s_a)$ which is plugged recursively back into the basic stochastic approximation relation; a

different technique for density estimation is employed, however.

In seeking an estimate of an unknown density function at some point one is lead to the work of Rosenblatt [32] and Parzen [28] on kernel estimators. A kernel function $W(\bullet)$ is a bounded integrable function with

$$(30)\qquad \int W(x)\,dx = 1 \quad .$$

An example is the triangular weight function

$$(31)\qquad W(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The empirical density function estimator at the point x is then given by

$$(32)\qquad f_n(x) = \frac{1}{nb_n} \sum_{j=1}^{n} W\left[\frac{x - X_j}{b_n}\right] \quad ,$$

where $\{b_n\}$ (called the "bandwidth" sequence) is a sequence of positive constants which tends to zero with increasing n; for example,

$$(33)\qquad b_n = b\,n^{-1/3} \,, \qquad b > 0 \quad .$$

We now define an estimator B_n of $f(s_a)$ using a kernel density estimator:

$$(34)\qquad B_n = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{b_j} W\left[\frac{\bar{S}_j - X_j}{b_j}\right]$$

and establish a new stochastic approximation process which uses the RM recursion formula (10) with B_n replacing A in

the gain sequence (11).

One advantage of the new density estimator (34) is that we are able to take twice as many steps as in Venter's method for the same sample size; this seems to permit faster convergence in small samples. Some computational experience with the new estimators shows them to be far superior to any other known non-parametric technique for quantile estimation. Almost sure convergence and asymptotic normality for the new procedure are established in Chapter II.

F. Scope of Research

The goal of this thesis is to investigate the application of the stochastic approximation techniques described in this Chapter to the problem of non-parametric quantile estimation in the hope of developing a practical method which is fairly robust with respect to the underlying distribution F(•). The chief disadvantage in using any stochastic approximation estimator - including Venter's procedure as well as the basic RM process - seems to be that in some cases the estimators are nowhere near s_a, even after as many as 20,000 steps. It is in this case that the RM process (10) has the worst convergence rate because reaching the immediate neighborhood of the true value may require an astronomical number of additional steps. Unless this unfortunate tendency can be overcome, stochastic approximation estimators cannot be recommended in practical applications.

Encouraging results have been achieved with the new estimator proposed here, particularly when it is combined with the maximum transformation technique and when some care

is taken in selecting the starting value, $\bar{s}_1$. When an entire set of quantiles is to be estimated a further improvement is possible. Since the quantiles are by definition ordered, a gross error in a single estimate can often be detected because the erroneous value is usually out of order with respect to the other estimates in the set. In this case alternate types of estimate can be used to replace the erroneous one, thus bypassing the lengthy path that the stochastic approximation process requires to reach the true quantile value. Assuming that only one or two of the set of estimates is in error, this approach should overcome the tendency of the stochastic approximation process to "blow up".

The thesis is organized as follows: in Chapter II, we establish the asymptotic properties of the new estimator and show it to be equivalent to the Venter process as $n \rightarrow \infty$. Chapter III describes some practical considerations relating to quantile estimation in finite samples of data using both order statistic and stochastic approximation estimators, while Chapter IV describes the results of an extensive digital computer simulation undertaken to determine the bias properties of the new estimator. Chapter V discusses the simultaneous estimation of an entire set of population quantiles and considers several techniques such as James-Stein estimation and isotonic regression to exploit the order relationships which are known to exist in such a set of estimates. Chapter VI discusses the estimation of functions of quantiles, in particular the estimation of the level of a test based on a given statistic and the estimation with the <u>same</u> simulation data of the power of the test. The last Chapter summarizes the work and discusses possible applications for the methods developed.

In summary, this thesis describes a method for estimating an entire set of quantiles with their corresponding densities for any statistic or other random quantity. The method is quite fast and uses a small fixed amount of memory; it is robust enough to be used as a basic building block in computer simulation programs.

G. Limitations of Research

In this thesis we deal only with non-parametric quantile estimators; substantial improvements are often possible if we know enough about the underlying distribution function $F(\bullet)$ to apply maximum likelihood or other parametric estimates. For example, if $F(\bullet)$ is the exponential distribution then

$$(35) \qquad \tilde{s}_a = -\bar{\mu}[X] \ln (1 - a)$$

(where $\bar{\mu}[X]$ denotes the sample mean) is the maximum likelihood estimator of s_a and is therefore asymptotically fully efficient. Clearly,

$$(36) \qquad E[\tilde{s}_a] = -\mu \ln (1 - a)$$
$$= s_a ,$$

so that $\tilde{s}_a$ is unbiased; furthermore,

$$(37) \qquad \mathrm{Var}[\tilde{s}_a] = \frac{1}{n} [\mu \ln (1 - a)]^2$$
$$= s_a^2 / n,$$

which is at most 65 % as large as the asymptotic non-parametric variance. As a approaches 0 or 1 the relative efficiency of the parametric estimator in this case becomes much greater.

This work is also limited to the consideration of continuous or partly continuous distributions. When the random variable X has a completely discrete distribution its a-quantile may not exist or may not be unique; to overcome this difficulty we may redefine the a-quantile as the solution of

$$(38) \qquad \inf_{s} F(s_a) \geq a,$$

which reduces to (1) in the continuous case. It is not at all clear, however, that the solution to (38) has any reasonable interpretation, particularly if X has only a few atoms.

The methods developed here have been investigated using only pseudorandom simulation data and this is typical of the proposed applications for the techniques. Real world data can certainly be used but the sample sizes required for reasonable results from stochastic approximation quantile estimation are so large that only in special cases will sufficient observations be available. It seems likely that the next-to-maximum transformation will prove more useful in dealing with real data than was found to be the case with the artificial samples used here since there is usually more difficulty with outliers in the former case. As Gaver and Lewis [12] point out the maximum transform will intensify any problems caused by outliers.

One final limitation of this work is that we consider

only samples with sequential independent observations; this will clearly not be the case for much real world data or for many kinds of simulation studies. We may be able to apply our methods in the simulation case by using the regenerative techniques of Iglehart [16] but the general problem of dependent observations is much more complex and is not considered further here.

Chapter II. ASYMPTOTIC PROPERTIES OF THE NEW ESTIMATOR

A. Definitions and Preliminaries

We wish to estimate the solution $x = s_a$ to

$$F(x) = a, \quad 0 < a < 1,$$

where $F(\bullet)$ is the distribution function of the random variable X. We assume:

F1. $F(x)$ has a derivative $f(x)$ which is continuous in some neighborhood of s_a with $f(s_a) = \beta > 0$.

F2. $F''(x)$ exists and is bounded in some neighborhood of s_a.

Note that (F1) is sufficient for s_a to exist and be unique.

A sequential estimation scheme is used with $\bar{s}_n$ the estimate of s_a at step n. The initial estimate $\bar{s}_1$ is chosen arbitrarily (or at random with $E[\bar{s}_1^2] < \infty$) and we apply the recursion

$$\text{(1)} \qquad \bar{s}_{n+1} = \bar{s}_n - a_n Y_n,$$

where Y_n is given by

$$(2) \qquad Y_n = -a \quad \text{if } X_n > \bar{s}_n ,$$

$$\qquad = 1 - a \quad \text{if } X_n \leq \bar{s}_n .$$

In (2) X_n is a random variable with distribution $F(\cdot)$ which is assumed independent of $\{\bar{s}_1 ; X_1 , \ldots , X_{n-1}\}$.

The gain sequence $\{a_n\}$ is given by

$$(3) \qquad a_n = 1 / n d_n ,$$

where d_n is essentially a "bounded" kernel density estimator (see Rosenblatt [32] or Parzen [28]):

$$(4) \qquad d_n = \text{Max} \left[C_1 n^{-L} , \text{Min} \{ B_n , C_2 \log(n+1) \} \right],$$

with $0 < L < 1/4$ and $0 < C_1 < C_2$. The estimator B_n is defined by

$$(5) \qquad B_n = \frac{1}{n} \sum_{j=1}^{n} w_j ,$$

$$(6) \qquad w_j = \frac{1}{b_j} W \left[\frac{\bar{s}_j - X_j}{b_j} \right] ,$$

where $\{b_n\}$ is a bandwidth sequence of positive constants satisfying

$$(7) \qquad b_n = O(n^{-g}) , \qquad 1/5 < g < 1/2 .$$

The function $W(\bullet)$ is called the kernel function; it is assumed to satisfy

W1. $W(u) \geq 0, \quad -\infty < u < \infty.$

W2. $\sup_{-\infty<u<\infty} W(u) = K < \infty.$

W3. $\int_{-\infty}^{\infty} W(u)\, du = 1.$

W4. $\lim_{|u|\to\infty} |uW(u)| = 0.$

Note that $W(\bullet)$ is a probability density under these assumptions.

In what follows, we show first that $\bar{s}_n \to s_a$ almost surely (abbreviated a.s.) and that $d_n \to \beta$ a.s.; then, using a theorem of Fabian [9], we develop the asymptotic distribution of $\bar{s}_n$. Throughout, $\{\Omega, \mathcal{S}, P\}$ will be a probability space and $\mathcal{B}_n = \sigma(\bar{s}_1; X_1, \ldots, X_{n-1}) \subset \mathcal{S}$ a sequence of σ-fields (i.e., the smallest σ-field with respect to which the indicated variables are measurable).

We begin by rewriting the basic relation (1) in the form

$$(8) \qquad \bar{s}_{n+1} = \bar{s}_n - T_n + U_n,$$

in which we define

$$T_n = a_n[F(\bar{s}_n) - a],$$

$$U_n = -a_n Z_n,$$

(9) $$Z_n = Y_n - E[Y_n \mid \mathcal{B}_n]$$

$$= Y_n - F(\bar{s}_n) + a \quad .$$

We note that $|Z_n| \leq 1$.

Since we will deal with sequences of the form (9), we begin by stating two lemmas relating to sequences of this type. Proofs may be found in Loeve [24].

__Lemma 1__ (Loeve) Let $\{V_n\}$ be a sequence of random variables with $\sum_{n=1}^{\infty} Var[V_n] < \infty$; then if $\sum_{n=1}^{\infty} E[V_n \mid V_1, \ldots, V_{n-1}]$ converges a.s., $\sum_{n=1}^{\infty} V_n$ converges a.s. to a random variable.

□

__Lemma 2__ (Loeve) If $c(n) \to \infty$ and $\sum_{n=1}^{\infty} \frac{1}{c(n)^2} Var[V_n] < \infty$ then

$$\frac{1}{c(n)} \sum_{k=1}^{n} \{V_k - E[V_k \mid V_1, \ldots, V_{k-1}]\} \to 0 \text{ a.s.}$$

□

B. Convergence of $\bar{s}_n$

The proofs in this Section follow the lines of Blum's

work [2]. In fact, the convergence of $\bar{s}_n$ follows at once from the bounds indicated by (4) (Fabian [9]) if we are willing to adopt a slightly different definition of B_n. Now we deal with the relation (8) and show

<u>Lemma 3</u> $\sum_{n=1}^{\infty} U_n$ converges a.s. to a random variable.

Proof:
Clearly,

$$\begin{aligned} \operatorname{Var}[U_n] &\le E[a_n^2 Z_n^2] \\ &\le \frac{1}{n^2} E[Z_n^2 / d_n^2] \\ &\le 1 / (n^{2-2L} C_1^2), \end{aligned}$$

so that $\sum_{n=1}^{\infty} \operatorname{Var}[U_n] < \infty$.

Now X_n is independent of $\{\bar{s}_1; X_1, \ldots, X_{n-1}\}$ and since these random variables uniquely determine d_{n-1} we have $E[Z_n/d_{n-1} \mid B_n] = 0$ a.s. Thus,

$$\begin{aligned} E[U_n \mid B_n] &= E[-a_n Z_n \mid B_n] + \frac{1}{n-1} E[Z_n/d_{n-1} \mid B_n] \\ &= E[\{1/(n-1)d_{n-1} - 1/nd_n\} Z_n \mid B_n] \\ &= \frac{1}{n(n-1)} E[\{(n-1)d_{n-1} - nd_n\} Z_n / d_{n-1} d_n \mid B_n]. \end{aligned}$$

Now we use the definition (4) of d_n to set an upper bound:

$$|E[U_n \mid \mathcal{B}_n]| \le \frac{1}{n(n-1)}\left[C_1^2\, n^{-L}(n-1)^{-L}\right]^{-1}$$

$$\bullet\; E[\;|nd_n - (n-1)d_{n-1}|\;|Z_n|\;\mid \mathcal{B}_n]$$

$$\le n^{L-1}(n-1)^{L-1}C_1^{-2}\,E[\;|nd_n - (n-1)d_{n-1}|\;\mid \mathcal{B}_n]$$

where we have used the fact that $|Z_n| \le 1$. The relationship

$$|\;\mathrm{Max}[a,b] - \mathrm{Max}[c,d]\;| \le \mathrm{Max}[\;|a-c|,\;|b-d|\;]$$

and the definition (4) then imply that

$$|nd_n - (n-1)d_{n-1}| \le \mathrm{Max}\;\{\;|C_1 n^{1-L} - C_1(n-1)^{1-L}|,$$

$$|nB_n - (n-1)B_{n-1}|,$$

$$|C_2 n\log(n+1) - C_2(n-1)\log n|\;\}.$$

Now the first term here approaches $C_1(1-L)n^{-L} + O(n^{-L-1})$ as $n \to \infty$ so in this case we have

$$|E[U_n \mid \mathcal{B}_n]| \le n^{L-1}(n-1)^{L-1}C_1^{-2}[C_1(1-L)n^{-L} + O(n^{-L-1})]$$

$$= O(n^{L-2})\text{ a.s.}$$

For the last term we get

$$C_2 n\log(n+1) - C_2(n-1)\log n = C_2\log(n+1) + C_2(n-1)\log(1+\tfrac{1}{n})$$

$$\le C_2\log(n+1)$$

so that

$$|E[U_n \mid \mathcal{B}_n]| \le n^{L-1}(n-1)^{L-1}C_1^{-2}\,C_2\log(n+1)$$

$$= O(n^{2L-2} \log n).$$

Finally we consider

$$|nB_n - (n-1)B_{n-1}| = w_n$$
$$\leq K / b_n$$
$$= O(n^g),$$

in view of (W2) and (7). Thus we conclude for this case that

$$|E[U_n | \mathcal{B}_n]| \leq n^{L-1} (n-1)^{L-1} C_1^{-2} K / b_n$$
$$= O(n^{2L+g-2}).$$

We thus have that $\sum_{n=1}^{\infty} |E[U_n | \mathcal{B}_n]|$ converges almost surely in all three cases because of the definitions of L (4) and g (7). An application of Lemma 1 then completes the proof.

□

<u>Lemma 4</u> (Blum) $\bar{s}_n$ converges a.s. to a random variable.

Proof:

Iterating (8) back to $\bar{s}_1$ yields

$$\bar{s}_{n+1} = \bar{s}_1 - \sum_{j=1}^{n} T_j + \sum_{j=1}^{n} U_j,$$

so that

$$\text{(10)} \qquad \bar{s}_{n+1} + \sum_{j=1}^{n} T_j = \bar{s}_1 + \sum_{j=1}^{n} U_j \text{ converges a.s.,}$$

in view of Lemma 3. Next we show

(11) $$\Pr\{\lim_{n\to\infty} \bar{s}_n = \infty\} = 0.$$

Suppose, for example, there exists a sample sequence $\{\bar{s}_n\}$ with $\lim_{n\to\infty} \bar{s}_n = \infty$; then $\bar{s}_n \leq s_a$ for only finitely many n so that $T_n = a_n[F(\bar{s}_n) - a] > 0$ when n is large enough. Thus $\lim_{n\to\infty} [\bar{s}_{n+1} + \sum_{j=1}^{n} T_j] \to \infty$ which occurs with probability zero by (10). This establishes (11) and we similarly show

(12) $$\Pr\{\lim_{n\to\infty} \bar{s}_n = -\infty\} = 0.$$

Now suppose the lemma is false; then there must exist sample sequences for which

(13) $$\begin{cases} \bar{s}_{n+1} + \sum_{j=1}^{n} T_j \text{ converges to a finite number} \\ \liminf_{n\to\infty} \bar{s}_n < \limsup_{n\to\infty} \bar{s}_n . \end{cases}$$

Letting $\{\bar{s}_n\}$ be such a sequence, we assume that $\limsup \bar{s}_n > s_a$ (a similar argument handles the case $\limsup \bar{s}_n \leq s_a$ for then $\liminf \bar{s}_n < s_a$ by (13)). We then choose numbers c and d such that $c > s_a$ and $\liminf \bar{s}_n < c < d < \limsup \bar{s}_n$. In view of (5)-(7), $a_n \to 0$; and since $\bar{s}_{n+1} + \sum_{j=1}^{n} T_j$

converges, we may choose N so that $N \le n < m$ implies

$$(14)\quad \begin{cases} a_n \le \dfrac{d-c}{2}, \\ \left|\bar{s}_m - \bar{s}_n + \sum_{j=n}^{m-1} T_j\right| \le \dfrac{d-c}{2}. \end{cases}$$

Now we select m and n with $N \le n < m$ such that

$$(15)\quad \begin{cases} \bar{s}_n < c, \\ \bar{s}_m > d, \\ c \le \bar{s}_j \le d \text{ for } n < j < m . \end{cases}$$

We may clearly do this. Thus,

$$(16)\qquad \bar{s}_m - \bar{s}_n \le \frac{d-c}{2} - \sum_{j=n}^{m-1} T_j \le \frac{d-c}{2} - T_n,$$

since $T_j = a_j[F(\bar{s}_j) - a] > 0$ for $\bar{s}_j \ge c > s_a$. Now if $\bar{s}_n >$ s_a we obtain

$$\bar{s}_m - \bar{s}_n \le \frac{d-c}{2},$$

in contradiction of (15) which implies $\bar{s}_m - \bar{s}_n > d - c$. If $\bar{s}_n < s_a$ we have

$$-T_n = a_n[a - F(\bar{s}_n)] \le a_n \le \frac{d-c}{2}$$

from (14); thus (16) becomes $\bar{s}_m - \bar{s}_n \le d - c$, which again

contradicts (15). This means no sequence $\{\bar{s}_n\}$ can satisfy (13), thus establishing the lemma in view of (11) and (12).

□

<u>Theorem 1</u> (Blum) $\bar{s}_n \to s_a$ a.s.

Proof:

We suppose $\Pr\{\lim_{n\to\infty} \bar{s}_n = S\} = 1$ as guaranteed by Lemma 4 and we also suppose that $\Pr\{S \neq s_a\} > 0$. Now we choose c and d with $s_a < c < d < \infty$ and $\Pr\{c < S < d\} > 0$. (Alternatively we take $-\infty < c < d < s_a$.) Then for every sample sequence $\{\bar{s}_n\}$ for which $\lim_{n\to\infty} \bar{s}_n = S$, $c < S < d$, we have $c < \bar{s}_n < d$ for almost all n. Lemma 3 and Lemma 4 show that

$$(17) \qquad \sum_{j=1}^{n} T_j = \sum_{j=1}^{n} a_j [F(\bar{s}_j) - a] \text{ converges;}$$

however, $F(\bar{s}_j) - a > F(c) - a > 0$ for almost all j so (17) must diverge because $a_j \geq \{j\, C_2 \log(j+1)\}^{-1}$; this follows from the definitions (3) and (4) and the fact that $C_2 > C_1$. Thus,

$$\sum_{j=1}^{n} a_j \geq \sum_{j=1}^{n} \{C_1 j \log(j+1)\}^{-1} = O[\log(\log n)]$$

This contradiction establishes the theorem.

□

C. Convergence of d_n

We begin by proving three preliminary Lemmas.

Lemma 5 Let $\{t_n(x)\}$ be a sequence of measurable functions uniformly continuous for every $n \geq N$ in some neighborhood of the point $X \in R$ with

(18) $$\lim_{n \to \infty} t_n(X) = t(X)$$

and $\{X_n\}$ a sequence of random variables with

(19) $$X_n \to X \text{ a.s.},$$

where $X \in R$ is a constant. Then

(20) $$t_n(X_n) \to t(X) \text{ a.s.}$$

Proof:
The convergence (18) implies that for each $\eta > 0$, whenever $n \geq N_1(\eta)$ we have

(21) $$|t_n(X) - t(X)| < \eta/2 .$$

The uniform continuity of $t_n(X)$ for $n \geq N$ likewise implies that given $\eta > 0$ there exists an $e > 0$ depending only on η such that

$$(22) \quad |X_n(\omega) - X| < e \Longrightarrow |t_n(X_n(\omega)) - t_n(X)| < \eta/2,$$

for each $\omega \in \Omega$. Combining (21) and (22) yields

$$(23) \quad |X_n(\omega) - X| < e \Longrightarrow |t_n(X_n(\omega)) - t(X)| < \eta .$$

Now by Egoroff's Theorem (19) implies that for each $\delta > 0$ there exists a set $A_\delta \subset S$ with $P(A_\delta) > 1 - \delta$ such that $X_n(\omega)$ converges uniformly in ω for every ω in A_δ. Evidently then if $n \geq N_2(e)$,

$$\omega \in A_\delta \Longrightarrow |X_n(\omega) - X| < e.$$

Now since e in (23) depends only on η, whenever $n \geq N(\eta) = \max[N_1(\eta), N_2(e)]$ we have

$$\omega \in A_\delta \Longrightarrow |t_n(X_n(\omega)) - t(X)| < \eta ,$$

which means that $t_n(X_n) \to t(X)$ uniformly on A_δ. Since δ is arbitrary, this means that $t_n(X_n) \to t(X)$ almost uniformly which implies (20) because of the equivalence of almost sure and almost uniform convergence (see Lukacs [25]).

□

<u>Lemma 6</u> Let $\{X_n\}$ be a sequence of bounded random variables with

$$(24) \quad X_n \longrightarrow 0 \text{ a.s.},$$

$$S_n = \frac{1}{n} \sum_{j=1}^{n} X_j \; .$$

Then $S_n \to 0$ a.s.

Proof:

Because of (24), given $e > 0$ there exists a set $A_e \subset S$ with $P(A_e) = 1$ such that

$$\omega \in A_e \implies |X_n(\omega)| < e/2$$

for all $n \geq N(e,\omega)$. Now for $t > 0$,

$$(25) \qquad |S_{N+t}(\omega)| = \left| \frac{1}{N+t} \sum_{j=1}^{N+t} X_j(\omega) \right|$$

$$\leq \frac{1}{N+t} \left| \sum_{j=1}^{N} X_j(\omega) \right| + \frac{1}{N+t} \left| \sum_{j=N+1}^{N+t} X_j(\omega) \right| \; .$$

Now we take

$$(26) \qquad C(N,\omega) = \sup_{n \leq N} |X_n(\omega)| < \infty ;$$

this follows from the hypothesis that $\{X_n\}$ is bounded, but the lemma will hold for any sequence satisfying (26). Now (25) becomes

$$|S_{N+t}(\omega)| \leq \frac{N}{N+t} C(N,\omega) + \frac{t}{N+t} \frac{e}{2}$$

$$\leq \frac{e}{2} + \frac{e}{2} = e$$

whenever we choose $t \geq T(e,\omega)$. Thus,

$$\omega \in A_e \implies |S_m(\omega)| < e$$

for all $m \geq M(e,\omega) = N + T$. Since $P(A_e) = 1$, we conclude that $S_n \to 0$ a.s.

□

<u>Lemma</u> <u>7</u> Under assumptions (F1) and (W1) through (W4) the function

$$t_n(x) = \frac{1}{b_n} \int_{\infty}^{\infty} W\left[\frac{x-y}{b_n}\right] dF(y) \tag{27}$$

is uniformly continuous in some neighborhood of $x = s_a$ for every $n \geq N$.

Proof:

Suppose in accordance with (F1) that the density $f(x)$ exists and is continuous for $x \in I = [s_a - \Delta, s_a + \Delta]$ for some $\Delta > 0$. Following Parzen [28] we may rewrite (27) in the form

$$t_n(x) - f(x) = \int_{|y| \leq \delta} [f(x-y) - f(x)]\, b_n^{-1} W(y b_n^{-1})\, dy$$

$$+ \int_{|y| > \delta} \frac{1}{b_n} W\left[\frac{x-y}{b_n}\right] dF(y)$$

$$- f(x) \int_{|y| > \delta} b_n^{-1} W(y b_n^{-1})\, dy$$

where $x \in I$ and δ is chosen such that $0 < \delta \leq \Delta$. Thus when $x \in I$,

$$|t_n(x) - f(x)| \leq \sup_{|y| \leq \delta} |f(x-y) - f(x)| \int_{|y| \leq \delta} W(u)\, du$$

$$+ \int_{|y|>\delta} \frac{|y|}{b_n} W\left[\frac{y}{b_n}\right] \frac{1}{|y|} dF(x-y)$$

$$+ f(x) \int_{|y|>\delta} b_n^{-1} W(yb_n^{-1})\, dy$$

$$|t_n(x)-f(x)| \leq \sup_{|y|\leq\delta} |f(x-y) - f(x)|$$

$$+ \frac{1}{\delta} \int_{z>\delta/b_n} \sup |z\, W(z)| \int dF(z)$$

$$+ f(x) \int_{|z|>\delta/b_n} W(z)\, dz .$$

Now given some $e > 0$ we may, by the continuity of $f(x)$ on I, choose a $\delta > 0$ such that the first term will be less than $e/3$. Having chosen δ we may then select N such that when $n \geq N$ (W4) implies that the second term will also be less than $e/3$. Finally, (W3) allows us to conclude that the last term will also be less than $e/3$ when n is large enough. We thus have that

$$\sup_{x \in I} |t_n(x) - f(x)| < e$$

when $n \geq N(e)$, i.e. $t_n(x)$ is uniformly continuous on I.

□

<u>Theorem 2</u> $d_n \to \beta$ a.s.

Proof:

In view of the bounds (4) it suffices to show that

$$(28) \qquad B_n \to \beta \quad \text{a.s.}$$

We first note that

$$w_n^*(y) = \frac{1}{b_n} W\left[\frac{y - X_n}{b_n}\right]$$

has a bounded variance whose bound is independent of y:

$$\mathrm{Var}[w_n^*(y)] \le \frac{1}{b_n^2} \int W^2\left[\frac{y - u}{b_n}\right] dF(u)$$

$$\le \frac{K^2}{b_n^2} \int dF(u) = \frac{K^2}{b_n^2},$$

which follows from (W2). Thus,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathrm{Var}[w_n] \le K^2 \sum_{n=1}^{\infty} (n\, b_n)^{-2}$$

which is finite by (7). Lemma 2 with $c_n = n$ then implies

(29) $$\frac{1}{n} \sum_{j=1}^{n} \{w_j - E[w_j \mid \mathcal{B}_j]\} \longrightarrow 0 \text{ a.s.}$$

Now

(30) $$B_n = \frac{1}{n} \sum_{j=1}^{n} w_j$$

$$= \frac{1}{n} \sum_{j=1}^{n} \{w_j - E[w_j \mid \mathcal{B}_j]\} + \frac{1}{n} \sum_{j=1}^{n} E[w_j \mid \mathcal{B}_j]$$

$$E[w_j \mid \mathcal{B}_j] = E\left[\frac{1}{b_j} W\left[\frac{\bar{s}_j - X_j}{b_j}\right] \,\middle|\, \mathcal{B}_j\right]$$

$$= \frac{1}{b_j} \int W\left[\frac{\bar{s}_j - y}{b_j}\right] dF(y)$$

$$= t_j(\bar{s}_j) \text{ a.s.,}$$

with $t_j(\bullet)$ given by (27). Now Parzen [28] has shown that (W1) - (W4) and (F1) imply

$$\lim_{n \to \infty} t_n(s_a) = f(s_a) = \beta .$$

Clearly $t_n(\bullet)$ is measurable and $t_n(s_a)$ is continuous for every n greater than some fixed N by Lemma 7 so we may apply Lemma 5 to assert

$$E[w_j \mid \mathcal{B}_j] \to \beta \quad \text{a.s.}$$

Now by (W2), $|E[w_j \mid \mathcal{B}_j]| \le K / b_j < \infty$ so that (26) is satisfied for $X_j = E[w_j \mid \mathcal{B}_j] - \beta$ and an application of Lemma 6 and (29) to the right-hand side of (30) establishes (28).

$\square$

D. Asymptotic Normality

We first state a Lemma due to Burkholder (see [3] for a proof) and then use it to obtain a result on the convergence of $\bar{s}_n$ in the quadratic mean.

<u>Lemma 8</u> (Burkholder) Let $\{X_n\}$ be a non-negative sequence of real numbers and $\{q_n\}$, $\{r_n\}$ real number sequences with $\liminf q_n = q > p > 0$ and $\limsup r_n = r > 0$ such that for every $n \ge N$

$$X_{n+1} \leq (1 - \frac{q_n}{n}) X_n + r_n / n^{p+1}.$$

Then

$$X_n \leq \frac{r}{q - p} n^{-p} + o(n^{-p}).$$

□

<u>Lemma 9</u> $E[(\bar{s}_n - s_a)^2] = O(n^{-1})$.

Proof:

In what follows we write $s^*_n = \bar{s}_n - s_a$. Expanding (8), we obtain

$$(31) \qquad s^*_{n+1} = s^*_n - a_n [F(\bar{s}_n) - a + Z_n].$$

If we expand $F(\bar{s}_n)$ in a Taylor's series about s_a we then get

$$(32) \qquad F(\bar{s}_n) - a = F(s_a) + (\bar{s}_n - s_a) f(s_a) + \delta(\bar{s}_n - s_a) - a$$

$$= \beta s^*_n + \delta(s^*_n),$$

where $\delta(x) = o(x)$ as $x \to 0$ because of (F2). We write δ_n for $\delta(s^*_n)$ in what follows. Substituting (32) into (31), squaring and simplifying yields

$$(33) \qquad s^{*2}_{n+1} = (1 - 2a_n \beta) s^{*2}_n - 2a_n \beta s^*_n (Z_n + \delta_n)$$

$$+ a_n^2 \, (\beta s_n^* + \delta_n + Z_n)^2$$

In order to apply Lemma 8 to (33) we define

$$q_n = 2 \, n \, a_n \, \beta \, (1 + Z_n / s_n^* + \delta_n / s_n^*)$$

$$\to \frac{2n\beta}{n\beta} \, (1 + o(1)) \text{ a.s.}$$

$$\to 2 \text{ a.s.},$$

where $\lim_{n \to \infty} \delta_n / s_n^* \to 0$ by the definition of $\delta(\bullet)$. We also take

$$r_n = n^2 a_n^2 \, (\beta s_n^* + \delta_n + Z_n)^2$$

$$= n^2 a_n^2 \, [F(\bar{s}_n) - a + Z_n]^2$$

$$\to Z_n^2 / \beta^2 \text{ a.s.}$$

$$\to a(1 - a) / \beta^2 \text{ a.s.}$$

We then rewrite (33) as

$$s_{n+1}^{*\,2} = (1 - \frac{q_n}{n}) \, s_n^{*\,2} + \frac{r_n}{n^2} \, ;$$

an application of Lemma 8 with c=2, p=1 then shows that $s_n^{*\,2} = O(n^{-1})$ a.s. and so we conclude

$$E[\, (\bar{s}_n - s_a)^2 \,] = O(n^{-1}).$$

□

We now state a specialization of a theorem due to

Fabian [9] to show the asymptotic normality of $\bar{s}_n$. The notation $I_{\{t\}}$ stands for the indicator function of the set $\{t\}$, i.e.

$$I_{\{t\}}(x) = 1 \quad x \in \{t\}$$
$$= 0 \quad x \notin \{t\}$$

<u>Lemma 10</u> (Fabian) Let B_n be a non-decreasing sequence of σ-fields, $B_n \subset S$. Let A_n, B_{n-1}, V_{n-1}, U_{n-1} and T_{n-1} be B_n-measurable random variables with

$$A_n \longrightarrow a \text{ a.s.,}$$

$$B_n \longrightarrow b \text{ a.s.,}$$

$$T_n \longrightarrow t \text{ a.s. or } E[(T_n - t)^2] \longrightarrow 0,$$

with $a,b,t \in R$. V_n satisfies

$$E[V_n \mid B_n] = 0 \text{ a.s.,}$$

$$C > E[V_n^2 - \sigma^2 \mid B_n] \longrightarrow 0 \text{ a.s.,}$$

(34)
$$\frac{1}{n}\sum_{j=1}^{n} E[I_{\{V_n^2 \geq ne\}}(V_n^2) \mid B_n] \longrightarrow 0 \text{ a.s.,}$$

for every $e > 0$, while U_n is defined by

$$U_{n+1} = \left[1 - \frac{A_n}{n}\right] U_n + \frac{1}{n} B_n V_n + n^{-3/2} T_n .$$

Then

$$n^{1/2} U_n \xrightarrow{L} N[\, t/(a - 1/2),\ \sigma^2 b^2] .$$

□

<u>Theorem</u> <u>3</u> $\bar{s}_n$ is asymptotically normal with mean s_a and variance $a(1-a)/n\beta^2$.

Proof:
To apply Fabian's theorem we use the Taylor's series expansion of $F(\bar{s}_n)$; putting (32) into (8) and simplifying we get

$$s^*_{n+1} = (1 - a_n\beta)\, s^*_n - a_n Z_n - a_n \delta_n .$$

Now we can take

$$A_n = \beta/d_n \;\text{-->}\; 1 \text{ a.s.},$$

$$B_n = -n\, a_n \;\text{-->}\; -\beta^{-1}, \text{ a.s.}$$

$$T_n = n^{3/2} a_n \delta_n$$

$$E[T_n^2] = n\, E[\delta_n^2 / d_n^2]$$

$$\text{-->}\; 0,$$

since $\delta_n^2 = o(n^{-1})$ by Lemma 9. Furthermore, we have

$$E[Z_n \mid B_n] = 0 \text{ a.s.},$$

$$E[Z_n^2 \mid B_n] = F(\bar{s}_n)[1 - F(\bar{s}_n)]$$

$$\text{-->}\; a(1 - a) \text{ a.s.},$$

while the convergence of (34) follows at once from the fact that Z_n is bounded. Thus we conclude from Lemma 10

(35) $$n^{1/2}(\bar{s}_n - s_a) \xrightarrow{L} N[\ 0,\ a(1-a)\beta^{-2}].$$

□

To show that d_n also has an asymptotically normal distribution we need a Central Limit Theorem for the sum of a sequence of dependent summands. For a proof, see Loeve [24], p. 377, Theorem C.

<u>Lemma 11</u> (Loeve) Let $\{X_n\}$ be a sequence of random variables with $S_n = \frac{1}{n}\sum_{j=1}^{n} X_j$. If

(i) $E[X_n | X_1, \ldots, X_{n-1}] = 0$ a.s.,

(ii) $Var[S_n] = \frac{1}{n^2}\sum_{j=1}^{n} E[X_j^2] = \sigma_n^2 < \infty$,

(iii) $\frac{1}{n^2}\sum_{j=1}^{n} E[\ |E\{X_j^2 | X_1, \ldots, X_{j-1}\} - E\{X_j^2\}|\ ] \longrightarrow 0$,

and (iv) for each $e > 0$,

$$\frac{1}{n^2}\sum_{j=1}^{n} E[I_{\{|X_k| \geq e\}}(X_k^2)] \longrightarrow 0,$$

then S_n has an asymptotically normal distribution with mean 0 and variance σ_n^2.

□

<u>Theorem 4</u> d_n has an asymptotically normal distribution with mean β and variance $O(n^{g-1})$.

Proof:

From (30) we have

$$B_n = \frac{1}{n}\sum_{j=1}^{n}\{w_j - E[w_j|\mathcal{B}_j]\} + \frac{1}{n}\sum_{j=1}^{n}E[w_j|\mathcal{B}_j],$$

where the second term converges a.s. to β. In order to apply Lemma 11 to the first term we define

$$v_k = w_k - E[w_k|\mathcal{B}_k].$$

Clearly,

$$E[v_k|\mathcal{B}_k] = 0.$$

$$(36) \qquad E[v_k^2|\mathcal{B}_k] = E[(w_k - E[w_k|\mathcal{B}_k])^2|\mathcal{B}_k]$$

$$= E[w_k^2|\mathcal{B}_k] - 2E[w_k t_k(\bar{s}_k)|\mathcal{B}_k]$$

$$+ E[t_k^2(\bar{s}_k)|\mathcal{B}_k],$$

where we have used the fact that

$$E[w_k|\mathcal{B}_k] = t_k(\bar{s}_k) \text{ a.s.}$$

from Theorem 2. Also

$$E[w_k^2|\mathcal{B}_k] = \frac{1}{\bar{b}_k^2}\int_{-\infty}^{\infty} W^2\left[\frac{\bar{s}_k - y}{\bar{b}_k}\right] dF(y)$$

$$\equiv T_k(\bar{s}_k).$$

Simplifying (36) then yields

$$E[v_k^2|\mathcal{B}_k] = T_k(\bar{s}_k) - t_k^2(\bar{s}_k)$$

$$\equiv \theta_k(\bar{s}_k).$$

Now Parzen [28] has shown that

$$\lim_{n\to\infty} \theta_n(s_a) = b_n^{-1} f(s_a) \int W^2(u)\, du;$$

we note that $\int W^2(u)\, du$ is finite by (W2) and (W3) but that the limit diverges because of the definition of b_n (7). The proof of Lemma 7 may be extended at once to show that $\theta_n(s_a)$ is continuous (at least for all n greater than some fixed N) so an application of Lemma 5 shows that

$$E[v_k^2 \mid \mathcal{B}_k] \to b_k^{-1} f(s_a) \int W^2(u)\, du \quad \text{a.s.}$$

Now we conclude

$$\begin{aligned} E[v_k^2] &= E[\, E[v_k^2 \mid \mathcal{B}_k]\,] \\ &\to b_k^{-1} f(s_a) \int W^2(u)\, du, \end{aligned}$$

so that

$$\frac{1}{n^2} \sum_{j=1}^{n} E[v_j^2] \to f(s_a) \int W^2(u)\, du \sum_{j=1}^{n} n^{-2} b_j^{-1} .$$

The summation clearly converges; if in fact $b_n = b\, n^{-g}$, $\frac{1}{5} < g < \frac{1}{2}$, an application of Euler's summation formula shows

$$\sigma_n^2 \to \frac{f(s_a) \int W^2(u)\, du}{b n^2} \sum_{j=1}^{n} j^g$$

$$\to \frac{f(s_a) \int W^2(u)\, du}{b\,(1+g)} \left[n^{g-1} + O(n^{g-2}) \right].$$

□

Chapter III. FINITE SAMPLE CONSIDERATIONS

In this Chapter we describe some methodological considerations in quantile estimation using both order statistic and stochastic approximation estimators. The emphasis throughout is on practical application of the techniques in finite samples of data rather than on the asymptotic theory of the first two Chapters.

It has long been known that the finite sample behavior of the basic stochastic approximation quantile estimators is seriously flawed from a practical viewpoint (Cochran and Davis [4]; Wetherill [36]; and Davis [6]). Since the problem of finite sample analysis of stochastic approximation estimators is analytically intractable we rely for the most part on digital simulation to examine the finite sample properties of our new estimator; it will be seen that most of the drawbacks have been overcome.

The asymptotic distributions asserted by (1.6) and (1.7) for order statistic estimators and by (1.15), (1.28) and (2.35) for the various stochastic approximation estimators may fail to describe the actual distribution of the estimator for some given n either because this actual distribution is markedly non-Gaussian in shape or because its mean and variance deviate appreciably from the theoretical values. In this Chapter we are for the most part concerned with the first difficulty, leaving the discussion of estimator bias and mean squared error for Chapter IV.

A. Order Statistic Estimators

1. Basic considerations

As pointed out in Chapter I, the order statistic quantile estimator $\hat{s}_n$ for the a-quantile is given by

$$\hat{s}_n = X_{(u)},$$

with $u = [\, a(n+1)\,]$. Unlike the stochastic approximation case, here we need not rely on the asymptotic normality of $\hat{s}_n$ to obtain a confidence interval on s_a; non-parametric confidence intervals may be constructed from the relationship (David [5])

$$(1) \qquad \Pr\{X_{(t)} \le s_a \le X_{(v)}\} = \sum_{i=t}^{v-1} \binom{n}{i} a^i (1-a)^{n-i}.$$

This formula may be evaluated using a table of the incomplete Beta function (see, for example, Kendall and Stuart [17]); however, direct use of the relation (1) is impractical and unnecessary for choosing the values of t and v for large sample sizes n since suitable values for given n and a may be obtained by using the normal approximation to the binomial random variable. For a 100 p % confidence interval we have

$$t = a(n+1) - \sqrt{a(1-a)n}\; u_p$$

and

$$v = a(n+1) + \sqrt{a(1-a)n}\ u_p,$$

where u_p is the upper $1 - \frac{1}{2}p$ significance point of a unit normal variate. To obtain a conservative interval, we round t down and v up to the nearest integer.

The quantile estimation problem may then be reduced to finding three order statistics $X_{(t)}$, $X_{(u)}$ and $X_{(v)}$; this does not require that the entire X sample be sorted nor need we save the entire sample. In fact, just a bit more than a n sample values (or $(1-a)n$ values for $a > 0.5$) must be stored. The three order statistics may then be found by applying Floyd and Rivest's SELECT algorithm [11] which requires an average amount of work proportional to n. This then represents a substantial computational advantage over the naive method of sorting the entire sample, as well as decreasing the memory requirements somewhat.

There remain, however, several serious shortcomings to the order statistic method. First, if more than just a single quantile must be estimated the memory requirements will probably increase drastically and the amount of work also increases quickly. For the simultaneous estimation of the 19 quantiles of Table I it will still be necessary to store the entire sample and the work needed to find the 57 order statistics of interest will be comparable to the effort required to sort the sample as a whole.

This may be shown to be the case by considering that the number of comparisons between observation values is a rough measure of the total amount of work required to sort a sample (or to find the order statistics of interest). The

SELECT algorithm [11] requires an average of about $n[1 + \min(a, 1-a)]$ comparisons to find $X_{(u)}$ so that finding the median requires about n/2 comparisons. Once the median is found, the upper sample quartile (i.e., 0.75 order statistic) must be found in a set of data which is only half as large as the original sample (this is a result of the sorting method employed); this requires n/8 comparisons, on the average. Proceeding in this manner, we find that determining all 57 order statistics will take about 15 n comparisons; a complete sort, on the other hand uses about 2 n ln n comparisons (see Knuth [19]). The advantage will then be with the complete sort for values of n less than 1500 and the amount of work will be about the same for $1500 < n < 10{,}000$.

Since order statistic estimation is not basically a sequential scheme, a second shortcoming of order statistic estimation arises when it is found that a larger sample is needed, perhaps because the estimates in a sample of size n are not precise enough or perhaps because more data become available. If one wishes to take advantage of the savings possible in storing only a n of the observations one must fix the value of n in advance. When a larger sample is to be investigated it will not in general be possible to find the exact order statistic of interest in the pooled sample unless all of the discarded data from the original sample can also be reviewed. Furthermore, the operation of the SELECT algorithm will still require an amount of time proportional to the new (larger) sample size.

2. Decreasing the storage - Payne's method

The most serious difficulty with order statistic estimators is the inescapable linear growth in storage requirements with sample size. For this reason, a technique due to Payne [29] may be considered. A value $m < n$ is first chosen; Payne shows that m may be proportional to $\sqrt{n}$. An array of size m is set aside and filled with the first m observations on X. The array is sorted and then, using (1), a confidence interval on s_a is obtained. Observations outside the confidence interval are discarded and new observations are obtained to fill the array. Any observation which does not fall within the confidence interval is counted toward the total number of observations but is not put into the array. When the array is again filled it is sorted in place and a new, narrower confidence interval is chosen. (The new interval is narrower in the sense that it is shorter than the earlier one, but it will have the same probability mass from (1) since it is based on a larger sample. Note that it will in general have more observations than the earlier interval.) The procedure is repeated until all the observations have been examined.

The main drawback to Payne's method is that if the initial confidence interval is not wide enough the technique may fail to cover the required order statistic when the entire sample has been examined. For this reason, the technique should probably be employed with extremely conservative confidence intervals - say 4-5 standard deviations - with the actual desired confidence interval

chosen at the final step. For example, to determine the median of a sample of 10^6 observations with very low probability of failure a total storage requirement of some 8000 observations should be ample.

The estimation of several quantiles by this so-called partial sorting method appears to involve a fairly complex algorithm, but the method should be useful for a small number of quantiles (say two or three) in fairly large samples of data. Although the method still requires memory which increases with sample size, the presence of more observations can often be handled by simply decreasing the coverage of the last confidence interval.

3. Approximate order statistics - Averaging

Another possible application of the order statistic method is to consider the X sample in sections of some fixed size, say 100 observations. We can then choose $Y_i = X_{(100a)}$ in section i. The final estimate could then be the average of the Y_i's or we may once again sort the Y sample and choose an appropriate order statistic as an estimate. If the second technique is adopted one may obtain yet another level of sections of the Y order statistics and then choose $Z_i = Y_{(100a')}$; we call this a "nested" method. Both the average and nested methods can be thought of as approximate order statistic methods since they do not find the actual order statistic in the entire sample but rather a value close to it.

The chief drawback to the averaging method is that there may be appreciable bias in the Y values if these are

drawn from samples small enough to be practical; Table II indicates some results for extreme quantiles from several

	Quantile		Bias for Sample of		
Distribution	Alpha	Value	100	1000	10000
Exponential	0.5	0.6931	-.0050	-.0005	$-5X10^{-5}$
	0.9	2.3026	-.0442	-.0045	-.0004
	0.99	4.6052	-.4175	-.0487	-.0049
	0.999	6.9077		-.4223	-.0491
Normal	0.5	0.0	-.0125	-.0013	-.0001
	0.9	1.2816	-.0320	-.0033	-.0003
	0.99	2.3263	-.1782	-.0206	-.0021
	0.999	3.0902		-.1361	-.0158
Uniform	0.5	0.5000	-.0045	-.0006	$-5X10^{-5}$
	0.9	0.9000	-.0089	-.0010	-.0001
	0.99	0.9900	-.0198	-.0020	-.0002
	0.999	0.9990		-.0010	-.0001
Cauchy	0.5	0.0	-.0159	-.0016	-.0002
	0.9	3.0777	-.0098	-.0010	-.0001
	0.99	31.820	-.0103	-.0010	-.0001
	0.999	318.31		-.0010	-.0001

Table II. Bias of the order statistic quantile estimator for various distributions. Note that these biases are for single order statistics; unbiased estimates of the median in the normal, uniform and Cauchy cases may be obtained by taking the usual sample median. Biases were evaluated analytically for the exponential and uniform distributions and by Gauss-Legendre quadrature for the normal and Cauchy distributions.

common distributions. The presence of bias means that the estimator will converge to the wrong value as larger and larger samples are obtained. Whether this asymptotic error is objectionable or not depends on pragmatic consideration of the total sample size available but it would certainly seem preferable to adopt an asymptotically unbiased scheme.

It should be pointed out that for the exponential distribution the bias is about 10 % of the true 0.99 quantile value for a sample of 100 observations and about 1 % when 1000 observations are considered. The normal distribution has similarly poor properties so that quite large sections may be required in these cases if bias is not to be a problem in the final approximate order statistic estimate.

Usually bias can be removed by using the jackknife technique (see Miller [27]) but since the order statistics are very non-linear functions of the observations the jackknifing eliminates bias only at the cost of a serious inflation of the variance. This inflation was found to be very bad for small samples by Goodman, Lewis and Robbins [14], where empirical evidence demonstrated that the mean square error of the jackknifed estimators was 50 % larger than for the ordinary order statistic method for samples of from 1000 to 10,000 observations. Moreover, implementation of a jackknife scheme is complicated by the requirement to sort not only the entire section but also a set of subsections.

4. Approximate order statistics - Nesting

If we use sections of size n in an approximate order statistic method and then choose

$$Y_i = X_{(u)}$$

with $u = [a(n+1)]$, then s_a has the same value as the a_Y-quantile of the Y values, where

$$a_Y = \Pr\{ Y \le s_a \}$$
$$= \Pr\{ X_{(u)} \le s_a \}$$
$$= \sum_{i=u}^{n} \binom{n}{i} a^i (1 - a)^{n-i} .$$

This is just a generalization of the two transformation methods of Section I.C. For a nested scheme, then, we accumulate a sample of n_Y Y observations and choose

$$Z_i = Y_{(u_Y)}$$

with $u_Y = [a_Y(n_Y+1)]$. The extension of this technique to higher levels of nesting is straightforward.

The price we must pay for this reduction in the storage requirements is an inflation of the asymptotic variance just as in the case of the maximum and next-to-maximum transforms; note that the averaging method involves no such inflation as long as the X sections are large enough for the asymptotic variance (1.5) to hold approximately. If Z_i were taken directly as an order statistic from an Y sample of $n_Y n$ observations we would have

$$\mathrm{Var}[Z_i] \doteq \frac{a(1-a)}{n_Y n f^2(s_a)} ;$$

with the nesting scheme, however,

$$\operatorname{Var}[Z_i] \doteq \frac{a_Y(1 - a_Y)}{n_Y f_Y^2(s_a)} ,$$

where

$$f_Y(s_a) = \binom{n}{u} u\, a^{u-1} (1 - a)^{n-u} f(s_a).$$

(See David [5].) Thus, the variance will be inflated by an approximate factor of

$$\frac{n\, a_Y(1 - a_Y)}{\binom{n}{u}^2 u^2 a^{2u-1} (1 - a)^{2(n-u)+1}} .$$

For example, if we estimate the 0.99 quantile by considering a Y sample generated by taking the 99th order statistic in X sections of 100 the variance of an estimate based on a Y order statistic will be 1.437 times the variance of an estimate taken from the X sample as a whole. Since $a_Y = 0.73576$ in this case, we may continue the nesting process by choosing $n_Y = 100$ in which case we take $Z_i = Y_{(74)}$; the variance will then be further inflated by a factor of 1.566 for an overall inflation of 2.242. We may obtain results with the same precision by considering a larger sample (assuming data is available); in the present case, we need a total X sample of 14,400 to obtain a variance equivalent to n = 10,000 in an untransformed case. The total storage requirements, however, are now just 244 observations - 100 for the X samples and 144 for the Y sample. Similarly, we may deal with a total X sample of 2,250,000 by using a triply nested scheme with 100 X, 100 Y and 225 Z observations, thus obtaining a variance equivalent to n = 1,000,000 in the unnested case.

The nested order statistic scheme results in the smallest asymptotic memory requirements - 115 ln n for repeated sections of 100 - but the increase in variance by a factor of about 1.5 per level is a very serious drawback. There is also the problem of determining the proper sample sizes and order statistics at each level - a problem which is most easily solved if the sample size can be specified in advance. The determination of the bias of the nested estimators and investigation of some reasonable way for finding confidence intervals are areas for further research, but the problem of variance inflation would seem to rule out these estimators unless a virtually unlimited amount of data is available.

	Asymptotic	n = 10,000		n = 10^6	
Method	Memory Size	Memory	Bias	Memory	Bias
Full Sort	n	10,000	-.0049	10^6	$-5X10^{-5}$
Censored	.01 n	130	-.0049	10,300	$-5X10^{-5}$
Payne's	$.8\sqrt{n}$	100	-.0049	8,000	$-5X10^{-5}$
Average	1000	1,000	-.0487	1,000	-.0487
Nested	115 ln n	200	-.0013	300	-.0063
		244	-.0079	425	-.0064

Table III. Comparison of various order statistic estimation methods for finding the 0.99 quantile. Bias values given are for the exponential distribution. Total samples of 14,400 and 2,250,000, respectively, are needed to give equivalent variance results in the nested method; memory and bias results for these larger samples are also given.

5. Summary

A summary of the order statistic quantile estimation methods discussed here appears in Table III; biases given are for the 0.99 quantile of the exponential distribution. Despite the conceptual simplicity and well-understood behavior of these estimators, we have shown then all to lack some desirable features. If we wish to estimate a set of quantiles based on a fairly large amount of data (say 100,000 observations) order statistic estimators are clearly inadequate.

B. Robbins - Monro Estimators

It should be mentioned at the outset that the basic Robbins-Monro (RM) process cannot be applied directly as a quantile estimation technique in any practical method since its properties depend so heavily on the unknown parameter $\beta = f(s_a)$, i.e. the value of the derivative of the unknown distribution function at the unknown quantile. The properties also depend to a lesser degree on the starting value $\bar{s}_1$ but the situation is not nearly so critical there. Both modifications to the basic RM process considered here overcome this difficulty by simultaneously obtaining an estimate of s_a and β; we thus investigate the RM process applied to a known distribution using the optimum step size $A = \beta$ in order to obtain results which should be better than those for methods which employ estimates of β.

1. Selecting the starting point

The first problem to be faced when dealing with RM quantile estimation is the selection of the initial guess, $\bar{s}_1$. The results of Hodges and Lehmann [15] indicate that the bias of the RM estimator is closely related to that of $\bar{s}_1$ so that starting with a value which is close to s_a is desirable. We must have $E[\bar{s}_1^2] < \infty$ in order to preserve

mean square convergence. One approach is to take a pilot sample with perhaps 1000 or 2000 observations and begin RM with an order statistic estimator; a second approach is to use a nested approximate order statistic estimator, as discussed in the previous section.

This latter approach is in fact adopted here; since we will for the most part be employing the maximum transform in this work, we begin all the stochastic approximation estimation procedures by choosing

$$X'_1 = \max \{ X_1, X_2, \ldots , X_v \}$$
$$X'_2 = \max \{ X_{v+1}, X_{v+2}, \ldots , X_{2v} \}$$
$$X'_3 = \max \{ X_{2v+1}, X_{2v+2}, \ldots , X_{3v} \}$$

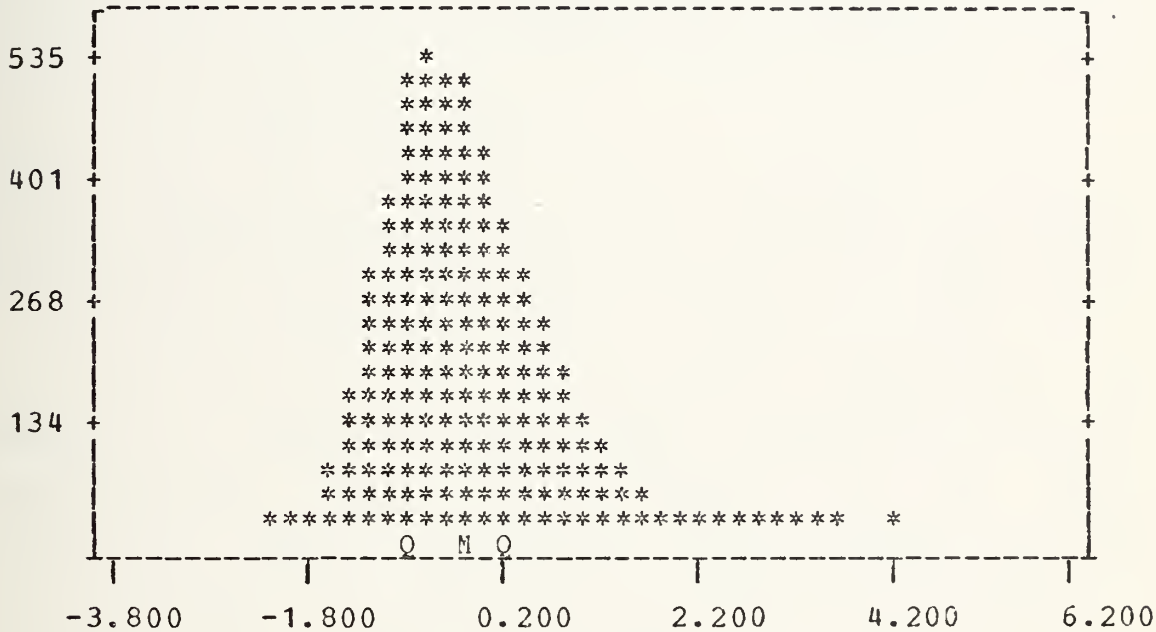
and then setting

$$\bar{s}_1 = X'_{(2)} \quad .$$

This procedure requires very little computer memory and turns out to be very convenient for the simultaneous estimation problem; it is adopted in other cases not employing the maximum transform in order to have an equivalent basis for comparison between stochastic approximation methods.

Throughout much of this work we deal with the problem of estimating the 0.99 quantile of the exponential distribution. This case was chosen because it is one in which the bias of the order statistic estimator in reasonable samples may be objectionable (see Tables II and III). The exponential distribution is also widely applied

as an empirical model for data and the 0.99 quantile is commonly used in statistical inference; thus, this case is typical of the contemplated application of our stochastic approximation estimators.



Mean	-1.121577E-01	Minimum	-2.041833E 00
Variance	6.673908E-01	Quartile	-6.886988E-01
Standard Deviation	8.169399E-01	Median	-2.153153E-01
Skewness	7.798659E-01	Quartile	3.552980E-01
Kurtosis	1.001406E 00	Maximum	4.213145E 00

Figure 1. Bias of the initial estimate $\bar{s}_1$ for the 0.99 quantile of the unit exponential distribution; v for maximum transformation is 56. True quantile value is 4.6052. Histogram sample size is 5000 observations on s^*_n.

The bias of the initial estimate $\bar{s}_1$ for the exponential 0.99 case is indicated by the histogram of Figure 1. The histograms for stochastic approximation quantile estimators in this Chapter display the bias of the estimators, i.e.

$$s^*_n = \bar{s}_n - s_a ,$$

rather than the estimator values themselves. In this Chapter, we use the term bias to refer to the entire distribution of s^*_n rather than to $E[s^*_n]$ as is usual. Data for Figure 1, as well as for the other histograms, was obtained by sampling pseudo-random numbers from various distributions; these were generated by the Naval Postgraduate School random number package LLRANDOM [21] and its extensions [31]. Note that the information of Figure 1 could have been obtained analytically, but the details would be messy.

The caption for each histogram in this Chapter indicates two sample sizes: one (the "X sample") for the total number of X observations from the underlying population used to compute the statistic (for example, s^*_n) whose distribution is displayed and the other (the "histogram sample") for the number of replications of this statistic used to compute the histogram and the sample summary statistics printed. Note that the X sample size will be larger than the indicated number of stochastic approximation steps taken because the X sample includes the 3 v values used for the starting point. Also, the number of steps taken in the stochastic approximation will be smaller than the corresponding sample size by a factor of v when the

maximum transform is used (or w for the next-to-maximum transform).

The letters "Q" printed below the histogram and above the scale indicate the location of the sample quartiles (including the median as the second quartile) while the letter "M" indicates the sample mean. The M may be printed instead of one of the Q's if they appear in the same column; this phenomenon occurs in Figure 1.

2. The basic RM process

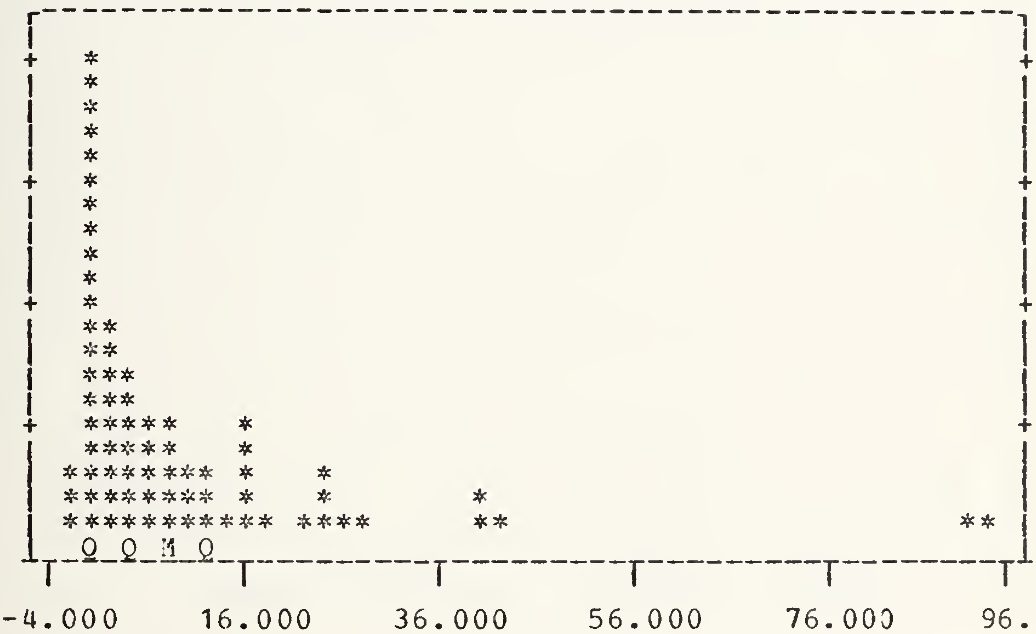
We begin our investigation of the distribution of s^*_n in the RM process by considering the untransformed RM estimator, i.e. one which takes a step with every sample value. We use the optimum step size $A = f(s_{0.99}) = 0.01$ for the exponential distribution. The results are shown in Figures 2 and 3 for s^*_{1121} and s^*_{5601}; the distributions are clearly grossly non-normal, despite the asymptotic normality indicated in Chapter I. Note that the appearance of Figure 3 does not suggest much of an improvement despite an additional 4480 X observations; the skewness and kurtosis of the estimator are, if anything, increasing with sample size.

An explanation of this behavior becomes clear if we consider the effect of the first observation, X_1. Because of the negative bias in $\bar{s}_1$ (see Figure 1), the probability that $X_1 > \bar{s}_1$ is slightly greater than 0.01; this means that about 1.5 % of the time the second quantile estimate is

$$\bar{s}_2 = \bar{s}_1 + 0.99 / (0.01 \times 1)$$

$$= \bar{s}_1 + 99.0 \; .$$

This is obviously much larger than the true quantile value of 4.60 so we expect that all of the observations on X will be less than $\bar{s}_n$ with high probability until the estimate has

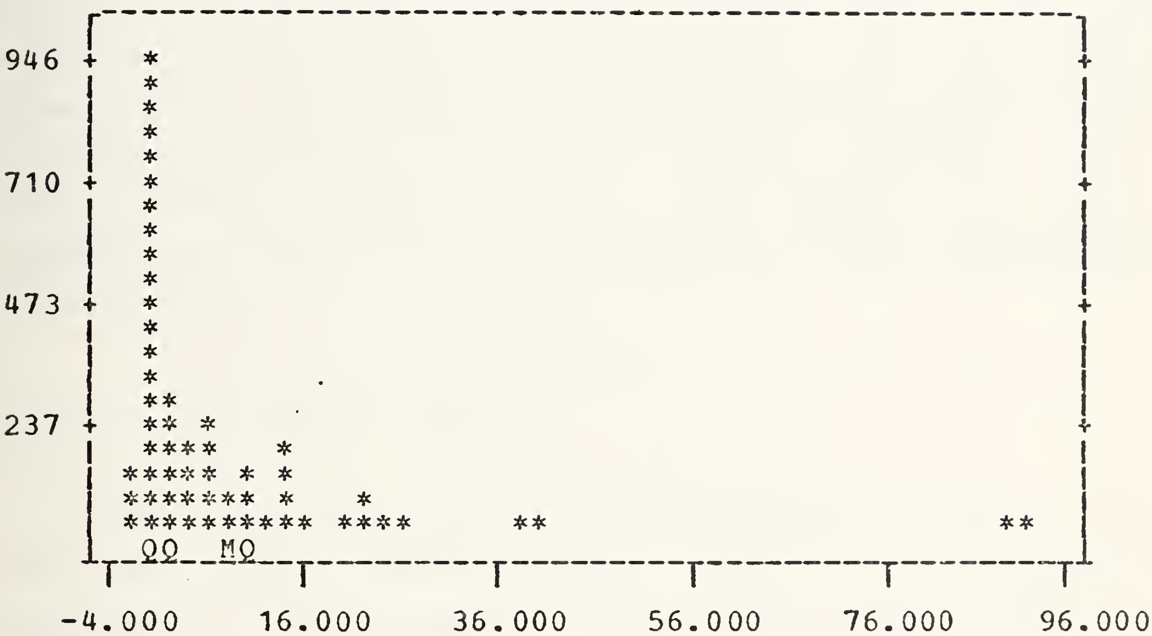


Mean	9.203441E 00	Minimum	-9.772558E-01
Variance	2.040441E 02	Quartile	1.099984E 00
Standard Deviation	1.428440E 01	Median	4.579803E 00
Skewness	3.593350E 00	Quartile	1.158600E 01
Kurtosis	1.623663E 01	Maximum	9.311420E 01

Figure 2. Bias of the RM stochastic approximation estimator s^*_{1121} for the 0.99 quantile of the exponential distribution. Maximum transform was not used. X sample = 1288 observations; histogram sample = 2500 replications of s^*_{1121}.

reached a reasonable level, perhaps 6.0. This in turn requires that the RM process take downward steps for about 90 units. These downward steps are proportional to 1 - a according to (1.10) and in this case are exactly equal to 1/n. The value of n such that

$$\sum_{i=2}^{n} i^{-1} = 90$$



Mean	7.983718E 00	Minimum	-3.477802E-01
Variance	1.966737E 02	Quartile	3.437433E-01
Standard Deviation	1.402404E 01	Median	2.965771E 00
Skewness	3.729717E 00	Quartile	9.932393E 00
Kurtosis	1.718553E 01	Maximum	9.143404E 01

Figure 3. Bias of the RM stochastic approximation estimator for the 0.99 quantile of the exponential distribution for an X sample of 5768 observations; maximum transform was not used. True value is 4.6052. Histogram based on 2500 observations on s^*_{5601}.

is about 2×10^{39} so that the RM process will in this case have 1.5 % of its distribution in the extreme right-hand tail at a substantial distance from the true quantile value for <u>any</u> reasonable sample size.

An additional 4 % of the quantile estimates will also move upwards a distance of 49.5 units after having taken the first step down, while 5 % and 8 %, respectively, will take the third and fourth steps upwards. Thus, nearly one-fifth of the time the RM process will be over 20 units from its starting point (and from the vicinity of the true value) after only four observations. This then accounts for the appearance of Figures 2 and 3; a similar situation exists with random samples from a wide variety of parent populations, i.e. it is not particular to the exponential distribution.

3. The gain sequence shift

What is needed is a way to decrease the size of the first few upward steps without changing the asymptotic behavior of the RM process. This can be done by using the gain sequence

$$(2) \qquad a_n = \frac{1}{(n+k)}$$

instead of the 1/n sequence of (1.11), where k is some positive constant, referred to hereafter as the shift constant. The proofs of Dvoretzky [7] and Sacks [33] allow for gain sequences of the form (2) and so we preserve the almost sure convergence and asymptotic normality of $\bar{s}_n$.

For the exponential 0.99 quantile case a k value of 98 would reduce the initial upwards step to a reasonable size

of 1 unit; from this point we need to move down a distance of only about 0.9 (on the average) to reach the true value of $s_{0.99}$. The n value such that

$$\sum_{i=100}^{n+99} i^{-1} = 0.9$$

is 146 so that there will be no difficulty in reaching the close proximity of the true value given a reasonable sample size.

Since the initial estimate $\bar{s}_1$ is actually based on a sample of 168 X observations, we adopt a shift constant k of 167; the resulting distribution of s^*_{1121} is shown in Figure 4. The data from the X population for this Figure are the same as in Figure 2 with which Figure 4 should be compared. Clearly the introduction of the shift constant has greatly improved the finite sample properties of the estimator.

Under more general conditions we wish to determine a gain sequence shift k such that the effects of a bad initial step can be reversed in a reasonable number of additional steps. Assuming that a > 0.5, the "bad" direction is upward and the initial step is $a / \beta(k+1)$, using the optimum step divisor $A = \beta$. Writing j for k+1 we must then find an n large enough that

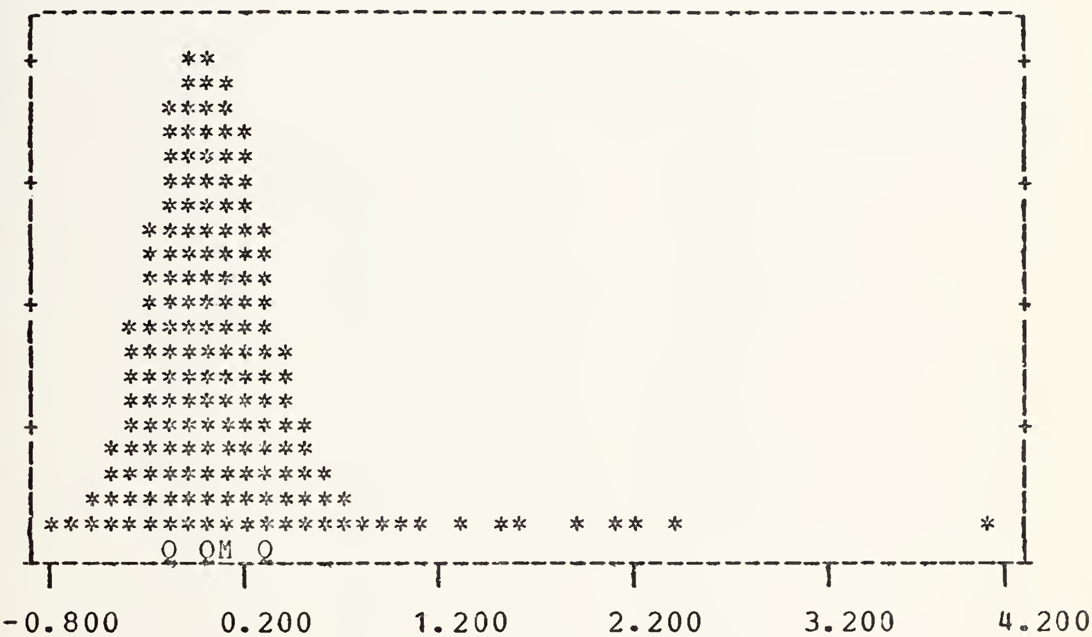
$$\sum_{i=1}^{n} \frac{1-a}{\beta(i+j)} > \frac{a}{\beta j} ,$$

or

$$\sum_{i=j+1}^{j+n} i^{-1} > \frac{a}{(1-a)j} .$$

Table IV shows values of n for various values of a and j. It is clear that using a shift constant of 100 to 200 may be useful for $0.01 < a < 0.99$.

Another interpretation of Table IV is also possible: the entries show the minimum number of additional



Mean	8.416826E-02	Minimum	-7.433958E-01
Variance	1.058326E-01	Quartile	-1.386166E-01
Standard Deviation	3.253192E-01	Median	6.168079E-02
Skewness	1.566202E 00	Quartile	2.729588E-01
Kurtosis	1.163162E 01	Maximum	4.016244E 00

Figure 4. Bias of the RM stochastic approximation estimator s^*_{1121} for the 0.99 quantile of the exponential distribution using a shifted gain sequence with k = 167. Maximum transform was not used. X sample was 1288 observations; histogram sample size = 2500.

observations needed to overcome an incorrect step upwards at stage j of the RM process. Note that as j increases this number of steps approaches a limit which is approximately a / (1 - a). This means that the RM process tends to remain in the vicinity of the true value s_a once it has reached it since here it will take a steps down on the average for each 1 - a steps upward.

j	Quantile Level, a				
(k+1)	0.75	0.900	0.990	0.999	Unit Step
1	30	12302	$2X10^{43}$	10^{434}	3
2	9	225	$2X10^{21}$	10^{217}	5
3	7	68	$8X10^{14}$	10^{145}	7
4	6	39	$2X10^{11}$	10^{109}	8
5	5	28	$2X10^{9}$	$3X10^{87}$	10
10	4	16	$2X10^{5}$	$2X10^{44}$	19
20	4	12	2874	$1X10^{23}$	36
50	4	10	316	$2X10^{10}$	87
100	4	10	170	$2X10^{6}$	173
200	4	10	129	29310	345
300	4	10	118	8095	517
500	4	10	110	3191	861
1000	4	10	105	1717	1720

Table IV. Number of additional observations required for a shifted stochastic approximation method to reverse an initial unfavorable step. The shift constant is one less than the entry in the first column. The entries may also be interpreted as the number of observations needed to reverse an incorrect step upward at step j. The last column gives the value of n satisfying $\sum_{i=j+1}^{j+n} i^{-1} > 1$.

We thus see that estimating the stochastic approximation starting point $\bar{s}_1$ by an order statistic method from an initial sample whose size is roughly proportional to $a / (1 - a)$ and then beginning the RM process with a shift constant $k = a / (1 - a)$ will avoid most of the serious instabilities of Figures 2 and 3. An interesting feature of this result is that it is distribution-free in the sense that the optimum step size multiplier $1/\beta$ does not appear in an explicit way. However, whether shifting the gain sequence will result in an effective estimation procedure depends on the bias of $\bar{s}_1$ as well as the properties of the random variable X whose quantile we are estimating.

For example, if the random variable X is widely dispersed it is quite possible that the RM process will take two or even more steps in the wrong direction. Since the harmonic series on which Table IV is based grows logarithmically the effect of several such incorrect steps may require many times the sample sizes indicated to overcome. The typical shape of the distribution of stochastic approximation quantile estimators is that of Figure 4; the long tail to the right is made up of estimation sequences which are in the process of correcting multi-step errors.

If there is an appreciable bias in $\bar{s}_1$ then a large shift constant may seriously impede the convergence of the estimator to the near proximity of s_a. The biases indicated

in Tables II and III in some cases are large enough to cause difficulties here and the order statistic estimators used to obtain the initial estimate $\bar{s}_1$ estimators are subject to considerable sampling variation. If the initial sample size for finding $\bar{s}_1$ is n_1, then on asymptotic grounds from (1.7) the initial variance is

$$\sigma_1^2 \doteq \frac{a(1-a)}{n_1 \beta^2}$$

which might be inflated somewhat if a nested scheme is used. Since $n_1 \doteq a / (1 - a)$, the initial standard deviation will be

$$\sigma_1 \doteq \frac{1-a}{\beta},$$

which is n_1 times the size of the first downward step. Thus if the initial estimate $\bar{s}_1$ is just one standard deviation high we need a sample of at least n observations to overcome this, where

$$\sum_{i=j}^{n+j} \frac{1-a}{\beta i} > \frac{1-a}{\beta}$$

or

$$(3) \qquad \sum_{i=j}^{n+j} i^{-1} > 1.$$

The last column of Table IV gives values of n satisfying

(3).

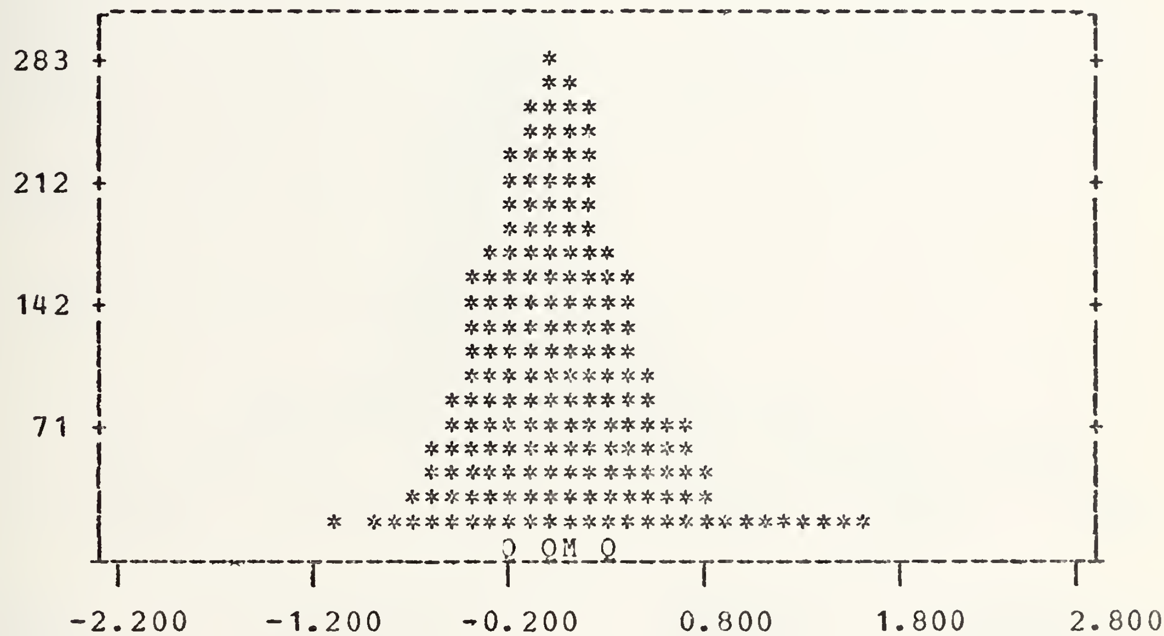
In a given case it is thus possible that both the bias and the sampling variation of $\bar{s}_1$ will combine to produce a starting point which is far from s_a. If this is so an unreasonably large sample may be needed to obtain a nearly Gaussian distribution for $\bar{s}'_n$ when a is close to 0 or 1. The long tail of Figure 4 is at least partially due to this phenomenon, especially in view of the skewed distribution of Figure 1.

4. Maximum and next-to-maximum transforms

The only way to overcome this problem is to transform the a values being used to values closer to 0.5; this of course can be done by means of the maximum or next-to-maximum transform methods of Chapter I. In the context of our present discussion, it is clear that these transform techniques work because the effect of steps in the wrong direction can be readily reversed. Examples of the maximum transform ($s^{*\prime}_{20}$) and next-to-maximum transform ($s^{*\prime\prime}_{6}$) used for estimating the 0.99 quantile of the exponential distribution are shown in Figures 5 and 6. The theoretical (asymptotic) variances of $\bar{s}'_n$ for these Figures are .1242 and .1848, respectively, which compare well with the observed values of .1431 and .1842. The distributions in both cases are normal or nearly so.

Examination of Figures 4, 5 and 6 (together with a

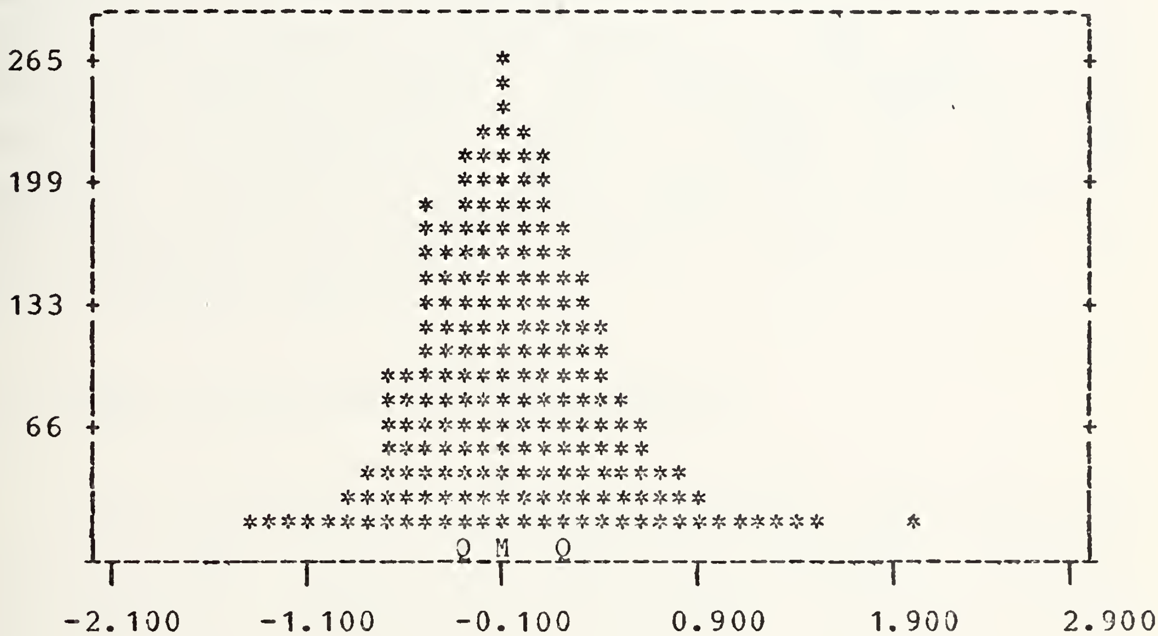
great deal of data from other distributions and quantiles) leads to the general conclusion that the distributional properties of the stochastic approximation estimator $\bar{s}_n$ are greatly improved by these transformation schemes. The



Mean	1.028941E-01	Minimum	-1.091178E 00
Variance	1.430671E-01	Quartile	-1.581392E-01
Standard Deviation	3.782421E-01	Median	8.505392E-02
Skewness	3.894951E-01	Quartile	3.338833E-01
Kurtosis	3.945213E-01	Maximum	1.622832E 00

Figure 5. Bias of the RM stochastic approximation estimator for the 0.99 quantile of the exponential distribution using the maximum transform with v = 56. X sample is 1232 observations; histogram based on 2500 replications of $s^{*\prime}_{20}$.

next-to-maximum method seems to result in a more nearly Gaussian shape (as measured by the sample coefficients of skewness and kurtosis) for the distribution and agrees more closely with the asymptotic variance, but both transform methods give quite satisfactory results even in relatively small samples.



Mean	-1.863602E-02	Minimum	-1.387260E 00
Variance	1.841508E-01	Quartile	-3.100567E-01
Standard Deviation	4.291279E-01	Median	-3.200531E-02
Skewness	2.201208E-01	Quartile	2.597713E-01
Kurtosis	2.572885E-01	Maximum	2.015632E 00

Figure 6. Bias of the RM stochastic approximation estimator for the 0.99 quantile of the exponential distribution using the next-to-maximum transform with w = 192. X sample is 1028 observations; histogram based on 2500 observations of $s^{*''}_{6}$.

A further advantage of the transform methods is that they involve less computational effort than does the untransformed (direct) technique. In fact the computation time for the untransformed case is over four times that for either transform method. Thus if the X sample is being generated by a pseudo-random process within the computer it may be more efficient computationally to use one of the transform methods despite the variance inflation which requires us to generate a larger X sample for the same estimate precision; the time saved in the estimation procedure may be sufficient to offset the generation time for the larger sample.

5. Direct application of the RM method

In the previous Subsection we used a fixed step size $A = \beta$, chosen so as to give the best asymptotic variance. As indicated earlier, the RM process cannot be applied optimally (i.e., with minimum asymptotic variance) in any real situation simply because we do not know the actual value of β. If a reasonable initial estimate of β can be found, however, it may be possible to use the RM process directly for quantile estimation.

This was done in the work of Goodman, Lewis and Robbins [14] and also by Yuguchi [38]. They used the same starting value as in the present work, but with a random A value given by

$$\bar{A} = \frac{X'_{(3)} - X'_{(1)}}{8(X'_{(2)} - X'_{(1)})(X'_{(3)} - X'_{(2)})} .$$

This $\bar{A}$ is used for <u>all</u> steps in the stochastic approximation

estimation process as opposed to the Venter method and the new method which use a dynamically changing A value. A second instance in which direct application of the RM process was attempted is given by Iglehart [16]; in this case a fixed estimate of $f(s_a)$ based on the empirical distribution function was used.

Now the convergence of $\bar{s}'_n$ to a limiting normal distribution with variance $O(n^{-1})$ requires that we have $A < 2\beta$ (Sacks [33]). This will not in general always be the case for $\bar{A}$ or for any other estimate of β. It is known that the convergence may be much worse for $A \geq 2\beta$; for example, when $A = 2\beta$ the variance is $O(\log n/n)$ (Major and Revesz [26]). Thus, the stochastic approximation process with a fixed gain sequence multiplier may result in very poor convergence properties even if the distribution does not blow up as in Figures 2 and 3.

In particular, the results of Yuguchi [38] indicate the presence of an $O(n^{-1/4})$ component in $MSE[\bar{s}'_n]$; also, the sample coefficients of skewness and kurtosis of the RM estimators increase with increasing sample size rather than decreasing as we would expect if the distributions were in fact approaching normality. The RM quantile estimators were also found to give "erratic results" by Iglehart [16] and he recommended that they not be used.

It is possible that these results could be improved if

a density estimator with better properties than $\bar{X}$ or the derivative of the empirical distribution function could be found. A possible candidate is just the kernel estimator of (1.32); see Rosenblatt [32] or Parzen [28]. We prefer to use a method which is guaranteed to have the minimum asymptotic variance, however, and so in Section III.C we turn to techniques which have this property.

6. Summary

The general conclusions of this Section are that the nested method for selecting $\bar{s}_1$ is sufficiently robust and that the maximum transform is a computationally and statistically effective technique for RM quantile estimation for well-behaved X populations. The next-to-maximum transformation and the gain sequence shift are also useful and may be necessary in some cases to increase the robustness of the RM process. Finally, the finite sample and asymptotic properties of methods using random values for the gain sequence divisor A will be much better if those values converge to the optimum value β rather than remaining fixed.

C. Venter's Estimator

With Venter's method we enter the realm of techniques which can be applied to real estimation problems, i.e. those in which β is unknown. General experience with the Venter estimator, however, shows that it is not very robust and often tends to blow up.

1. Choice of parameters

The first question to be addressed in a practical implementation of this stochastic approximation method is the choice of the finite difference sequence $\{c_n\}$, which from (1.22) is given by

$$(4) \qquad c_n = c\, n^{-r}, \qquad 0.25 < r < 0.50.$$

In order to avoid the necessity of computing n^{-r} at each step of the estimation process (this requires a logarithm and an exponential to be calculated) we adopt instead the sequence defined recursively by

$$(5) \qquad \begin{cases} e_1 = 1, \\ e_{n+1} = \left(1 - \frac{e_n^3}{3}\right) e_n . \end{cases}$$

This sequence requires only elementary arithmetic operations and may be generated about 100 times faster than the sequence (4).

The properties of $\{e_n\}$ may be readily found. First we note that $e_n > 0$ and that $e_{n+1} < e_n$ for all n, i.e. the sequence is bounded below and monotone decreasing. At stage n suppose that

$$e_n = n^{-1/3} + o\,(n^{-4/3});$$

then using (5) we have

$$e_{n+1} = \left(1 - \frac{1}{3n}\right) n^{-1/3} + o\left(n^{-4/3}\right)$$

$$= (n+1)^{-1/3} + o\left(n^{-4/3}\right).$$

Thus taking $c_n = c\,e_n$ results in a Venter process with $r = 1/3$; Venter's proof [20] allows for gain sequences of this form.

Selection of the modulating constant c is the next problem. Intuitively it seems that c should be larger when the X population is more widely dispersed in the vicinity of s_a; thus $c = 1/\beta$ would be a reasonable choice except that β is usually unknown. We might thus decide to estimate β from the same initial sample as $\bar{s}_1$ and so use a random value for c or else choose a reasonably robust fixed value for c.

It turns out that the behavior of the Venter quantile estimator is bad regardless of the value chosen for c. The selection of c, however, does not seem to influence the estimation process as much as the bounding process (1.25) or (1.29). Venter's convergence proof required that the estimate A_n of β be restricted to the interval (a^*, b^*) [37] while Fabian [9] showed that we may take

$$a^* = C_1\, n^{-L}$$

(6)

$$b^* = C_2 \ln (n+1).$$

It has been found empirically that in most applications only the lower bound a^* is essential, though the upper bound improves the estimates somewhat. Following the discussion of the previous Section we can understand the function of the lower bound as limiting the size of the steps which we allow the Venter process to take.

We may generate the bounds (6) by using the $\{e_n\}$ sequence (5) with the multiplier C_1 for a^* and the sequence $\{H_n\}$ defined by

$$(7) \qquad H_n = \sum_{i=1}^{n} i^{-1}$$

with the multiplier C_2 for b^*. It is well known that

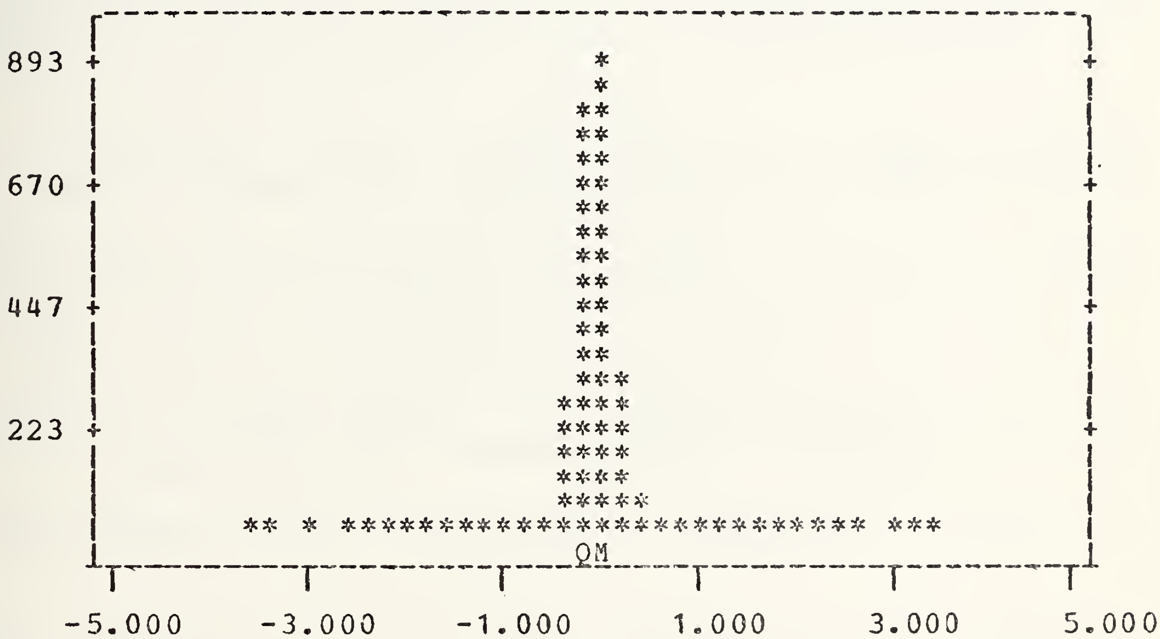
$$H_n = \ln n + \gamma + O(\, n^{-1} \,),$$

where $\gamma = 0.57722$ is Euler's constant; this approach is about 20 times faster than computing the logarithm directly but still preserves the asymptotic behavior required, for example, in the proof of Theorem 1 in Chapter II.

2. Simulation results

Considerable simulation effort was devoted to investigating optimum values for c, C_1 and C_2; in general, it was found that the Venter estimator is not very robust when random values are used and that it is difficult to select fixed values which give good results in a variety of applications. Figure 7 shows a typical example of the

Venter estimator with $c = C_1 = 1$ and $C_2 = 2$ applied to the 0.99 quantile of the exponential distribution. It was found that increasing the value of C_1 decreased the spread of the estimator somewhat while altering the value of C_2 seems to have little effect on the distribution of $\bar{s}'_n$.



Mean	3.523847E-02	Minimum	-3.523700E 00
Variance	1.692490E-01	Quartile	-1.163054E-01
Standard Deviation	4.113988E-01	Median	2.083635E-02
Skewness	7.713655E-01	Quartile	1.608515E-01
Kurtosis	2.362407E 01	Maximum	3.517703E 00

Figure 7. Bias of the Venter stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution based on an X sample of 5768 observations. Maximum transform with v = 56 used. Histogram sample = 2500 observations of $s^{*'}_{101}$.

Increasing C_1 does improve the distributional properties of the Venter quantile estimator but only at the cost of introducing considerable bias into the estimation process. In fact, the Venter estimator seems to be particularly bias-prone. In pseudo-random sampling experiments in which several quantiles from normal, uniform, exponential and gamma populations were estimated it was found that the Venter estimators had biases which were from 50 to 1000 times as high as those of the RM estimators.

A further drawback to this method may be seen in Figure 8 which displays the density estimate A'_n obtained in the same sampling experiment as the quantile estimates of Figure 7 (the notation A'_n indicates that the estimate is based on a maximum transform scheme). The negative estimate values for $\beta = f(s_a)$ are quite common for the Venter procedure, but they prevent us from obtaining any reasonable estimate of the variance of $\bar{s}'_n$. We denote $\mathrm{Var}[\bar{s}'_n]$ by σ^2_n and based on the asymptotic theory we estimate this variance by

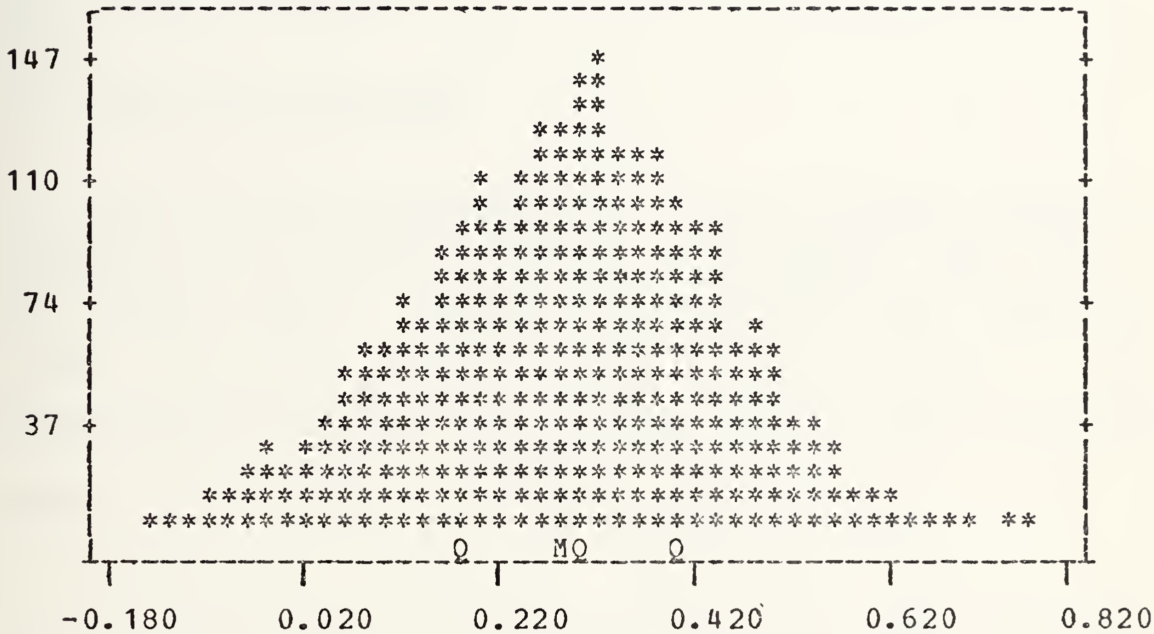
$$(8) \qquad \hat{\sigma}^2_n = \frac{v\check{a}_n(1-\check{a}_n)}{A'^2_n}$$

where v is the size of the maximum transform sample. Normally, the larger A'_n is the less variable is $\bar{s}'_n$ but when $A'_n < 0$ we can say very little about σ^2_n.

Note that the appearance of Figure 8 is quite Gaussian and that the mean of A'_n is very close to the theoretical value for the exponential 0.99 quantile

$$\beta' = va^{v-1} f(s_a)$$

$$= 0.3222 .$$



Mean	2.979725E-01	Minimum	-1.202695E-01
Variance	2.326651E-02	Quartile	1.942998E-01
Standard Deviation	1.525337E-01	Median	3.045059E-01
Skewness	-9.599537E-02	Quartile	4.029205E-01
Kurtosis	-2.614073E-01	Maximum	7.663841E-01

Figure 8. Venter estimator A'_{100} of $f(s_{0.99})$ for the exponential distribution based on the same experiment as Figure 7. True value is 0.3222. X sample = 5768 observations; histogram based on 2500 observations of A'_{100}.

The distribution of $\hat{A}'_n$ thus agrees with the asymptotic results of Venter [37] but the negative values are unacceptable for the determination of confidence intervals or for assessing the variability of $\bar{\xi}'_n$.

D. The New Estimator

1. Choice of parameters

To use the new estimator of Section I.E and Chapter II we must first decide on a number of parameters, just as in the Venter case. These decisions include the choice of a kernel function $W(\bullet)$ and a bandwidth sequence $\{b_n\}$ as well as the specification of the bounding method (2.4), analogous to the interval (a^*,b^*) for the Venter process.

Considerable experience with density estimators, both in this thesis and in [23], indicates that the triangular weight function

$$(9) \qquad W_t(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

gives results comparable to those of smoother kernels with some saving in computational efficiency. Other kernels investigated include the uniform

$$W_u(x) = \begin{cases} 1 & \text{if } -1/2 \le x \le 1/2 \\ 0 & \text{otherwise} \end{cases}$$

which is somewhat unstable and subject to bias, as well as the smoother quadratic weight function

$$W_2(x) = \begin{cases} 1.5\ (1 - x^2) & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the exponential weight function

$$W_e(x) = \frac{1}{2} e^{-|x|} .$$

All of these functions clearly satisfy assumptions (W1) to (W4) of Chapter II and so are admissible for stochastic approximation quantile estimation.

For the bandwidth sequence we again adopt the $\{e_n\}$ sequence used for the Venter case. Selection of $b_n = b\ e_n$ satisfies (2.7) with $g = 1/3$; once again, the savings in computation time make the use of the $\{e_n\}$ sequence very attractive. As an alternative we might use the sequence $\{e'_n\}$ based on the recursion

$$e'_1 = 1$$

$$e'_{n+1} = \left(1 - \frac{e'^2_n}{2} \right) e'_n$$

which may be shown to be $O(n^{-1/2})$. Since excellent results were obtained with $\{e_n\}$ this other sequence has not been investigated.

Selection of the bandwidth multiplier b must take into account the spread of the random variables. If too small a value is used it is unlikely that any X observations will fall close enough to the $\bar{s}_n$ values to make a contribution to

the density estimate. (Recall that

$$(10) \qquad w_n = \frac{1}{b_n} W_t \left[\frac{\bar{s}_n - X_n}{b_n} \right]$$

will be positive only if $|\bar{s}_n - X_n| < b_n$.) On the other hand, if b is too large it is possible that B_n will be unable to increase fast enough in a small sample to reach very large values of β.

Practical experience with the method shows it to be quite robust with respect to the choice of b; most of the work reported in this Chapter and in Chapter IV was done with a fixed b value of 1. In data where the observations are more widely dispersed than those considered here, it may be desirable to use a random value for b. If the nested method is used for finding $\bar{s}_1$ a convenient b value to use is

$$b = X'_{(3)} - X'_{(1)} \;;$$

using this value guarantees that further X' observations will be within a single bandwidth b_n of $\bar{s}'_n$.

The lower bound on the sequential estimate A_n of β in the Venter process was absolutely essential since the Venter technique sometimes results in negative A_n values. If these values were used, steps in the wrong direction would be taken and so a lower bound on the value of A_n must be established. For the new process all of the increments to

the density estimator B_n are positive, so that once a positive estimate is obtained we need not be concerned with this type of behavior. We may assure that the $\bar{s}_n$ estimator will be fairly stable by setting the initial value of B_n, which we call B_0, to a positive value: either some random a priori estimate of β or else a fixed number. The larger the bandwidth sequence multiplier b is, the smaller the value for B_0 we want to use. We thus set B_0 = 1/b whether b is fixed or random.

As mentioned above, we adopt here the fixed values $b = B_0 = 1$, i.e. we use the estimate B_n given by

$$B_n = \frac{1}{n} \left[1 + \sum_{j=1}^{n} w_j \right]$$

where w_j is given by (10). Note that this is equivalent to a lower bound with $C_1 = 1$ and $L = 1$; although this does not satisfy the requirements of (2.4) the results in all cases investigated so far do not seem to call for a more stringent method.

For an upper bound we again adopt the $\{H_n\}$ sequence used in the Venter case, using a C_2 value of 1. Although the upper bound makes very little difference in most cases it seems prudent to use it to avoid any possible instability in the early phases of the estimation process.

2. The basic stochastic approximation algorithm

A succinct description of the estimation process may now be given by setting forth its three phases as follows. (Note that the same basic method holds for both untransformed and maximum transformed estimators.) For notational simplicity, we write "m" for B_0 in the algorithm.

1. <u>Initialize.</u> Obtain the initial estimate $\bar{s}_1$ and the bandwidth multiplier m and initialize:

   $s = \bar{s}_1$; $f = 1 / m$;

   $n = 1$; $b = m$; $h = 1$.

2. <u>Update.</u> For each new X observation update the estimates as follows:

   a. <u>Density</u> Set $t = |s - X|$. If $t < b$ increase

   $$f = f + \frac{(b-t)}{b^2} .$$

   b. <u>Quantile</u> If $X \le s$ set $y = a-1$ otherwise set $y = a$. If $f > h \cdot n$ set $d = h \cdot n$ otherwise set $d = f$ (this is the upper bound operation). Finally adjust s according to

   $s = s + y / d$.

   c. <u>Constants</u> Update the constants for the next phase:

   $h = h + 1/n$; $n = n + 1$;

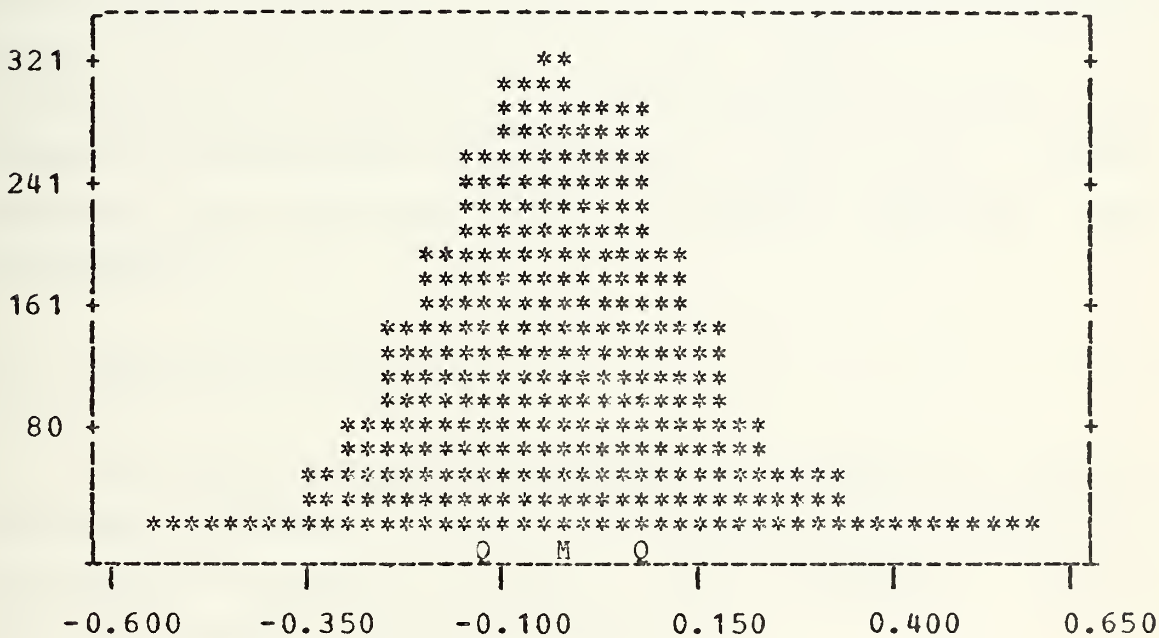
   $b = (1 - b^3 / 3m^3) b$.

3. <u>Results.</u> The final estimate of the a-quantile is s. An estimate of Var[s] is given by

   $$Var[s] = (n-1)\ a\ (1-a) / f^2,$$

while $f/(n-1)$ is an estimate of $f(s_a)$.

The process thus requires us to store just five variable values (s, f, n, b and h) and a pair of fixed values (a and m). After the k*th* X value has been used in step 2, s has the value $\bar{s}_{k+1}$, f is $k\,B_k$, n is k+1, b is e_{k+1}



Mean	-1.247959E-02	Minimum	-5.472403E-01
Variance	2.434824E-02	Quartile	-1.185350E-01
Standard Deviation	1.560392E-01	Median	-1.727819E-02
Skewness	1.905374E-01	Quartile	8.935070E-02
Kurtosis	2.595820E-02	Maximum	5.563519E-01

Figure 9. Bias of the stochastic approximation estimator for the 0.99 quantile of the exponential distribution using kernel density estimators. Total X sample = 5768 observations; maximum transform with v = 56 was used. Histogram based on 2500 observations of $s^{*\prime}_{101}$.

and h is H_{k+1}.

To carry out the maximum (minimum) transform with sections of size v, we use the value $a' = a^v$ ($a' = (1-a)^v$) in steps 2.b and 3 and carry out step 2 only for each of the section maxima (minima). The estimate of $f(s_a)$ in step 3 is then $f/[va^{v-1}(n-1)]$ {or $f/[v(1-a)^{v-1}(n-1)]$ for the minimum case}. Here we will require one more constant (v) to be stored as well as two more variables which keep track of the number of observations considered so far in the current section and the value of the maximum (minimum) value encountered.

3. Simulation results

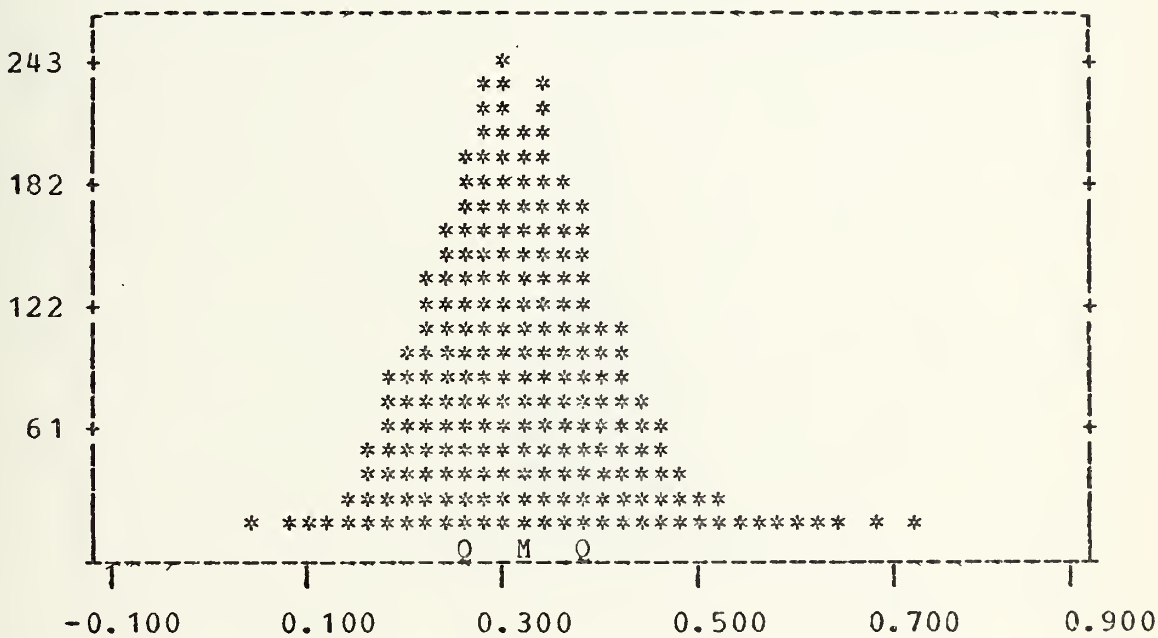
An example of the new stochastic approximation quantile estimator applied to the 0.99 quantile of the exponential distribution appears in Figure 9. The asymptotic variance for this maximum transformed case is

$$\text{(11)} \qquad \mathrm{Var}[\bar{s}'_n] = \frac{a'(1-a')}{n\{va^{v-1}f(s_a)\}^2}$$

$$= \frac{2.3615}{n}$$

or 0.02362 for n = 100. This corresponds quite closely to the observed value of 0.02435 and the shape of the histogram also appears reasonably Gaussian. We thus conclude that the asymptotic theory is a generally acceptable description of the behavior of the new stochastic approximation quantile estimation scheme for moderately large samples.

Comparing this distribution to that of the corresponding Venter estimate (Figure 7) we see that the new method results in an estimator whose properties are much more reasonable; the observed mean bias is less for the new estimator while the variance is smaller by a factor of 7. The distribution also appears much more Gaussian and the sample coefficients of skewness and kurtosis are smaller in

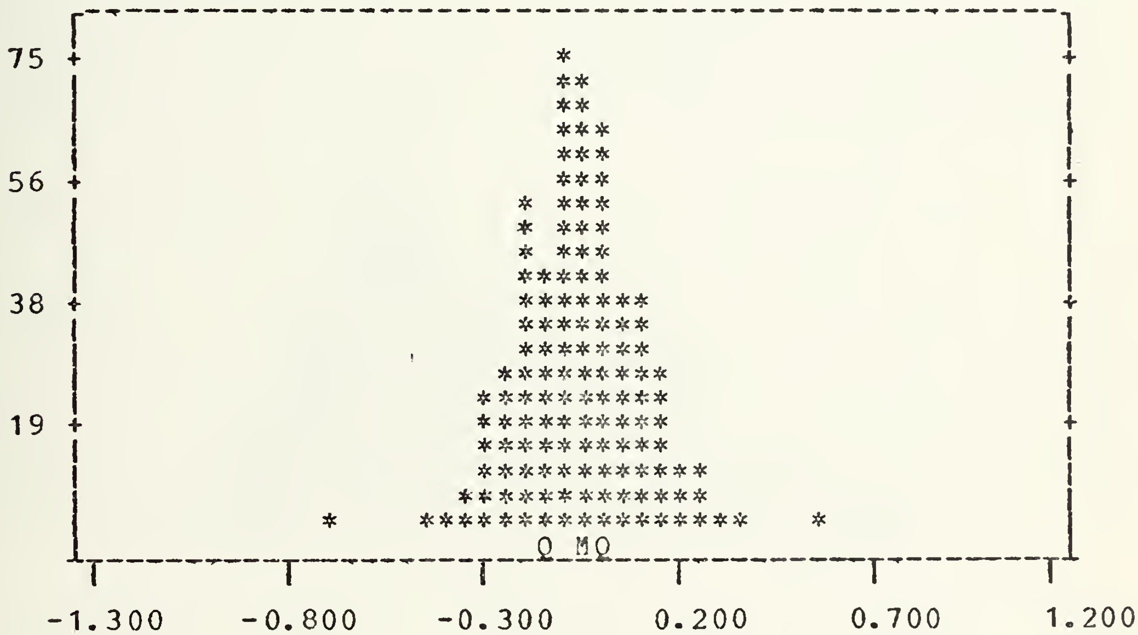


Mean	3.263939E-01	Minimum	5.802400E-02
Variance	7.812276E-03	Quartile	2.653056E-01
Standard Deviation	8.838707E-02	Median	3.204337E-01
Skewness	3.807945E-01	Quartile	3.818382E-01
Kurtosis	3.405670E-01	Maximum	7.326112E-01

Figure 10. Estimator B'_{100} of $va^{v-1} f(s_a)$ for the 0.99 quantile of the exponential distribution based on 5768 X observations. True value is 0.322. Histogram sample = 2500 realizations of B'_{100}.

Figure 9. We conclude that the new procedure is decidedly better than the Venter technique for quantile estimation.

Figure 10 shows the distribution of the density estimate B'_n (or f/(n-1) from the algorithm) which was obtained at the same time as the data of Figure 9. Once again the distribution appears approximately Gaussian and



Mean	-3.969568E-02	Minimum	-6.638889E-01
Variance	2.327731E-02	Quartile	-1.466389E-01
Standard Deviation	1.525691E-01	Median	-4.307795E-02
Skewness	1.872554E-01	Quartile	5.155754E-02
Kurtosis	8.735478E-01	Maximum	5.740843E-01

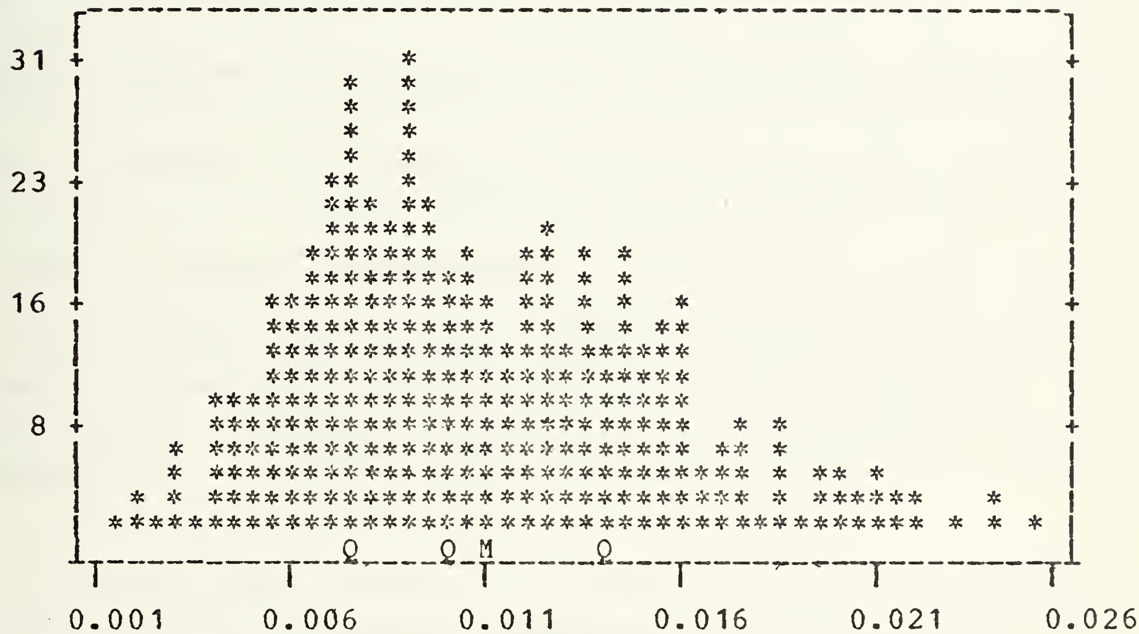
Figure 11. Bias of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution based on an X sample of 5768 observations. Maximum transformation was <u>not</u> used in this case. Histogram sample size = 2500 observations on s^*_{5601}.

the observed mean of 0.3264 is quite close to the theoretical value of 0.3222. On asymptotic grounds from Theorem 4 the variance should be

$$\mathrm{Var}[B'_n] = \frac{\beta' \int W^2(u)\,du}{(1+g)} \quad n^{g-1}$$

$$= \frac{1}{2}\,\beta' \; n^{-2/3}$$

$$= 7.478 \times 10^{-3}$$



Mean	1.108336E-02	Minimum	1.552738E-03
Variance	2.068462E-05	Quartile	7.708944E-03
Standard Deviation	4.548036E-03	Median	1.035528E-02
Skewness	5.468071E-01	Quartile	1.414721E-02
Kurtosis	-6.513911E-02	Maximum	2.519264E-02

Figure 12. Density estimate B_{5600} for the 0.99 quantile of the exponential distribution; based on an X sample of 5600 observations without maximum transform. Actual value is 0.01. Histogram sample is 2500 observations of B_{5600}.

which is very close to the observed value of 7.812×10^{-3}. Also there are no negative values of B'_n so that all of them are admissible as variance estimators.

The new estimator was also applied to the 0.99 quantile of the exponential distribution without using the maximum transform; the results appear in Figure 11. Clearly the new process is far more stable than either the RM or Venter methods; the distribution of $\bar{S}_{5600}$ is very nearly normal with an observed variance (0.02328) close to the asymptotic value (0.01768). The density estimate B_{5600} for this case is shown in Figure 12; the mean is close to the true value of 0.01 while the observed variance of 2.07×10^{-5} is also close to the asymptotic value of 1.59×10^{-5} although the distribution is skewed to the right and does not appear Gaussian.

Despite the results of Figures 11 and 12 we still prefer to use the maximum transformed version of the new process both because it is computationally faster and because its finite sample properties are generally superior, especially for quantiles more extreme than the 0.99. It is nevertheless encouraging to find the new process sufficiently stable to avoid the very heavy tails displayed by the untransformed RM estimator (see Figures 2 and 3).

4. The stability of the new estimator

An explanation of the stability displayed in Figure 11 follows if we consider the role of the variable f in the algorithmic description of the new method given above. Recall that $f = n B_n$, i.e. it is the divisor in the basic stochastic approximation recurrence relation. Now f will increase at each step when we use the triangular kernel function only as long as the latest X_n observation is close to $\bar{s}_n$. If f does not increase, however, the size of the steps taken by the process will remain the same; we thus have an analog to the accelerated process of Kesten [18] where the step size remains constant until we have straddled the true value by taking steps in both directions.

The new method is an improvement on Kesten's technique because the step size adjustment here is made for each X observation. Instead of determining that the estimator $\bar{s}_n$ is in the vicinity of s_a by looking at the changes in step direction we examine directly the relationship between X_n and $\bar{s}_n$. For example, if $\bar{s}_1$ is a long ways from s_a so that none of the X observations are near $\bar{s}_j$ for small j values then the process will take steps of size $1/B_0 = 1$ until it

reaches a point where $\bar{s}_n$ is close to the latest observation value X_n. Once the $\bar{s}_n$ values are close to the X_n observations the w_n terms added to f will be positive and so the step size will decrease.

5. Confidence intervals

The final area to be investigated here is that of applying the new estimation procedure to the determination of confidence intervals on s_a. To obtain a 100 p % confidence interval on s_a we use

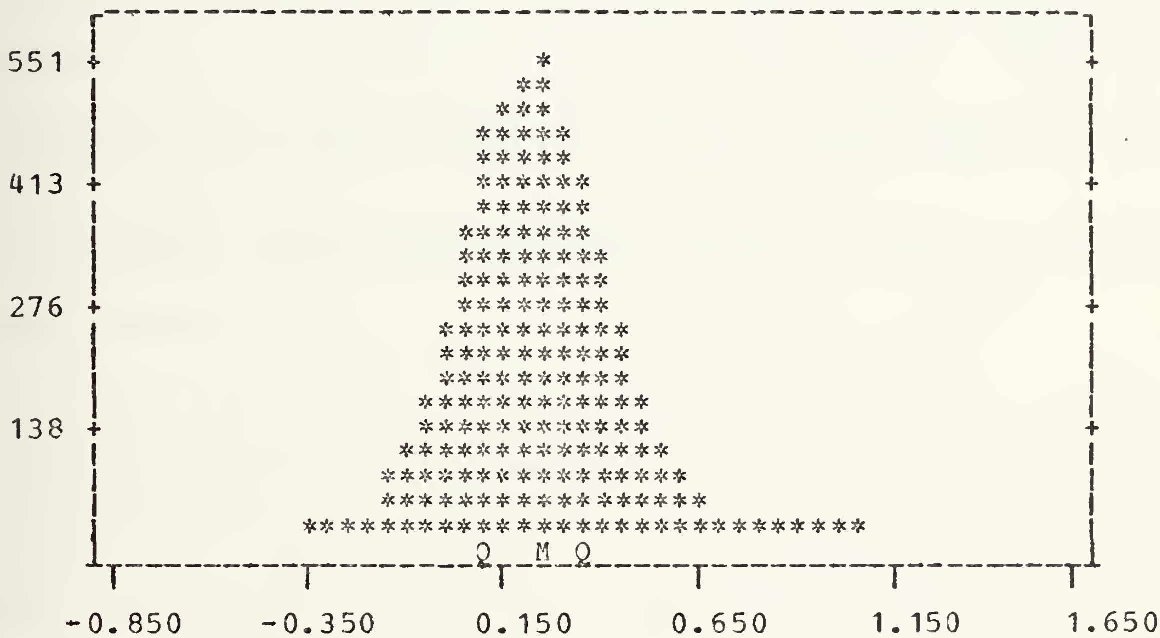
$$\text{(12)} \qquad \bar{s}_{n+1} \pm \sqrt{\frac{a(1-a)}{n}}\; B_n^{-1}\, u_p$$

where u_p is the upper 1 - p/2 point of a standard normal random variable. It would be possible to establish the asymptotic properties of confidence intervals estimated in this way following the work of Sielken [34]; this has not been done here.

To investigate the finite sample properties of the confidence intervals in the exponential 0.99 case, however, further simulation experiments were undertaken. Based on 10,000 replications the coverage of the confidence interval (12) for various p values was as follows:

p	Actual Coverage
0.90	0.8777 ± 0.0033
0.95	0.9265 ± 0.0026
0.99	0.9755 ± 0.0015

The data of Figure 13 show the distribution of the upper 95 % confidence limit (with the mean of 4.605165 subtracted) for a sample of 5768 X observations. On asymptotic grounds, the expected value for this limit should be 0.25271 which corresponds very well to the observed mean of 0.25623.



Mean	2.562305E-01	Minimum	-3.441450E-01
Variance	3.547240E-02	Quartile	1.279681E-01
Standard Deviation	1.883412E-01	Median	2.511141E-01
Skewness	2.462817E-01	Quartile	3.741887E-01
Kurtosis	2.908053E-01	Maximum	1.082688E 00

Figure 13. Value of the upper 95 % confidence limit for the 0.99 quantile of the unit exponential distribution; the true value of 4.605165 has been subtracted from each observation. Estimated by stochastic approximation from X samples of 5768. Histogram sample size = 2500.

6. Summary

The new estimator has been used to estimate all the quantiles in Table I for random variables from the uniform, normal, exponential, gamma and Cauchy distributions. So much data was obtained that it would be impractical to attempt to display it all here; the results were, with few exceptions, in general agreement with those shown here for the exponential 0.99 case. Serious irregularities were noted in the Cauchy case; these were due to the infinite variance of the initial estimate $\bar{s}_1$. When the Cauchy experiment was repeated with the fixed starting value $\bar{s}_1 = 0$, however, reasonable agreement with the asymptotic theory was obtained.

The other major limitation found was in using the maximum transform for the estimation of extreme quantiles from distributions whose densities do not approach zero in one or both tails; examples include the uniform distribution and the left-hand tail of the exponential distribution. In these cases the transformed density β' is very large - 31.90 for the 0.01 quantile of the exponential distribution, for example - and it requires very large X samples for the value of B'_n to increase sufficiently to obtain good estimates of β'. The resulting $\bar{s}_n$ values have distributions which agree with the asymptotic theory, but the too-small density estimates result in confidence intervals which are much too wide. In other words, in this

situation the point estimator of s_a is satisfactory but the density estimate (and hence the confidence interval) is relatively poor.

We conclude this Chapter with the observation (based on the above digital simulation experience) that the new method overcomes most of the limitations of stochastic approximation techniques for quantile estimation. The asymptotic theory appears to be an adequate description of the behavior of the estimators in samples large enough to give reasonable variances and we are confident enough of the distribution of $\bar{s}_n$ that we may use the estimate of $f(s_a)$ for the construction of confidence intervals.

Chapter IV. BIAS AND MEAN SQUARED ERROR

In the previous Chapter we examined the problem of finite sample performance of stochastic approximation quantile estimates by investigating the <u>distribution</u> of the difference $s^*_n = \bar{s}_n - s_a$, which we refer to hereafter as the bias of the estimator. Considering the distribution of s^*_n was done because simply looking at its expected value is not sufficient if one is to explain the extremely poor performance of some stochastic approximation quantile estimators. As illustrated by Figures 2 and 3, this poor performance is characterized by very heavy tails and exceptionally wide dispersion of s^*_n. By using the maximum transform, however, and the new technique of Section III.D, we were able to overcome these drawbacks and obtain estimates $\bar{s}_n$ whose distribution is approximately Gaussian.

Bias is usually taken to be $E[s^*_n]$ and once the problem of extremely large deviations has been overcome it is necessary to look at bias in this average sense. This is because one facet of the poor performance of stochastic approximation quantile estimators is that convergence of $E[\bar{s}_n]$ to the true value s_a is very slow as measured empirically even though the estimates are asymptotically

unbiased. In fact, Yuguchi [38] found empirical evidence that the rate of convergence of the bias is $O(n^{-1/4})$ for the stochastic approximation estimator proposed by Goodman, Lewis and Robbins [14]. This compares with $O(n^{-1})$ for the order statistic case.

We examine this question here for the new estimator through simulation because no analytical results are available or easily obtained. Our goal is to determine whether the bias converges as $n^{-1/2}$ as indicated by the theory or whether the rate of convergence is slower, as indicated by Yuguchi [38]. By developing a model for the convergence of the bias, we will be able to compare stochastic approximation estimators with order statistic estimators; we may also be able to use techniques such as the jackknife [27] to reduce the bias in situations where it is significant.

A. Description of the Model

In a general statistical problem, if T_n is an estimator of the fixed but unknown parameter θ based on a sample of size n then we have for the mean squared error of T_n

$$(1) \qquad MSE[T_n] = E[(T_n - \theta)^2]$$

$$= \{E[T_n - \theta]\}^2 + \mathrm{Var}[T_n],$$

where the first term is due to estimator bias and the second to sampling variation. Now it may be that T_n converges weakly (i.e., in distribution) to a random variable T (which is often normal) and also that $MSE[T_n] \to M$. (In either event we may have T_n suitably normalized, e.g. $n T_n \to T$.) We may thus choose either MSE[T] or M as a measure of the expected error of the estimator. Hodges and Lehmann [15] point out that $MSE[T] \leq M$ and that strict inequality is possible.

For the stochastic approximation quantile estimation problem, the result of Lemma 9 in Chapter II implies that

$$(2) \qquad n\, MSE[\bar{s}_n] \to M \geq 0$$

while the asymptotic normality result of Theorem 3 shows

$$(3) \qquad n\, MSE[S] = \frac{a(1-a)}{\beta^2},$$

where $\bar{s}_n$ is weakly convergent to S. Now similar results exist for the basic RM process [7] as well as for the Venter method [36]; the asymptotic variance (3) is the same in all three cases as long as we select $A = \beta$ for RM. Thus to assess the practical utility of any given stochastic approximation method which has a suitably Gaussian distribution we attempt to measure the value of M which results when we sample from a population with known properties, e.g. independent and identically distributed exponential variates.

Hodges and Lehmann [15] have found for a linear model of the RM process that the mean square error components of (1) result from a bias term related to the squared error of the initial estimate and a variance term related to the asymptotic variance. The quantile estimation problem does not satisfy the hypotheses of the Hodges and Lehmann model but those authors state that some Monte Carlo experimentation has indicated that their results are fairly robust. We thus begin our analysis of stochastic approximation quantile estimation with the assumption that the differences between methods will be due to differing estimator bias.

In view of (2), we have that the bias of $\bar{s}_n$ is $O(n^{-1/2})$ and we adopt the model

$$(4) \qquad E[s^*_n] = r_0 + r_1 n^{-1/2} + r_2 n^{-1} + o(n^{-1}),$$

in accordance with the Hodges and Lehmann results. (Recall that $s^*_n = \bar{s}_n - s_a$.) We recognize that (4) must be empirically validated before it can be applied in a specific case. Despite the Hodges and Lehmann result, it is possible that terms of other orders (such as $n^{-3/4}$ or $n^{-1/2}/\log n$) may be present. Nevertheless, this model provides a convenient means of assessing the relative bias of different stochastic approximation estimators.

One possible objection to (4) can be raised based on the results of Yuguchi [38] who found that there was a

significant $n^{-1/4}$ term in the bias of stochastic approximation quantile estimators. Following Goodman, Lewis and Robbins [14], Yuguchi used the basic RM process with a fixed random divisor $\bar{A}$. The problem with this approach is that due to sampling variation $\bar{A}$ will sometimes be larger than 2β and then, according to the results of Sacks [33] and Major and Revesz [26], the convergence of $\bar{s}_n$ to s_a may be much slower than the $n^{-1/2}$ implied by (1.13). Lemma 9 guarantees that this situation will not exist with the new estimator; however, it is prudent to see whether the simulation results show bias terms of a lower order than $n^{-1/2}$ and also to compare the model (4) with alternative schemes.

The estimation of r_0, r_1 and r_2 from specific realizations of $\{\bar{s}_n\}$ is a difficult problem because of the high degree of autocorrelation within any stochastic approximation process, i.e. between $\bar{s}_n$ and $\bar{s}_{n+1}$. The general design problem of assessing the model (4) with dependence has not been addressed. To overcome this strong dependence we generate n independent realizations of the process:

$$\begin{array}{ll} 1: & s^*_1 \\ 2: & s^*_1, s^*_2 \\ & \cdots \\ n: & s^*_1, s^*_2, \ldots, s^*_{n-1}, s^*_n \end{array}$$

and select as our sample the final estimate value in each realization. The result is a sample $\{s^*_1, s^*_2, \ldots, s^*_n\}$ of independent random variables; note that a total of n different starting values and $\frac{n(n-1)}{2}$ observations of X from the parent population are required to obtain each independent sample. If we are using the maximum transform (as we will be throughout this Chapter) each new $s^{*\prime}_n$ value will be based on v observations of X so that the total X sample will consist of $\frac{vn(n-1)}{2}$ values. We repeat this scheme to obtain m independent $\{s^{*\prime}_n\}$ samples; $s^{*\prime}_{k;i}$ will denote the bias of $\bar{s}'_k$ in the $i\underline{th}$ independent sample.

The evaluation of a specific stochastic approximation method with respect to bias will then consist of estimating the value of r_1 subject to some sort of validation effort. (Note that (1) and (2) imply that $r_0 = 0$.) We then obtain the required estimates $\bar{r}_0$, $\bar{r}_1$ and $\bar{r}_2$ by generalized least squares from the linear model

$$(5) \qquad s^*_n = r_0 + r_1 n^{-1/2} + r_2 n^{-1} + v_n; \quad n = 1,2,\ldots,$$

where the v_n's are mutually independent random variables with

$$(6) \qquad E[v_n] = 0, \qquad Var[v_n] = \sigma^2/n .$$

In this formulation, σ^2 is unknown and is also to be estimated; one criterion of the adequacy of the model (4) will then be how closely we approach the asymptotic value

$$\sigma^2 = \frac{a(1-a)}{\beta^2} .$$

B. A Variance Reduction Scheme

When using the new estimator with the maximum transform to estimate $s_{0.99}$ for the unit exponential distribution, one finds that the bias is about -0.007 for X samples of size 7000 (i.e., about 125 maximum transformed steps with v = 56). The asymptotic standard deviation in this case is 0.137 from (3.11). Thus to determine the bias for each maximum transformed step to within a sampling variation equal to one-tenth of the absolute value of the bias requires a total of

$$m = \left[\frac{0.137}{0.0007} \right]^2 \doteq 38,500$$

replications of the independent $\{s^{*\prime}_n\}$ sequences of the previous Section.

The amount of work required by this naive approach leads us to investigate methods of reducing the sampling

variation of s^*_n without changing its expected value. The classical simulation techniques of variance reduction represent an obvious means of doing this; for more details on these methods see Gaver and Thompson [13]. The approach we adopt here is to define a control variate P_n which is a statistic computed from the same X sample used to find s^*_n and which is highly correlated with s^*_n. The technique can be applied with or without the maximum transform.

In general we choose as our control variate P_n a statistic whose distribution (or at least whose moments) we can find. As our estimate of the bias $E[s^*_n]$ we then use

$$(7) \qquad s^+_n = s^*_n + P_n - E[P_n],$$

where $E[P_n]$ is known. Clearly

$$E[s^+_n] = E[s^*_n]$$

$$Var[s^+_n] = Var[s^*_n] + Var[P_n] + 2\ Covar[s^*_n, P_n],$$

so that if P_n is negatively correlated with s^*_n there may be a decrease in the variance. One way to insure that there will be such a decrease is to use instead the value

$$(8) \qquad s^+_n = s^*_n + \Pi_n\{P_n - E[P_n]\}$$

where the constant Π_n is chosen to minimize $Var[s^+_n]$. Note that the estimate (8) is also a variance reduced estimate

using $\pi_n P_n$ as a control variate so we are justified in using the same symbol s^+_n as for the estimate (7).

As Gaver and Thompson [13] show, the optimum value of π_n is given by

$$\pi_n = -\operatorname{Covar}[s^*_n, P_n] / \operatorname{Var}[P_n];$$

the resulting variance of s^+_n is then given by

$$(9) \qquad \begin{aligned} \operatorname{Var}[s^+_n] &= \operatorname{Var}[s^*_n] - \operatorname{Covar}[s^*_n, P_n] / \operatorname{Var}[P_n] \\ &= \operatorname{Var}[s^*_n]\,(1 - \rho^2_n), \end{aligned}$$

where ρ_n is the correlation between s^*_n and P_n. Thus if P_n is highly correlated with s^*_n we may expect substantial improvement in the variance of our final result.

Of course we will not in general know $\operatorname{Covar}[s^*_n, P_n]$ (although $\operatorname{Var}[P_n]$ will sometimes be known) and so we are unable to choose the optimum value for π_n; we may estimate the optimum, however, by using

$$(10) \qquad \hat{\pi}_n = \frac{-\sum_{i=1}^{m} (s^*_{n;i} - \bar{\mu}[s^*_n])(P_{n;i} - E[P_n])}{\operatorname{Var}[P_n]}$$

where $\bar{\mu}[s^*_n]$ is the mean of the m realizations of s^*_n. When we use the $\hat{\Pi}_n$ value given by (10) in (8), however, the resulting s^+_n is no longer an unbiased estimate of $E[s^*_n]$, although as Gaver and Thompson [13] point out we expect the bias to decrease with increasing m.

It has been found that the values of Π_n do not change very much with n, at least not when n is moderately large. Since by the design of the simulation experiment $s^*_{j;i}$ and $s^*_{k;i}$ are based on disjoint X samples for $j \neq k$, the value

$$(11) \qquad \hat{\hat{\Pi}}_n = (\hat{\Pi}_{n-1} + \hat{\Pi}_{n+1}) / 2$$

will be independent of s^*_n and P_n and therefore it will not cause s^+_n to be biased. Furthermore, for large m values it should be close enough to Π_n to allow for a close approach to the variance reduction (9).

The foregoing analysis applies no matter which control variate P_n we choose. The art in control variate variance reduction lies in choosing a suitable P_n; a good choice will be easy to compute sequentially from the X sample, will have

known moments and will be highly correlated with s^*_n. One such choice for the stochastic approximation quantile estimation problem is to use an estimate of the s_a-percentile, i.e. we take

(12) $$P_n = \{\text{Number of X values} \le s_a\} \,/\, n .$$

Since we are performing a synthetic sampling experiment, s_a is known and from the definition of s_a we conclude that $n\,P_n$ has a binomial distribution with parameters a and n. Furthermore,

$$E[P_n] = a$$

$$Var[P_n] = a\ (1 - a)\ /\ n.$$

Now if the observed value of P_n is greater than a we expect the X values in the sample to be larger than usual and consequently the value of $\bar{s}_n$ to be larger than s_a. This conjectured positive relationship between P_n and $\bar{s}_n$ (or, equivalently, s^*_n) is borne out in sampling experiments; what is surprising is the very high correlation coefficient observed between these two random variables in many applications. For example, in the case of the 0.99 quantile of the exponential distribution we observe correlations as high as 0.90 for moderate values of n; this results in variance reductions of about 80 % based on (9). This in turn leads to confidence intervals on $E[s^*_n]$ which are just

40 % as wide as those obtained using the uncontrolled s^*_n values.

A plot of a joint simulation sample of p_n and $s^{*\prime}_n$ for the exponential 0.99 quantile is shown in Figure 14. The X sample in this case was 5768 observations which corresponds to 100 maximum transform steps (v = 56). A total of 2500 replications were generated to produce this plot. The computer program used to produce Figure 14 is typical of the software tools developed in the course of this research; other examples include the histogram Figures of Chapter III and the histogram plots of Section IV.D.

C. Regression Analysis

For the purposes of analysis we adopt the general bias model

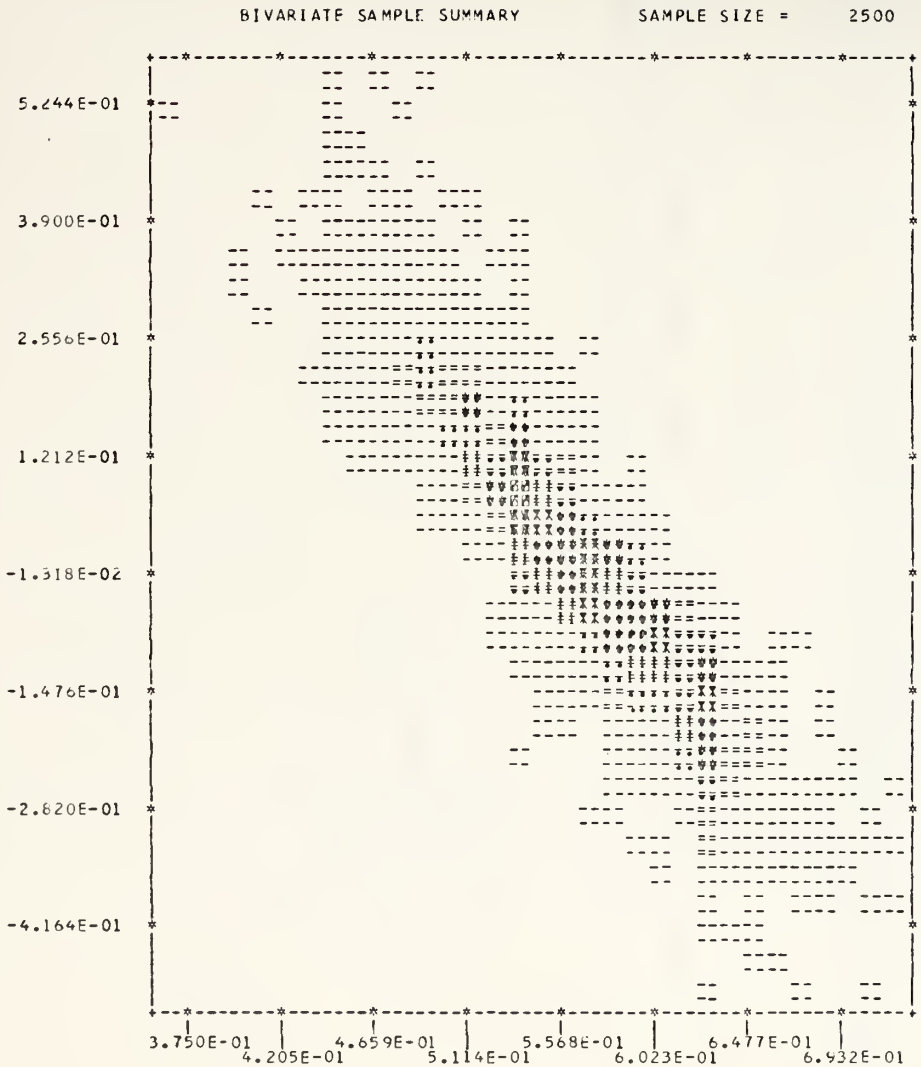
$$(13) \qquad E[s^*_n] = \sum_{j=0}^{k} r_j \, g_j(n),$$

where $g_0(n) = 1$ for all n and $g_j(n)$, $j \geq 1$, is some function of n; for example

$$g_j(n) = n^{-j/2}; \quad j = 1, 2,$$

corresponds to the model (4).

To estimate the r_j's in (13), we obtain a set of m_n independent realizations of s^*_n for n = L, L+1, ..., N and then use generalized least squares with the relation



KEY

SYMBOL PRINTED		-	=	‡	≑	ǂ	♦	•	X	✱	Ж	₮	▤	▥	▦	■
NO. OBSERVATIONS	0	9	14	19	24	29	34	39	44	49	54	59	64	69	74	79

MEASURES OF ASSOCIATION

COVARIANCE	-7.375896E-03
CORRELATION COEFFICIENT	-0.892803
SPEARMAN'S RANK CORRELATION COEF.	-0.907612

TESTS FOR EQUIDISTRIBUTION

	TEST STATISTIC	NORMALIZED STATISTIC
KOLMOGOROV-SMIRNOV TEST	0.992000	-70.01056
WALD-WOLFOWITZ RUNS TEST	25	
MANN-WHITNEY TEST	6246490	61.16231
WILCOXSON TEST	9372740	61.16231
SIEGEL-TUKEY TEST	6257690	0.12618

UNIVARIATE STATISTICS

	X SAMPLE	Y SAMPLE
MEAN	5.687999E-01	-7.793173E-03
MEDIAN	5.656562E-01	-1.269960E-02
VARIANCE	2.614161E-03	2.610874E-02
STD DEV	5.112887E-02	1.615820E-01
RANGE	3.636363E-01	1.075222E 00
SKEWNESS	-1.260945E-01	1.640709E-01
KURTOSIS	-3.474045E-02	8.385277E-02
MAXIMUM	7.272726E-01	5.580232E-01
MINIMUM	3.636363E-01	-5.171928E-01

FIGURE 14. JOINT DISTRIBUTION PLOT FOR S^*_{100} (Y-AXIS) AND P_{100} (X-AXIS) FOR THE EXPONENTIAL 0.99 QUANTILE. MAXIMUM TRANSFORM WITH V = 56 WAS USED.

$$(14) \qquad s^*_{n;i} = \sum_{j=0}^{k} r_j g_j(n) + v_{n;i}; \qquad n = L, L+1, \ldots, N; \quad i = 1, 2, \ldots, m_n.$$

As before, we assume that the $v_{n;i}$'s are independent random variables with zero mean and variance proportional to $1/n$; we choose L large enough that we may invoke the asymptotic distribution of s^*_n to claim a normal distribution for v_n. This will allow us to apply the ususal F and t tests in the regression.

To apply generalized least squares to (14) we multiply the relation by $\sqrt{n}$; the random errors in the transformed equation are now independent with zero mean and common variance σ^2. We express this transformed relationship in the compact form

$$(15) \qquad \mathbf{s} = \mathbf{G}\,\mathbf{r} + \mathbf{v},$$

where boldface lower case letters represent vectors and upper case ones, matrices. We define

$$\mathbf{s} = \begin{bmatrix} s_L \\ s_{L+1} \\ \cdots \\ s_N \end{bmatrix}; \qquad s_n = \sqrt{n} \begin{bmatrix} s^*_{n;1} \\ s^*_{n;2} \\ \cdots \\ s^*_{n;m_n} \end{bmatrix};$$

$$n = L, \ldots N;$$

$$G = \begin{bmatrix} G_L \\ G_{L+1} \\ \cdots \\ G_N \end{bmatrix} ; \quad G_n = \sqrt{n} \begin{bmatrix} 1 & g_1(n) & \cdots & g_k(n) \\ 1 & g_1(n) & \cdots & g_k(n) \\ & \cdot & \cdot \; \cdot \; \cdot & \\ 1 & g_1(n) & \cdots & g_k(n) \end{bmatrix}$$

$$n = L, \ldots, N;$$

Note that G_n has m_n identical rows.

$$r = \begin{bmatrix} r_0 \\ r_1 \\ \cdots \\ r_k \end{bmatrix} ;$$

$$v = \begin{bmatrix} v_L \\ v_{L+1} \\ \cdots \\ v_N \end{bmatrix} ; \quad v_n = \sqrt{n} \begin{bmatrix} v_{n;1} \\ v_{n;2} \\ \cdots \\ v_{n;m_n} \end{bmatrix} ;$$

$$n = L, \ldots, N.$$

The least squares estimate of r is then

$$(16) \qquad \bar{r} = (G^T G)^{-1} G^T s ,$$

while an estimate of σ^2 from the residual sum of squares is given by the well-known relationship

$$(17) \qquad \hat{\sigma}^2 = \frac{s^T s - \bar{r}^T G^T s}{M - (K+1)} ,$$

where M is defined as the total number of s^*_n observations, i.e.

$$M = \sum_{n=L}^{N} m_n .$$

Some straightforward analysis then establishes that

$$(18) \qquad G^T G = \sum_{n=L}^{N} G_n^T G_n$$

$$= [g_{ij}] ; \qquad i,j=0,1,\ldots,k;$$

where the general element of the matrix is given by

$$g_{ij} = \sum_{n=L}^{N} n \; m_n \; g_i(n) \; g_j(n) \; ;$$

note that $G^T G$ depends only on the model selected and not at all on the observed s^*_n values. We also have that

$$(19) \qquad G^T s = \sum_{n=L}^{N} G_n^T s_n$$

$$= [y_j]; \qquad j=0,1,\ldots,k .$$

In this case the general term is

$$y_j = \sum_{n=L}^{N} [n \; g_j(n) \sum_{i=1}^{m} s^*_{n;i} \;]$$

$$= \sum_{n=L}^{N} n \; m_n \; g_j(n) \; \bar{\mu}[s^*_n],$$

where $\bar{\mu}[s^*_n]$ is the mean of the m_n observations of s^*_n.

Finally,

$$(20) \qquad s^T s = \sum_{n=L}^{N} s_n^T s_n$$

$$= \sum_{n=L}^{N} n \sum_{i=1}^{m} s^{*2}_{n;i}$$

$$= \sum_{n=L}^{N} n\, m_n\, \bar{\mu}_2[s^*_n]$$

where $\bar{\mu}_2[s^*_n]$ is the sample second moment of the s^*_n observations.

As indicated above we expect $\sqrt{n}\, v_{n;i}$ to be normally distributed, or approximately so. Thus it will be reasonable to use F-tests to test the significance of the regression and also to compute multiple correlation coefficients as long as the transformed equation (15) contains a constant term. This will be the case only if one of the functions $g_j(n)$ is equal to $1 / \sqrt{n}$ for some j. We will then also require the value

$$(21) \qquad D = \sum_{n=L}^{N} \sqrt{n}\, m_n\, \bar{\mu}[s^*_n]$$

for use in the analysis of variance table in the regression.

We may thus accumulate data for the regression by recording m_n, $\bar{\mu}[s^*_n]$ and $\bar{\mu}_2[s^*_n]$ for the n values of interest. The necessary regression values are computed by means of (18)-(21) and may then be used to estimate r and σ^2 according to (16) and (17). This means that we may deal

with arbitrarily large values of m_n with a relatively modest (and fixed) amount of memory. Furthermore, we may estimate the parameters for several models with the same simulation output values.

When we substitute the control variate estimate s^+_n for s^*_n in this analysis we obtain random errors v^+_n which still have zero mean but whose variance properties are unknown. From (9) we have

$$\mathrm{Var}[s^+_n] = \mathrm{Var}[s^*_n]\ (1 - \rho^2_n)$$

$$\doteq \frac{a(1-a)}{n\ \beta^2}\ (1 - \rho^2_n)$$

so that $\mathrm{Var}[s^+_n]$ decreases at least as quickly as $1/n$. We adopt the hypothesis, then, that $\mathrm{Var}[s^+_n] = O(n^{-1})$, recognizing that we will have to validate the conjecture based on the simulation output. The constant of proportionality in this case will be less than σ^2 because of the variance reduction obtained through the use of s^+_n. Since the control variate P_n will have a distribution close to normality for moderately large values of n, we expect the distribution of v^+_n to be once again approximately Gaussian.

D. Simulation and Regression Results

A summary of the output from a simulation in which $s_{0.99}$ for the exponential distribution was estimated appears in Table V. The estimation used the algorithm of Section

III.D.2 and the maximum transform with v = 56 (see Figure 9 in Chapter II for an example of the distribution of $s^{*\prime}_{101}$ in this case). Values of n (i.e., number of steps) ranging from 1 to 150 were investigated with m_n = 40,000 replications per step. A regression using all this data will thus have 6,000,000 degrees of freedom.

The first question we address here is whether the observed variances are adequately described by our assumption of σ^2/n or not. The variance of $s^{*\prime}_n$ is asymptotically 2.361 / n (see (2.11)) but the order of the variance of $s^{+\prime}_n$ is in general unknown. A simple linear regression on the data of Table V shows, however, that

$$\mathrm{Var}[s^{*\prime}_n] = -0.00003 + \frac{2.53712}{n} .$$

$$\mathrm{Var}[s^{+\prime}_n] = -0.00322 + \frac{0.87762}{n} .$$

The regressions are both significant (respective F-ratios are 141,550 and 12,311) and neither suffers from lack of fit; we thus conclude that our assumption of a 1 / n factor in the variances of s^*_n and s^+_n is justified. The relative sizes of the coefficients of the 1/n terms indicate that the control variate scheme results in a 35 % variance reduction in this case; in fact, 0.87762/2.53712 = 0.346.

n	X Sample Size	$\bar{\mu}[s^{*\prime}_n]$	$\bar{\sigma}^2[s^{*\prime}_n]$	$\bar{\mu}[s^{+\prime}_n]$	$\bar{\sigma}^2[s^{+\prime}_n]$
1	168	-0.11274	0.65112	-0.11274	0.65112
2	224	-0.04441	0.55387	-0.04338	0.45698
3	280	-0.01716	0.45439	-0.01847	0.32622
4	336	-0.01009	0.39033	-0.01015	0.25298
5	392	-0.00954	0.34222	-0.00719	0.20001
6	448	-0.00673	0.30674	-0.00708	0.17032
7	504	-0.00691	0.27737	-0.00625	0.14684
8	560	0.00241	0.25076	-0.00059	0.12566
9	616	0.00133	0.23395	-0.00124	0.11364
10	672	-0.00245	0.21860	-0.00450	0.10212
11	728	-0.00734	0.20184	-0.00742	0.09279
12	784	-0.00821	0.18674	-0.00506	0.08458
13	840	-0.00710	0.17652	-0.00840	0.07819
14	896	-0.00600	0.16646	-0.00623	0.07073
15	952	-0.00662	0.15746	-0.00603	0.06600
16	1008	-0.00553	0.14916	-0.00604	0.06108
17	1064	-0.00872	0.14253	-0.00854	0.05872
18	1120	-0.00928	0.13442	-0.00835	0.05400
19	1176	-0.00391	0.12725	-0.00760	0.05001
20	1232	-0.00760	0.12209	-0.00777	0.04712
21	1288	-0.00901	0.11687	-0.00887	0.04490
22	1344	-0.00596	0.11210	-0.00872	0.04184
23	1400	-0.01158	0.10711	-0.00964	0.04074
24	1456	-0.00847	0.10286	-0.00921	0.03823
25	1512	-0.00955	0.10043	-0.00744	0.03684

Table V. Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform ($v = 56$) were used.

n	X Sample Size	$\bar{\mu}[s^{*'}_n]$	$\bar{\sigma}^2[s^{*'}_n]$	$\bar{\mu}[s^{+'}_n]$	$\bar{\sigma}^2[s^{+'}_n]$
26	1568	-0.00889	0.09470	-0.00936	0.03448
27	1624	-0.00865	0.09208	-0.00894	0.03272
28	1680	-0.00670	0.08803	-0.00878	0.03116
29	1736	-0.00897	0.08668	-0.00880	0.03046
30	1792	-0.00807	0.08380	-0.00852	0.02841
31	1848	-0.00914	0.08040	-0.01007	0.02738
32	1904	-0.00938	0.07866	-0.00931	0.02660
33	1960	-0.01160	0.07589	-0.00981	0.02526
34	2016	-0.00945	0.07496	-0.00881	0.02488
35	2072	-0.01082	0.07106	-0.00923	0.02275
36	2128	-0.01093	0.06944	-0.01005	0.02256
37	2184	-0.00918	0.06846	-0.00906	0.02181
38	2240	-0.00832	0.06662	-0.00806	0.02086
39	2296	-0.00976	0.06461	-0.01057	0.01993
40	2352	-0.00931	0.06278	-0.00918	0.01965
41	2408	-0.00927	0.06100	-0.00872	0.01875
42	2464	-0.00945	0.06044	-0.00918	0.01823
43	2520	-0.00915	0.05885	-0.00964	0.01793
44	2576	-0.01071	0.05782	-0.00982	0.01717
45	2632	-0.01146	0.05545	-0.00977	0.01640
46	2688	-0.00819	0.05434	-0.00822	0.01603
47	2744	-0.00989	0.05362	-0.00901	0.01599
48	2800	-0.00689	0.05316	-0.00902	0.01545
49	2856	-0.01275	0.05165	-0.01001	0.01491
50	2912	-0.00889	0.05052	-0.00960	0.01434

Table V. (Continued) Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform ($v = 56$) were used.

n	X Sample Size	$\bar{\mu}[s^{*'}_n]$	$\bar{\sigma}^2[s^{*'}_n]$	$\bar{\mu}[s^{+'}_n]$	$\bar{\sigma}^2[s^{+'}_n]$
51	2968	-0.01104	0.04972	-0.00994	0.01394
52	3024	-0.01072	0.04852	-0.00934	0.01369
53	3080	-0.00788	0.04734	-0.00798	0.01326
54	3136	-0.00847	0.04653	-0.00956	0.01297
55	3192	-0.00839	0.04571	-0.00910	0.01238
56	3248	-0.01078	0.04513	-0.00971	0.01232
57	3304	-0.00999	0.04454	-0.00924	0.01203
58	3360	-0.00985	0.04361	-0.00992	0.01167
59	3416	-0.00778	0.04257	-0.00907	0.01140
60	3472	-0.00843	0.04242	-0.00854	0.01136
61	3528	-0.00739	0.04161	-0.00875	0.01108
62	3584	-0.00581	0.04053	-0.00818	0.01057
63	3640	-0.00891	0.04038	-0.00854	0.01049
64	3696	-0.00919	0.03911	-0.00831	0.01007
65	3752	-0.00965	0.03870	-0.00856	0.00992
66	3808	-0.00841	0.03795	-0.00863	0.00968
67	3864	-0.00825	0.03750	-0.00882	0.00958
68	3920	-0.00753	0.03665	-0.00829	0.00926
69	3976	-0.00879	0.03666	-0.00840	0.00937
70	4032	-0.00731	0.03593	-0.00906	0.00893
71	4088	-0.01065	0.03508	-0.00916	0.00879
72	4144	-0.00814	0.03477	-0.00777	0.00872
73	4200	-0.00876	0.03449	-0.00845	0.00841
74	4256	-0.00906	0.03344	-0.00848	0.00823
75	4312	-0.00932	0.03330	-0.00941	0.00807

Table V. (Continued) Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform ($v = 56$) were used.

n	X Sample Size	$\bar{\mu}[s^{*'}_n]$	$\bar{\sigma}^2[s^{*'}_n]$	$\bar{\mu}[s^{+'}_n]$	$\bar{\sigma}^2[s^{+'}_n]$
76	4368	-0.00798	0.03328	-0.00758	0.00794
77	4424	-0.00903	0.03299	-0.00872	0.00795
78	4480	-0.00863	0.03210	-0.00828	0.00777
79	4536	-0.00820	0.03166	-0.00756	0.00751
80	4592	-0.00962	0.03082	-0.00853	0.00732
81	4648	-0.00921	0.03091	-0.00816	0.00731
82	4704	-0.00949	0.03085	-0.00777	0.00717
83	4760	-0.00662	0.03082	-0.00744	0.00725
84	4816	-0.00911	0.02989	-0.00807	0.00692
85	4872	-0.00711	0.02934	-0.00771	0.00679
86	4928	-0.00773	0.02907	-0.00782	0.00669
87	4984	-0.00815	0.02899	-0.00823	0.00662
88	5040	-0.00794	0.02860	-0.00836	0.00643
89	5096	-0.00846	0.02844	-0.00823	0.00648
90	5152	-0.00765	0.02811	-0.00736	0.00634
91	5208	-0.00767	0.02742	-0.00796	0.00612
92	5264	-0.00778	0.02732	-0.00773	0.00607
93	5320	-0.00662	0.02703	-0.00710	0.00608
94	5376	-0.00714	0.02662	-0.00778	0.00602
95	5432	-0.00786	0.02626	-0.00743	0.00578
96	5488	-0.00800	0.02632	-0.00791	0.00577
97	5544	-0.00789	0.02584	-0.00804	0.00569
98	5600	-0.00735	0.02541	-0.00704	0.00554
99	5656	-0.00828	0.02512	-0.00798	0.00549
100	5712	-0.00811	0.02535	-0.00744	0.00555

Table V. (Continued) Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform ($v = 56$) were used.

n	X Sample Size	$\bar{\mu}[s^{*\prime}_n]$	$\bar{\sigma}^2[s^{*\prime}_n]$	$\bar{\mu}[s^{+\prime}_n]$	$\bar{\sigma}^2[s^{+\prime}_n]$
101	5768	-0.00796	0.02534	-0.00747	0.00543
102	5824	-0.00758	0.02427	-0.00747	0.00521
103	5880	-0.00579	0.02421	-0.00699	0.00513
104	5936	-0.00662	0.02404	-0.00718	0.00512
105	5992	-0.00628	0.02405	-0.00756	0.00512
106	6048	-0.00754	0.02367	-0.00761	0.00501
107	6104	-0.00690	0.02326	-0.00692	0.00490
108	6160	-0.00764	0.02328	-0.00748	0.00486
109	6216	-0.00787	0.02296	-0.00729	0.00481
110	6272	-0.00776	0.02271	-0.00710	0.00476
111	6328	-0.00723	0.02263	-0.00698	0.00471
112	6384	-0.00856	0.02194	-0.00695	0.00450
113	6440	-0.00719	0.02212	-0.00710	0.00459
114	6496	-0.00794	0.02200	-0.00697	0.00444
115	6552	-0.00691	0.02140	-0.00681	0.00443
116	6608	-0.00687	0.02177	-0.00731	0.00439
117	6664	-0.00690	0.02124	-0.00726	0.00429
118	6720	-0.00634	0.02106	-0.00690	0.00422
119	6776	-0.00795	0.02066	-0.00728	0.00417
120	6832	-0.00711	0.02082	-0.00721	0.00415
121	6888	-0.00578	0.02053	-0.00653	0.00414
122	6944	-0.00684	0.02069	-0.00652	0.00409
123	7000	-0.00696	0.02058	-0.00723	0.00404
124	7056	-0.00644	0.02007	-0.00670	0.00397
125	7112	-0.00606	0.02019	-0.00725	0.00391

Table V. (Continued) Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform (v = 56) were used.

n	X Sample Size	$\bar{\mu}[s^{*'}_n]$	$\bar{\sigma}^2[s^{*'}_n]$	$\bar{\mu}[s^{+'}_n]$	$\bar{\sigma}^2[s^{+'}_n]$
126	7168	-0.00605	0.01991	-0.00641	0.00387
127	7224	-0.00779	0.01968	-0.00712	0.00383
128	7280	-0.00720	0.01934	-0.00695	0.00377
129	7336	-0.00660	0.01922	-0.00661	0.00375
130	7392	-0.00597	0.01929	-0.00713	0.00370
131	7448	-0.00751	0.01891	-0.00649	0.00369
132	7504	-0.00681	0.01903	-0.00679	0.00360
133	7560	-0.00851	0.01885	-0.00722	0.00362
134	7616	-0.00595	0.01874	-0.00671	0.00361
135	7672	-0.00651	0.01842	-0.00614	0.00353
136	7728	-0.00601	0.01831	-0.00640	0.00347
137	7784	-0.00658	0.01829	-0.00663	0.00345
138	7840	-0.00646	0.01795	-0.00636	0.00339
139	7896	-0.00635	0.01793	-0.00693	0.00336
140	7952	-0.00634	0.01762	-0.00634	0.00328
141	8008	-0.00608	0.01773	-0.00631	0.00333
142	8064	-0.00630	0.01761	-0.00618	0.00326
143	8120	-0.00617	0.01724	-0.00640	0.00323
144	8176	-0.00701	0.01737	-0.00635	0.00322
145	8232	-0.00702	0.01702	-0.00638	0.00312
146	8288	-0.00735	0.01706	-0.00640	0.00313
147	8344	-0.00559	0.01695	-0.00602	0.00307
148	8400	-0.00637	0.01680	-0.00621	0.00315
149	8456	-0.00614	0.01704	-0.00610	0.00310
150	8512	-0.00607	0.01668	-0.00575	0.00304

Table V. (Continued) Estimated bias and variance of the improved stochastic approximation estimator for the 0.99 quantile of the exponential distribution. Algorithm of Section III.D.2 and maximum transform (v = 56) were used.

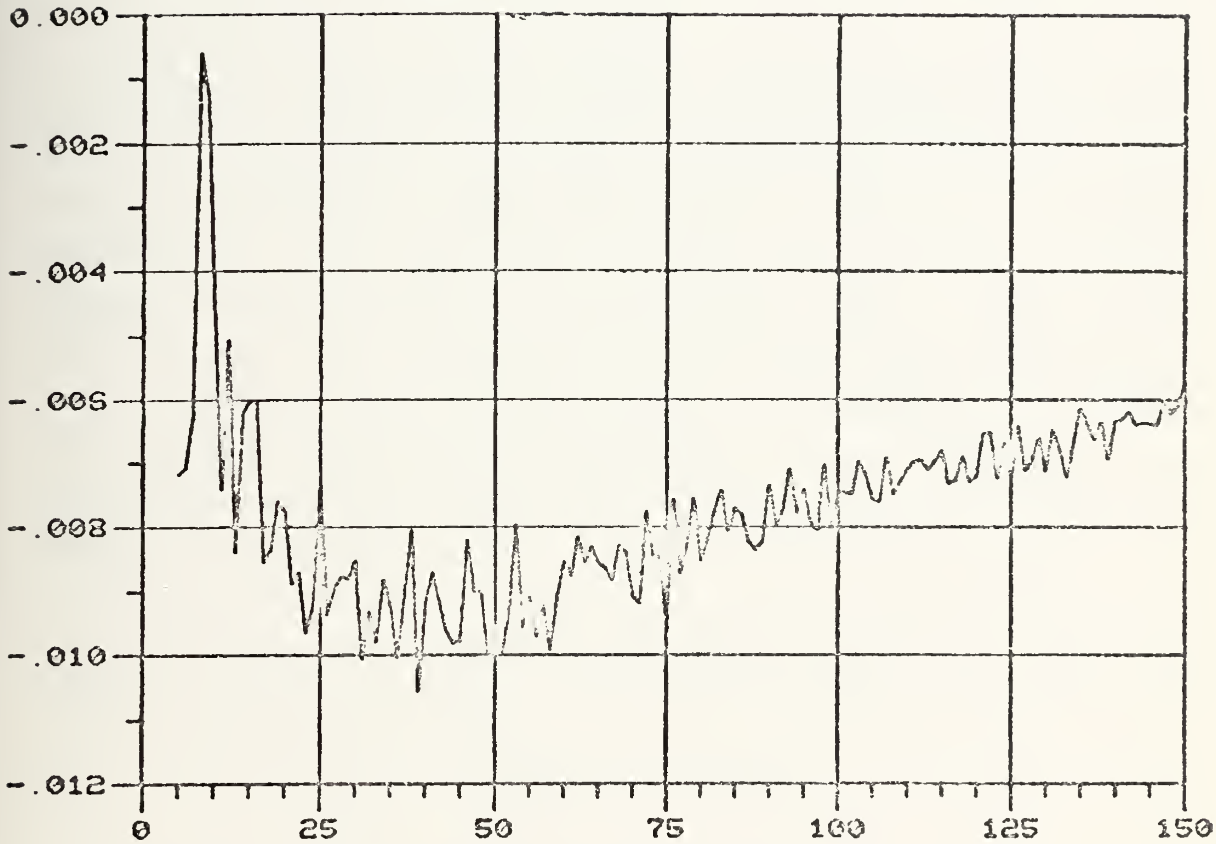


Figure 15. Expected bias of the stochastic approximation estimator $\bar{s}_n$ for the 0.99 quantile of the exponential distribution (Y-axis) vs. step number n (X-axis).

1. Order of the bias

We now proceed to direct consideration of the bias estimates in Table V; the control variate estimates of the bias are plotted in Figure 15 where it may be seen that there is a definite decreasing trend. The rate of decrease appears to be very slow, however; furthermore, there are marked irregularities in the first few steps. Since we are for the most part interested in the large sample behavior of the stochastic approximation quantile estimators we suppress the initial instability by including in the regression only the estimate values from steps greater than 50 (i.e., X samples larger than 2912).

Carrying out a linear regression using the model (5) results in the estimates

$$\begin{aligned} \bar{r}_0 &= 0.00264 \pm 0.00174, \\ \bar{r}_1 &= -0.14103 \pm 0.03330, \qquad (22) \\ \bar{r}_2 &= 0.39692 \pm 0.15633; \end{aligned}$$

the second figure given is the standard deviation of the estimate. Assuming that the errors in (5) are approximately normally distributed, we compute the following analysis of variance table:

Source	Sum of Squares	Degrees of Freedom	Mean Square
Constant	21,985.10983	1	21,985.1098
r (0)	1.28580	1	1.2858
r (1), r (2)	29.01798	1	29.0180
Regression	30.30378	2	15.1519
Explained	22,015.41361	3	7,338.4712
Pure Error	2,245,341.29442	4,039,899	0.5558
Lack of Fit	47.68185	98	0.4865
Residual	2,245,388.97627	4,039,997	0.5558
Total	2,267,404.38989	4,040,000	0.5612

The regression is significant as measured by the F-ratio of 27.2618 which is significant at the 0.999 level. The ratio of the sum of the squared deviations about the regression line ("pure error") to the squared deviations between the fitted and mean biases ("lack of fit") is 0.8754 which is not significant at the 0.9 level; we thus conclude that the fitted line adequately describes the data of Table V. Note that our hypothesis that $r_0 = 0$ is certainly consistent with these results although the F-ratio of 22.568 will not allow us to reject the r_0 term as not significant in the regression.

One problem encountered in most of the regressions carried out on this data is the high degree of multicollinearity in the $G^T G$ matrix when more than just a few terms of the form $g_j(n) = n^{-j/p}$ are included in the model. The result of this multicollinearity is considerable

variability in the $\bar{r}_j$ estimates as measured by the standard errors as well as some irregularities in the analysis of variance. This is one reason that so much data had to be accumulated for this experiment.

Discriminating between the model (5) and a model such as

$$\text{(23)} \qquad E[s^*_n] = r_1 n^{-1/4} + r_2 n^{-1/2} + r_3 n^{-3/4}$$

also requires a great many observations on s^*_n. The results of a regression using (23) are

$$\bar{r}_1 = 0.03821 \pm 0.02146,$$

$$\bar{r}_2 = -0.34324 \pm 0.13249,$$

$$\bar{r}_3 = 0.46642 \pm 0.20350.$$

The $\bar{r}_1$ coefficient estimate is thus just significant at the 0.9 level while $\bar{r}_2$ is significant at the 0.99 level. An analysis of variance indicates that the $n^{-1/4}$ term contributes 1.763 to the regression sum of squares while the other two terms contribute 28.565. Although neither an F-test nor a t-test will allow us to reject the low order term as not statistically significant this regression provides convincing evidence that the order of the bias is

in fact $n^{-1/2}$ as indicated by the theory. This is certainly a considerable improvement over the results of Yuguchi [38].

A regression was also carried out using the model

$$|E[s^*_n]| = c_1 n^{-r}, \tag{24}$$

to attempt a direct verification of the order of the bias. (24) can be handled as a linear model by using a logarithmic transform on the data. It is apparent from Figure 14 that higher order terms have an important effect on the bias; therefore a power series in n^{-r} was also fitted using nonlinear regression. Unfortunately, the results were too unstable to be of much use. To minimize the effect of higher order terms, then, we include in the regression only data from the later steps. The resulting estimates are:

Lowest Step in Regression	$\bar{c}_1$	$\bar{r}$
1	0.00912	−0.04258 ± 0.03146
20	0.02002	−0.21628 ± 0.01374
50	0.04305	−0.38227 ± 0.01562
100	0.06525	−0.46887 ± 0.04677

Based on these results we conclude that the data of Table V display a definite $n^{-1/2}$ trend and that the evidence does not seem to warrant the assumption of a lower order of bias.

One way to explore more fully the effect of the initial starting point on the bias of the stochastic approximation

quantile estimator is to begin the procedure with $\bar{s}_1$ fixed at some value of interest instead of using random values as in Table V. This has been done for values of $\bar{s}_1$ between 0 and 9 (corresponding to initial biases from -4.5 to 4.5). We then carry out a regression using the model

$$E[s^*_n] = r_1 n^{-1/2} + r_2 n^{-1} ;$$

the resulting estimates $\bar{r}_1$ are plotted in Figure 16 and summarized in Table VI.

We conclude from Figure 16 that the bias of the initial estimate plays a significant role in determining the asymptotic bias of the stochastic approximation quantile estimator. This is in general agreement with the results of Hodges and Lehmann [15]; although the relationship of Figure 16 is clearly not linear, the asymptotic bias apparently increases with increasing deviations in $\bar{s}_1$. There is insufficient data here to investigate the relationship more fully, but the quadratic fit

$$\bar{r}_1 = -0.112 + 0.023\ \bar{s}_1 - 0.004\ \bar{s}_1^2$$

plotted in Figure 16 seems to describe the data fairly well.

$\bar{s}_1$	$\bar{r}_1$	Standard Error
0.0	-0.09672	0.00652
0.5	-0.12741	0.01783
1.0	-0.07937	0.01741
1.5	-0.13922	0.01699
2.0	-0.11091	0.01641
2.5	-0.10804	0.01582
3.0	-0.09765	0.01508
3.5	-0.09152	0.01460
4.0	-0.10550	0.01430
4.5	-0.07405	0.01406
4.605	-0.07898	0.00503
5.0	-0.07215	0.01432
5.5	-0.12150	0.01694
6.0	-0.09135	0.01602
6.5	-0.09890	0.01694
7.0	-0.15344	0.01799
7.5	-0.16990	0.01883
8.0	-0.19678	0.01981
8.5	-0.19446	0.02061
9.0	-0.26129	0.02151

Table VI. Estimated coefficients for the $O(n^{-1/2})$ term in the bias of the stochastic approximation estimator for the 0.99 quantile of the exponential distribution as a function of the initial starting point, $\bar{s}_1$. Estimated by linear regressions which included 1000 replications of steps 50 to 150 of the stochastic approximation process.

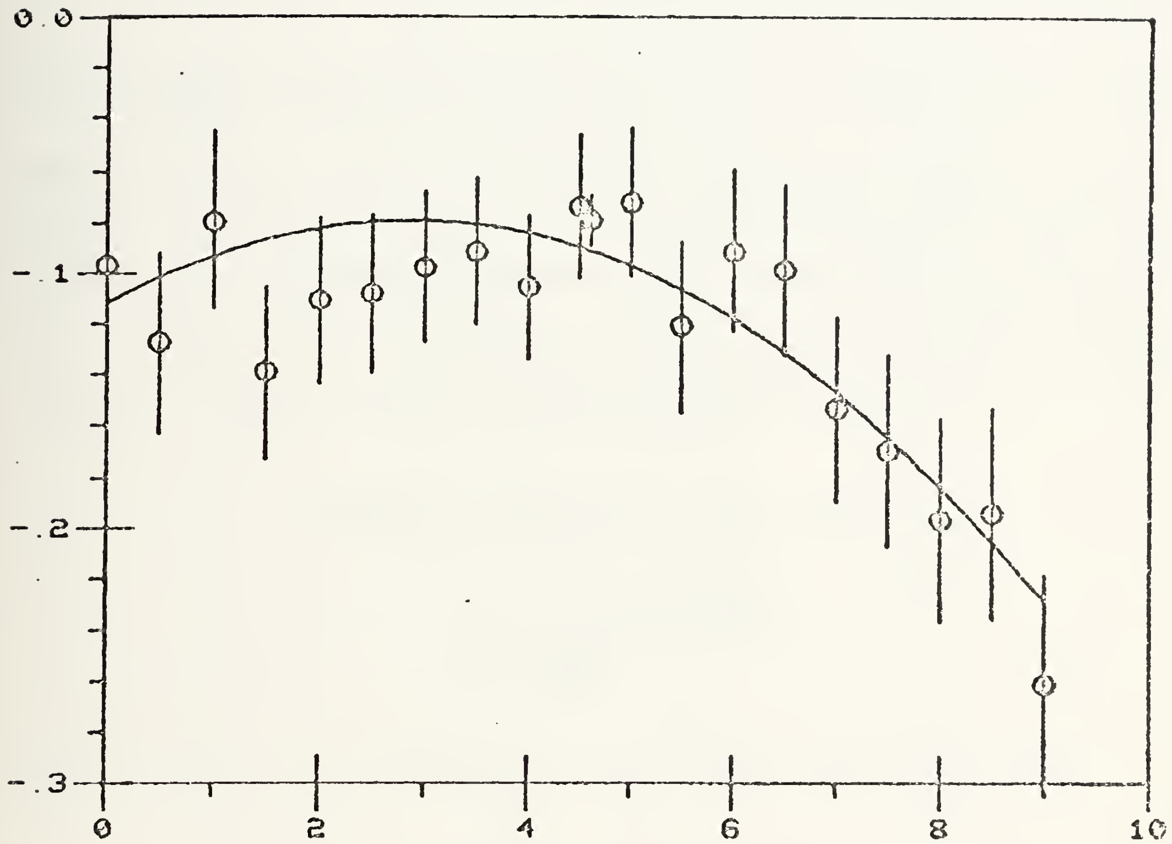


Figure 16. Estimated coefficient of the $n^{-1/2}$ term in the bias of the stochastic approximation estimator for the 0.99 quantile of the exponential distribution (Y-axis) vs. the bias of the initial starting point ξ_1. The vertical lines represent two estimated standard deviations about the estimated coefficients.

2. Comparison with order statistics

The presence of the $n^{-1/2}$ bias term puts stochastic approximation quantile estimators at a disadvantage when compared with order statistic estimators whose bias is $O(n^{-1})$. The data of Table V, however, indicate that the stochastic approximation biases are quite small as compared with the estimator variance. The net effect of the bias, then, will be to inflate the asymptotic mean squared error slightly. Based on (1) and (22) we have

$$\begin{aligned} MSE[\bar{s}'_n] &= Var[\bar{s}'_n] + \frac{r_1^2}{n} + O(n^{-1}) \\ &\to \frac{2.361 + 0.020}{n} \\ &\to \frac{2.381}{n}, \end{aligned}$$

which should be compared with the order statistic case:

$$\begin{aligned} MSE[s_{(n+2)v}] &= Var[s_{(n+2)v}] + o(n^{-1}) \\ &\to \frac{1.768}{n}. \end{aligned}$$

(Recall that the order statistic estimator will be based on the entire X sample and not just on the section maxima.) Most of the asymptotic difference between the two quantile estimators is thus due to the variance inflation (1.15) which accompanies the use of the maximum transform.

A comparison between finite sample order statistic and stochastic approximation quantile estimators is presented in Table VII and plotted in Figures 17 and 18; Figure 17

n	X Sample Size	Stochastic Approximation Bias	Stochastic Approximation MSE	Order Statistic Bias	Order Statistic MSE
1	168	-0.11274	0.66383	0.09898	0.64880
2	224	-0.04338	0.55576	-0.11408	0.40349
3	280	-0.01847	0.45473	0.10862	0.40317
4	336	-0.01015	0.39043	-0.04269	0.28267
5	392	-0.00719	0.34227	0.11125	0.29365
6	448	-0.00708	0.30679	-0.00538	0.21912
7	504	-0.00625	0.27741	-0.08772	0.18704
8	560	-0.00059	0.25076	0.01754	0.17985
9	616	-0.00124	0.23395	-0.05390	0.15483
10	672	-0.00450	0.21862	0.03305	0.15315
11	728	-0.00742	0.20189	-0.02982	0.13265
12	784	-0.00506	0.18676	0.04423	0.13382
13	840	-0.00840	0.17659	-0.01181	0.11646
14	896	-0.00623	0.16650	0.05269	0.11917
15	952	-0.00603	0.15750	0.00217	0.10412
16	1008	-0.00604	0.14920	-0.04070	0.09583
17	1064	-0.00854	0.14260	0.01334	0.09440
18	1120	-0.00835	0.13449	-0.02630	0.08670
19	1176	-0.00760	0.12730	0.02247	0.08656
20	1232	-0.00777	0.12215	-0.01437	0.07935
21	1288	-0.00887	0.11695	0.03007	0.08009
22	1344	-0.00872	0.11217	-0.00431	0.07332
23	1400	-0.00964	0.10720	-0.03493	0.06944
24	1456	-0.00921	0.10294	0.00427	0.06827
25	1512	-0.00744	0.10049	-0.02466	0.06444

Table VII. Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Stochastic Approximation Bias	Stochastic Approximation MSE	Order Statistic Bias	Order Statistic MSE
26	1568	-0.00936	0.09479	0.01169	0.06399
27	1624	-0.00894	0.09216	-0.01573	0.06022
28	1680	-0.00878	0.08811	0.01816	0.06032
29	1736	-0.00880	0.08676	-0.00788	0.05661
30	1792	-0.00852	0.08387	0.02386	0.05714
31	1848	-0.01007	0.08050	-0.00093	0.05350
32	1904	-0.00931	0.07875	-0.02372	0.05131
33	1960	-0.00981	0.07598	0.00526	0.05079
34	2016	-0.00881	0.07504	-0.01658	0.04855
35	2072	-0.00923	0.07115	0.01082	0.04841
36	2128	-0.01005	0.06954	-0.01014	0.04614
37	2184	-0.00906	0.06854	0.01583	0.04630
38	2240	-0.00806	0.06668	-0.00431	0.04401
39	2296	-0.01057	0.06472	0.02037	0.04442
40	2352	-0.00918	0.06286	0.00099	0.04212
41	2408	-0.00872	0.06108	-0.01715	0.04069
42	2464	-0.00918	0.06053	0.00583	0.04044
43	2520	-0.00964	0.05894	-0.01170	0.03895
44	2576	-0.00982	0.05791	0.01027	0.03893
45	2632	-0.00977	0.05555	-0.00668	0.03740
46	2688	-0.00822	0.05441	0.01436	0.03757
47	2744	-0.00901	0.05370	-0.00206	0.03600
48	2800	-0.00902	0.05324	-0.01757	0.03504
49	2856	-0.01001	0.05175	0.00223	0.03474
50	2912	-0.00960	0.05062	-0.01284	0.03372

Table VII. (Continued) Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Stochastic Approximation Bias	MSE	Order Statistic Bias	MSE
51	2968	-0.00994	0.04982	0.00620	0.03360
52	3024	-0.00934	0.04861	-0.00844	0.03252
53	3080	-0.00798	0.04740	0.00991	0.03256
54	3136	-0.00956	0.04662	-0.00434	0.03144
55	3192	-0.00910	0.04579	0.01336	0.03161
56	3248	-0.00971	0.04522	-0.00050	0.03046
57	3304	-0.00924	0.04462	-0.01371	0.02973
58	3360	-0.00992	0.04371	0.00309	0.02956
59	3416	-0.00907	0.04265	-0.00979	0.02879
60	3472	-0.00854	0.04249	0.00647	0.02874
61	3528	-0.00875	0.04168	-0.00611	0.02792
62	3584	-0.00818	0.04060	0.00964	0.02798
63	3640	-0.00854	0.04046	-0.00264	0.02713
64	3696	-0.00831	0.03918	0.01263	0.02728
65	3752	-0.00856	0.03877	0.00064	0.02640
66	3808	-0.00863	0.03803	-0.01087	0.02583
67	3864	-0.00882	0.03758	0.00373	0.02573
68	3920	-0.00829	0.03671	-0.00752	0.02512
69	3976	-0.00840	0.03673	0.00666	0.02511
70	4032	-0.00906	0.03601	-0.00436	0.02446
71	4088	-0.00916	0.03516	0.00944	0.02453
72	4144	-0.00777	0.03483	-0.00135	0.02386
73	4200	-0.00845	0.03456	-0.01174	0.02343
74	4256	-0.00848	0.03351	0.00151	0.02330
75	4312	-0.00941	0.03339	-0.00868	0.02283

Table VII. (Continued) Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Stochastic Approximation Bias	Stochastic Approximation MSE	Order Statistic Bias	Order Statistic MSE
76	4368	-0.00758	0.03334	0.00422	0.02278
77	4424	-0.00872	0.03307	-0.00577	0.02228
78	4480	-0.00828	0.03216	0.00681	0.02229
79	4536	-0.00756	0.03172	-0.00299	0.02177
80	4592	-0.00853	0.03089	0.00928	0.02185
81	4648	-0.00816	0.03098	-0.00034	0.02129
82	4704	-0.00777	0.03091	-0.00964	0.02093
83	4760	-0.00744	0.03088	0.00219	0.02085
84	4816	-0.00807	0.02996	-0.00695	0.02046
85	4872	-0.00771	0.02940	0.00461	0.02043
86	4928	-0.00782	0.02913	-0.00437	0.02002
87	4984	-0.00823	0.02905	0.00693	0.02005
88	5040	-0.00836	0.02867	-0.00190	0.01961
89	5096	-0.00823	0.02851	0.00915	0.01969
90	5152	-0.00736	0.02817	0.00047	0.01922
91	5208	-0.00796	0.02749	-0.00795	0.01892
92	5264	-0.00773	0.02738	0.00274	0.01886
93	5320	-0.00710	0.02708	-0.00554	0.01853
94	5376	-0.00778	0.02668	0.00493	0.01853
95	5432	-0.00743	0.02632	-0.00323	0.01817
96	5488	-0.00791	0.02638	0.00703	0.01822
97	5544	-0.00804	0.02591	-0.00101	0.01784
98	5600	-0.00704	0.02546	-0.00881	0.01760
99	5656	-0.00798	0.02518	0.00114	0.01752
100	5712	-0.00744	0.02540	-0.00656	0.01726

Table VII. (Continued) Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Stochastic Approximation Bias	Stochastic Approximation MSE	Order Statistic Bias	Order Statistic MSE
101	5768	-0.00747	0.02540	0.00320	0.01723
102	5824	-0.00747	0.02433	-0.00438	0.01694
103	5880	-0.00699	0.02426	0.00519	0.01695
104	5936	-0.00718	0.02410	-0.00228	0.01664
105	5992	-0.00756	0.02411	0.00711	0.01669
106	6048	-0.00761	0.02373	-0.00026	0.01636
107	6104	-0.00692	0.02330	-0.00744	0.01615
108	6160	-0.00748	0.02333	0.00169	0.01610
109	6216	-0.00729	0.02301	-0.00539	0.01587
110	6272	-0.00710	0.02276	0.00358	0.01585
111	6328	-0.00698	0.02268	-0.00340	0.01560
112	6384	-0.00695	0.02198	0.00541	0.01562
113	6440	-0.00710	0.02217	-0.00148	0.01535
114	6496	-0.00697	0.02205	0.00717	0.01540
115	6552	-0.00681	0.02144	0.00037	0.01511
116	6608	-0.00731	0.02182	-0.00627	0.01493
117	6664	-0.00726	0.02129	0.00217	0.01489
118	6720	-0.00690	0.02111	-0.00439	0.01469
119	6776	-0.00728	0.02071	0.00391	0.01468
120	6832	-0.00721	0.02087	-0.00257	0.01446
121	6888	-0.00653	0.02057	0.00560	0.01448
122	6944	-0.00652	0.02073	-0.00080	0.01424
123	7000	-0.00723	0.02063	-0.00705	0.01409
124	7056	-0.00670	0.02011	0.00091	0.01404
125	7112	-0.00725	0.02024	-0.00527	0.01387

Table VII. (Continued) Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Stochastic Approximation Bias	MSE	Order Statistic Bias	MSE
126	7168	-0.00641	0.01995	0.00258	0.01385
127	7224	-0.00712	0.01973	-0.00353	0.01367
128	7280	-0.00695	0.01938	0.00419	0.01367
129	7336	-0.00661	0.01927	-0.00185	0.01347
130	7392	-0.00713	0.01934	0.00576	0.01350
131	7448	-0.00649	0.01895	-0.00021	0.01329
132	7504	-0.00679	0.01907	-0.00605	0.01315
133	7560	-0.00722	0.01890	0.00138	0.01311
134	7616	-0.00671	0.01879	-0.00440	0.01296
135	7672	-0.00614	0.01846	0.00293	0.01295
136	7728	-0.00640	0.01835	-0.00278	0.01278
137	7784	-0.00663	0.01834	0.00443	0.01279
138	7840	-0.00636	0.01800	-0.00122	0.01261
139	7896	-0.00693	0.01798	0.00590	0.01265
140	7952	-0.00634	0.01766	0.00031	0.01245
141	8008	-0.00631	0.01777	-0.00518	0.01232
142	8064	-0.00618	0.01765	0.00179	0.01230
143	8120	-0.00640	0.01728	-0.00363	0.01216
144	8176	-0.00635	0.01741	0.00324	0.01216
145	8232	-0.00638	0.01706	-0.00213	0.01200
146	8288	-0.00640	0.01710	0.00465	0.01202
147	8344	-0.00602	0.01699	-0.00066	0.01186
148	8400	-0.00621	0.01684	-0.00588	0.01175
149	8456	-0.00610	0.01707	0.00076	0.01172
150	8512	-0.00575	0.01671	-0.00440	0.01160

Table VII. (Continued) Comparison of order statistic and stochastic approximation estimators for the 0.99 quantile of the exponential distribution.

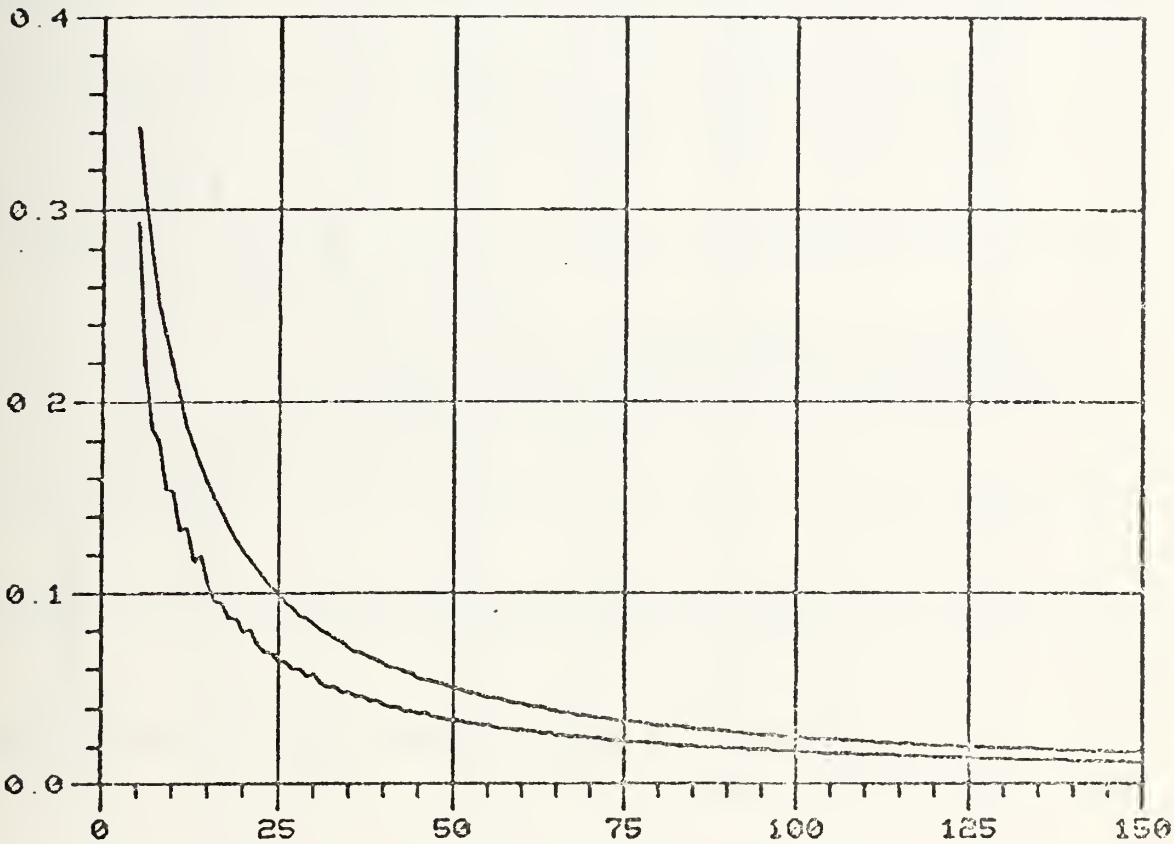


Figure 17. Mean squared error of the order statistic estimator (lower curve) and the stochastic approximation estimator (upper curve) for the 0.99 quantile of the exponential distribution vs. the number of stochastic approximation steps.

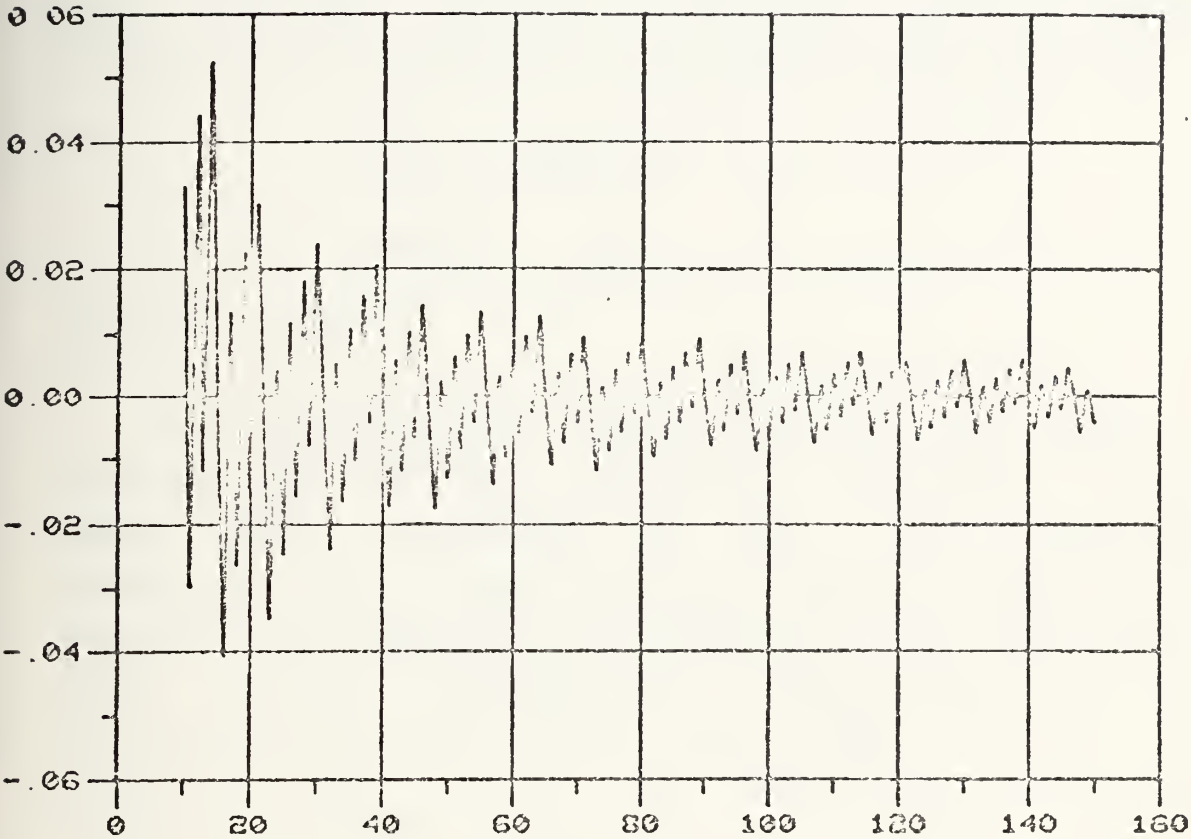


Figure 18. Bias of the order statistic estimator for the 0.99 quantile of the exponential distribution; the same horizontal scale as in Figure 17 is used.

displays the mean squared errors of the two estimators while Figure 18 is a plot of the bias of the order statistic estimator. Values for the stochastic approximation estimator were obtained from the simulation data of Table V while the order statistic values were computed from the formulas for the exponential distribution (see David [5])

$$E[\hat{s}_n] = \sum_{i=n-u+1}^{n} i^{-1} ,$$

$$Var[\hat{s}_n] = \sum_{i=n-u+1}^{n} i^{-2} ,$$

where $u = [a(n+1)]$.

The characteristic jagged appearance of Figure 18 reflects the truncation inherent in calculating u; it also makes direct comparison of bias terms difficult. Nevertheless it is clear that the stochastic approximation estimators are generally less biased than the corresponding order statistic estimators for X samples smaller than 3500 observations while the biases are roughly the same for samples of from 3500 to 5500 observations. For larger samples, the asymptotic advantage of the order statistic estimators begins to assert itself and we find that the stochastic approximation estimators are for the most part more biased. The mean squared error plot (Figure 17) merely confirms the asymptotic superiority of the order statistic estimator in terms of variance. Note that even when the stochastic approximation estimator is more biased this does not seem to have much influence on the mean squared error.

In practice the approximate order statistic estimators of Section III.A are often used in order to conserve computer memory; the problem with these techniques is that they may introduce an objectionable bias into the estimates (see Tables II and III). If stochastic approximation

quantile estimators were used in an approximate design, however, Table VI shows that for sample sizes small enough to be practical for the order statistic estimators the bias will be smaller for the stochastic approximation case, although the variance will be greater. This trading of bias for variance is also seen when the jackknife ([27], [38]) is applied to the order statistic estimators.

Of course there is no need to carry out a section averaging or nesting procedure with stochastic approximation quantile estimators; this is necessitated in the order statistic case because the requirement to store and sort an entire section imposes an upper limit on permissible section size. The fixed memory size for the stochastic approximation estimator, however, means that we may reduce both bias and variance by considering larger X samples directly without sectioning the data. In a practical sense, then, the stochastic approximation estimates are less biased than the corresponding order statistic estimators for very large data samples.

E. Higher Moments and Distribution of $\bar{s}_n$

Besides the significant $O(n^{-1/4})$ term in the bias, another disturbing result of Yuguchi's thesis [38] was the apparent increase in the coefficients of skewness and kurtosis of $\bar{s}'_n$ with increasing values of n. The coefficient of skewness of a random variable X is

$$\gamma_1 = E[(X - \mu)^3] / \sigma^3$$

where $\mu = E[X]$ and $\sigma^2 = Var[X]$. γ_1 is zero for any symmetric random variable, e.g. normal. The coefficient of kurtosis (sometimes called excess kurtosis) we define as

$$\gamma_2 = E[(X - u)^4] / \sigma^4 - 3;$$

γ_2 is also zero for a normal random variable.

If $\bar{s}'_n$ converges weakly (i.e. in distribution) to a normal random variable it is desirable from a practical point of view for $\gamma_1(\bar{s}_n)$ and $\gamma_2(\bar{s}_n)$ both to approach zero as n increases. Of course, weak convergence (or even almost sure convergence) does not imply convergence in p<u>th</u> mean, $p > 1$, so that γ_1 and γ_2 need not even approach a finite limit; an example is provided by Figures 2 and 3 where the RM estimator converges in quadratic mean and in distribution but apparently not in third or fourth means.

This problem does not occur for the new estimator, however. The sample means, variances and coefficients of skewness and kurtosis are tabulated in Table VIII for one-fourth of the data from Table VI, i.e. 10,000 independent replications of $s^{*\prime}_n$ for the exponential 0.99 quantile using the maximum transform with $v = 56$. The third and fourth central moments were not obtained for the remaining 30,000 observations for each n value in Table VI in order to save computer time; the data that was collected

n	X Sample Size	Mean	Variance	Skewness	Kurtosis
1	168	-0.11274	0.65112	0.72596	0.93864
2	224	-0.05081	0.54749	0.62491	1.31966
3	280	-0.01345	0.45878	0.46303	0.68834
4	336	-0.01079	0.38981	0.49609	0.61954
5	392	-0.01409	0.34421	0.44722	0.50377
6	443	-0.00117	0.30619	0.46074	0.48721
7	504	-0.00101	0.28681	0.46856	0.54782
8	560	-0.00628	0.25554	0.44417	0.51575
9	616	-0.00112	0.23298	0.38892	0.33607
10	672	-0.00347	0.21493	0.44614	0.50744
11	728	-0.00430	0.20665	0.45871	0.81734
12	784	0.00011	0.18656	0.34032	0.24865
13	840	-0.00320	0.17232	0.38481	0.45498
14	896	-0.00479	0.16871	0.36231	0.36680
15	952	-0.00115	0.15853	0.37063	0.61999
16	1008	-0.00776	0.14934	0.36837	0.45296
17	1064	-0.01065	0.13933	0.26925	0.26946
18	1120	-0.00710	0.13890	0.35880	0.32412
19	1176	-0.00797	0.12758	0.33897	0.53789
20	1232	-0.01016	0.12054	0.33085	0.42341
21	1288	-0.00948	0.11434	0.33706	0.32702
22	1344	-0.00987	0.11342	0.35054	0.36447
23	1400	-0.01199	0.10640	0.31868	0.31489
24	1456	-0.00879	0.10344	0.29160	0.27693
25	1512	-0.00898	0.09892	0.33703	0.40463

Table VIII. Sample moments for 10,000 realizations of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Mean	Variance	Skewness	Kurtosis
26	1568	-0.00785	0.09412	0.31614	0.62028
27	1624	-0.01234	0.09181	0.27344	0.32623
28	1680	-0.01041	0.08664	0.24567	0.39986
29	1736	-0.00649	0.08490	0.25753	0.15673
30	1792	-0.00879	0.08245	0.21642	0.15205
31	1848	-0.00996	0.08109	0.28204	0.14137
32	1904	-0.00900	0.08013	0.25671	0.20375
33	1960	-0.00595	0.07646	0.25526	0.18383
34	2016	-0.00698	0.07306	0.23054	0.11701
35	2072	-0.01030	0.07136	0.28635	0.31507
36	2128	-0.00765	0.06857	0.19870	0.08330
37	2184	-0.01157	0.06731	0.28060	0.27422
38	2240	-0.00767	0.06548	0.27972	0.26658
39	2296	-0.00703	0.06339	0.24303	0.14686
40	2352	-0.00493	0.06313	0.24346	0.16414
41	2408	-0.00963	0.06112	0.24595	0.19981
42	2464	-0.00833	0.06057	0.24375	0.23653
43	2520	-0.01114	0.05890	0.22788	0.23471
44	2576	-0.01300	0.05687	0.29007	0.49698
45	2632	-0.00895	0.05617	0.30989	0.37786
46	2688	-0.01024	0.05525	0.25985	0.21686
47	2744	-0.00915	0.05378	0.25192	0.18870
48	2800	-0.00799	0.05225	0.22194	0.29575
49	2856	-0.00881	0.05202	0.28420	0.51976
50	2912	-0.00671	0.05032	0.24716	0.12429

Table VIII. (Continued) Sample moments for 10,000 realizations of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Mean	Variance	Skewness	Kurtosis
51	2968	-0.01012	0.04907	0.24310	0.28737
52	3024	-0.01059	0.04704	0.23455	0.15058
53	3080	-0.00681	0.04792	0.21635	0.12167
54	3136	-0.00419	0.04660	0.21268	0.11278
55	3192	-0.01063	0.04529	0.18625	0.24244
56	3248	-0.00697	0.04626	0.23597	0.13575
57	3304	-0.00725	0.04409	0.18121	0.12925
58	3360	-0.00981	0.04296	0.20815	0.11342
59	3416	-0.00738	0.04314	0.23661	0.13994
60	3472	-0.01049	0.04116	0.23520	0.35075
61	3528	-0.00903	0.04072	0.19669	0.06457
62	3584	-0.00984	0.04113	0.12989	0.15193
63	3640	-0.00832	0.03954	0.18986	0.13455
64	3696	-0.00730	0.03929	0.20019	0.19601
65	3752	-0.00947	0.03800	0.15993	0.12546
66	3808	-0.00937	0.03806	0.17441	0.03580
67	3864	-0.00703	0.03837	0.19802	0.23513
68	3920	-0.00721	0.03660	0.15186	0.14438
69	3976	-0.00788	0.03624	0.16858	-0.04834
70	4032	-0.00718	0.03656	0.20713	0.14707
71	4088	-0.00822	0.03535	0.21347	0.25922
72	4144	-0.00846	0.03435	0.19317	0.15211
73	4200	-0.00973	0.03501	0.16768	0.07293
74	4256	-0.00830	0.03363	0.16786	0.02438
75	4312	-0.01061	0.03401	0.19544	0.11400

Table VIII. (Continued) Sample moments for 10,000 realizations of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Mean	Variance	Skewness	Kurtosis
76	4368	-0.00740	0.03339	0.23192	0.40494
77	4424	-0.00650	0.03280	0.17281	0.14084
78	4480	-0.00837	0.03264	0.14122	0.00701
79	4536	-0.00812	0.03185	0.21379	0.12254
80	4592	-0.00781	0.03163	0.18469	0.08219
81	4648	-0.01054	0.03183	0.20911	0.24620
82	4704	-0.00963	0.03032	0.18384	0.18129
83	4760	-0.00795	0.02969	0.18287	0.19079
84	4816	-0.00731	0.02947	0.15839	0.16455
85	4872	-0.00968	0.02950	0.17036	0.00225
86	4928	-0.01065	0.02851	0.17953	0.11817
87	4984	-0.00693	0.02897	0.16184	0.05916
88	5040	-0.00924	0.02800	0.15865	0.11181
89	5096	-0.00709	0.02854	0.18523	0.15899
90	5152	-0.00830	0.02816	0.17175	0.20298
91	5208	-0.00901	0.02795	0.19455	0.03783
92	5264	-0.00654	0.02701	0.13790	0.09320
93	5320	-0.00688	0.02680	0.13519	0.07674
94	5376	-0.00843	0.02677	0.11591	0.02051
95	5432	-0.00797	0.02617	0.14022	0.10174
96	5488	-0.00751	0.02623	0.14651	0.18201
97	5544	-0.00658	0.02546	0.13555	0.08283
98	5600	-0.00448	0.02596	0.11760	0.07470
99	5656	-0.00572	0.02508	0.13814	0.06399
100	5712	-0.00439	0.02514	0.17241	0.14293

Table VIII. (Continued) Sample moments for 10,000 realizations of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution.

n	X Sample Size	Mean	Variance	Skewness	Kurtosis
101	5768	-0.00581	0.02448	0.15044	0.14362
102	5824	-0.00637	0.02504	0.15366	0.11828
103	5880	-0.00872	0.02453	0.15101	0.05956
104	5936	-0.00564	0.02385	0.17769	0.11094
105	5992	-0.00838	0.02428	0.13391	0.13484
106	6048	-0.00903	0.02366	0.13756	0.07061
107	6104	-0.00825	0.02338	0.15209	0.06478
108	6160	-0.00581	0.02292	0.13397	0.10493
109	6216	-0.00541	0.02304	0.13187	0.10158
110	6272	-0.00810	0.02254	0.14520	0.00261
111	6328	-0.00507	0.02230	0.17181	0.10032
112	6384	-0.00959	0.02188	0.15548	0.07619
113	6440	-0.00836	0.02209	0.11682	0.00358
114	6496	-0.00758	0.02165	0.14474	0.06447
115	6552	-0.00714	0.02170	0.14750	0.07069
116	6608	-0.00212	0.02154	0.14713	0.10814
117	6664	-0.00851	0.02154	0.13412	0.03059
118	6720	-0.00606	0.02113	0.15224	0.03194
119	6776	-0.00462	0.02115	0.15435	0.07507
120	6832	-0.00603	0.02020	0.13705	-0.00582
121	6888	-0.00690	0.02029	0.11964	-0.06280

Table VIII. (Continued) Sample moments for 10,000 realizations of the stochastic approximation quantile estimator for the 0.99 quantile of the exponential distribution.

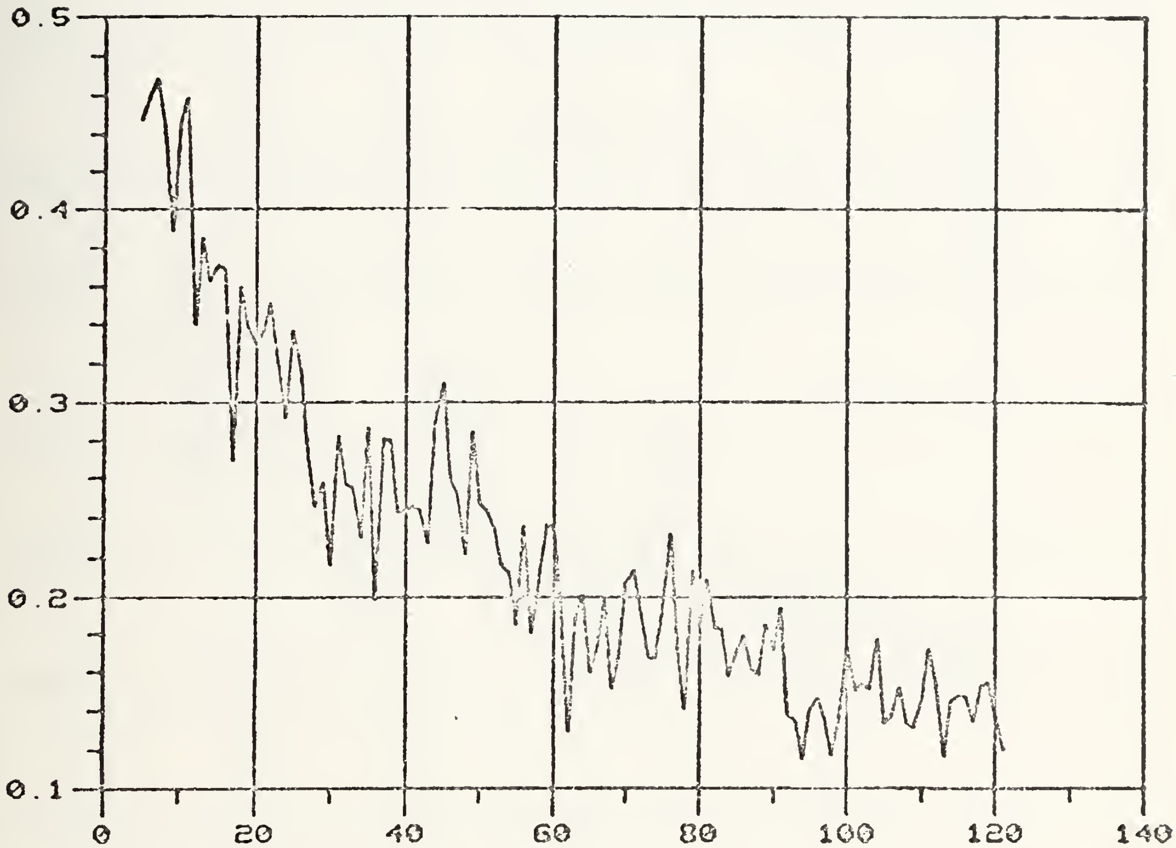


Figure 19. Coefficient of skewness of the stochastic approximation estimator for the 0.99 quantile of the exponential distribution vs. stochastic approximation step number.

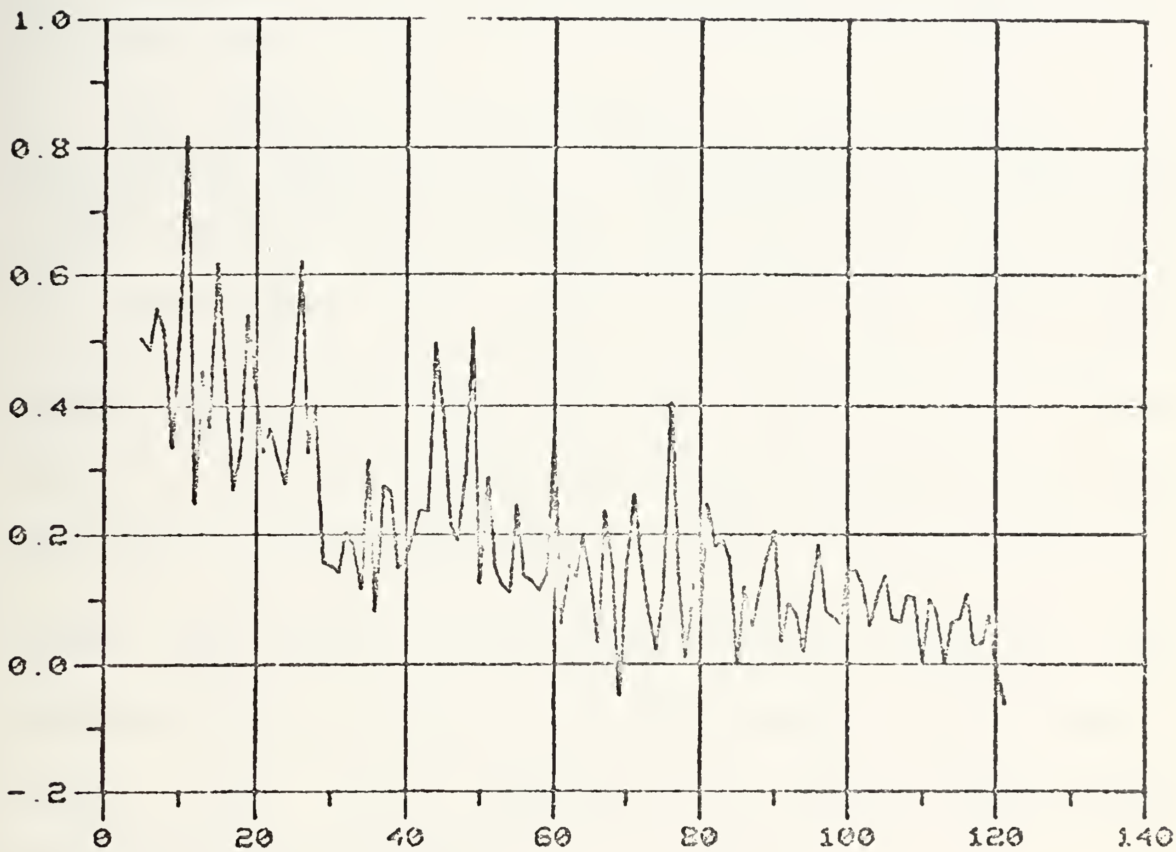


Figure 20. Coefficient of kurtosis of the stochastic approximation estimator for the 0.99 quantile of the exponential distribution vs. stochastic approximation step number.

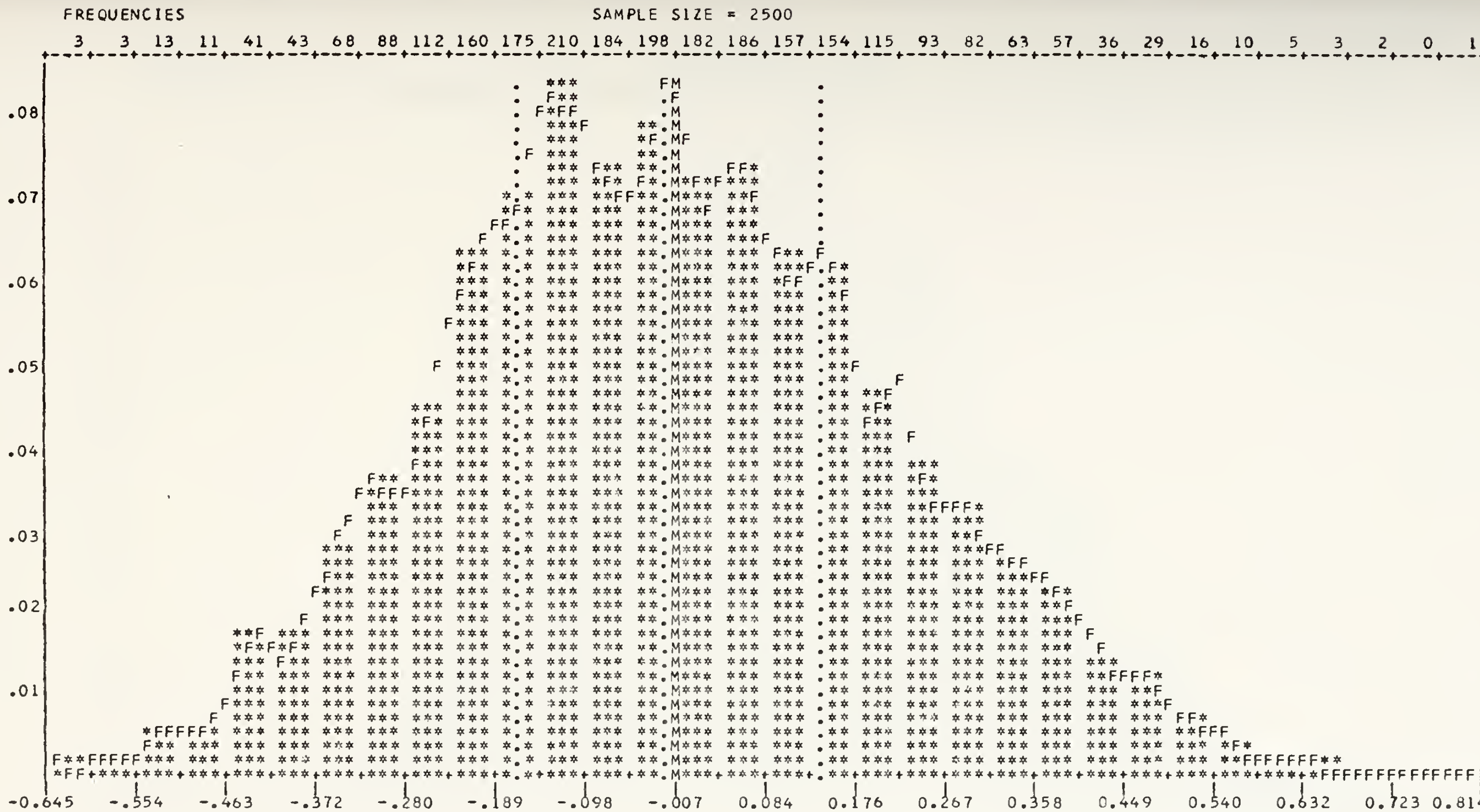
clearly supports the conjecture that $\bar{s}'_n$ converges in the fourth mean and that both γ_1 and γ_2 rapidly approach zero. See Figure 19 for a plot of the skewness and Figure 20 for the kurtosis.

The generally positive kurtosis values indicate that confidence intervals for the mean based on the asymptotic normal theory will be slightly too narrow since the tails of the distribution of $\bar{s}'_n$ will be heavier than those for the normal case; this is confirmed empirically in Section III.D.5. The positive skewness values probably derive from the shape of the distribution of the starting value $\bar{s}_1$, which from Figure 1 is markedly skewed to the right. Note that neither γ_1 nor γ_2 is great enough for X samples larger that 3000 observations to cause objectionable departures from normality.

Figures 21 through 23 allow us to examine the convergence of $s^{*\prime}_n$ in distribution more directly. These histograms were computed from samples of 2500 replications of $s^{*\prime}_{50}$, $s^{*\prime}_{100}$ and $s^{*\prime}_{150}$ respectively. In this case the replications were <u>not</u> independent; this enables one to gauge the progress of a specific $\{\bar{s}'_n\}$ sequence. The F's plotted on the histograms are a kernel estimate of the underlying

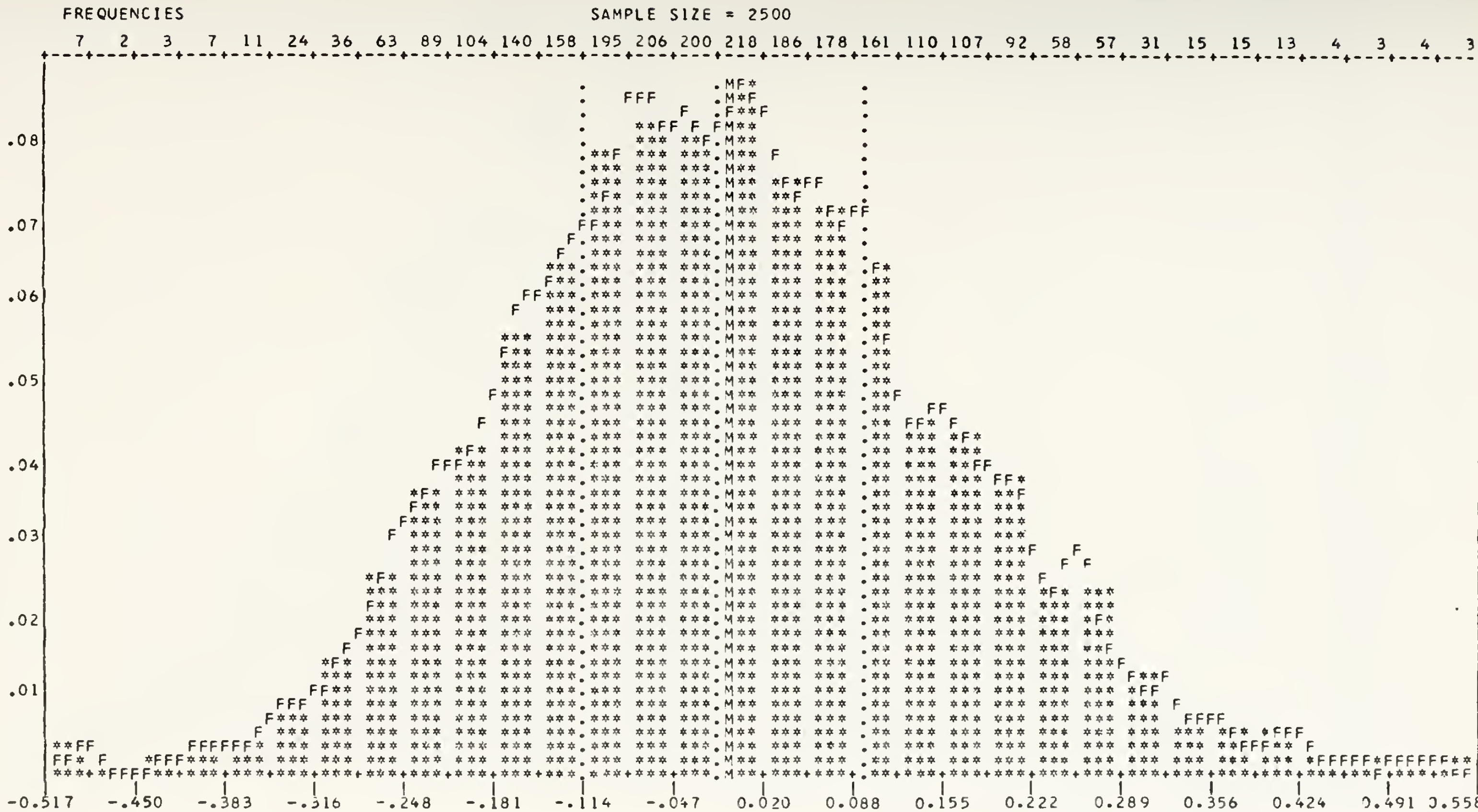
density of the $s^{*\prime}_n$ population; such density estimates have been found to give better insight into the nature of the underlying distribution than does the histogram alone.

In general, the histograms reinforce the conjecture that $\bar{s}'_n$ is converging rapidly to normality; in all three cases, the density has a definitely Gaussian shape which is slightly skewed to the right, the degree of skewness decreasing with increasing n. The sample extrema and range also decrease in a satisfactory manner.



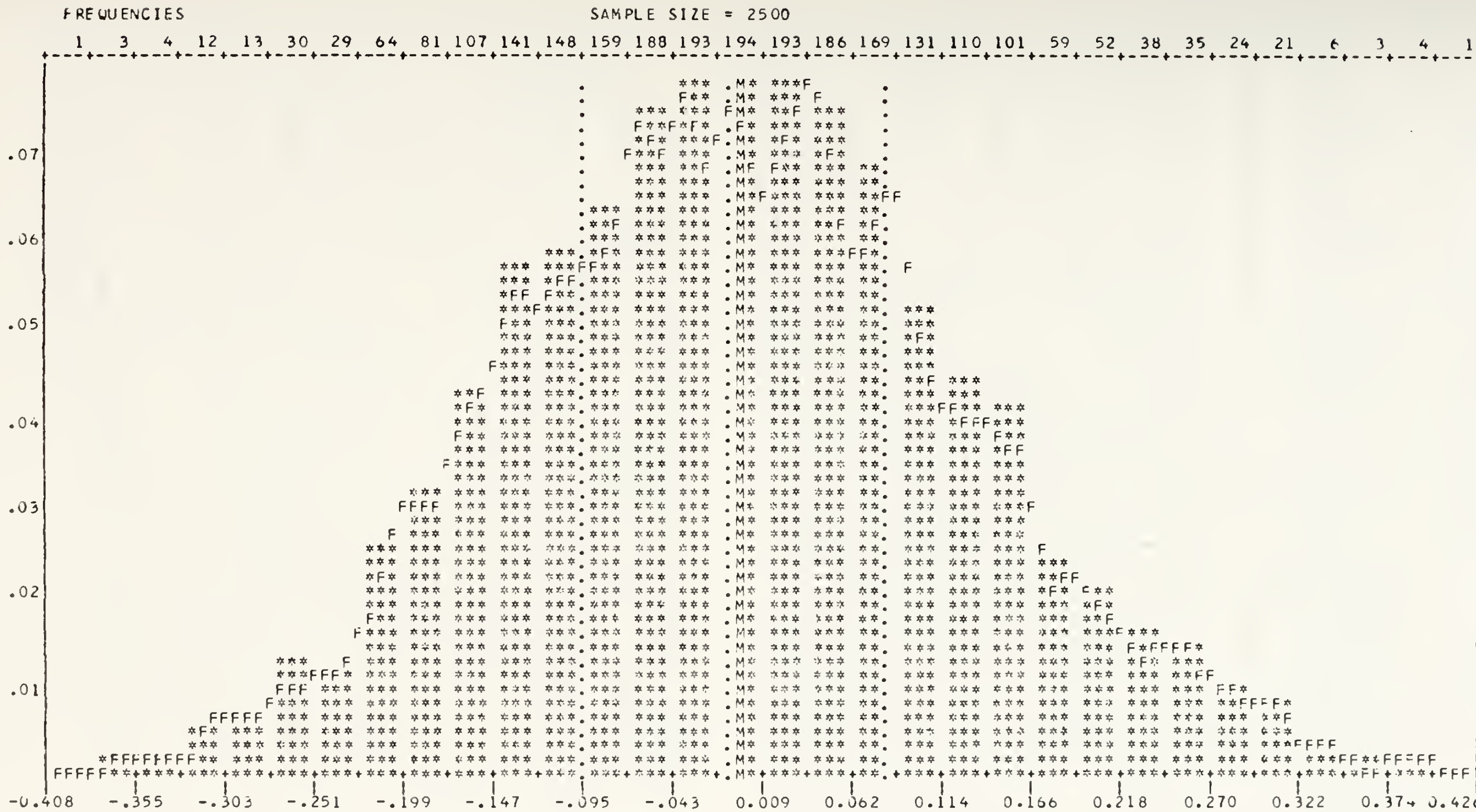
CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION		
MEAN	-8.358367E-03	VARIANCE	5.026086E-02	M3	2.405728E-03	MINIMUM		-6.452217E-01
MEDIAN	-1.905441E-02	STD DEV	2.241893E-01	M4	7.080279E-03	.10 QUANTILE		-2.910881E-01
TRIMEAN	-1.517487E-02	COEF VAR	2.682214E 01	SKEWNESS	2.135019E-01	.25 QUANTILE	(HINGE)	-1.653376E-01
MIDMEAN	-1.809308E-02	MEAN DEV	1.804244E-01	KURTOSIS	-1.972103E-01	.50 QUANTILE	(MEDIAN)	-1.905441E-02
MIDRANGE	8.438778E-02	RANGE	1.459219E 00	BETA1	2.402843E-03	.75 QUANTILE	(HINGE)	1.427469E-01
		MIDSPREAD	3.080845E-01	BETA2	7.075012E-03	.90 QUANTILE		2.949877E-01
						MAXIMUM		8.139973E-01

FIGURE 21. DISTRIBUTION OF THE BIAS OF THE STOCHASTIC APPROXIMATION ESTIMATOR $\bar{S}'_{50}$ FOR THE 0.99 QUANTILE OF THE EXPONENTIAL DISTRIBUTION. X SAMPLE SIZE WAS 2912 OBSERVATIONS.



CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	-7.793173E-03	VARIANCE	2.610873E-02	M3	6.921641E-04	MINIMUM	-5.171928E-01
MEDIAN	-1.269960E-02	STD DEV	1.615819E-01	M4	2.102156E-03	.10 QUANTILE	-2.117186E-01
TRIMEAN	-1.179504E-02	COEF VAR	2.073376E 01	SKEWNESS	1.640707E-01	.25 QUANTILE (HINGE)	-1.178966E-01
MIDMEAN	-1.245294E-02	MEAN DEV	1.284919E-01	KURTOSIS	8.385277E-02	.50 QUANTILE (MEDIAN)	-1.269960E-02
MIDRANGE	2.041817E-02	RANGE	1.075222E 00	BETA1	6.913340E-04	.75 QUANTILE (HINGE)	9.611559E-02
		MIDSPREAD	2.140121E-01	BETA2	2.100428E-03	.90 QUANTILE	2.046356E-01
						MAXIMUM	5.580292E-01

FIGURE 22. DISTRIBUTION OF THE BIAS OF THE STOCHASTIC APPROXIMATION ESTIMATOR $\bar{S}'_{100}$ FOR THE 0.99 QUANTILE OF THE EXPONENTIAL DISTRIBUTION. X SAMPLE SIZE WAS 5712 OBSERVATIONS.



CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION		
MEAN	-4.239768E-03	VARIANCE	1.700265E-02	M3	3.217466E-04	MINIMUM		-4.075422E-01
MEDIAN	-7.963181E-03	STD DEV	1.303942E-01	M4	8.264876E-04	.10 QUANTILE		-1.689816E-01
TRIMEAN	-7.327318E-03	COEF VAR	3.075504E 01	SKEWNESS	1.451239E-01	.25 QUANTILE	(HINGE)	-9.592819E-02
MIDMEAN	-6.958816E-03	MEAN DEV	1.045342E-01	KURTOSIS	-1.410723E-01	.50 QUANTILE	(MEDIAN)	-7.963181E-03
MIDRANGE	9.418011E-03	RANGE	8.339205E-01	BETA1	3.213605E-04	.75 QUANTILE	(HINGE)	8.254528E-02
		MIDSPREAD	1.784735E-01	BETA2	8.258584E-04	.90 QUANTILE		1.638727E-01
						MAXIMUM		4.263783E-01

FIGURE 23. DISTRIBUTION OF THE BIAS OF THE STOCHASTIC APPROXIMATION ESTIMATOR $\bar{S}'_{150}$ FOR THE 0.99 QUANTILE OF THE EXPONENTIAL DISTRIBUTION. X SAMPLE SIZE WAS 3512 OBSERVATIONS.

Chapter V. JOINT ESTIMATION OF A SET OF QUANTILES

In this Chapter we address the problem of obtaining estimates for several different quantiles from the same X population based on a single sample $X_1,\ldots,X_n$. This problem is one of considerable practical interest since one usually wishes to estimate more than just a single extreme quantile in data analysis or simulation studies. The problem also constitutes the primary area of application for the new stochastic approximation methods described in this work; as long as only one quantile is to be estimated the order statistic techniques of Section III.A can be quite modest in terms of both computation time and memory but they are completely impractical when dealing with ten or more quantiles at a time.

The major development in this Chapter is a computer program which is capable of providing estimates of the moments and quantiles of an arbitrary population given only sequential independent observations on the random variable. The total computer memory requirement (besides the code for the program) is just 150 memory cells per random variable. As Lewis [22] points out, there is often a requirement in statistical sampling experiments or systems simulation studies to collect simultaneous estimates on 30 or more random quantities; the FORTRAN subprogram QUANT given in the Appendix represents a way to do this with a reasonable amount of memory. The subroutine could thus be used directly in Lewis' COMPSTAT package [22] at a considerable saving in memory.

Non-parametric Quantile Estimation Through Stochastic Approximation

A. An Estimation Algorithm

Our basic approach to joint quantile estimation is to employ the nested design of Table I of Chapter I with the algorithm for the new stochastic approximation method given in Subsection IV.D.2. The main complication is that we must now provide a data structure to accommodate all of our set of estimates as well as the other information required to find the respective section maxima and minima.

We assume that the population median is to be estimated along with the a_j and $(1-a_j)$ quantiles, j=1,...,top. The quantile estimates are to be kept in array s with the median estimate in s[0], the a_j quantile in s[2j-1] and the $(1-a_j)$ quantile in s[2j]. A second array f is also required; f[k] will contain the density estimate corresponding to the quantile estimate in s[k].

Each quantile estimate also requires the five values n, b, m, h and a to be stored, just as in the single quantile algorithm of Section III.D.2; we may use the same value for each of these variables for both the a_j and the $(1-a_j)$ quantiles in this case, however, so we can save some memory storage here. Since we will be applying the maximum transform, we also require arrays u, max and min; u[j] will contain the size of the sample section considered so far for the a_j quantile, max[j] the largest value in the section and min[j] the smallest. One final array v will contain the v values for the maximum transform for each quantile. Since we use a nested method for determining the respective maxima here, a v[j] value of 2 means that the section for the a_j

quantile is twice as large as the section for the a_{j-1} quantile.

The values in the a and v arrays must be precomputed and will remain fixed throughout the estimation process. The remaining arrays must be initialized at the beginning of the algorithm just as in Subsection IV.D.2. In the ALGOL description below we suppress the initialization steps as they tend to obscure the operation of the method. We give an ALGOL-like description both because ALGOL is the standard language for setting forth algorithms and also because the result is more easily understood than a FORTRAN program. A FORTRAN implementation is given in the Appendix.

```
comment This first section carries out the stochastic
         approximation process for the median.
         The algorithm updates the various
         stochastic approximation arrays given
         the single input observation X;
t := | s[0] - X|;
if t < b[0] then f[0] := f[0] + (b[0] - t) / b[0]²;

comment upper bound on divisor;
nh := n[0] * h[0];
if f[0] > nh then d := nh else d := f[0];
s[0] := s[0] + y / d;
h[0] := h[0] + 1 / n[0];
n[0] := n[0] + 1;
b[0] := ( 1 - b[0]³/3m[0]³ ) * b[0];

comment here we pass the X values one at a time outwards
         to the other quantiles;
max[1] := X;    min[1] := X;
j := 1;    k := 1;
while j ≤ top do
```

```
begin
comment first we update the current max and min values;
if j > 1 then
     begin
     if u[j] = 0 then
          begin
          max[j] := max[j-1];    min[j] := min[j-1]
          end
     else
          begin
          if max[j] < max[j-1] then max[j] := max[j-1];
          if min[j] > min[j-1] then min[j] := min[j-1]
          end
     end;
u[j] := u[j] + 1;
comment determine if the current section is complete;
if u[j] ≠ v[j] then j := top + 1 else
     begin
     u[j] := 0;

     comment this section is for the alpha[j] quantile;
     t := |s[k] - max[j]|;
     if t < b[j] then f[k] := f[k] + (b[j]-t)/b[j]²;
     if max[j] ≤ s[k] then y := a[j]-1 else y := a[j];
     nh := n[j] * h[j];
     if f[k] > nh then d := nh else d := f[k];
     s[k] := s[k] + y / d;

     comment this  section  is  for  the  (1-alpha[j])
               quantile;
     t := |s[k+1] - min[j]|;
     if t < b[j] then f[k+1] := f[k+1] +
                                        (b[j]-t)/b[j]²;
     if min[j] ≤ s[k+1] then y := -a[j]
                        else y := 1 - a[j];
     if f[k+1] > nh then d := nh else d := f[k+1];
```

```
        s[k+1] := s[k+1] + y / d;

        comment here we update the constants for the
                alpha[j] quantile;
        h[j] := h[j] + 1 / n[j];    n[j] := n[j] + 1;
        b[j] := ( 1 - b[j]^3 / 3m[j]^3 ) * b[j];
        j := j + 1;    k := k + 2
        end
    end;
```

The introduction of an initialization section makes the algorithm somewhat more complex, but even greater difficulty ensues when we combine all the arrays into a single data structure (which is also an array) as we do in the FORTRAN subroutine QUANT which is listed in the Appendix. This use of a single array has the advantage, however, that we may now accumulate quantile estimates on several different random variables as long as each one is allocated its own estimation array.

We may incorporate the next-to-maximum transform into this scheme by adding yet another array nextmax to our algorithm (or an extra set of memory locations into the single array as has been done in QUANT, for example.) The section maximum update steps in the algorithm now become

```
  if max[j-1] > nextmax[j] then
    if max[j-1] > max[j] then
        begin
        nextmax[j] := max[j];
        max[j] := max[j-1];
        if nextmax[j-1] > nextmax[j] then
                nextmax[j] := nextmax[j-1]
        end
    else nextmax[j] := max[j-1];
```

A second array nextmin with a similar update sequence will also be required for the lower quantiles. The stochastic approximation operations will then be carried out using the values in nextmax[j] and nextmin[j].

Either version of the joint estimation algorithm requires that we have available fairly large samples of data. In order to obtain varaince estimates for the most extreme quantiles we need a minimum of 4v+3 observations, i.e. a total of 2691 for the maximum transform design of Table I and 6147 for the next-to-maximum transform design. This emphasizes the point that stochastic approximation quantile estimation is a large sample technique.

B. Reordering Techniques

The first discrepancy noted when using Monte Carlo methods to investigate the performance of the algorithm of Section A is that the resulting quantile estimates are sometimes not in the proper order. In what follows, we assume that we are to estimate the $a(1), a(2), \ldots, a(m)$ quantiles, where $a(i) < a(j)$ for $i < j$. Since s_a satisfies $F(s_a) = a$ and since every distribution function $F(\cdot)$ is monotone, we must have

$$(1) \qquad s_{a(i)} \leq s_{a(j)} \quad \text{for } a(i) < a(j)$$

with strict inequality when $F(\cdot)$ is continuous. In any event, if the joint estimates $\bar{s}_{a(i)}(n) > \bar{s}_{a(j)}(n)$ result from a sample $X_1, \ldots, X_n$ from the parent population we

clearly have an error for which we should make some adjustment.

This adjustment may be made only after the final set of estimates is obtained or it may be carried out dynamically throughout the estimation process whenever any of the set of quantile estimates violates the relation (1). It turns out that the dynamic readjustment of the estimates can materially improve the overall precision of the final estimates, where we adopt as a measure of this precision the total squared error of the set of m quantile estimates, i.e.

$$T_{mn} = \sum_{j=1}^{m} [\bar{s}_{a(j)}(n) - s_{a(j)}]^2 .$$

The expected value of T_{mn} is just the sum of the mean squared errors of the individual quantile estimates. None of the readjustment processes considered here changes the asymptotic distribution of $\bar{s}_n$ since none of them will be used if the set of quantile estimates satisfies (1); the almost sure convergence of $\bar{s}_n$ implies that the order relationship (1) will hold almost surely for any sequence of joint estimates. A reduction in the value of $E[T_{mn}]$ thus represents a decrease in the bias of the individual estimates $\bar{s}_{a(j)}(n)$ rather than a change in the asymptotic

variance.

One way to reduce the expected value of T_{mn} for $m \geq 3$ is to employ the James-Stein estimation process (for an explananation with examples see Efron and Morris [8]). Briefly, the idea is to decrease the value of each $\bar{s}_{a(j)}(n)$ slightly [the amount depends on the actual variance of $\bar{s}_{a(j)}(n)$] so as to move the estimate closer to the surface of the m-dimensional hypersphere on which the point $[s_{a(1)},\ldots,s_{a(m)}]$ lies.

The set $\{\bar{s}_{a(1)},\ldots,\bar{s}_{a(m)}\}$ of quantile estimates does not exactly satisfy the requirements for the James-Stein adjustment since we do not know the precise theoretical variances. Furthermore, although some perturbation of the order of the set of estimates occurs, the adjusted set does not in general satisfy (1). The James-Stein technique was applied dynamically (using estimated variances) during the stochastic approximation joint quantile estimation procedure and it was found to make the properties of the extreme quantile estimates materially worse. We thus reject this method of adjustment.

The most straightforward of the methods that has been found to reduce the expected value of T_{mn} in some cases is

simply to adjust any of the $\bar{s}_{a(j)}(n)$ values which fall outside of the interval $[X_{(1)}, X_{(n)}]$ back to the nearest boundary of the interval. It is quite easy to keep track of the sample extrema $X_{(1)}$ and $X_{(n)}$ since the process requires only two additional memory cells; the subprogram QUANT in the Appendix was designed with this capability.

From (3.1), the probability that the sample range $[X_{(1)}, X_{(n)}]$ covers the a-quantile is just

$$\Pr\{X_{(1)} \le s_a \le X_{(n)}\} = \sum_{i=1}^{n-1} \binom{n}{i} a^i (1-a)^{n-i}$$

$$= 1 - a^n - (1-a)^n ;$$

thus the adjustment is more likely to reduce the bias of $\bar{s}_n$ as the sample size increases. Since the initial estimate is based on a sample of size $3v$, where $a^v \doteq 0.5$, the probability that the interval for the first maximum transformed estimate $\bar{s}'_2$ contains s_a is approximately 0.875; this follows because the interval is $[X_{(1)}, X_{(4v)}]$ and $a^{4v} \doteq 0.0625$. The probability that the interval for $\bar{s}'_n$ contains s_a is similarly $1 - [\frac{1}{2}]^{n+1}$, which rapidly approaches 1.0.

For reasons of practical utility we prefer to carry this so-called extremum adjustment (as well as the other adjustment methods) only when the value of the most extreme quantile estimate $\bar{s}_{a(m)}$ changes; in the maximum transformed case this will occur for each v[m] observations on X. This not only decreases the amount of time devoted to the adjustment process but it also can be done very conveniently in the algorithm. In subroutine QUANT, the call to subroutine CHECK near the end of the quantile estimation loop is an invocation of the order adjustment method.

As can be seen in Table IX the extremum adjustment apparently helps slightly in the small sample exponential case but there seems to be very little basis for adopting this method in general. Furthermore, the extremum adjustment will have no effect on violations of (1) unless both quantile estimates lie outside $[X_{(1)}, X_{(n)}]$ in the same direction. We thus seek a general technique for dealing with estimates which are in reverse order.

Such a technique arises from considering the problem of estimating the means μ_1 and μ_2 of two independent normal random variables X_1 and X_2 with respective known variances σ_1^2 and σ_2^2. If we have a single pair of realizations x_1 and x_2, the maximum likelihood estimators $\bar{\mu}_1$ and $\bar{\mu}_2$ arise

from minimizing the quadratic form

$$Q(\bar{\mu}_1, \bar{\mu}_2) = \sum_{i=1}^{2} \left[\frac{x_i - \bar{\mu}_i}{\sigma_i} \right]^2 ;$$

the result is clearly $\bar{\mu}_i = x_i$, $i=1,2$. If we know a priori, however, that $\mu_1 \geq \mu_2$ and it happens that $x_1 < x_2$, we must solve the quadratic programming problem

$$\min \quad Q(\bar{\mu}_1, \bar{\mu}_2)$$

subject to

$$\bar{\mu}_1 \geq \bar{\mu}_2$$

in order to obtain the maximum likelihood estimators. The required minimum occurs at

$$\bar{\mu}_1 = \bar{\mu}_2 = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} .$$

Note that this is just a weighted average of the x_i's, the weights being chosen as $1/\sigma_i^2$; in a sense, the weight w_i for x_i is just a measure of the precision of x_i as an estimate of μ_i.

The foregoing discussion is an example of so-called "isotonic" regression techniques (the term isotonic means "order preserving"). These techniques are applicable in situations far more complex than our present simple

requirement that $\bar{s}_{a(1)} \le \bar{s}_{a(2)} \le \dots \le \bar{s}_{a(m)}$; for more sophisticated applications as well as a summary of the basic theory see Barlow et al [1]. The isotonic adjustment technique for the situation where $\bar{s}_{a(i)}(n) > \bar{s}_{a(i+1)}(n)$ is then to use as estimates for both quantiles the same value, namely

$$\frac{w_i \bar{s}_{a(i)}(n) + w_{i+1} \bar{s}_{a(i+1)}(n)}{w_i + w_{i+1}} \;;$$

the weights used here are just the reciprocals of the estimated variances, i.e.

$$(2) \qquad w_i = \frac{n\ B_n^2}{a(i)[1 - a(i)]}$$

where B_n is the density estimate for $f(s_{a(i)})$. The value of n in (2) may change with a(i) depending on the maximum transform scheme used, if any.

The main complication here is that the entire set of m quantiles must be ordered rather than just adjacent pairs of estimates; thus, if it is found that $\bar{s}_{a(i+2)} < \bar{s}_{a(i+1)}$ after the adjustment of the previous paragraph is made it will be necessary to set <u>all three</u> of the estimates to the same value which is now

$$\frac{w_i \bar{s}_{a(i)}(n) + w_{i+1} \bar{s}_{a(i+1)}(n) + w_{i+2} \bar{s}_{a(i+2)}(n)}{w_i + w_{i+1} + w_{i+2}} .$$

We have now created a block $\{\bar{s}_{a(i)}, \bar{s}_{a(i+1)}, \bar{s}_{a(i+2)}\}$ of estimates whose values are equal; if this constant value is not in the proper order with respect to some other adjacent block of estimates it will then become necessary to coalesce the two blocks together in the same fashion.

An algorithm for manipulating the blocks in this manner was developed by Kruskal [20]; it is also given by Barlow et al [1]. This so-called "up-and-down blocks" algorithm has also been implemented using the weights (2) for the data structure used by the QUANT subroutine. The resulting FORTRAN program is called CHECK and is listed in the Appendix.

A possible extension to the isotonic adjustment is to adjust the density estimates B_n at the same time that the quantile estimates are adjusted. There is of course no reason to suppose that the densities will also be in order, but it seems reasonable that if all the quantile estimates in a block have the same value that all the corresponding density estimates should also be constant. This may be accomplished using the same weights as used for updating the $\bar{s}_{a(i)}$ values. Alternatively, we may adjust each B_n so that the estimated variance calculated by (3.8) for each estimator in the block is the same. Recalling that we chose $w_i = 1/\sigma_i^2$, the variance of the block average in block b is given by

$$V_b = \frac{\sum w_i^2 \sigma_i^2}{(\sum w_i)^2}$$

$$= \frac{\sum w_i}{(\sum w_i)^2}$$

$$= (\sum w_i)^{-1} .$$

The adjusted density estimates are then given by

$$B_n(i) = \sqrt{\frac{a(i)[1-a(i)]}{n(i)} V_b} , \tag{3}$$

where $n(i)$ is the sample size for the $a(i)$-quantile. This second scheme was in fact investigated; Monte Carlo results for both the isotonic adjustment technique and the isotonic technique with density modification are given in Table IX.

It is apparent from Table IX that the isotonic adjustment method greatly improves the expected total squared error of the set of quantile estimates. The decrease is over 50 % for both the normal and exponential cases. The density adjustment, however, does not improve $E[T_{mn}]$ nearly as much if, indeed, it improves it at all.

One difficulty encountered in using the isotonic adjustment technique is that if one of the extreme quantile estimates (say $\bar{s}_{0.995}$) is out of order with respect to an estimate on the other extreme (e.g., $\bar{s}_{0.995} < \bar{s}_{0.05}$) then

Adjustment Method	Normal Distribution		Exponential Distribution	
	n=6720	n=67,200	n=6720	n=67,200
Unmodified	1.9348	0.0444	7.8323	0.5722
	(0.5396)	(0.0328)	(1.7157)	(0.4473)
James-Stein	2.5162	6.7694	*	*
	(0.0561)	(0.0607)		
Extrema	1.9347	0.0444	7.8300	0.5723
	(0.5396)	(0.0328)	(1.7158)	(0.4473)
Isotonic	0.9262	0.0105	3.5547	0.2131
	(0.0362)	(0.0007)	(1.0650)	(0.1332)
Isotonic (Density)	1.5109	0.0492	6.2388	*
	(0.4205)	(0.0132)	(1.3962)	
Limited Reorder	1.5188	0.0197	2.0328	0.1343
	(0.5766)	(0.0075)	(1.0749)	(0.0501)
Limited Reorder (Density)	1.4605	0.0612	*	*
	(0.4939)	(0.0426)		

Table IX. Mean of the total squared error T_{mn} for the m = 19 quantiles of Table I using various reordering methods to adjust for estimates which are out of order. Values are the mean of 100 replications of each T_{mn} statistic; numbers in parentheses are the estimated standard deviations of the given estimates of $E[T_{mn}]$. Asterisks (*) denote experiments that were not conducted.

all of the intervening quantile estimates will be set to the same value even though they may be close to their correct values. The extent to which this may occur depends on the parent population but it is likely to be a problem since the extreme quantile estimates will be the most variable, especially for moderate sample sizes.

One way to overcome this difficulty is to use a "limited" reorder scheme in which each estimate is checked with respect to those immediately adjacent. If it is found, for example, that

$$\bar{s}_{a(i)} < \bar{s}_{a(i-1)}$$

but that

$$\bar{s}_{a(i-1)} < \bar{s}_{a(i+1)},$$

then we discard the old estimate $\bar{s}_{a(i)}$ and set

$$\bar{s}_{a(i)} = \frac{w_{i-1}\bar{s}_{a(i-1)} + w_{i+1}\bar{s}_{a(i+1)}}{w_{i-1} + w_{i+1}} .$$

If the estimates $\bar{s}_{a(i-1)}$ and $\bar{s}_{a(i+1)}$ are also out of order we merely carry out the usual isotonic regression adjustment.

The limited reorder adjustment may also be applied with the density adjustment (3) used in the isotonic case. The results from Table IX indicate that this method shows some promise but it does not appear to be generally as good as the isotonic case. Once again, the density adjustment does

not seem to be useful.

The results in Table IX show a substantial reduction in $E[T_{mn}]$ when we adopt the isotonic adjustment; as mentioned previously, this is an indication of a reduction in the bias of the $\bar{s}_{a(i)}(n)$'s. It is possible that this bias reduction will now make the stochastic approximation estimators more competitive with order statistic estimators. Direct computation shows for the order statistic case, however, that the total squared error for a sample of 6720 exponential variates is 0.2907 and it is 0.0285 for 67,200 observations. Thus, a better reordering method is needed to obtain comparable bias results. Even though it is possible that the stochastic approximation estimators can be further improved, we will be unable to improve the order statistic estimators any further in this way since none of the reordering methods are applicable in this case.

Our conclusion then is that the isotonic adjustment is a robust and flexible method for reducing the expected total squared error of a set of stochastic approximation quantile estimates and that simultaneous adjustment of the step size parameters is not indicated. The limited reorder adjustment may be better in some applications; more work could be done in this area.

Chapter VI. FUNCTIONS OF QUANTILES

In this Chapter we investigate the question of whether our methods can be adapted to the joint estimation of an unknown quantile and some random function of that quantile. Of course, one case in which we already know that this can be done is the estimation of $\beta = f(s_a)$ using a kernel estimator since this density estimate is used directly in the quantile estimation process. We first determine what kinds of functions we may use in this joint estimation procedure and we then give an example which is of practical use in statistical simulation studies.

A. Sufficient Conditions for Convergence

Given a sample $X_1,\ldots,X_n$ from a population with distribution function $F(\cdot)$ satsifying (F1) and (F2) we obtain the corresponding a-quantile estimates $\bar{s}_1,\bar{s}_2,\ldots,\bar{s}_{n+1}$. At each stage of the process we also have a random vector Y_n (possibly empty) which we use to compute the value of the known function $P_n(\bar{s}_n,X_n,Y_n)$; we are then interested in the properties of

$$(1) \qquad p_n = \frac{1}{n} \sum_{i=1}^{n} P_i(\bar{s}_i,X_i,Y_i) .$$

Of course more general formulations involving several previous X_n or $\bar{s}_n$ values are possible but since our emphasis throughout this work is on methods which conserve storage we limit ourselves to the formulation (1).

We approach (1) in the same way as we proved Theorem 2 in Chapter II; first, however, we must redefine the sequence of σ-fields $\mathcal{B}_n = \sigma(\bar{s}_1;X_1,\ldots,X_{n-1};Y_1,\ldots,Y_{n-1})$ to include the Y variables. Then we write

$$(2) \qquad t_n(\bar{s}_n) = E[P_n(\bar{s}_n,X_n,Y_n) \mid \mathcal{B}_n] .$$

Expanding (1) we have

$$(3) \qquad p_n = \frac{1}{n} \sum_{i=1}^{n} \{P_i(\bar{s}_i,X_i,Y_i) - t_i(\bar{s}_i)\} + \frac{1}{n} \sum_{i=1}^{n} t_i(\bar{s}_i) .$$

The first term in (3) will approach 0 almost surely according to Lemma 2 if we have

$$(4) \qquad \mathrm{Var}[P_n(\bar{s}_n,X_n,Y_n)] = o(n),$$

since then $\sum_{n=1}^{\infty} n^{-2}\,\mathrm{Var}[P_n(\bar{s}_n,X_n,Y_n)]$ will converge.

The second term in (3) will converge a.s. according to Lemma 5 as long as $t_n(\bullet)$ is measurable and uniformly continuous for every $n \geq N$; in this case we have

$t_n(s_a) \to t(s_a)$ and so

$$t_n(\bar{s}_n) \to t(s_a) \text{ a.s.}$$

In view of Lemma 6 we have thus proved:

<u>Theorem 5</u> As long as $P_n(\bar{s}_n, X_n, Y_n)$ satisfies (4) and $t_n(\bullet)$ given by (2) is measurable and uniformly continuous then

$$p_n \to t(s_a) \text{ a.s.,}$$

where p_n is given by (1).

□

B. Applications

In a statistical simulation study we may generate sufficient pseudo-random samples of X_n to obtain a satisfactory estimate $\bar{s}_n$ of the a-quantile and then repeat the experiment and compute p_n using the final quantile estimate value, i.e. we calculate

$$p'_n = \frac{1}{n} \sum_{i=1}^{n} P_i(\bar{s}_{n+1}, X_i, Y_i) .$$

This value should have a lower bias than p_n (at least in the first few terms) since it is based on a more correct

estimate of s_a. We may also use the p'_n estimate with a fixed data sample which is recorded on a storage medium which allows re-examination of the data, e.g. magnetic tape.

If the X values are difficult to generate, however, it may become prohibitively expensive to repeat the entire experiment from the beginning to take advantage of the presumably lower bias of p'_n. It may also be impossible to repeat the early X values if the source of the data is a real-time system of some sort, for example. In these cases we prefer to use the dynamic estimate p_n in order to conserve memory storage requirements.

The basic application envisioned for this technique is the estimation of empirical distribution functions and percentiles (see the next Section). It may also be used for estimating density values from other distributions, i.e. we take

$$P_n(\bar{s}_n, Y_n) = \frac{1}{b_n} W\left[\frac{\bar{s}_n - Y_n}{b_n}\right].$$

Evidently then $p_n \longrightarrow f_Y(s_a)$ in this case as long as the distribution function $F_Y(\bullet)$ of the Y population satsifies (F1) (see Lemma 7). This same method may be readily extended to the estimation of joint density functions.

C. Power and Level of a Test

As an application of our method we consider the statistical simulation problem of estimating the level of a statistical test and then determining the power of the test against various alternative hypotheses at the chosen level. Suppose, then, that we have a simple hypothesis H_0 and a finite set of simple alternate hypotheses $H_1, \ldots, H_m$. The test statistic T is proposed for testing H_0; the (unknown) distribution of T under H_j will be denoted by $F_j(\bullet)$, $j=0,1,\ldots,m$. We assume that $F_0(\bullet)$ satsifies (F1) and (F2) and that each of the $F_j(\bullet)$, $j=1,\ldots,m$, satisfies (F1).

We wish to determine a level T_a for the test statistic T such that the probability of a Type I error in testing H_0 will be a. Assuming that the test region is $T \leq T_a$, the test level is the solution to

$$\Pr\{T \leq T_a \mid H_0\} = 1 - a,$$

or

$$F_0(T_a) = 1 - a,$$

i.e. T_a is the 1 - a quantile of $F_0(\bullet)$. It is straightforward to extend this to other test regions.

Realizations of the statistic T are now obtained by

sampling sets $X_1^0, X_2^0, \ldots, X_k^0$ of k values each from a population satisfying H_0; in the simulation context, these samples are generated by a pseudo-random number generator. The value T_n^0 is then computed from the n<u>th</u> $\{X^0\}$ sample and may be used to obtain a new stochastic approximation estimator of T_a using the algorithm of Chapter III (or Chapter V if several different values of a are of interest). We denote this n<u>th</u> sequential estimate of T_a by $\hat{T}_n$.

Now suppose that in addition to the $\{X^0\}$ sample we have samples from populations satsifying H_j, $j=1,\ldots,m$; we denote such a sample by $\{X^j\}$. Note that it may be very easy to generate such samples given the basic $\{X^0\}$ sample; if, for example, the null hypothesis involves $E[X^0] = 0$ while H_j requires $E[X^j] = \mu_j \neq 0$ then each $\{X^j\}$ sample may be generated by adding an appropriate constant to $\{X^0\}$. From each $\{X^j\}$ sample, then, we compute the statistic T, denoting the n<u>th</u> realization by T_n^j.

The power of the test based on T is just the probability that under H_j the statistic T fails the test, i.e.

$$(5) \qquad p^j = \Pr\{\, T > T_a \mid H_j \}$$
$$= 1 - F_j(T_a)\ .$$

Note that the power defined by (5) is one minus the T_a-percentile of $F_j(\bullet)$. According to Theorem 5 we may then use as an estimate of p^j

$$\bar{p}_n^j = \frac{1}{n} \sum_{i=1}^{n} P_i(T_i, T_i^j)$$

as long as $P_i(T_i, T_i^j)$ has the correct properties. In fact, if we choose

$$(6) \qquad P_i(T_i, T_i^j) = \begin{cases} 0 \text{ if } T_i^j \le T_i \\ 1 \text{ if } T_i^j > T_i \end{cases}$$

then we have

$$\mathrm{Var}[\, P_i(T_i, T_i^j)\,] \le \frac{1}{4} = o(n)$$

and

$$(7) \qquad E[\, P_i(T_i, T_i^j) \mid B_i\,] = 1 - F_j(T_i) \text{ a.s.}$$

Now (F1) guarantees that $F_j(\bullet)$ will be continuous in some

closed neighborhood of s_a and since $F_j(\bullet)$ is bounded it will be uniformly continuous there. Thus,

$$\bar{p}^j_n \to 1 - F_j(T_a) = p^j \text{ a.s.}$$

Note that (7) does not require that T^0_i and T^j_i be independent; in fact if we are able to use the $\{X^0\}$ sample to generate $\{X^j\}$ they will certainly not be. A degree of positive correlation between T^0_i and T^j_i, moreover, may actually improve the estimate $\bar{p}^j_n$. If T^0_i is large then T_i will also be large; however, T^j_i is also large in this case so that the tendency will be for (6) to add an appropriate value to $\bar{p}^j_n$.

Since we are usually interested in very small probabilities of Type I error, we will generally have the probability of error, a, very small. Hence, it will most often be necessary to use the maximum transform to estimate T_a. In this case we continue to accumulate $P_i(T_i, T^j_i)$ terms even though the value of T_i has not changed since the previous step. This does not change the analysis to any great extent; we are merely adding a binomial random variable to the sum instead of a Bernoulli as before.

It is not hard to show, using Lemma 11 and following the lines of the proof of Theorem 4, that $\bar{p}_n^j$ has an asymptotically normal distribution. In fact,

$$(8) \qquad \bar{p}_n^j \xrightarrow{L} N\left[p^j, \frac{p^j(1 - p^j)}{n} \right].$$

Some empirical investigation of this method has been carried out using the FORTRAN subprogram POWER listed in the Appendix. The example chosen was the estimation of the power of the t-test. The statistic is

$$t_n = \frac{z + d}{s_z}$$

where z is a zero-mean normal random variable, d a constant and s_z an independent estimate of Var[z] based on n degrees of freedom. When d is zero t_n has Student's t-distribution with n degrees of freedom while t_n has a non-central t-distribution when $d \neq 0$.

The quantiles of both the central and non-central t-distributions may be readily approximated so that the results of the joint estimation procedure can be checked. The null hypothesis is $H_0 : d = 0$ while the alternate hypotheses are $H_j : d = d_j \neq 0$. Because of the time required to carry out the simulation no attempt was made to determine the order of the bias or to verify the asymptotic distribution (8); the results for several different n values, however, were in good agreement with theory.

Chapter VII. SUMMARY AND CONCLUSIONS

A. Main Results

The main contribution of this research is the development of a practical sequential quantile estimation method which can be applied even for extreme quantiles. Both the asymptotic and finite sample properties of this new method have been shown to be comparable to those of the order statistic method which is the most commonly used non-parametric technique for estimating quantiles; the new method requires only a small, fixed amount of memory for its implementation, however, and is thus superior to the order statistic estimator for large samples of data.

Monte Carlo experience with the new estimation method shows that it is quite robust with respect to the underlying distribution of the random variable whose quantile is to be estimated. Use of the maximum transform of Goodman, Lewis and Robbins [14] allows the method to be applied even for extreme quantiles without the grossly unstable finite sample behavior which has characterized most attempts at stochastic approximation quantile estimation; see, for example, Wetherhill [36], Cochran and Davis [4] or Iglehart [16]. Since the method also provides an estimate of the variance of the quantile estimate, confidence intervals on the quantile may be computed. This is a <u>sine qua non</u> of good simulation practice. The technique thus qualifies as a flexible building block for use in data analysis or simulation computer programs. Because of the modest memory requirements it may be used in such programs for dealing with more than a single random variable.

Extension of the scheme to the estimation of a set of quantiles allows further improvement of the results by taking advantage of the known order relations in the set of quantiles; the resulting reduction in the bias may be substantial. Furthermore an entire set of quantiles such as the 19 values considered in this thesis provides an excellent characterization of the distribution of a random variable X; this information may be much more meaningful than just the moments of X, especially for highly skewed or outlier-prone data.

The development of a technique for the simultaneous estimation of both the level and power of a statistical test is also a useful contribution. When carrying out such statistical estimation experiments Monte Carlo methods are generally applied for a wide range of test sample sizes. The overall savings can be substantial since use of the simultaneous estimation method results in a saving for each test investigated.

All of the algorithms described in this thesis have been implemented as FORTRAN subroutines; some of these are particularly flexible and are listed in the Appendix. Subroutine QUANT implements the joint quantile estimation algorithm of Chapter V while subroutine CHECK implements the isotonic adjustment algorithm of Chapter V. Subroutine POWER is for the simultaneous level/power estimation technique of Section VI.C while QOUT and PWROUT print out the estimates accumulated by QUANT and POWER, respectively. Specific details of the data structures and algorithms employed may be found in the comments which accompany the subroutine listings.

Sample output from subroutine QUANT is also included in the Appendix; the input data in this case was a pseudorandom

sample from the exponential distribution. The accuracy of the results may be judged by comparing them with the true values which are also listed in the Appendix. A sample of the application of subroutine POWER is also included; the input was the t-test experiment data described in Section VI.C. Once again the true values are also listed for comparison.

B. Proposed Applications

As has been mentioned several times, Monte Carlo simulation is the primary application envisioned for the improved stochastic approximation quantile estimator developed in this work. The large samples of data required to obtain reasonable results from the procedure are easily obtained in a simulation experiment; further, the experiment can be designed so that the sequential X observations are independent and have a continuous distribution. The inevitable development of larger and faster computers will make the techniques even more valuable as larger simulation experiments become possible. Finally, in simulation work we usually wish to obtain estimates of high precision so that the magnitude of the bias encountered in some order statistic methods is often unsatisfactory.

The algorithm of Section V.A could profitably be employed as a part of a large-scale simulation package (even though the implementation given in subroutine QUANT is for independent use). An example is the COMPSTAT program of Lewis [22] which was designed to allow the user to employ Monte Carlo methods to investigate statistical distribution problems; a large part of COMPSTAT is concerned with providing summary data on the statistics generated by the user and subroutine QUANT is ideal for that purpose.

The method is not as readily applicable to more general systems simulation studies (e.g., queueing problems) because sequential observations are often not independent in this case. If one is interested in steady-state behavior, however, the regenerative simulation techniques of Iglehart [16] can be used to generate independent replications. Since these regenerative techniques tend to be fairly specific to the problem at hand some care must be exercised in using the improved stochastic approximation quantile estimator here.

The question of independence is also an important one in deciding whether the new quantile method can be applied in a general data analytic role with "real world" data. A more important consideration here, though, is whether sufficient observations are available; subroutine QUANT, for example, requires a minimum of 2691 data points and this number will be much larger if the next-to-maximum transform is used. Given the memory size of modern-day computers, however, it is reasonable to accommodate arrays of up to 5000 observations in core storage; it will then be possible to use one of the order statistic methods of Section III.A directly on the sample. Since the order statistic estimators avoid the maximum transform variance inflation of the stochastic approximation estimators they should be used when it is possible to do so.

Two cases in which enough data will be available are real time systems and large data bases; in both cases obtaining information for system management is a topic of considerable current interest (see Gaver and Lewis [12]). In fact, so much data may be available in these instances that order statistic estimators cannot be applied because of memory restrictions. The modest memory requirement for subroutine QUANT would make it ideal for dynamically

accumulating data in a real time system; for example, estimating job execution time parameters in a computer operating system can be done very easily without the necessity for saving a complete record of all the job times on some external storage medium as is usually done.

Gaver and Lewis [12] give an example of applying stochastic approximation quantile estimation in large data bases. They suggest that the next-to-maximum transform be used and that sample maxima which deviate too far from the quantile estimate be subjected to verification by the original source of the data as an automatic error correction device. In this application the density estimates provided by the improved method should be useful for deciding just when the maximum is "too far" from the quantile estimate. When working with data base information, however, care must be exercised that the data is sufficiently continuous to allow application of stochastic approximation.

C. Areas for Further Study

Three general areas in which more work could profitably be done suggest themselves: improving and refining the stochastic approximation quantile estimation procedure given here, investigating the performance of the procedure when it is applied to other kinds of data than those considered for this thesis and extending the procedure to handle more general kinds of inputs.

The basic method set forth in Section III.D could perhaps be improved if a better kernel function or a better bandwidth sequence were chosen. There is the danger that a combination of density estimation parameters may be nearly optimal in one application and yet lack the robustness displayed by the present choices; a practical choice must

also be fairly rapid computationally. A specific combination may be evaluated by using the regression methods of Chapter IV to estimate the $n^{-1/2}$ bias component in a set of independent realizations of s^*_n, $n = 1,2,\ldots$; some investigation of the distributional properties of the estimator along the lines of Section IV.E would also be indicated.

The joint estimation method could also possibly be improved by a better reordering scheme; the limited reorder technique, for example, shows some promise here. Once again a new adjustment method should be fairly robust, not disturb the distributional properties of the individual estimators and be computationally fast. A more careful comparison with the order statistic case might also be carried out here.

The data used in the testing of the improved stochastic approximation quantile estimator was all from fairly well-behaved distributions and the resulting estimates were also well-behaved; the performance of the method in the face of outlier-contaminated data should also be investigated. The idea of Gaver and Lewis [12] for the possible rejection of section maxima as outliers based on quantiles estimated from the next-to-maxima would be a good place to begin this investigation. General use of the method as presented in this thesis on real world data might also disclose shortcomings which might be overcome by using other kernel functions or by changing the starting values.

It would also be interesting to determine the effect of using the stochastic approximation algorithm on data samples from discrete distributions or from an autocorrelated process of some sort. Although convergence in these cases

is not guaranteed one has the feeling that the results ought not to be too bad for samples which are not too extreme.

One criticism of the stochastic approximation estimator is that it is sensitive to the order in which the sample is obtained; a determination of the effect of the order of the original sample on the final estimate might disclose how robust the procedure is in this case. Note that the process of reordering a sample may be used to introduce dependencies into the data, if desired.

Finally we turn to extending the theory behind the stochastic approximation method to include X samples from populations more general than those allowed in Chapter II, e.g. those with weak dependencies of some sort or those which are discrete. Almost nothing has appeared in the literature on these questions but weakening some of the assumptions of Chapter II would provide a powerful extension to the method presented here.

```
C      SUBROUTINE  QUANT  (Q, X, N)
C
C      THIS SUBROUTINE IS DESIGNED TO PERFORM SEQUENTIAL DATA
C         ANALYSIS ON A SET OF INPUT DATA WHEN ALL THE DATA CANNOT
C         BE EXAMINED AT ONE TIME.
C
C      THE DATA IS MADE AVAILABLE TO THE SUBROUTINE IN THE INPUT ARRAY
C         X, WHICH CONTAINS N OBSERVATION VALUES.
C
C      THE ARRAY Q IS USED TO ACCUMULATE ESTIMATES.  POSITIONS 1 - 7
C         IN THIS ARRAY ARE USED FOR MOMENT ESTIMATES AS FOLLOWS:
C
C         Q(1)  ==>  TOTAL SAMPLE SIZE FOR ESTIMATES ACCUMULATED SO FAR
C         Q(2)  ==>  SUM OF ALL OBSERVATIONS
C         Q(3)  ==>  SUM OF SQUARED DEVIATIONS ABOUT SAMPLE MEAN.
C                    CORRECTED WHEN MEAN ESTIMATE CHANGES.
C         Q(4)  ==>  SUM OF CUBED DEVIATIONS.
C         Q(5)  ==>  SUM OF FOURTH POWERS OF DEVIATIONS.
C         Q(6)  ==>  MINIMUM OBSERVATION IN SAMPLE
C         Q(7)  ==>  MAXIMUM OBSERVATION IN SAMPLE
C
C      THE REMAINING 143 VALUES IN THE ARRAY ARE USED FOR THE ESTIMATION
C         OF A SET OF 19 QUANTILES OF THE PARENT POPULATION BY A
C         STOCHASTIC APPROXIMATION SCHEME.  THE 6 ARRAY ENTRIES
C         FOR THE MEDIAN ESTIMATE ARE AS FOLLOWS:
C
C         Q(8)  ==>  MEDIAN ESTIMATE
C         Q(9)  ==>  MEDIAN DENSITY ESTIMATE
C         Q(10) ==>  NUMBER OF STOCHASTIC APPROXIMATION STEPS
C         Q(11) ==>  BANDWIDTH CONSTANT
C         Q(12) ==>  UPDATE CONSTANT FOR BANDWIDTH SEQUENCE
C         Q(13) ==>  H(N) VALUE FOR UPPER BOUND
C         Q(14) ==>  NOT USED
C         Q(15) ==>  NOT USED
C
C      THERE ARE 15 ARRAY ENTRIES USED FOR THE ESTIMATION OF EACH
C         PAIR OF REMAINING QUANTILES; THESE ENTRIES ARE AS FOLLOWS:
C
C         Q(15J + 1)   ==>  ALPHA-QUANTILE ESTIMATE
C         Q(15J + 2)   ==>  ALPHA-QUANTILE DENSITY ESTIMATE
C         Q(15J + 3)   ==>  (1-ALPHA)-QUANTILE ESTIMATE
C         Q(15J + 4)   ==>  (1-ALPHA)-QUANTILE DENSITY ESTIMATE
C         Q(15J + 5)   ==>  SIZE OF SECTION FOR MAX/MIN TRANSFORM
C         Q(15J + 6)   ==>  MAXIMUM OBSERVATION SO FAR IN CURRENT SECTION
C         Q(15J + 7)   ==>  NOT USED
C         Q(15J + 8)   ==>  MINIMUM OBSERVATION SO FAR IN CURRENT SECTION
C         Q(15J + 9)   ==>  NOT USED
C         Q(15J +10)   ==>  NUMBER OF STEPS FOR ALPHA-QUANTILE
```

```
C          Q(15J +11)  ==>  BANDWIDTH CONSTANT FOR ALPHA-QUANTILE
C          Q(15J +12)  ==>  UPDATE CONSTANT FOR BANDWIDTH SEQUENCE
C          Q(15J +13)  ==>  BANDWIDTH CONSTANT FOR (1-ALPHA)-QUANTILE
C          Q(15J +14)  ==>  UPDATE CONSTNAT FOR BANDWIDTH SEQUENCE
C          Q(15J +15)  ==>  H(N) VALUE FOR UPPER BOUND
C
C
C     THE INPUT ARRAY MUST BE INITIALIZED TO ZEROS BEFORE DATA CAN BE
C        ACCUMULATED.  ESTIMATED PARAMETERS MAY BE PRINTED OUT BY THE
C        SUBROUTINE QOUT ONCE 2691 OBSERVATIONS HAVE BEEN ACCUMULATED.
C
C     THIS SUBROUTINE INVOKES THE SUBROUTINE SORT3 TO PRODUCE THE
C        INITIAL QUANTILE ESTIMATES AND THE SUBROUTINE CHECK
C        WHICH PERFORMS REORDERING OF THE QUANTILE ESTIMATES IF
C        REQUIRED.
C
C     PROGRAMMER:   D.W. ROBINSON
C
C     DATE:         25 APRIL 1975
C
C
      SUBROUTINE  QUANT  (Q, X, N)
C
      DOUBLE PRECISION AVG, SUM, SUM2, SUM3, SUM4
      INTEGER  HI
C
      DIMENSION  Q(150), X(N), A(9), S(9)
C
      DATA       A(1), A(2), A(3), A(4), A(5), A(6), A(7), A(8), A(9)
     *             / 0.75, 0.4782969, 0.487675, 0.4921860, 0.5679762,
     *               0.5896012, 0.5704073, 0.5103427, 0.5105145 /
C
      DATA       S(1), S(2), S(3), S(4), S(5), S(6), S(7), S(8), S(9)
     *             / 1.0, 7.0, 2.0, 2.0, 1.0, 2.0, 2.0, 3.0, 2.0 /
C
C
C     UPDATE MOMENT ESTIMATES
      AN = N
      SIZE = Q(1) + AN
C --- MEAN
      SUM = 0.0D0
      DO 10 I=1,N
      SUM = SUM + X(I)
   10 CONTINUE
      TOTAL = SUM + Q(2)
      AVG   = (SUM + Q(2)) / SIZE
C --- CORRECT CENTRAL MOMENT ESTIMATES FOR NEW MEAN
      IF ( Q(1) .LT. 1.0) GO TO 15
```

```
C     = AVG - Q(2) / Q(1)
C2    = C * C * Q(1)
Q(5) = Q(5) - C * (4. * Q(4) - C * (6. * Q(3) - 3. * C2))
Q(4) = Q(4) - C * (3. * Q(3) - C2 - C2)
Q(3) = Q(3) - C2
GO TO 20
C --- INITIALIZE MAX/MIN VALUES
   15 Q(6) = X(1)
      Q(7) = X(1)
C --- ADD TO CENTRAL MOMENT ESTIMATES
   20 Q(2) = TOTAL
      SUM2 = 0.D0
      SUM3 = 0.D0
      SUM4 = 0.D0
      XBAR = AVG
      DO 30 I=1,N
      DIFF = X(I) - XBAR
      DIF2 = DIFF * DIFF
      SUM2 = SUM2 + DIF2
      SUM3 = SUM3 + DIF2 * DIFF
      SUM4 = SUM4 + DIF2 * DIF2
      IF ( Q(6) .GT. X(I) ) Q(6) = X(I)
      IF ( Q(7) .LT. X(I) ) Q(7) = X(I)
   30 CONTINUE
      Q(3) = Q(3) + SUM2
      Q(4) = Q(4) + SUM3
      Q(5) = Q(5) + SUM4
C
C     OBTAIN NEW QUANTILE ESTIMATES
      IP = 1
      IF ( Q(13) .GT. 0.0 ) GO TO 60
C --- ACCUMULATE INITIAL THREE OBSERVATIONS TO SERVE AS INITIAL MEDIAN
C ---     ESTIMATE
      NO = Q(1) + 1.0
      Q(1) = SIZE
   40 Q(NO+7) = X(IP)
      NO = NO + 1
      IP = IP + 1
      IF ( NO .GT. 3 ) GO TO 50
      IF ( IP .GT. N ) RETURN
      GO TO 40
C --- INITIALIZE Q ARRAY ENTRIES FOR MEDIAN
   50 CALL SORT3 ( Q(8) )
      Q(11) = Q(10) - Q(8)
      IF ( Q(11) .LE. 0.0 ) Q(11) = 1.0
      Q(8)  = Q(9)
      Q(9)  = 1. / Q(11)
      Q(10) = 1.0
```

```
      Q(12) = 0.3333333 / Q(11) ** 3
      Q(13) = 1.0
      IF ( IP .GT. N ) RETURN
C
C
C --- QUANTILE ESTIMATION LOOP BEGINS HERE
   60 Q(1) = SIZE
      DO 130 I=IP,N
C
C --- UPDATE MEDIAN ESTIMATE
      T     = ABS ( Q(8) - X(I) )
      IF ( T .LT. Q(11) ) Q(9) = Q(9) + (Q(11) - T) / Q(11) ** 2
      Y     = 0.5
      IF ( X(I) .LT. Q(8) ) Y = -0.5
      D     = Q(9)
      UP    = Q(10) * Q(13)
      IF ( D .GT. UP ) D = UP
      Q(8)  = Q(8) + Y / D
      Q(13) = Q(13) + 1.0 / Q(10)
      Q(10) = Q(10) + 1.0
      Q(11) = (1.0 - Q(12) * Q(11) ** 3) * Q(11)
C
C --- PASS LATEST X VALUES OUTWARDS FOR OTHER QUANTILES
      Q(21) = X(I)
      Q(23) = X(I)
      J     = 15
      K     =  1
      GO TO 100
C
   70 IF ( Q(J+5) .GT. 0.0 ) GO TO 80
      Q(J+6) = Q(J-9)
      Q(J+8) = Q(J-7)
      GO TO 90
   80 IF ( Q(J+6) .LT. Q(J-9) ) Q(J+6) = Q(J-9)
      IF ( Q(J+8) .GT. Q(J-7) ) Q(J+8) = Q(J-7)
   90 Q(J+5) = Q(J+5) + 1.0
      IF ( Q(J+5) .NE. S(K) ) GO TO 130
      Q(J+5) = 0.0
  100 IF ( Q(J+15) .GT. 0.0 ) GO TO 110
C
C --- INITIALIZE Q ARRAY ENTRY FOR K-TH QUANTILE
C --- FIRST SAVE THE FIRST THREE OBSERVATIONS
      Q(J+10) = Q(J+10) + 1.0
      NJ = Q(J+10)
      Q(J+NJ) = Q(J+6)
      Q(J+NJ+10) = Q(J+8)
      IF ( NJ .LT. 3 ) GO TO 120
C --- NOW FIND STARTING VALUES FOR STOCHASTIC APPROXIMATION
```

```
      CALL SORT3 ( Q(J+1) )
      BJ1      = Q(J+3) - Q(J+1)
      IF ( BJ1 .LE. 0.0 ) BJ1 = 1.0
      Q(J+1)   = Q(J+2)
      Q(J+2)   = 1.0 / BJ1
C
      CALL SORT3 ( Q(J+11) )
      Q(J+3)   = Q(J+12)
      BJ2      = Q(J+13) - Q(J+11)
      IF ( BJ2  .LE. 0.0 ) BJ2  = 1.0
      Q(J+4)   = 1.0 / BJ2
C
      Q(J+10) = 1.0
      Q(J+11) = BJ1
      Q(J+12) = 0.3333333 / BJ1 ** 3
      Q(J+13) = BJ2
      Q(J+14) = 0.3333333 / BJ2 ** 3
      Q(J+15) = 1.0
      GO TO 120
C
C --- UPDATE ESTIMATES FOR THE K-TH QUANTILE PAIR
  110 T      = ABS ( Q(J+1) - Q(J+6) )
      IF (T .LT. Q(J+11)) Q(J+2) = Q(J+2) + (Q(J+11) - T) / Q(J+11) ** 2
      Y      = A(K)
      IF ( Q(J+6) .LE. Q(J+1) ) Y = Y - 1.0
      D      = Q(J+2)
      UP     = Q(J+10) * Q(J+15)
      IF ( D .GT. UP ) D = UP
      Q(J+1) = Q(J+1) + Y / D
C
      T      = ABS ( Q(J+3) - Q(J+8) )
      IF (T .LT. Q(J+13)) Q(J+4) = Q(J+4) + (Q(J+13) - T) / Q(J+13) ** 2
      Y      = - A(K)
      IF ( Q(J+8) .GT. Q(J+3) ) Y = Y + 1.0
      D      = Q(J+4)
      IF ( D .GT. UP ) D = UP
      Q(J+3) = Q(J+3) + Y / D
C
      Q(J+15) = Q(J+15) + 1.0 / Q(J+10)
      Q(J+10) = Q(J+10) + 1.0
      Q(J+11) = (1.0 - Q(J+12) * Q(J+11) ** 3) * Q(J+11)
      Q(J+13) = (1.0 - Q(J+14) * Q(J+13) ** 3) * Q(J+13)
C
C --- PASS VALUES OUT TO OTHER QUANTILES
  120 J      = J + 15
      K      = K + 1
      IF ( K .LT. 10 ) GO TO 70
C --- REACHED FURTHEST QUANTILE.  CHECK ORDER OF ESTIMATES.
```

```
      IF ( Q(150) .LE. 1.0 ) GO TO 130
      CALL CHECK ( Q )
C
C --- END OF QUANTILE LOOP
  130 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CHECK (Q)
C
C     THIS SUBROUTINE ADJUSTS THE 19 QUANTILE ESTIMATES IN THE Q-VECTOR
C        BY ISOTONIC REGRESSION TECHNIQUES SO THAT THEY ARE IN
C        INCREASING ORDER.
C
C     KRUSKAL'S UP AND DOWN BLOCKS ALGORITHM IS USED.
C
C     FOR A DETAILED DESCRIPTION OF THE DATA STRUCTURE OF THE Q-VECTOR,
C        SEE THE INTRODUCTION TO SUBROUTINE QUANT.
C
C     PROGRAMMER:   D. W. ROBINSON
C
C     DATE:         27 APRIL 1975
C
C
      DOUBLE PRECISION  VSUM, WSUM
      INTEGER   TOP, ACT, HIGH
      LOGICAL   ORD
C
      DIMENSION  Q(150), APRIME(19), J(19), NUM(19)
      DIMENSION  VAL(19), AVG(19), W(19), TOP(19)
C
      DATA  APRIME(1),  APRIME(2),  APRIME(3),  APRIME(4),  APRIME(5),
     *      APRIME(6),  APRIME(7),  APRIME(8),  APRIME(9),  APRIME(10),
     *      APRIME(11), APRIME(12), APRIME(13), APRIME(14), APRIME(15),
     *      APRIME(16), APRIME(17), APRIME(18), APRIME(19)
     *           / 4.001770, 4.001712, 4.080919, 4.079041,
     *             4.075324, 4.000977, 4.002432, 4.007551,
     *             5.333333, 4.000000, 5.333333, 4.007551,
     *             4.002432, 4.000977, 4.075324, 4.079041,
     *             4.080919, 4.001712, 4.001770  /
C
      DATA  J(1),  J(2),  J(3),  J(4),  J(5),  J(6),  J(7),  J(8),
     *      J(9),  J(10), J(11), J(12), J(13), J(14), J(15), J(16),
     *      J(17), J(18), J(19)
     *        / 138, 123, 108, 93, 78, 63, 48, 33, 18, 8, 16, 31, 46,
     *          61, 76, 91, 106, 121, 136 /
C
      DATA  NUM(1),  NUM(2),  NUM(3),  NUM(4),  NUM(5),  NUM(6),
     *      NUM(7),  NUM(8),  NUM(9),  NUM(10), NUM(11), NUM(12),
     *      NUM(13), NUM(14), NUM(15), NUM(16), NUM(17), NUM(18),
     *      NUM(19)
     *          / 145, 130, 115, 100, 85, 70, 55, 40, 25, 10, 25, 40,
     *            55, 70, 85, 100, 115, 130, 145 /
C
C --- FIRST ADJUST ESTIMATES SO THEY LIE WITHIN THE SAMPLE EXTREMA
C ---    AND CHECK TO SEE IF THEY ARE ALREADY IN ORDER
```

```
      ORD = .TRUE.
      IF ( Q(138) .LT. Q(6) ) Q(138) = Q(6)
      IF ( Q(138) .GT. Q(7) ) Q(138) = Q(7)
      VAL(1) = Q(138)
      DO 10 I=2,19
      JI = J(I)
      IF ( Q(JI) .LT. Q(6) ) Q(JI) = Q(6)
      IF ( Q(JI) .GT. Q(7) ) Q(JI) = Q(7)
      VAL(I) = Q(JI)
      IF ( Q(JI) .LT. Q(J(I-1)) ) ORD = .FALSE.
   10 CONTINUE
C
C --- QUIT IF ESTIMATES IN ORDER
      IF (ORD) RETURN
C
C --- FIRST WE INITIALIZE 19 ESTIMATION BLOCKS, ONE FOR EACH QUANTILE
C ---    TOP(I) POINTS TO THE HIGHEST QUANTILE INCLUDED IN BLOCK I
C ---    AVG(I) IS THE AVERAGE VALUE OF THE QUANTILES IN BLOCK I
C ---    W(I) IS THE WEIGHT ATTACHED TO THE I-TH QUANTILE ESTIMATE IN
C ---          COMPUTING THE BLOCK AVERAGE
      DO 20 I=1,19
      AVG(I) = VAL(I)
      W(I) = APRIME(I) * Q(J(I)+1) ** 2 / (Q(NUM(I)) - 1.0)
      TOP(I) = I
   20 CONTINUE
C
C --- NOW WE INITIALIZE .  SET THE NUMBER OF BLOCKS TO 19 AND SET THE
C ---    FIRST BLOCK ACTIVE.
      HIGH = 19
      ACT = 1
C
C
C --- SEE IF ACTIVE BLOCK  IS IN ORDER UPWARDS
   30 IF ( TOP(ACT) .EQ. 19)          GO TO 40
      IF ( AVG(ACT) .GT. AVG(ACT+1)) GO TO 110
C --- SEE IF ACTIVE BLOCK IS IN ORDER DOWNWARDS
   40 IF ( ACT  .EQ.  1)              GO TO 150
      IF ( AVG(ACT) .GT. AVG(ACT-1)) GO TO 150
C
C --- ACTIVE BLOCK NOT IN ORDER DOWNWARDS.  POOL THE BLOCKS.
   50 TOP(ACT-1) = TOP(ACT)
      HIGH = HIGH - 1
      IF ( ACT .GT. HIGH) GO TO 70
C --- SHIFT THE HIGHER BLOCKS DOWN
      DO 60 I=ACT,HIGH
      TOP(I) = TOP(I+1)
      AVG(I) = AVG(I+1)
   60 CONTINUE
```

```
   70 ACT = ACT - 1
C --- FIND NEW AVERAGE FOR POOLED BLOCK
      WSUM = 0.
      VSUM = 0.
      IHI = TOP(ACT)
      ILO = 1
      IF ( ACT .NE. 1) ILO = TOP(ACT-1) + 1
      DO 80 I=ILO,IHI
      WSUM = WSUM + W(I)
      VSUM = VSUM + W(I) * VAL(I)
   80 CONTINUE
      AVG(ACT) = VSUM / WSUM
      GO TO 30
C
C --- SEE IF ACTIVE BLOCK IS IN ORDER DOWNWARDS
   90 IF ( ACT  .EQ.  1)              GO TO 100
      IF ( AVG(ACT) .LT. AVG(ACT-1))  GO TO 50
C --- SEE IF ACTIVE BLOCK IS IN ORDER UPWARDS
  100 IF ( TOP(ACT) .EQ. 19)          GO TO 150
      IF ( AVG(ACT) .LT. AVG(ACT+1))  GO TO 150
C
C --- ACTIVE BLOCK NOT IN ORDER WITH HIGHER BLOCK. POOL THE TWO BLOCKS.
  110 TOP(ACT) = TOP(ACT+1)
      HIGH = HIGH - 1
      ILO = ACT + 1
      IF ( ILO .GT. HIGH ) GO TO 130
C --- SHIFT HIGHER BLOCKS DOWN
      DO 120 I=ILO,HIGH
      TOP(I) = TOP(I+1)
      AVG(I) = AVG(I+1)
  120 CONTINUE
C --- FIND NEW AVERAGE FOR POOLED BLOCK.
  130 WSUM = 0.
      VSUM = 0.
      IHI = TOP(ACT)
      ILO = 1
      IF ( ACT .NE. 1) ILO = TOP(ACT-1) + 1
      DO 140 I=ILO,IHI
      WSUM = WSUM + W(I)
      VSUM = VSUM + W(I) * VAL(I)
  140 CONTINUE
      AVG(ACT) = VSUM / WSUM
      GO TO 90
C
C --- ACTIVE BLOCK IS IN ORDER WITH HIGHER & LOWER BLOCKS.
C ---     MAKE NEXT HIGHER BLOCK ACTIVE
  150 IF ( TOP(ACT) .EQ.  19 ) GO TO 160
      ACT = ACT + 1
```

```
      GO TO 100
C
C --- FINISHED. MOVE UPDATED ESTIMATES INTO Q-ARRAY.
  160 N = 1
      DO 170 I=1,19
      Q(J(I)) = AVG(N)
      IF ( I .EQ. TOP(N)) N = N + 1
  170 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE  QOUT  (Q)
C
C     THIS SUBROUTINE CONVERTS THE DATA ACCUMULATED IN THE INPUT
C        ARRAY Q BY SUBROUTINE QUANT INTO ESTIMATES OF POPULATION
C        MOMENTS AND QUANTILES AND PRINTS OUT THE ESTIMATES.
C
C     AT LEAST 2691 OBSERVATIONS MUST BE ACCUMULATED BEFORE RESULTS
C        CAN BE PRINTED OUT.
C
C     FOR A DETAILED DESCRIPTION OF THE DATA STRUCTURE OF THE
C        Q ARRAY, SEE THE INTRODUCTION TO SUBROUTINE QUANT.
C
C     PROGRAMMER:  D.W. ROBINSON
C
C     DATE:        17 JANUARY 1975
C
C
      DIMENSION Q(130), ALPHA(19), APRIME(19), FAC(19), J(19), NUM(19)
C
      DATA  ALPHA(1),  ALPHA(2),  ALPHA(3),  ALPHA(4),  ALPHA(5),
     *      ALPHA(6),  ALPHA(7),  ALPHA(8),  ALPHA(9),  ALPHA(10),
     *      ALPHA(11), ALPHA(12), ALPHA(13), ALPHA(14), ALPHA(15),
     *      ALPHA(16), ALPHA(17), ALPHA(18), ALPHA(19)
     *           / .001, .002, .005, .010, .020, .025, .050, .100,
     *             .250, .500, .750, .900, .950, .975, .980, .990,
     *             .995, .998, .999 /
C
      DATA  APRIME(1),  APRIME(2),  APRIME(3),  APRIME(4),  APRIME(5),
     *      APRIME(6),  APRIME(7),  APRIME(8),  APRIME(9),  APRIME(10),
     *      APRIME(11), APRIME(12), APRIME(13), APRIME(14), APRIME(15),
     *      APRIME(16), APRIME(17), APRIME(18), APRIME(19)
     *           / 0.4998894, 0.4998930, 0.4950180, 0.4951320,
     *             0.4953577, 0.4999389, 0.4998481, 0.4995288,
     *             0.4330127, 0.5000000, 0.4330127, 0.4995288,
     *             0.4998481, 0.4999389, 0.4953577, 0.4951320,
     *             0.4950180, 0.4998930, 0.4998894  /
C
      DATA  FAC(1),  FAC(2),  FAC(3),  FAC(4),  FAC(5),  FAC(6),
     *      FAC(7),  FAC(8),  FAC(9),  FAC(10), FAC(11), FAC(12),
     *      FAC(13), FAC(14), FAC(15), FAC(16), FAC(17), FAC(18),
     *      FAC(19)
     *         / 2.911978E-03, 5.820084E-03, 1.557471E-02, 3.103675E-02
     *,          6.162230E-02, 7.074851E-02, 1.391442E-01, 2.688109E-01
     *,          1.0, 1.0,1.0, 2.688109E-01, 1.391442E-01, 7.074851E-02
     *,          6.162230E-02, 3.103675E-02, 1.557471E-02, 5.820084E-03
     *,          2.911978E-03 /
C
      DATA  J(1),  J(2),  J(3),  J(4),  J(5),  J(6),  J(7),  J(8),
```

```
     *      J(9),  J(10), J(11), J(12), J(13), J(14), J(15), J(16),
     *      J(17), J(18), J(19)
     *         / 138, 123, 108, 93, 78, 63, 48, 33, 18, 8, 16, 31, 46,
     *           61, 76, 91, 106, 121, 136 /
C
      DATA  NUM(1),  NUM(2),  NUM(3),  NUM(4),  NUM(5),  NUM(6),
     *      NUM(7),  NUM(8),  NUM(9),  NUM(10), NUM(11), NUM(12),
     *      NUM(13), NUM(14), NUM(15), NUM(16), NUM(17), NUM(18),
     *      NUM(19)
     *           / 145, 130, 115, 100, 85, 70, 55, 40, 25, 10, 25, 40,
     *             55, 70, 85, 100, 115, 130, 145 /
C
      DATA NOUT /6/
C
C     TEST FOR SUFFICIENT OBSERVATIONS
      IF ( Q(1) .LT. 2691. ) RETURN
C
C     MOMENT ESTIMATES
      N     = Q(1)
      XMEAN = Q(2) / Q(1)
      XVAR  = Q(3) / (Q(1) - 1.0)
      XDEV  = SQRT(XVAR)
      XM3   = Q(4) * Q(1) / ((Q(1) - 1.0) * (Q(1) - 2.0))
      CSKEW = XM3 / (XVAR * XDEV)
      XM4   = (((Q(1) - 2.) * Q(1) + 3.) * Q(5)
     1          -   (6. * Q(1) - 9.) * Q(3) ** 2 / Q(1))
     2             / (Q(1) * (Q(1) - 1.0) * (Q(1) - 2.0))
      CKURT = XM4 / (XVAR * XVAR) - 3.0
      XINT  = 1.960 * SQRT ( XVAR / Q(1))
      XLO   = XMEAN - XINT
      XHI   = XMEAN + XINT
C
      WRITE(NOUT,10) N, XMEAN, XLO, XHI, XVAR, XDEV, XM3, CSKEW,
     *               XM4, CKURT, Q(6), Q(7)
   10 FORMAT('1',28X,'DATA SUMMARY FOR A SAMPLE OF ',I7,' OBSERVATIONS'/
     2      /'0',25X,'MEAN',1PE14.5,4X,'95% CONFIDENCE INTERVAL',E14.5,
     *             '   TO',E14.5/
     3       '0',21X,'VARIANCE',E14.5,9X,'STANDARD DEVIATION',E14.5/
     4       '0',11X,'3RD CENTRAL MOMENT',E14.5,4X,'COEFFICIENT OF SKEWN
     *ESS'             ,E14.5/
     5       '0',11X,'4TH CENTRAL MOMENT',E14.5,4X,'COEFFICIENT OF KURTO
     *SIS'             ,E14.5/
     6       '0',22X,'MINIMUM',E14.5,20X,'MAXIMUM',E14.5///)
C
C     QUANTILE ESTIMATES
      WRITE(NOUT,20)
   20 FORMAT('0 ALPHA',6X,'QUANTILE',6X,'EST STD DEV',5X,'95 PER CENT CO
     *NFIDENCE INTERVAL',6X,'DENSITY',6X,'XFORMED DENSITY',3X,'SAMPLE SI
```

```
     *ZE'/)
C
      DO 40 I=1,19
      QUANT = Q(J(I))
      DEV = APRIME(I) * SQRT ( Q(NUM(I)) - 1.0 ) / Q(J(I)+1)
      DENST = Q(J(I)+1) / (Q(NUM(I)) - 1.0)
      DENS  = DENST * FAC(I)
      QLO = QUANT - 1.959963 * DEV
      QHI = QUANT + 1.959963 * DEV
      N = Q(NUM(I)) - 1.0
      WRITE(NOUT,30) ALPHA(I), QUANT, DEV, QLO, QHI, DENS, DENST, N
   30 FORMAT('0',F6.3, 1P3E16.6, '  TO',E15.5,2E16.6, 5XI8)
   40 CONTINUE
      RETURN
      END
```

```
C       SUBROUTINE  POWER  ( P, Q, T, M )
C
C       THIS SUBROUTINE ACCUMULATES ESTIMATES OF THE POWER OF A
C          STATISTICAL TEST UNDER EACH OF M ALTERNATE HYPOTHESES AT
C          LEVELS CORRESPONDING TO ESTIMATED QUANTILES OF THE DISTRIBUTION
C          OF THE TEST UNDER A NULL HYPOTHESIS.
C
C       THE QUANTILE ESTIMATES ARE ACCUMULATED IN THE INPUT ARRAY Q BY
C          SUBROUTINE QUANT; SEE ITS LISTING FOR DETAILS.  AT LEAST
C          2019 OBSERVATIONS MUST BE ACCUMULATED IN THE Q ARRAY BEFORE
C          POWER ESTIMATES CAN BE OBTAINED.
C
C       THE INPUT ARRAY P MUST BE INITIALIZED TO ZEROES BEFORE POWER IS
C          CALLED THE FIRST TIME.  THERE ARE 20 ARRAY ELEMENTS IN THE P
C          ARRAY FOR EACH ALTERNATE HYPOTHESIS TO BE TESTED, I.E. 20*M
C          ELEMENTS IN ALL.  FOR HYPOTHESIS J WE HAVE
C
C            P( 20(J-1) + 1 )    ==>  NUMBER OBSERVATIONS FOR HYPOTHESIS J
C            P( 20(J-1) + I+1 )  ==>  PERCENTILE FOR I-TH QUANTILE ESTIMATE
C
C       THE INPUT ARRAY T CONTAINS THE LATEST OBSERVATIONS ON THE TEST
C          STATISTIC; T(J) IS AN OBSERVATION UNDER THE J-TH ALTERNATE
C          HYPOTHESIS.
C
C       ESTIMATES IN THE P ARRAY MAY BE PRINTED OUT BY SUBROUTINE PWROUT.
C
C
C       PROGRAMMER:  D.W. ROBINSON
C
C       DATE:          9 MAY 1975
C
C
        SUBROUTINE POWER ( P, Q, T, M )
C
        DIMENSION P(20), Q(150), T(M), J(20)
C
        DATA  J(1),  J(2),  J(3),  J(4),  J(5),  J(6);  J(7),  J(8),
     *        J(9),  J(10), J(11), J(12), J(13), J(14), J(15), J(16),
     *        J(17), J(18), J(19), J(20)
     *          / 0, 138, 123, 108, 93, 78, 63, 48, 33, 18, 8, 16, 31,
     *            46, 61, 76, 91, 106, 121, 136  /
C
C --- SEE IF Q ARRAY IS SET UP YET
        IF ( Q(1) .LT. 2019.0 ) RETURN
C
C --- UPDATE FREQUENCIES
        LIM = 20 * M
        DO 10 I=1,LIM,20
```

```
      P(I) = P(I) + 1.0
   10 CONTINUE
C
C --- UPDATE POWER ESTIMATES
      DO 30 I=2,20
      QUANT = Q(J(I))
      INX = I
      DO 20 K=1,M
      IF ( T(K) .LE. QUANT ) P(INX) = P(INX) + 1.0
      INX = INX + 20
   20 CONTINUE
   30 CONTINUE
C
      RETURN
      END
```

```
      SUBROUTINE PWROUT ( Q, P, M )
C
C     THIS SUBROUTINE PRINTS OUT THE PERCENTILE/POWER ESTIMATES
C        ACCUMULATED IN THE P ARRAY BY SUBROUTINE POWER.
C
C     FOR MORE INFORMATION SEE THE LISTING FOR SUBROUTINE POWER.
C
C
      INTEGER HI, OUT
      DIMENSION Q(150), P(20), J(19), ALPHA(19), PWR(10)
C
      DATA  ALPHA(1),  ALPHA(2),  ALPHA(3),  ALPHA(4),  ALPHA(5),
     *      ALPHA(6),  ALPHA(7),  ALPHA(8),  ALPHA(9),  ALPHA(10),
     *      ALPHA(11), ALPHA(12), ALPHA(13), ALPHA(14), ALPHA(15),
     *      ALPHA(16), ALPHA(17), ALPHA(18), ALPHA(19)
     *        / 0.001, 0.002, 0.005, 0.010, 0.020, 0.025, 0.050, 0.100,
     *          0.250, 0.500, 0.750, 0.900, 0.950, 0.975, 0.980, 0.990,
     *          0.995, 0.998, 0.999 /
C
      DATA  J(1),  J(2),  J(3),  J(4),  J(5),  J(6),  J(7),  J(8),
     *      J(9),  J(10), J(11), J(12), J(13), J(14), J(15), J(16),
     *      J(17), J(18), J(19)
     *        / 138, 123, 108, 93, 78, 63, 48, 33, 18, 8, 16, 31,
     *          46, 61, 76, 91, 106, 121, 136  /
C
      DATA  OUT /6/
C
C --- SEE IF Q ARRAY IS SET UP
      IF ( Q(1) .LT. 2019.0 ) RETURN
C
C --- NUMBER BLOCKS TO PRINT
      NL = (M - 1) / 10 + 1
      LO = 1
      HI = 10
      INX1 = 1
      DO 60 K = 1,NL
C
C --- PRINT HEADING
      IF ( MOD(K,2) .EQ. 1 ) WRITE(OUT,10)
   10 FORMAT('1')
      IF ( HI .GT. M ) HI = M
      WRITE(OUT,20) (I,I=LO,HI)
   20 FORMAT(//'0 ALPHA',6X,'QUANTILE',7X,10('H',I2,6X))
C
C --- COMPUTE AND PRINT ESTIMATES
      LIM = HI - LO + 1
      DO 50 I=1,19
      INX = INX1
```

```
      DO 30 L=1,LIM                                                     PO 00490
      PWR(L) = P(INX+I) / P(INX)                                        PO 00500
      INX = INX + 20                                                    PO 00510
   30 CONTINUE                                                          PO 00520
      WRITE(OUT,40) ALPHA(I), Q(J(I)), (PWR(L),L=1,LIM)                 PO 00530
   40 FORMAT('0',F6.3,1PE16.6,0P10F9.4)                                 PO 00540
   50 CONTINUE                                                          PO 00550
      LO = LO + 10                                                      PO 00560
      HI = HI + 10                                                      PO 00570
      INX1 = INX1 + 200                                                 PO 00575
   60 CONTINUE                                                          PO 00580
C                                                                       PO 00590
      RETURN                                                            PO 00600
      END                                                               PO 00610
```

```
      SUBROUTINE  TRUE (XMOM, FINV, F, Q, D, N)
C
C     THIS SUBROUTINE COMPUTES AND PRINTS THE TRUE VALUES WHICH
C        ARE TO BE ESTIMATED BY SUBROUTINE QUANT.  THE OUTPUT FORMAT OF
C        SUBROUTINE QOUT IS USED.
C
C     THE SUBROUTINE INPUTS ARE:
C
C        XMOM    A REAL*4 VECTOR OF THE MEAN (XMOM(1) ) AND THE
C                   FIRST THREE CENTRAL MOMENTS OF THE PARENT
C                   POPULATION.
C
C        FINV    THE INVERSE DISTRIBUTION FUNCTION OF THE POPULATION.
C
C        F       THE DENSITY FUNCTION.
C
C        Q       AN OUTPUT ARRAY (REAL*4) INTO WHICH THE TRUE
C                   QUANTILE VALUES ARE PLACED
C
C        D       AN OUTPUT ARRAY (REAL*4) INTO WHICH THE QUANTILE
C                   ESTIMATE STANDARD DEVIATIONS ARE PLACED.
C
C     PROGRAMMER: D.W. ROBINSON
C
C     DATE:       10 DECEMBER 1974
C
      DIMENSION  FAC(19), APRIME(19), ALPHA(19), XMOM(4), Q(19), D(19)
      DIMENSION  SIZE(19)
C
      DATA  APRIME / .4894855, .4896573, .4295927, .4303988, .4320238,
     *               .5078140, .5123250, .5217031, .2500000, .5000000,
     *               .7500000, .4782969, .4876750, .4921860, .5679762,
     *               .5696012, .5704073, .5103427, .5105145 /
C
      DATA  ALPHA / .001, .002, .005, .010, .020, .025, .050, .100,
     *              .250, .500, .750, .900, .950, .975, .980, .990,
     *              .995, .998, .999 /
C
      DATA  FAC / 2.911978E-03, 5.820084E-03, 1.557471E-02, 3.103675E-02
     *,           6.162230E-02, 7.074851E-02, 1.391442E-01, 2.688109E-01
     *,           1.0, 1.0,1.0, 2.688109E-01, 1.391442E-01, 7.074851E-02
     *,           6.162230E-02, 3.103675E-02, 1.557471E-02, 5.820084E-03
     *,           2.911978E-03 /
C
      DATA  SIZE / 1., 2., 6., 12., 2*24., 48., 96., 3*672., 96., 48.,
     *             2*24., 12., 6., 2., 1. /
C
```

```
      DATA   NOUT /6/
C
C     COMPUTE VALUES RELATED TO MOMENTS
      XDEV = SQRT(XMOM(2))
      CSKEW = XMOM(3) / (XMOM(2) * XDEV)
      CKURT = XMOM(4) / (XMOM(2) * XMOM(2)) - 3.0
C
      WRITE(NOUT,10) XMOM(1), XMOM(2), XDEV, XMOM(3), CSKEW, XMOM(4),
     *                CKURT
   10 FORMAT('1',28X,'ACTUAL POPULATION PARAMETER VALUES TO BE ESTIMATED
     *'/
     2          '0',25X,'MEAN',1PE14.5/
     3          '0',21X,'VARIANCE',E14.5,9X,'STANDARD DEVIATION',E14.5/
     4          '0',11X,'3RD CENTRAL MOMENT',E14.5,4X,'COEFFICIENT OF SKEWN
     *ESS'                ,E14.5/
     5          '0',11X,'4TH CENTRAL MOMENT',E14.5,4X,'COEFFICIENT OF KURTO
     *SIS'                ,E14.5////)
C
C     FIND QUANTILE VALUES
      WRITE(NOUT,20)
   20 FORMAT('0 ALPHA',6X,'QUANTILE',6X,'EST STD DEV',5X,'95 PER CENT CO
     *NFIDENCE INTERVAL',6X,'DENSITY',6X,'XFORMED DENSITY'/)
C
      DO 40 I=1,19
      QUANT = FINV (ALPHA(I) )
      DENS  = F (QUANT)
      DENST = DENS / FAC(I)
      QDEV  = SQRT ( APRIME(I) * (1. - APRIME(I)) / (SIZE(I) * N)  )
     *                      / DENST
      QHI   = QUANT + 1.960 * QDEV
      QLO   = QUANT - 1.960 * QDEV
      WRITE(NOUT,30) ALPHA(I), QUANT, QDEV, QLO, QHI, DENS, DENST
   30 FORMAT('0',F6.3, 1P3E16.6, '  TO',E15.6,2E16.6)
      Q(I) = QUANT
      D(I) = QDEV
   40 CONTINUE
      RETURN
      END
```

DATA SUMMARY FOR A SAMPLE OF 67210 OBSERVATIONS

MEAN	1.00153E 00	95% CONFIDENCE INTERVAL	9.93959E-01	TO	1.00910E 00
VARIANCE	1.00268E 00	STANDARD DEVIATION	1.00134E 00		
3RD CENTRAL MOMENT	1.99165E 00	COEFFICIENT OF SKEWNESS	1.98368E 00		
4TH CENTRAL MOMENT	8.84217E 00	COEFFICIENT OF KURTOSIS	5.79499E 00		
MINIMUM	7.80821E-06	MAXIMUM	1.02374E 01		

ALPHA	QUANTILE	EST STD DEV	95 PER CENT CONFIDENCE INTERVAL			DENSITY	XFORMED DENSITY	SAMPLE SIZE
0.001	4.744714E-04	2.437855E-04	-3.339024E-06	TO	9.522818E-04	6.062730E-01	2.081997E 02	97
0.002	3.142301E-03	2.674477E-04	2.618114E-03	TO	3.666488E-03	7.750574E-01	1.331695E 02	197
0.005	6.374318E-03	3.549338E-04	5.678661F-03	TO	7.069975E-03	8.890089E-01	5.708029E 01	597
0.010	1.038808E-02	5.263209E-04	9.356514E-03	TO	1.141965E-02	8.439158E-01	2.719086E 01	1197
0.020	2.080646E-02	6.801810E-04	1.947333E-02	TO	2.213959E-02	9.166383E-01	1.487511E 01	2397
0.025	2.610381E-02	7.949311E-04	2.454577E-02	TO	2.766184E-02	9.088042E-01	1.284556E 01	2397
0.050	5.270571E-02	1.093987E-03	5.056154E-02	TO	5.484989E-02	9.179223E-01	6.596915E 00	4797
0.100	1.058710E-01	1.538805E-03	1.028550E-01	TO	1.088870E-01	8.907039E-01	3.313497E 00	9598
0.250	2.864336E-01	2.225945E-03	2.820708E-01	TO	2.907963E-01	7.503926E-01	7.503926E-01	67204
0.500	6.900608E-01	3.851132E-03	6.825127E-01	TO	6.976088E-01	5.008115E-01	5.008115E-01	67207
0.750	1.374121F 00	6.676197E-03	1.361035E 00	TO	1.387205E 00	2.501921E-01	2.501921E-01	67204
0.900	2.302994E 00	1.368417E-02	2.276173E 00	TO	2.329814E 00	1.001609E-01	3.726075E-01	9598
0.950	3.030518E 00	2.052224E-02	2.990294E 00	TO	3.070740E 00	4.893203E-02	3.516642E-01	4797
0.975	3.719206E 00	3.113391E-02	3.658184E 00	TO	3.780227E 00	2.320418E-02	3.279812E-01	2397
0.980	3.940490E 00	3.028188E-02	3.881138E 00	TO	3.999841E 00	2.058921E-02	3.341195E-01	2397
0.990	4.642439E 00	4.635442E-02	4.551585E 00	TO	4.733292E 00	9.582054E-03	3.087326E-01	1197
0.995	5.307442E 00	6.414485E-02	5.181720E 00	TO	5.433163E 00	4.919164E-03	3.158432E-01	597
0.998	6.224498E 00	1.078273E-01	6.013160E 00	TO	6.435835E 00	1.922401E-03	3.303049E-01	197
0.999	6.827236E 00	1.439586E-01	6.545082E 00	TO	7.109389E 00	1.026688E-03	3.525741E-01	97

ACTUAL POPULATION PARAMETER VALUES TO BE ESTIMATED

MEAN	1.00000E 00		
VARIANCE	1.00000E 00	STANDARD DEVIATION	1.00000E 00
3RD CENTRAL MOMENT	2.00000E 00	COEFFICIENT OF SKEWNESS	2.00000E 00
4TH CENTRAL MOMENT	9.00000E 00	COEFFICIENT OF KURTOSIS	6.00000E 00

ALPHA	QUANTILE	EST STD DEV	95 PER CENT CONFIDENCE INTERVAL			DENSITY	XFORMED DENSITY
0.001	1.000487E-03	1.479486E-04	7.105086E-04	TO	1.290466E-03	9.990000E-01	3.430657E 02
0.002	2.001977E-03	2.093026E-04	1.591744E-03	TO	2.412210E-03	9.980000E-01	1.714753E 02
0.005	5.012535E-03	3.211857E-04	4.383009E-03	TO	5.642056E-03	9.950000E-01	6.388562E 01
0.010	1.005033E-02	4.549727E-04	9.158578E-03	TO	1.094207E-02	9.900000E-01	3.189766E 01
0.020	2.020269E-02	6.455637E-04	1.893738E-02	TO	2.146799E-02	9.800000E-01	1.590333E 01
0.025	2.531778E-02	7.518618E-04	2.384413E-02	TO	2.679143E-02	9.750000E-01	1.378121E 01
0.050	5.129324E-02	1.072934E-03	4.919029E-02	TO	5.339619E-02	9.500000E-01	6.827450E 00
0.100	1.053605E-01	1.546120E-03	1.023301E-01	TO	1.083909E-01	9.000000E-01	3.348079E 00
0.250	2.876821E-01	2.261356E-03	2.832499E-01	TO	2.921143E-01	7.499999E-01	7.499999E-01
0.500	6.931472E-01	3.916781E-03	6.854703E-01	TO	7.008241E-01	5.000000E-01	5.000000E-01
0.750	1.386294E 00	6.784067E-03	1.372997E 00	TO	1.399590E 00	2.500000E-01	2.500000E-01
0.900	2.302585E 00	1.391508E-02	2.275311E 00	TO	2.329858E 00	1.000000E-01	3.720088E-01
0.950	2.995732E 00	2.038575E-02	2.955776E 00	TO	3.035687E 00	5.000000E-02	3.593394E-01
0.975	3.688878E 00	2.932258E-02	3.631406E 00	TO	3.746349E 00	2.500004E-02	3.533648E-01
0.980	3.912021E 00	3.163255E-02	3.850020E 00	TO	3.974020E 00	2.000005E-02	3.245586E-01
0.990	4.605165E 00	4.504210E-02	4.516883E 00	TO	4.693447E 00	1.000005E-02	3.222002E-01
0.995	5.298306E 00	6.391525E-02	5.173032E 00	TO	5.423580E 00	5.000055E-03	3.210368E-01
0.998	6.214591E 00	1.044402E-01	6.009888E 00	TO	6.419293E 00	2.000034E-03	3.436437E-01
0.999	6.907709E 00	1.477937E-01	6.618033E 00	TO	7.197384E 00	1.000046E-03	3.434250E-01

EXPONENTIAL RANDOM VARIABLE

DATA SUMMARY FOR A SAMPLE OF 10000 OBSERVATIONS

MEAN	3.11150E-04	95% CONFIDENCE INTERVAL	-2.01607E-02	TO	2.07830E-02
VARIANCE	1.09095E 00	STANDARD DEVIATION	1.04449E 00		
3RD CENTRAL MOMENT	-2.63083E-02	COEFFICIENT OF SKEWNESS	-2.30880E-02		
4TH CENTRAL MOMENT	4.03228E 00	COEFFICIENT OF KURTOSIS	3.87986E-01		
MINIMUM	-4.81230E 00	MAXIMUM	4.86073E 00		

ALPHA	QUANTILE	EST STD DEV	95 PER CENT CONFIDENCE INTERVAL			DENSITY	XFORMED DENSITY	SAMPLE SIZE
0.001	-3.560387E 00	2.305142E-01	-4.012185E 00	TO	-3.108587E 00	1.904003E-03	6.538521E-01	11
0.002	-3.148242E 00	1.257510E-01	-3.394709E 00	TO	-2.901774E 00	4.537407E-03	7.796126E-01	26
0.005	-2.826620E 00	7.541674E-02	-2.974434E 00	TO	-2.678805E 00	1.102360E-02	7.077886E-01	86
0.010	-2.591060E 00	6.598794E-02	-2.720393E 00	TO	-2.461725E 00	1.760407E-02	5.672010E-01	175
0.020	-2.131455E 00	4.581636E-02	-2.221253E 00	TO	-2.041656E 00	3.541069E-02	5.746408E-01	354
0.025	-2.025303E 00	4.433684E-02	-2.112201E 00	TO	-1.938404E 00	4.240016E-02	5.993082E-01	354
0.050	-1.692453E 00	2.939543E-02	-1.750067E 00	TO	-1.634839E 00	8.873361E-02	6.377100E-01	711
0.100	-1.304002E 00	2.315917E-02	-1.349393E 00	TO	-1.258611E 00	1.535949E-01	5.713865E-01	1425
0.250	-6.774128E-01	1.366422E-02	-7.041941E-01	TO	-6.506314E-01	3.169902E-01	3.169902E-01	9994
0.500	2.240059E-03	1.328269E-02	-2.379351E-02	TO	2.827363E-02	3.764862E-01	3.764862E-01	9997
0.750	6.790532E-01	1.413222E-02	6.513546E-01	TO	7.067518E-01	3.064928E-01	3.064928E-01	9994
0.900	1.296035E 00	2.137692E-02	1.254136E 00	TO	1.337933E 00	1.664005E-01	6.190245E-01	1425
0.950	1.707108E 00	2.879910E-02	1.650663E 00	TO	1.763554E 00	9.057099E-02	6.509149E-01	711
0.975	2.084478E 00	3.624368E-02	2.013442E 00	TO	2.155514E 00	5.186806E-02	7.331329E-01	354
0.980	2.175792E 00	3.635389E-02	2.104539E 00	TO	2.247044E 00	4.462766E-02	7.242128E-01	354
0.990	2.458925E 00	5.130028E-02	2.358378E 00	TO	2.559471E 00	2.264427E-02	7.295955E-01	175
0.995	2.801413E 00	7.835460E-02	2.647840E 00	TO	2.954985E 00	1.061028E-02	6.812506E-01	86
0.998	3.286646E 00	3.237893E-01	2.652031E 00	TO	3.921261E 00	1.762206E-03	3.027804E-01	26
0.999	3.590726E 00	1.516731E-01	3.293451E 00	TO	3.888000E 00	2.893721E-03	9.937304E-01	11

ALPHA	QUANTILE	H 1	H 2	H 3	H 4	H 5	H 6	H 7	H 8	H 9	H10
0.001	-3.560387E 00	0.0016	0.0004	0.0003	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.002	-3.148242E 00	0.0024	0.0009	0.0004	0.0003	0.0	0.0	0.0	0.0	0.0	0.0
0.005	-2.826620E 00	0.0063	0.0016	0.0005	0.0003	0.0	0.0	0.0	0.0	0.0	0.0
0.010	-2.591060E 00	0.0117	0.0025	0.0008	0.0003	0.0003	0.0	0.0	0.0	0.0	0.0
0.020	-2.131455E 00	0.0232	0.0094	0.0019	0.0005	0.0003	0.0001	0.0	0.0	0.0	0.0
0.025	-2.025303E 00	0.0269	0.0117	0.0028	0.0009	0.0003	0.0003	0.0	0.0	0.0	0.0
0.050	-1.692453E 00	0.0501	0.0207	0.0074	0.0015	0.0004	0.0003	0.0001	0.0	0.0	0.0
0.100	-1.304002E 00	0.0997	0.0421	0.0182	0.0058	0.0011	0.0003	0.0003	0.0001	0.0	0.0
0.250	-6.774128E-01	0.2489	0.1299	0.0568	0.0227	0.0086	0.0023	0.0005	0.0003	0.0001	0.0
0.500	2.240059E-03	0.4994	0.3275	0.1823	0.0869	0.0356	0.0134	0.0038	0.0011	0.0003	0.0003
0.750	6.790532E-01	0.7467	0.5794	0.4083	0.2461	0.1260	0.0565	0.0222	0.0075	0.0023	0.0004
0.900	1.296035E 00	0.8980	0.7880	0.6366	0.4702	0.3021	0.1671	0.0783	0.0329	0.0123	0.0033
0.950	1.707108E 00	0.9465	0.8845	0.7662	0.6148	0.4496	0.2863	0.1582	0.0732	0.0314	0.0120
0.975	2.084478E 00	0.9742	0.9381	0.8696	0.7440	0.5907	0.4265	0.2716	0.1482	0.0693	0.0293
0.980	2.175792E 00	0.9785	0.9453	0.8824	0.7701	0.6193	0.4578	0.2970	0.1713	0.0802	0.0357
0.990	2.458925E 00	0.9891	0.9650	0.9231	0.8381	0.7077	0.5525	0.3890	0.2414	0.1309	0.0596
0.995	2.801413E 00	0.9940	0.9805	0.9493	0.8924	0.7913	0.6490	0.4946	0.3375	0.2045	0.1052
0.998	3.286646E 00	0.9979	0.9937	0.9808	0.9525	0.8978	0.8014	0.6703	0.5197	0.3706	0.2305
0.999	3.590726E 00	0.9987	0.9971	0.9892	0.9694	0.9280	0.8587	0.7484	0.6089	0.4511	0.3097

ACTUAL POPULATION PARAMETER VALUES TO BE ESTIMATED

MEAN	0.0		
VARIANCE	1.05556E 00	STANDARD DEVIATION	1.02740E 00
3RD CENTRAL MOMENT	0.0	COEFFICIENT OF SKEWNESS	0.0
4TH CENTRAL MOMENT	3.78827E 00	COEFFICIENT OF KURTOSIS	4.00002E-01

ALPHA	QUANTILE	EST STD DEV	95 PER CENT CONFIDENCE INTERVAL			DENSITY	XFORMED DENSITY
0.001	-3.579397E 00	6.401971E-03	-3.591945E 00	TO	-3.566849E 00	2.274120E-03	7.809537E-01
0.002	-3.273146E 00	4.572511E-03	-3.282107E 00	TO	-3.264183E 00	4.499882E-03	7.731649E-01
0.005	-2.860933E 00	2.874612E-03	-2.866567E 00	TO	-2.855299E 00	1.095093E-02	7.031229E-01
0.010	-2.539482E 00	2.094300E-03	-2.543587E 00	TO	-2.535377E 00	2.118521E-02	6.825847E-01
0.020	-2.204699E 00	1.544232E-03	-2.207725E 00	TO	-2.201672E 00	4.035557E-02	6.548059E-01
0.025	-2.093020E 00	1.460285E-03	-2.095882E 00	TO	-2.090158E 00	4.944884E-02	6.989383E-01
0.050	-1.729132E 00	1.099625E-03	-1.731286E 00	TO	-1.726976E 00	9.130675E-02	6.562024E-01
0.100	-1.327726E 00	8.454153E-04	-1.329383E 00	TO	-1.326069E 00	1.621312E-01	6.031424E-01
0.250	-6.876179E-01	5.425389E-04	-6.886812E-01	TO	-6.865545E-01	3.079287E-01	3.079287E-01
0.500	0.0	4.899504E-04	-9.603023E-04	TO	9.603023E-04	3.937298E-01	3.937298E-01
0.750	6.876179E-01	5.425389E-04	6.865545E-01	TO	6.886812E-01	3.079287E-01	3.079287E-01
0.900	1.327726E 00	8.454153E-04	1.326069E 00	TO	1.329383E 00	1.621312E-01	6.031424E-01
0.950	1.729132E 00	1.099625E-03	1.726976E 00	TO	1.731286E 00	9.130675E-02	6.562024E-01
0.975	2.093020E 00	1.460285E-03	2.090158E 00	TO	2.095882E 00	4.944884E-02	6.989383E-01
0.980	2.204690E 00	1.544208E-03	2.201663E 00	TO	2.207716E 00	4.035621E-02	6.548961E-01
0.990	2.539471E 00	2.094240E-03	2.535365E 00	TO	2.543575E 00	2.118582E-02	6.826043E-01
0.995	2.860920E 00	2.874516E-03	2.855286E 00	TO	2.866553E 00	1.095130E-02	7.031463E-01
0.998	3.273071E 00	4.571784E-03	3.264111E 00	TO	3.282031E 00	4.500598E-03	7.732878E-01
0.999	3.579220E 00	6.399393E-03	3.566676E 00	TO	3.591763E 00	2.275037E-03	7.312685E-01

ACTUAL RESULTS FOR T-TEST ON A SAMPLE WITH 20 OBSERVATIONS

ALPHA	QUANTILE	H 1	H 2	H 3	H 4	H 5	H 6	H 7	H 8	H 9	H10
0.001	-3.579397E 00	0.0010	0.0002	0.0001	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0
0.002	-3.273146E 00	0.0020	0.0005	0.0001	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0
0.005	-2.860933E 00	0.0050	0.0014	0.0003	0.0001	0.0000	0.0000	0.0000	0.0	0.0	0.0
0.010	-2.539482E 00	0.0100	0.0030	0.0008	0.0002	0.0000	0.0000	0.0000	0.0	0.0	0.0
0.020	-2.204699E 00	0.0200	0.0066	0.0019	0.0004	0.0001	0.0000	0.0000	0.0	0.0	0.0
0.025	-2.093020E 00	0.0250	0.0085	0.0025	0.0006	0.0001	0.0000	0.0000	0.0	0.0	0.0
0.050	-1.729132E 00	0.0500	0.0189	0.0061	0.0016	0.0004	0.0001	0.0000	0.0000	0.0	0.0
0.100	-1.327726E 00	0.1000	0.0428	0.0155	0.0047	0.0012	0.0003	0.0000	0.0000	0.0000	0.0
0.250	-6.876179E-01	0.2500	0.1316	0.0590	0.0223	0.0071	0.0019	0.0004	0.0001	0.0000	0.0000
0.500	0.0	0.5000	0.3274	0.1855	0.0899	0.0368	0.0127	0.0036	0.0009	0.0002	0.0000
0.750	6.876179E-01	0.7500	0.5909	0.4151	0.2549	0.1349	0.0608	0.0232	0.0074	0.0020	0.0004
0.900	1.327726E 00	0.9000	0.8007	0.6578	0.4877	0.3199	0.1827	0.0897	0.0376	0.0133	0.0040
0.950	1.729132E 00	0.9500	0.8875	0.7829	0.6372	0.4682	0.3049	0.1734	0.0852	0.0358	0.0128
0.975	2.093020E 00	0.9750	0.9376	0.8665	0.7535	0.6031	0.4355	0.2790	0.1565	0.0761	0.0318
0.980	2.204690E 00	0.9800	0.9485	0.8865	0.7840	0.6420	0.4770	0.3160	0.1841	0.0934	0.0409
0.990	2.539471E 00	0.9900	0.9719	0.9324	0.8595	0.7464	0.5982	0.4344	0.2814	0.1607	0.0802
0.995	2.860920E 00	0.9950	0.9848	0.9605	0.9111	0.8262	0.7026	0.5499	0.3895	0.2464	0.1377
0.998	3.273071E 00	0.9980	0.9934	0.9810	0.9531	0.8992	0.8102	0.6850	0.5341	0.3783	0.2406
0.999	3.579220E 00	0.9990	0.9965	0.9893	0.9717	0.9351	0.8695	0.7685	0.6350	0.4830	0.3338

ACTUAL RESULTS FOR T-TEST ON A SAMPLE WITH 20 OBSERVATIONS

BIBLIOGRAPHY

[1] Barlow, R.E., and others, Statistical Inference under Order Restrictions, Wiley, 1972.

[2] Blum, J.R., "Approximation Methods which Converge with Probability One," Annals of Mathematical Statistics, v. 25, p. 382-386, 1954.

[3] Burkholder, D.L., "On a Class of Stochastic Approximation Procedures", Annals of Mathematical Statistics, v. 27, p. 1044-1059, 1956.

[4] Cochran, W.G., and Davis, M., "The Robbins-Monro Method for Estimating the Median Lethal Dose," Journal of the Royal Statistical Society, Series B, v. 27, p. 28-44, 1965.

[5] David, H.A., Order Statistics, Wiley, 1970.

[6] Davis, M., "Comparison of Sequential Bioassays in Small Samples," Journal of the Royal Statistical Society, Series B, v. 33, p. 78-87, 1971.

[7] Dvoretzky, A., "On Stochastic Approximation," Proceedings of the Third Berkely Symposium on Mathematical Statistics and Probability, v. 1, p. 39-55, 1956.

[8] Efron, B., and Morris, C., Data Analysis using Stein's Estimator and its Generalizations, Rand Corporation Technical Report R-1394-OEO, 1974.

[9] Fabian, V., "On Asymptotic Normality in Stochastic Approximation," Annals of Mathematical Statistics, v. 39, p. 1327-1332, 1968.

[10] Fabian, V., "Stochastic Approximation," Optimizing Methods in Statistics, J.S. Rustagi (ed.), p. 439-470, Academic Press, 1971.

[11] Floyd, R.W., and Rivest, R.L., "The Algorithm SELECT -- for Finding the ith Smallest of n Elements [M1]," ACM Algorithm 489, Communications of the ACM, v. 18, p. 173, 1975.

[12] Gaver, D.P., and Lewis, P.A.W., Statistical Approaches to Managing Manpower Data Bases, Naval Postgraduate School Technical Report NPS-55Gv55Lw-74021, February 1974.

[13] Gaver, D.P., and Thompson, G.L., Programming and Probability Models in Operations Research, Brooks-Cole, Monterey, California, p. 582-590, 1973.

[14] Goodman, A.S., Lewis, P.A.W., and Robbins, H.E., Simultaneous Estimation of Large Numbers of Extreme Quantiles in Simulation Experiments, Naval Postgraduate School Technical Report NPS-55Lw-71122A, 1971.

[15] Hodges, J.L., Jr. and Lehmann, E.L., "Two Approximations to the Robbins-Monro Process," Proceedings of the Third Berkely Symposium on Mathematical Statistics and Probability, v. 1, p. 95-104, 1956.

[16] Iglehart, D.L., Simulating Stable Stochastic Systems, VI: Quantile Estimation, Control Analysis Corporation Technical Report No. 86-15, December 1974.

[17] Kendall, M.G., and Stuart, A., The Advanced Theory of Statistics, Volume 2, Hafner, New York, p. 517-518, 1961.

[18] Kesten, H., "Accelerated Stochastic Approximation," Annals of Mathematical Statistics, v. 29, p. 41-59, 1958.

[19] Knuth, D.E., The Art of Computer Programming, Volume III: Sorting and Searching, Addison-Wesley, 1973.

[20] Kruskal, J.B., "Nonmetric Multidimensional Scaling: A Numerical Method," Psychometrika, v. 29, p. 115-129, 1964.

[21] Learmonth, G.P., and Lewis, P.A.W., Naval Postgraduate School Random Number Package LLRANDOM, Naval Postgraduate School Technical Report NPS-55Lw-73061A, June 1973.

[22] Lewis, P.A.W., Large-Scale Computer-Aided Statistical Mathematics, Naval Postgraduate School Technical Report NPS-55Lw-72111A, November 1972.

[23] Lewis, P.A.W., Liu, L.H., Robinson, D.W., and Rosenblatt, M., "Empirical Sampling Study of a Goodness of Fit Statistic for Density Function Estimation," Submitted for publication, 1975.

[24] Loeve, M., Probability Theory, 3rd ed., D. Van Nostrand, 1963.

[25] Lukacs, E., Stochastic Convergence, Heath, 1968.

[26] Major, P., and Revesz, P., "A Limit Theorem for the Robbins-Monro Approximation," Zeitschrift fur Wahrscheinlichkeitstheorie, v. 27, p. 79-86, 1973.

[27] Miller, R.G., "The Jackknife - A Review," Biometrika, 61, p. 1-15, April 1974.

[28] Parzen, E., "On Estimation of a Probability Density Function and Mode," Annals of Mathematical Statistics, v. 33, p. 1065-1076, 1962.

[29] Payne, W.H., "Partial Sorting: A Large Vector Technique and Its Application," International Journal of Computer and Information Sciences, v. 2, p. 141-156, 1973.

[30] Robbins, H.E., and Monro, S., "A Stochastic Approximation Method," Annals of Mathematical Statistics, v. 22, p. 400-407, 1951.

[31] Robinson, D.W., and Lewis, P.A.W., Generating Gamma and Cauchy Random Variables: An Extension to the Naval Postgraduate School Random Number Package, Naval Postgraduate School Technical Report NPS-72Ro-75041, April 1975.

[32] Rosenblatt, M., "Remarks on some Nonparametric Estimates of a Density Function," Annals of Mathematical Statistics, v. 27, p. 832-835, 1956.

[33] Sacks, J., "Asymptotic Distribution of Stochastic Approximation Procedures," Annals of Mathematical Statistics, v. 29, p. 373-405, 1958.

[34] Sielken, R.L., Some Stopping Times for Stochastic Approximation Procedures, Florida State University Technical Report M217, 1971.

[35] Sielken, R.L., "Stopping Times for Stochastic Approximation Procedures," Zeitschrift fur Wahrscheinlichkeitstheorie, v. 26, p. 67-75, 1973.

[36] Wetherill, G.B., "Sequential Analysis of Quantal Response Curves," Journal of the Royal Statistical Society, Series B, v. 25, p. 1-48, 1963.

[37] Venter, J.H., "An Extension of the Robbins-Monro Procedure," Annals of Mathematical Statistics, v. 38, p. 181-190, 1967.

[38] Yuguchi, T.G., "Quantile Estimation Using the Maximum Transformation, Stochastic Approximation and the Jackknife," M.S. Thesis, Naval Postgraduate School, Monterey, California, March 1973.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department of Operations Research and Administrative Science, Code 55 Naval Postgraduate School Monterey, California 93940	1
4.	Computer Science Group, Code 72 Naval Postgraduate School Monterey, California 93940	1
5.	Professor D.R. Barr (Code 55Bn) Professor G.H. Bradley (Code 55Bz) Professor G.G. Brown (Code 55Zr) Professor D.P. Gaver, Jr. (Code 55Gv) Professor J.K. Hartman (Code 55Hh) Professor P.A.W. Lewis (Code 55Lw) Professor W.M. Raike (Code 55Rj) Professor B.O. Shubert (Code 55Sy) Professor M.U. Thomas (Code 55To) Department of Operations Research and Administrative Science Naval Postgraduate School Monterey, California 93940	1 1 1 1 1 3 1 1 1
6.	Professor U.R. Kodres (Code 72Kr) Professor V.M. Powers (Code 72Pw) Computer Science Group Naval Postgraduate School Monterey, California 93940	1 1
7.	Mr. D.W. Robinson 836 Loudan Lane Newtown Square, Pennsylvania 19073	1
8.	Professor W.G. Cochran Statistics Department Harvard University 2 Divinity Avenue Cambridge, Massachusetts 02138	1
9.	Professor D.R. Cox Department of Mathematics Imperial College 53 Princes Gate London, S.W. 7, England	1
10.	Dr. M.A. Crane Control Analysis Corporation 800 Welch Road Palo Alto, California 94304	1
11.	Professor M. Davis Department of Biostatistics Johns Hopkins University 615 North Wolfe Street Baltimore, Maryland 21205	1

12. Mr. M. Denicoff (Code 437) 1
Office of Naval Research
Arlington, Virginia 22217

13. Professor V. Fabian 1
Department of Statistics and Probability
Michigan State University
Wells Hall
East Lansing, Michigan 48823

14. Dr. D.C. Hoaglin 1
National Bureau of Economic Research, Incorporated
575 Technology Square
Cambridge, Massachusetts 02139

15. Professor D.L. Iglehart 1
Department of Operations Research
Stanford University
Stanford, California 94305

16. Professor H. Kesten 1
Department of Mathematics
Cornell University
Ithaca, New York 14850

17. Professor D.E. Knuth 1
Department of Computer Science
Stanford University
Stanford, California 94305

18. Professor E.L. Lehmann 1
Department of Statistics
University of California
Berkeley, California 94720

19. Professor G. Marsaglia 1
Department of Computing Sciences
McGill University
Montreal, Province of Quebec
Canada

20. Dr. B. Mcdonald 1
Office of Naval Research
Arlington, Virginia 22217

21. Professor H.E. Robbins 1
Department of Mathematical Statistics
Columbia University
New York, New York 10027

22. Professor M. Rosenblatt 1
Department of Mathematics
University of California, San Diego
La Jolla, California 92037

23. Professor J. Sacks 1
Department of Mathematics
Northwestern University
Evanston, Illinois 60201

24. Mr. G.S. Shedler 1
Department K53, Building 282
IBM Research Laboratory
Monterey and Cottle Roads
San Jose, California 95193

25. Professor R.L. Sielken, Jr. 1
Institute of Statistics
Texas A & M University
College Station, Texas 77843

26. Dr. T. Varley (Code 434) 1
Office of Naval Research
Arlington, Virginia 22217

27. Professor J.H. Venter 1
Department of Statistics
Potchefstroom University
Potchefstroom, SouthAfrica

28. Professor G.B. Wetherill 1
School of Mathematics, Bath University
Claverton Down
Bath Ba2 7ay, England

161383

Thesis
R6381 Robinson
c.1 Non-parametric quantile estimation through stochastic approximation.