STATISTICAL INFERENCE FOR BOUNDS OF RANDOM VARIABLES

BY

PETER COOKE

TECHNICAL REPORT NO. 32
APRIL 3, 1979

PREPARED UNDER GRANT
DAAG29-77-G-0031
FOR THE U.S. ARMY RESEARCH OFFICE

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

STATISTICAL INFERENCE FOR BOUNDS OF RANDOM VARIABLES

By

PETER COOKE

TECHNICAL REPORT NO. 32

APRIL 3, 1979

Prepared under Grant DAAG29-77-G-0031
For the U.S. Army Research Office

Herbert Solomon, Project Director

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 271.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Statistical Inference for Bounds of Random Variables

by Peter Cooke
Statistics Department,
University of New South Wales, Sydney.

Summary

Robson and Whitlock (1964) considered point estimation and confidence limits for the upper bound of a random variable when the bound was known to be a truncation point. However, their approach to the point estimation problem failed to produce an estimator with smaller mean squared error than the largest order statistic from a random sample. In this paper we will construct point estimators of the bounds of random variables which are substantially better estimators than the extreme order statistics for many classes of random variables, including those whose distributions are truncated at one or both ends. We will also construct confidence limits and tests of hypotheses for bounds. The main results are large sample results.

1. Introduction

Suppose $X_1, X_2, \ldots, X_n$ are independent random variables, each with absolutely continuous cumulative distribution function $F(x)$, where $F(x) \in (0,1)$ only for $x \in (\phi, \theta)$. The interval (ϕ, θ) is sometimes referred to as the support of the distribution function F; see, for example, Feller (1966). Let $Y_1 \leq Y_2 \leq \ldots \leq Y_n$ be the order statistics based on $X_1, X_2, \ldots, X_n$. We will construct a point estimator of θ when θ is known to be finite and ϕ is unknown. Of course the result which follow also apply when ϕ is known and,

in particular, when $\phi = -\infty$, the only assumption required then being that

$$\int_{-\infty}^{\theta} x^2 dF(x) < \infty .$$

We will also construct confidence limits for θ and large sample tests of hypotheses about θ. No results for lower bounds will be proved since these can easily be derived from the upper board results.

2. Point Estimation of θ.

With no information about the form of F and, in particular, with no information about the shape of the upper tail of $f(x) = F'(x)$, the statistic chosen to estimate θ seems likely to be Y_n. In an attempt to improve on the estimator Y_n in the sense of reducing its mean squared error when θ was known to be a truncation point of the upper tail of $f(x)$, Robson and Whitlock (1964) applied a modification of Quenoulli's (1956) bias reduction technique to Y_n. This led to the family of estimators

$$T_n^{(k)} = \sum_{i=0}^{k} (-1)^i \binom{k+1}{i+1} Y_{n-i} , \qquad k = 1,2,\ldots,n-1 ,$$

where removal of the leading bias term gave $T_n^{(1)}$, removal of the next term gave $T_n^{(2)}$ and so on. However, it was found that $T_n^{(1)}$ had the same asymptotic mean squared error as Y_n, while the mean squared error of $T_n^{(k)}$ increased with k. The author has been able to show that for the other types of upper bound considered in this paper the mean squared error of $T_n^{(k)}$ also increases with k, but that $T_n^{(1)}$ has smaller asymptotic mean squared error than Y_n. Thus we will compare the mean squared error of the estimator derived here with that of

$$T_n^{(1)} = Y_n + (Y_n - Y_{n-1}) = 2Y_n - Y_{n-1} .$$

The term $Y_n - Y_{n-1}$ attempts to correct the bias in Y_n and of course does so in a sensible way since Y_n underestimates θ and $Y_n - Y_{n-1}$ is nonnegative with probability one .

The random variable Y_n has distribution function $F^n(y)$ and mean

$$E(Y_n) = \int_{\phi}^{\theta} y dF^n(y) = \theta - \int_{\phi}^{\theta} F^n(y)dy , \qquad (1)$$

on integrating by parts. Thus we can write

$$\theta = E(Y_n) + \int_{\phi}^{\theta} F^n(y)dy$$

and this suggests the estimator

$$Y_n + \int_{Y_1}^{Y_n} \hat{F}^n(y)dy$$

where $\hat{F}(y)$ is the empirical distribution function based on the order statistics $Y_1, Y_2, \ldots, Y_n$; that is,

$$\hat{F}(y) = \begin{cases} 0 & , \quad y < Y_1 \\ \frac{i}{n} & , \quad y_i \leq y < Y_{i+1}, \; i = 1,2,\ldots,n-1 \\ 1 & , \quad y \geq Y_n \end{cases} \qquad (2)$$

Now

$$\int_{Y_1}^{Y_n} \hat{F}^n(y)dy = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n (Y_{i+1} - Y_i)$$

$$= Y_n - \sum_{i=1}^{n} \left\{\left(\frac{i}{n}\right)^n - \left(\frac{i-1}{n}\right)^n\right\} Y_i = Y_n - \sum_{i=0}^{n-1} \left\{ \left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n \right\} Y_{n-i}$$

and so the suggested estimator is

$$2Y_n - \sum_{i=0}^{n-1} \left\{\left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n \right\} Y_{n-i} .$$

However, since we will only investigate the case in which n is large, we will consider the estimator

$$\hat{\theta}_n = 2Y_n - (1-e^{-1}) \sum_{i=0}^{n-1} e^{-i} Y_{n-i} \tag{3}$$

and compare its performance with that of $T_n^{(1)}$. The asymptotic efficiency of $T_n^{(1)}$ relative to $\hat{\theta}_n$ will be defined to be

$$\lim_{n \to \infty} \frac{E(\hat{\theta}_n-\theta)^2}{E(T_n^{(1)}-\theta)^2} .$$

When estimating ϕ, the equation

$$E(Y_1) = \phi + \int_{\phi}^{\theta} [1-F(y)]^n dy$$

suggests the estimator

$$Y_1 - \int_{Y_1}^{Y_n} \{1 - \hat{F}(y)\}^n \, dy = 2Y_1 - \sum_{i=1}^{n} \left\{ \left(1 - \frac{i-1}{n}\right)^n - \left(1 - \frac{i}{n}\right)^n \right\} Y_i$$

and so, for large n, we will consider the estimator

$$\hat{\phi}_n = 2Y_1 - (e-1) \sum_{i=1}^{n} e^{-i} Y_i .$$

It should be noted at this stage that although we will only investigate properties of estimators based on continuous random variables, in the case of discrete variables, arguments like those above lead to similar estimators to $\hat{\theta}_n$ and $\hat{\phi}_n$. For example, suppose a random variable with distribution function F can take only the integer values $\phi < \phi + 1 < \ldots < \theta$ with positive probability. As above, suppose $Y_1, Y_2, \ldots, Y_n$ are the order statistics based on a random sample of size n from F. Then

$$E(Y_n) = \phi F^n(\phi) + \sum_{y=\phi+1}^{\theta} y\{F^n(y) - F^n(y-1)\} = \theta+1 - \sum_{y=\phi}^{\theta} F^n(y)$$

and so the suggested estimator of θ is

$$Y_n - 1 + \sum_{y=Y_1}^{Y_n} \hat{F}^n(y) .$$

Now

$$\sum_{y=Y_1}^{Y_n} \hat{F}^n(y) = 1 + \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n (Y_{i+1} - Y_i)$$

since the left hand side has $Y_{i+1} - Y_i$ terms equal to $\left(\frac{i}{n}\right)^n$, $i = 1,2,\ldots,n-1$ and $\hat{F}^n(Y_n) = 1$. Thus the estimator of θ is exactly as in the continuous case.

3. Asymptotic Efficiency

According to Gnedenko (1943), when θ is finite there are only two possible nondegenerate limiting distributions of normalized values of Y_n. In this paper we will only derive results for the case

$$F^n(y) \sim \exp\left\{-\left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu}\right\} \quad \text{as} \quad n \to \infty ,$$

for which the necessary and sufficient condition is that for $c > 0$,

$$\lim_{y \to 0-} \frac{1-F(cy+\theta)}{1-F(y+\theta)} = c^{1/\nu}$$

where $u_n = F^{-1}\left(1 - \frac{1}{n}\right)$.

In this section we will discuss the asymptotic efficiency of $T_n^{(1)}$ relative to $\hat{\theta}_n$ when F satisfies the above condition. First we require expressions for means, variances and covariances of order statistics. The following expressions are not difficult to obtain:

$$E(Y_{n-i}) = n\binom{n-1}{i} \int_0^1 F^{-1}(x)x^{n-i-1}(1-x)^i dx, \; 0 \le i \le n-1 , \tag{4}$$

$$E(Y_{n-i}|Y_{n-j}=y)=(n-j-1)\binom{n-j-2}{i-j-1}\int_0^1 F^{-1}\{xF(y)\}x^{n-i-1}(1-x)^{i-j-1}dx, 0\le j<i\le n-i. \tag{5}$$

If

$$F^n(y) \sim \exp\left\{-\left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu}\right\} \quad \text{as} \quad n \to \infty ,$$

then $F^{-1}(x) \sim \theta - (\theta-u_n)(-\log x^n)^{\nu}$ as $n \to \infty$ and, using (4), we find that for i small

$$E(Y_{n-i}) \sim \theta - (\theta-u_n)\frac{\Gamma(\nu+i+1)}{\Gamma(i+1)} \quad \text{as} \quad n \to \infty \tag{6}$$

and, from (5), when i is small and $i > j \geq 0$,

$$E(Y_{n-i}|Y_{n-j}=y) \sim \theta - \frac{(\theta-u_n)}{\Gamma(i-j)}\int_0^{\infty}\left\{z+\left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu}\right\}^{\nu} e^{-z} z^{i-j-1}dz \text{ as } n \to \infty. \tag{7}$$

It is convenient to write

$$\mathrm{Cov}(Y_{n-i},Y_{n-j}) = E(Y_{n-i}-\theta)\{Y_{n-j}-E(Y_{n-j})\}$$

$$= n\binom{n-1}{j}\int_{\phi}^{\theta}\{y-E(Y_{n-j})\}E(Y_{n-i}-\theta|Y_{n-j}=y)F^{n-j-1}(y)\{1-F(y)\}^j\, dF(y)$$

from which, using (6) and (7), we obtain

$$\mathrm{Cov}(Y_{n-i},Y_{n-j}) \sim \frac{(\theta-u_n)^2}{\Gamma(j+1)\Gamma(i-j)}\int_0^{\infty}\left\{t^{\nu+j}-\frac{\Gamma(\nu+j+1)}{\Gamma(j+1)}t^j\right\}e^{-t}\int_0^{\infty}(z+t)^{\nu}e^{-z}z^{i-j-1}dzdt$$

as $n \to \infty$ and finally, for $i > j \geq 0$ and i small,

$$\mathrm{Cov}(Y_{n-i},Y_{n-j}) \sim (\theta-u_n)^2\frac{\Gamma(\nu+j+1)}{\Gamma(j+1)}\left\{\frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)}-\frac{\Gamma(\nu+i+1)}{\Gamma(i+1)}\right\} \text{ as } n \to \infty. \tag{8}$$

Similar calculations to those above give

$$\mathrm{Var}(Y_{n-i}) \sim (\theta-u_n)^2\left\{\frac{\Gamma(2\nu+i+1)}{\Gamma(i+1)}-\frac{\Gamma^2(\nu+i+1)}{\Gamma^2(i+1)}\right\} \text{ as } n \to \infty \tag{9}$$

for $i \geq 0$ and i small. Thus (8) also holds for $i = j$.

From (6), (8) and (9) we now have

$$(\theta-u_n)^{-2}E(T_n^{(1)}-\theta)^2 \sim \frac{(2\nu^2-\nu+1)}{(\nu+1)}\Gamma(2\nu+1) \quad \text{as} \quad n \to \infty. \tag{10}$$

We will now find an expression for $E(\hat{\theta}_n-\theta)^2$. From (3) and (6) we obtain

$$(\theta-u_n)^{-1}E(\hat{\theta}_n-\theta) \sim \{(1-e^{-1})^{-\nu}-2\}\Gamma(\nu+1) \quad \text{as} \quad n \to \infty \tag{11}$$

and, using (3), (8), (9) and (11),

$$(\theta-u_n)^{-2}E(\hat{\theta}_n-\theta)^2 \sim 4\Gamma(2\nu+1)+4\Gamma^2(\nu+1)(1-e^{-1})^{-\nu}+\Gamma(2\nu+1)(1-e^{-1})^2(1-e^{-2})^{-2\nu-1}$$

$$-\Gamma^2(\nu+1)(1-e^{-1})^{-2\nu}-4\Gamma(\nu+1)(1-e^{-1})\sum_{i=0}^{\infty} e^{-i}\frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)}$$

$$+2(1-e^{-1})^2\sum_{i=1}^{\infty} e^{-i}\frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)}\sum_{j=0}^{i-1} e^{-j}\frac{\Gamma(\nu+j+1)}{\Gamma(j+1)} \quad \text{as} \quad n\to\infty \quad . \tag{12}$$

Thus the asymptotic efficiency of $T_n^{(1)}$ relative to $\hat{\theta}_n$ is given by

$$\eta(\nu) = \lim_{n\to\infty}\frac{E(\hat{\theta}_n-\theta)^2}{E(T_n^{(1)}-\theta)^2}$$

which equals the ratio of the expressions on the right hand sides of (12) and (11). Some values of $\eta(\nu)$ are given to three significant figures in Table 1.

Table 1: Values of the asymptotic efficiency of $T_n^{(1)}$ relative to $\hat{\theta}_n$.

ν	1/5	1/4	1/3	1/2	1
$\eta(\nu)$	1.16	1.17	1.17	1.08	.666

The value $\nu = 1$ corresponds to densities $f(x)$ which are truncated at θ ; that is, $0 < f(\theta) < \infty$. Thus $\hat{\theta}_n$ provides a solution to the problem of Robson and Whitlock, though it is slightly inferior to $T_n^{(1)}$ for values of ν other than $\nu = 1$. However, in the next section we will improve on $\hat{\theta}_n$ to produce an estimator with smaller asymptotic mean squared error than an improved $T_n^{(1)}$ for all values of ν considered in Table 1. The value $\nu = \frac{1}{2}$ corresponds to densities $f(x)$ with $f(\theta)$ equal to zero or infinity, but with $f'(\theta)$ nonzero and finite and, in

general, $\nu = \frac{1}{k+1}$ corresponds to a density which is zero or infinite at θ and whose first finite, nonzero derivative at θ is its kth derivative. The derivatives mentioned here are of course all left derivatives at θ.

4. Improving on $\hat{\theta}_n$.

Suppose now that the value of ν corresponding to the upper tail of the density f is known, though the form of the function is unknown. This situation seems likely to occur in practical problems, since for example one might know that he is sampling from a distribution which is truncated at θ, so $\nu = 1$. In this case it does not seem to be possible in general, by considering a function of Y_n alone, to improve on the estimator Y_n either in the sense of reducing the order of magnitude of its bias or of reducing its mean squared error. For example, if we consider estimators proportional to Y_n, then the constant of proportionality which minimizes the mean squared error of the 'estimator' is

$$c(\alpha) = 1 + \Gamma(\alpha+1)\theta^{-1}(\theta-u_n) + \theta^{-2}\, O[(\theta-u_n)^2]$$

which is a function of θ.

On the other hand, the estimator

$$\tilde{T}_n^{(1)} = Y_n + \nu^{-1}(Y_n - Y_{n-1})$$

has bias of order of magnitude $(\theta-u_n)^2$ compared with order $\theta-u_n$ for $T_n^{(1)}$. Using (6) and (8) we easily find

$$(\theta-u_n)^{-2}E(\tilde{T}_n^{(1)}-\theta)^2 \sim \Gamma(2\nu+1) \quad \text{as} \quad n \to \infty$$

and

$$(\theta-u_n)^{-2}E(Y_n-\theta)^2 \sim \Gamma(2\nu+1) \quad \text{as} \quad n \to \infty \quad ,$$

so that bias reduction in $T_n^{(1)}$ is achieved at the expense of increased asymptotic mean squared error except for $\nu = 1$ where there is no change. The estimator

$$\tilde{\theta}_n = Y_n + \{1-(1-e^{-1})^{-\nu}\}^{-1} \{Y_n-(1-e^{-1})\sum_{i=0}^{n-1} e^{-i}Y_{n-i}\}$$

also has bias of order of magnitude $(\theta-u_n)^2$ compared with order $\theta-u_n$ for $\hat{\theta}_n$. Some values of

$$\eta_1(\nu) = \lim_{n \to \infty} \frac{E(\tilde{\theta}_n-\theta)^2}{E(\tilde{T}_n^{(1)}-\theta)^2} ,$$

the asymptotic efficiency of $\tilde{T}_n^{(1)}$ relative to $\tilde{\theta}_n$, are given to three significant figures in Table 2.

Table 2: Values of the asymptotic efficiency of $\tilde{T}_n^{(1)}$ relative to $\tilde{\theta}_n$.

ν	1/5	1/4	1/3	1/2	1
$\eta_1(\nu)$	.680	.683	.689	.700	.731

However, our main aim is to construct estimators with mean squared error as small as possible. When ν is known, the estimator of θ of the form

$$\bar{T}_n^{(1)} = Y_n + c_1(\nu)(Y_n-Y_{n-1})$$

with smallest mean squared error is the one for which $c_1(\nu) = 1/2\nu$ and, for this choice of $c_1(\nu)$, we can show that

$$(\theta-u_n)^{-2} E(\bar{T}_n^{(1)}-\theta)^2 \sim \frac{\Gamma(2\nu+2)}{2(1+\nu)} \quad \text{as} \quad n \to \infty .$$

The estimator of the form

$$\bar{\theta}_n = Y_n + c_2(\nu) \{Y_n - (1-e^{-1})\sum_{i=0}^{n-1} e^{-i}Y_{n-i} \}$$

with smallest mean squared error is the one for which $c_2(\nu)=a_1(\nu)/a_2(\nu)$, where

$$a_1(\nu) = \Gamma(\nu+1)(1-e^{-1}) \sum_{i=0}^{\infty} e^{-i} \frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)} - \Gamma(2\nu+1)$$

and

$$a_2(\nu) = \Gamma(2\nu+1)\{1+(1-e^{-1})^2(1-e^{-2})^{-2\nu-1}\}-2\Gamma(\nu+1)(1-e^{-1}) \sum_{i=0}^{\infty} e^{-i} \frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)}$$

$$+ 2(1-e^{-1})^2 \sum_{i=1}^{\infty} e^{-i} \frac{\Gamma(2\nu+i+1)}{\Gamma(\nu+i+1)} \sum_{j=0}^{i-1} e^{-j} \frac{\Gamma(\nu+j+1)}{\Gamma(j+1)} .$$

The asymptotic efficiency of $\overline{T}_n^{(1)}$ relative to $\overline{\theta}_n$ for the above choices of $c_1(\nu)$ and $c_2(\nu)$ is given by

$$\eta_2(\nu) = \lim_{n \to \infty} \frac{E(\overline{\theta}_n-\theta)^2}{E(\overline{T}_n^{(1)}-\theta)^2} .$$

Some values of $\eta_2(\nu)$ are given to three significant figures in Table 3.

Table 3: Values of $c_2(\nu)$ and the asymptotic efficiency of $\overline{T}_n^{(1)}$ relative to $\overline{\theta}_n$.

ν	1/5	1/4	1/3	1/2	1
$c_2(\nu)$	6.42	5.11	3.80	2.49	1.18
$\eta_2(\nu)$	.819	.824	.831	.844	.877

For practical purposes the simplicity of the estimators $T_n^{(1)}$ and $\overline{T}_n^{(1)}$ makes them attractive. $T_n^{(1)}$ could be used when ν is unknown, though $\hat{\theta}_n$ is a much better estimator if θ happens to be a truncation point. $\overline{T}_n^{(1)}$ could be used when ν is known since even though it is always inefficient relative to $\overline{\theta}_n$, its efficiency doesn't drop below 81% for the values of ν considered here.

5. Confidence Limits For θ .

If

$$F^n(y) \sim \exp\left\{-\left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu}\right\} \quad \text{as} \quad n \to \infty ,$$

then

$$\lim_{n \to \infty} P\left(\frac{\theta-Y_n}{\theta-u_n} \leq x\right) = 1 - \exp(-x^{1/\nu}) \quad \text{for} \quad x \geq 0$$

and it is not difficult to prove that $(\theta - T_n^{(1)})/(\theta-u_n)$ also has a non-degenerate limiting distribution which is not a function of θ . This leads us to consider $(\theta-Y_n)/(\theta-T_n^{(1)})$, or equivalently, $(\theta-Y_n)/(Y_n-Y_{n-1})$ as the basis for constructing confidence limits for θ .

When

$$F^n(y) \sim \exp\left\{-\left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu}\right\} \quad \text{as} \quad n \to \infty$$

we find

$$\lim_{n \to \infty} P\left[\frac{\theta - Y_n}{Y_n-Y_{n-1}} \leq x\right] = \left[\frac{x}{1+x}\right]^{1/\nu} , \quad 0 \leq x < \infty \tag{13}$$

and hence, whatever the values of θ,ϕ or any other parameters of the distribution F ,

$$\lim_{n \to \infty} P[Y_n+\{(\tfrac{\alpha}{2})^{-\nu}-1\}^{-1}\{Y_n-Y_{n-1}\} \leq \theta \leq Y_n+\{(1-\tfrac{\alpha}{2})^{-\nu}-1\}^{-1}\{Y_n-Y_{n-1}\}]=1-\alpha,$$

$$\lim_{n \to \infty} P[\theta \geq Y_n + \{\alpha^{-\nu}-1\}^{-1}\{Y_n-Y_{n-1}\}] = 1-\alpha$$

and

$$\lim_{n \to \infty} P[\theta \leq Y_n + \{(1-\alpha)^{-\nu}-1\}^{-1} \{Y_n-Y_{n-1}\}] = 1 - \alpha .$$

When θ is a truncation point $\nu = 1$ and the upper bound for θ in the last statement equals $Y_n + (\alpha^{-1}-1)(Y_n-Y_{n-1})$, which is Robson and Whitlock's approximate upper confidence bound for θ.

6. Large Sample Tests of Hypotheses About θ.

Consider the hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1 (< \theta_0)$ and the test with rejection region for $H_0 : (\theta_0-Y_n)/(Y_n-Y_{n-1}) \geq \{(1-\alpha)^{-\nu}-1\}^{-1}$. If we denote the power function of this test by $\beta_n(\theta)$ and, if

$$F^n(y) \sim \exp\left\{ - \left(\frac{\theta-y}{\theta-u_n}\right)^{1/\nu} \right\} \quad \text{as} \quad n \to \infty ,$$

from (13) we have $\lim_{n \to \infty} \beta_n(\theta_0) = \alpha$, whatever the values of ϕ or any other parameters of the distribution F. Thus, when n is large the test will have size approximately equal to α. Also, we find the following asymptotic expressions for $\beta_n(\theta)$:

$$\beta_n(\theta) \sim 1-(1-\alpha)\{(1-\alpha)^{-\nu}-1\}^{1/\nu.(1-1/\nu)} \left(\frac{\theta_0-\theta}{\theta-u_n} \right)^{1/\nu.(1-1/\nu)} \nu^{1/\nu-1}\Gamma(1/\nu).$$

$$\exp\left[-\{(1-\alpha)^{-\nu}-1\}\left\{\frac{\theta_0-\theta}{\theta-u_n}\right\}^{1/\nu}\right] \quad \text{as} \quad n \to \infty \quad \text{for} \quad \theta < \theta_0 , \tag{14}$$

$$\beta_n(\theta) \sim \alpha \exp\left\{ - \left(\frac{\theta-\theta_0}{\theta-u_n} \right)^{1/\nu} \right\} \quad \text{as} \quad n \to \infty \quad \text{for} \quad \theta > \theta_0 . \tag{15}$$

These expressions are not necessarily good representations of $\beta_n(\theta)$ for θ near θ_0, except when $\nu = 1$. However, for θ not near θ_0 and n large, $|(\theta_0-\theta)/(\theta-u_n)|$ will be large, so that both expressions are decreasing functions of $(\theta-\theta_0)/(\theta-u_n)$. Thus, if $(\theta-\theta_0)/(\theta-u_n)$ is an increasing function

of θ, the above test is a test of $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ with size approximately α and power function a decreasing function of θ when n is large.

For a test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1 (> \theta_0)$, if we use the rejection region for $H_0 : (\theta_0 - Y_n)/(Y_n - Y_{n-1}) \leq (\alpha^{-\nu}-1)^{-1}$ and denote the power function by $\beta_n^*(\theta)$, using (13) we find $\lim_{n \to \infty} \beta_n^*(\theta_0) = \alpha$ whatever the values of the parameters of F other than θ. We find the following asymptotic expressions for $\beta_n^*(\theta)$:

$$\beta_n^*(\theta) \sim \alpha(\alpha^{-\nu}-1)^{1/\nu.(1-1/\nu)} \left(\frac{\theta_0-\theta}{\theta-u_n} \right)^{1/\nu.(1-1/\nu)} \nu^{1/\nu-1} \Gamma(1/\nu) \ .$$

$$\exp\left[-\left\{ (\alpha^{-\nu}-1) \left(\frac{\theta_0-\theta}{\theta-u_n} \right)^{1/\nu} \right\} \right] \quad \text{as} \quad n \to \infty \quad \text{for } \theta < \theta_0, \tag{16}$$

$$\beta_n^*(\theta) \sim 1 - (1-\alpha) \exp\left[- \left(\frac{\theta-\theta_0}{\theta-u_n} \right)^{1/\nu} \right] \quad \text{as} \quad n \to \infty \quad \text{for } \theta > \theta_0 \ . \tag{17}$$

As in the previous case, except when $\nu = 1$, the expressions in (16) and (17) are not necessarily good representations of the power function when θ is near θ_0. Also, if $(\theta-\theta_0)/(\theta-u_n)$ is an increasing function of θ, the test is a test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ with size approximately α and power function increasing in θ when n is large.

References

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. 2 New York: Wiley.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. 44, 423-54.

Quenoulli, M.H. (1956). Notes on bias in estimation. Biometrika, 43, 353-60.

Robson, D.S. and Whitlock, J.H. (1964). Estimation of a truncation point, Biometrika 51, 33-39.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 32

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): STATISTICAL INFERENCE FOR BOUNDS OF RANDOM VARIABLES

5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): PETER COOKE

8. CONTRACT OR GRANT NUMBER(s): DAAG29-77-G-0031

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Statistics
Stanford University
Stanford, CA 94305

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P-14435-M

11. CONTROLLING OFFICE NAME AND ADDRESS:
U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

12. REPORT DATE: APRIL 3, 1979

13. NUMBER OF PAGES: 14

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

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:
The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents. This report partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 271.

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

Asymptotic efficiency; Confidence limits; Order statistics; Power function; Truncation point.

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

PLEASE SEE REVERSE SIDE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

STATISTICAL INFERENCE FOR BOUNDS OF RANDOM VARIABLES

Robson and Whitlock (1964) considered point estimation and confidence limits for the upper bound of a random variable when the bound was known to be a truncation point. However, their approach to the point estimation problem failed to produce an estimator with smaller mean squared error than the largest order statistic from a random sample. In this paper we will construct point estimators of the bounds of random variables which are substantially better estimators than the extreme order statistics for many classes of random variables, including those whose distributions are truncated at one or both ends. We will also construct confidence limits and tests of hypotheses for bounds. The main results are large sample results.

#271/32 Author: Peter Cooke