ON WEISSMAN'S METHOD OF ESTIMATING LARGE PERCENTILES

Dennis D. Boos

Inst. of Statistics Mimeo Series #1369

ABSTRACT

On Weissman's Method of Estimating Large Percentiles

Weissman (1978, JASA) suggested percentile estimators $\hat{\eta}$ based on the joint limiting distribution of the k largest order statistics. The present work extends the method to censored situations and investigates the consistency and asymptotic distribution of $\hat{\eta}$ under two different limiting schemes. Mean squared error calculations indicate that $\hat{\eta}$ is often an improvement over the usual sample percentile estimators.

1. INTRODUCTION

Weissman (1978) suggested a method of estimating the large (or small) percentiles of a distribution using extreme value theory. His approach is attractive because it is asymptotically nonparametric as long as the domain of attraction of the underlying distribution is known. Although the method is applied to estimation of specific quantiles, it may also be viewed as a smoothing technique for the tails of the quantile process. The purpose of this paper is to extend Weissman's work to censored samples, derive confidence intervals, address the consistency issue, and identify those situations where improvement over raw percentile methods is possible.

Weissman's results can be summarized as follows. Let $X_1,\ldots,X_n$ be a sample from the distribution function F and let $X_{1n} \geq X_{2n} \geq \ldots \geq X_{nn}$ be the order statistics labeled from largest to smallest. Suppose that there are sequences $a_n > 0$ and b_n such that

$$P((X_{1n}-b_n)/a_n \leq x) \to G(x) \text{ as } n\to\infty \tag{1.1}$$

for all x in the support of G. For convenience, consider only the case $G(x) = \exp(-\exp(-x))$ since results for the other two possible limiting distributions are similar. Then for k fixed, Theorem 2 of Weissman (1978) yields

$$\left(\frac{X_{1n}-b_n}{a_n},\ldots,\frac{X_{kn}-b_n}{a_n}\right) \xrightarrow{d} M_k\ , \tag{1.2}$$

where M_k is the k-dimensional extremal variate with density

$$\psi_k(x_1,\ldots,x_k) = \exp\left[-\exp(-x_k) - \sum_{i=1}^{k} x_i\right] \quad x_1 \geq x_2 \geq \cdots \geq x_k .$$

Thus, $(X_{1n},\ldots,X_{kn})$ has approximately the density $\psi_k((x_1-b_n)/a_n,\ldots,(x_k-b_n)/a_n)/a_n^k$ and a_n and b_n can be estimated by maximum likelihood to get $\hat{a}_n = \bar{X}_{kn} - X_{kn}$ and $\hat{b}_n = \hat{a}_n \ell n\, k + X_{kn}$, where $\bar{X}_{kn} = k^{-1}\sum_{i=1}^{k} X_{in}$. Let $\eta_{1-c/n}$ be the upper $(1-c/n)$th quantile of F. (I use the terms "quantile" and "percentile" interchangeably although technically the .9th quantile is the 90th percentile.) Then, if (1.1) holds, $(\eta_{1-c/n} - b_n)/a_n \to -\ell n\, c$ for each $c > 0$ and a natural quantile estimator is $\hat{\eta}_{1-c/n} = \hat{a}(-\ell n\, c) + \hat{b}_n = \hat{a}_n \ell n(k/c) + X_{kn}$.

Weissman suggested $\hat{\eta}$ for the situation where only the k largest order statistics are available or as an alternative to raw percentile methods. This investigation is aimed at the latter context and was motivated by the practical problem explained in Section 2. Section 3 extends the Weissman theory to include censoring of the extremal variate and gives confidence intervals for η_p. Section 4 discusses consistency of $\hat{\eta}$ under two different limiting schemes. Section 5 reports some Monte Carlo results which suggest values of (n,k,c) for $\hat{\eta}$ to be a true improvement over raw percentile methods. In Section 6, a numerical example illustrates the procedures. Section 7 is a summary.

2. A MOTIVATING EXAMPLE

While preparing a paper (Boos, 1981) on minimum distance estimation, it became necessary to use Monte Carlo methods to estimate the upper percentiles of a statistic which involved minimization in several

dimensions. The minimizations were fairly costly so that I allowed myself only N = 1000 replications. A packaged program (see Dickey, 1981) processed the Monte Carlo results by placing the output in 800 bins rather than printing out the N exact observations. However, in some situations I did not specify a large enough range in the program, so that a few of the largest observations were censored. Likewise, in other cases I did not trust the largest observations because certain convergence criteria had not been met. Thus, I had situations of Type I censoring and Type II censoring (trimming). The grouping effect of the bins was small compared to the standard error of percentile estimators. Unfortunately, these latter standard errors for the 95th-99th sample percentile estimators were higher than desired and no more observations could be taken. The extreme value approach turned out to be a useful alternative. See Section 6 for a specific example.

3. INFERENCE FROM THE EXTREMAL VARIATE

In this section I want to extend Weissman's (1978) theory to include censoring of the extremal variate. Suppose that $(X_1,X_2,\ldots,X_k)$ is an extremal variate with location and scale parameters μ and σ, i.e., having density $\psi_k((x_1-\mu)/\sigma,\ldots,(x_k-\mu)/\sigma)/\sigma^k$.

Type I censoring. Suppose that we can only observe those values of $(X_1,\ldots,X_k)$ which are $\leq x_0$. Then, if $X_{r-1} > x_0$ and $X_r \leq x_0$, the likelihood is

$$L = \exp\left\{-\exp\left(-\left(\frac{X_k-\mu}{\sigma}\right)\right) - \sum_{i=r}^{k}\left(\frac{X_i-\mu}{\sigma}\right) - (r-1)x_0\right\}/\sigma^{k-r+1}(r-1)! \quad \infty \geq x_0 \geq X_r \geq \ldots \geq X_k . \tag{3.1}$$

Solving the likelihood equations gives

$$\hat{\sigma} = \left(\sum_{i=r}^{k} X_i + (r-1)x_0 - kX_k \right) / (k-r+1), \quad \hat{\mu} = \hat{\sigma}\, \ell n\, k + X_k \,. \tag{3.2}$$

Type II censoring. Here, $(X_1,\ldots,X_{r-1})$ are again unavailable but r is not random. The likelihood is just the marginal density of the remaining random variables and is the same as (3.1) except that x_0 is replaced by X_r. Likewise, $\hat{\sigma}$ and $\hat{\mu}$ are given by (3.2) with x_0 replaced by X_r.

The next results apply only to Type II censoring. Weissman (1978, Theorem 3) showed that the spacings $D_i = (X_i - X_{i+1})/\sigma$ are independent exponentials with mean i^{-1} and also independent of X_k. Writing $X_i/\sigma = \sum_{j=i}^{k} D_j$, one can easily verify that $\hat{\sigma} = \sigma T_{k-r}/(k-r+1)$, where T_{k-r} is a standard gamma random variable with parameter $k - r$ which is independent of X_k. For the situation described in the introduction, Weissman's estimator of the pth quantile η_p is given by $\hat{\eta}_p = \hat{\sigma}\, \ell n(k/c) + X_k$. (Recall that $c = n(1-p)$ where n is the sample size, and $a_n = \sigma$, $b_n = \mu$.) Then using the above expression for $\hat{\sigma}$ and Weissman's (1978) expressions (2.8) and (2.9) for the mean and variance of X_k, we have

$$E\,\hat{\eta}_p = \left[\left(\frac{k-r}{k-r+1} \right) \ell n(k/c) + \gamma - \sum_{j=1}^{k-1} j^{-1} \right] \sigma + \mu$$

and (3.3)

$$\mathrm{Var}\,\hat{\eta}_p = \left[\frac{k-r}{(k-r+1)^2} \left(\ell n(k/c) \right)^2 + \frac{\pi^2}{6} - \sum_{j=1}^{k-1} j^{-2} \right] \sigma^2 \,,$$

where $\gamma = .5772 \ldots$ is Euler's constant. If $(X_{1n},\ldots,X_{kn})$ is *approximately* an extremal variate, then (2.3) with $\sigma = a_n$ and $\mu = b_n$ should give reasonable approximations to the mean and variance of $\hat{\eta}_p$. Moreover, since the distribution of $W = (\hat{\eta}_p - (-\sigma \ell n\, c + \mu))/\hat{\sigma}$ is free of parameters, confidence intervals can be constructed for $-\sigma \ell n\, c + \mu \approx \eta_p$. For the case $r = 1$, $c = 1$, $2 \leq k \leq 30$, the percentiles of W may be obtained from a table in Weissman (1978). In general I suggest Pearson curve approximations to the distribution of W (see Solomon and Stephens, 1979) since the moments can be calculated using the fact that

$$W = \ell n(k/c) + \left[\left(\frac{X_k - \mu}{\sigma}\right) + \ell n\, c\right] \cdot (k-r+1)/T_{k-r}$$

and X_k and T_{k-r} are independent. The first four moments of $(X_k - \mu)/\sigma$ are given by

$$a_1 = \gamma - S_{1k}$$

$$a_2 = S_{2\infty} - S_{2k} + a_1^2$$

$$a_3 = 2(S_{3\infty} - S_{3k}) + 3a_1 a_2 - 2a_1^3$$

$$a_4 = 6(S_{4\infty} - S_{4k}) + 4a_1 a_3 - 12a_1^2 a_2 + 3a_2^2 + 6a_1^4 ,$$

where $\gamma = .5772 \ldots$ is Euler's constant and $S_{ik} = \sum_{j=1}^{k-1} j^{-i}$. Since T_{k-r} is a gamma, $E(T_{k-r})^{-\ell} = 1/[(k-r-1)(k-r-2)\ldots(k-r-\ell)]$. At $r = 1$, $c = 1$, the skewness and kurtosis of W are $\sqrt{\beta_1} = -.409$ and $\beta_2 = 3.317$ at $k = 100$, $\sqrt{\beta_1} = -.331$ and $\beta_2 = 3.207$ at $k = 150$, and $\sqrt{\beta_1} = -.286$ and $\beta_2 = 3.154$ at $k = 200$. Clearly, the

distribution of W is approaching normality as $k \to \infty$. Since $\gamma - S_{1k} \approx -\ell n\, k$ and $S_{2\infty} - S_{2k} \approx k^{-1}$, we have for $k \gg r$

$$E\, W = \ell n(k/c) + [\gamma - S_{1k} + \ell n\, c]\left[\frac{k-r+1}{k-r-1}\right]$$

$$\approx 0$$

and

$$\mathrm{Var}\, W = \left[S_{2\infty} - S_{2k} + \frac{[\gamma - S_{1k} + \ell n c]^2}{k-r-1}\right]\left[\frac{(k-r+1)^2}{(k-r-1)(k-r-2)}\right]$$

$$\approx [1 + (\ell n(k/c))^2]/k\ .$$

Thus, for large k an approximate $(1-u) \times 100\%$ confidence interval for η_p is given by

$$\hat{\eta}_p \pm \hat{\sigma}\left(\frac{1 + (\ell n(k/c))^2}{k}\right)^{\frac{1}{2}} z_{1-u/2}, \tag{3.4}$$

where z_α is the αth quantile of the standard normal.

4. CONSISTENCY

I want to consider the limiting behavior of $\hat{\eta}_{1-c/n} = \hat{a}_n \ell n(k/c) + X_{kn}$ under Type II censoring where $\hat{a}_n = [\sum_{i=r}^{k} X_{in} + (r-1)X_{rn} - kX_{kn}]/(k-r+1)$. In terms of the original sample $X_1, \ldots, X_n$, $r-1$ observations are censored on the right. But since we have relabeled the upper k order statistics in reverse order, the censoring of $X_{1n} \geq X_{2n} \geq \ldots \geq X_{kn}$ to get $X_{rn} \geq \ldots \geq X_{kn}$ may be said to be on the left.

The first lemma is for the $k > 0$ *fixed* situation described by (1.2).

LEMMA 4.1. If (1.1) holds and $k > 0$, $c > 0$, and $r \geq 1$ are fixed, then

a) $\hat{a}_n/a_n \xrightarrow{d} T_{k-r}/(k-r+1)$

and

b) $(\hat{\eta}_{1-c/n} - \eta_{1-c/n})/a_n \xrightarrow{d} T_{k-r}\, \ell n(k/c)/(k-r+1) + m_k + \ell n\, c$,

where T_{k-r} is a gamma random variable with parameter $k-r$ which is independent of m_k having density $\exp\{-\exp(-x)-kx\}/(k-1)!$.

The proof follows directly from (1.1) since $\hat{a}_n$ and $\hat{\eta}$ are linear combinations of $(X_{rn},\ldots,X_{kn})$.

When c is fixed, $\eta_{1-c/n}$ converges as $n \to \infty$ to the supremum of the support of F (typically $+\infty$). Thus, let us define consistency to mean that

$$\hat{\eta}_{1-c/n} - \eta_{1-c/n} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty . \tag{4.1}$$

Lemma 4.1 says that (4.1) holds iff $a_n \to 0$ as $n \to \infty$. This notion of of consistency may be a little stringent since $\eta_{1-c/n}$ is harder to estimate as n gets large.

EXAMPLES. If $F = \Phi$, the standard normal, then $a_n = (2 \log n)^{-\frac{1}{2}}$ (see Galambos, 1978, p. 65) and (4.1) holds. If F is a Weibull with parameter $\frac{1}{2}$, i.e., $F(x) = 1 - \exp(-x^{\frac{1}{2}})$, $x \geq 0$, then $a_n = 2(\ell n n + 1)$ and (4.1) does not hold. If F is a chi-squared distribution with four degrees of freedom, than $a_n \to 2$ and (4.1) does not hold. If F has exponential tails, i.e., $F(x) = 1 - \exp(-(x-d_1)/d_2)$ for some constants d_1 and d_2 for all x sufficiently large, then $a_n \to d_2$ and (4.1) does not hold.

The above examples suggest that (4.1) is not adequate, especially since the convergence in (1.1) is well known to be faster for the exponential than for the normal (see Serfling, 1980, p. 90, or Galambos, 1978, Sec. 2.11). An alternative definition of consistency which is satisfied by all of the above examples is

$$\frac{\hat{\eta}_{1-c/n} - \eta_{1-c/n}}{\eta_{1-c/n}} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty .$$

A second asymptotic approach is to let k and c grow with the sample size n. Then we may apply standard L-statistic theory (see Serfling, 1980, Ch. 8) to get

LEMMA 4.2. Let $k/n \to \alpha_1$, $c/n \to 1 - p$, $r/n \to \alpha_2$ as $n \to \infty$ where $0 < 1 - \alpha_1 < 1 - \alpha_2 \leq 1$ and $0 < p < 1$. If $\int |x| dF(x) < \infty$ and $\eta_{1-\alpha_1}$ and $\eta_{1-\alpha_2}$ are unique quantiles, then

a) $$\hat{a}_n \xrightarrow{wp1} a_\infty = \frac{1}{\alpha_1 - \alpha_2} \left[\int_{1-\alpha_1}^{1-\alpha_2} \eta_t dt + \alpha_2 \eta_{1-\alpha_2} - \alpha_1 \eta_{1-\alpha_1} \right]$$

and

b) $$\hat{\eta}_{1-c/n} \xrightarrow{wp1} \eta_\infty = a_\infty \, \ell n(\alpha_1/(1-p)) + \eta_{1-\alpha_1} .$$

Moreover, if $\int [F(x)(1-F(x))]^{\frac{1}{2}} dx < \infty$ and $F'(\eta_{1-\alpha_1}) > 0$ and $F'(\eta_{1-\alpha_2}) > 0$, then

c) $$n^{\frac{1}{2}}(\hat{\eta}_{1-c/n} - \eta_\infty) \xrightarrow{d} \text{Normal } (0, \text{Var } IC(X_1)) ,$$

where $IC(\cdot)$ is the influence curve of $\hat{\eta}_{1-c/n}$ given by

$$IC(x) = [\alpha_1 - \alpha_2]^{-1} [IC_1(x) + \alpha_2 IC_2(x) - \alpha_1 IC_3(x)] + IC_3(x) ,$$

with

$$IC_1(x) = [I(x \le \eta_{1-\alpha_1}) - (1-\alpha_1)]\eta_{1-\alpha_1} + [I(x > \eta_{1-\alpha_2}) - \alpha_2]\eta_{1-\alpha_2}$$
$$+ xI(\eta_{1-\alpha_1} \le x \le \eta_{1-\alpha_2}) - \int_{1-\alpha_1}^{1-\alpha_2} \eta_t dt ,$$

$$IC_2(x) = [1 - \alpha_2 - I(x \le \eta_{1-\alpha_2})]/F'(\eta_{1-\alpha_2}) ,$$

$$IC_3(x) = [1-\alpha_1 - I(x \le \eta_{1-\alpha_1})]/F'\eta_{1-\alpha_1}) .$$

Note that $\hat{a}_n$ is just a trimmed mean plus a linear combination of X_{rn} and X_{kn}. The asymptotic variance in c) is fairly complicated but may be a more accurate approximation in finite samples than that provided by Lemma 4.1(b) (see Tables 3 and 4).

Examples. Suppose that F has exponential tails, i.e., $\eta_t = d_1 - d_2 \ell n(1-t)$ for all t sufficiently large. Then $a_\infty = d_2$ and $\eta_\infty = \eta_p$. Thus $\hat{\eta}_p = \hat{\eta}_{1-c/n}$ is strongly consistent in the usual sense under the conditions of Lemma 4.2. Suppose that $F(x) = (1+\exp(-x))^{-1}$, the standard logistic. Then F has approximately exponential tails and the asymptotic bias $\eta_\infty - \eta_p$ divided by $\sigma_p = [p(1-p)]^{\frac{1}{2}}/f(\eta_p)$ is given in Table 1.

---insert Table 1 here---

Note that $p(1-p)/f^2(\eta_p)n$ is the asymptotic variance of the sample pth quantile. Standardization by σ_p helps compare across distributions and the square of each entry multiplied by n allows one to assess the importance of the bias. For example, at $p = .95$, $\alpha_1 = .2$, and $p = .95$, $(.0025)^2 1000 = .006$ suggests that bias is not

important. However, at $p = .99$, $(.0133)^2 1000 = .18$ suggests that bias is beginning to have an affect since it is 18% as large as the asymptotic variance of the 99th sample percentile. The ratio $k/c \approx 4\text{-}5$ tends to give the smallest bias except for the two zeros which occur when $k/c = 1$ and $\hat{\eta}_p$ reduces to the usual sample pth quantile.

---insert Table 2 here---

Table 2 contains similar results for the Weibull, $F(x) = 1 - \exp(-x^{\frac{1}{2}})$, $x \geq 0$. Since the tails of F are not close to exponential, it is not surprising that the bias is generally larger than in Table 1. Here the ratio $k/c \approx 8\text{-}10$ appears to give the smallest bias. Of course, the *variance* of $\hat{\eta}$ will decrease as k gets large, so that a balance must be struck between the contribution of bias to mean squared error (MSE) and the variance contribution.

Tables 1 and 2 are based on Lemma 4.2. However, either Lemma 4.1 (actually (3.3) with $\sigma = a_n$ and $\mu = b_n$) or Lemma 4.2 can be used to calculate the bias, variance, and MSE of $\hat{\eta}$. Tables 3 and 4 illustrate these calculations. Table 3 is for $p = .99$,

---insert Table 3 here---

$\alpha_2 = 0$ (no censoring) and $F = \Phi$, the standard normal. The "truth" is supplied by Monte Carlo results explained in the next section. Clearly, the $k/n \to \alpha_1$ approach of Lemma 4.2 gives much better results than the k fixed approach of Lemma 4.1. A partial explanation may be the slow convergence of (1.2) for the normal and/or the inaccuracy of using $a_n = (2\ell n\ n)^{-\frac{1}{2}}$ and $b_n = (2\ell n\ n)^{\frac{1}{2}} - \frac{1}{2}[\log \log n + 4\pi]/[2\log n]^{\frac{1}{2}}$. Table 4 lists similar calculations for the chi-squared distribution with 4 degrees of freedom. Here, the k fixed approach does better, but

still not quite as good as the L-statistic approach. It is interesting that the asymptotic theory which produced the estimators $\hat{\eta}_p$ does not give as good an approximation in finite samples as the L-statistic approach. The last columns of Tables 3 and 4 may be compared to 13.94 and 524, the respective asymptotic variances of the 99th sample percentiles. These last comparisons illustrate the kinds of reductions in MSE that can be made by using $\hat{\eta}$. More comparisons are given in the next section.

5. MONTE CARLO RESULTS

The goal of this section is to indicate empirically some values of n, k, and $p = 1 - c/n$ which can be used in practical situations. The distributions studied were normal, t distributions with 3 and 8 degrees of freedom, and chi-squared distributions with 1, 4, and 8 degrees of freedom. These distributions were motivated by the fact that most statistics of interest are approximately normal or chi-squared distributed. A secondary motivation for their use was the availability of the location swindle (see Gross, 1973) to reduce Monte Carlo variance. Sample sizes studied were $n = 50$, 100, and 500 and the Monte Carlo replication size was $N = 5000$. The estimated standard errors for the bias calculations in Table 5 are in the range $\pm$ .01 to $\pm$ .09. For MSE the maximum relative standard error (s.e./nMSE) for each column is given in parentheses at the bottom of each column. Overall these relative standard errors ranged from .010 to .024. All random variables were generated from normal deviates produced by the Super-Duper generator of Marsaglia *et al.* (1976).

For comparison purposes I computed the four percentile estimators listed as options in SAS routine UNIVARIATE (see Chilko (1979), p. 429) and a fifth estimator (called AV) motivated by the usual definition of the sample median. If $X_{(1)} \leq \ldots \leq X_{(n)}$ are the order statistics of the sample in ascending order and $[\cdot]$ is the greatest integer function, this last estimator is

$$AV = \tfrac{1}{2}(X_{(np)} + X_{(np+1)}) \quad \text{if } np \text{ is an integer}$$

$$= X_{([np]+1)} \quad \text{otherwise .}$$

The best of the SAS methods was the default option on UNIVARIATE which I shall call LINT for linear interpolation. LINT is actually obtained by linear inverse interpolation of the empirical distribution function and defined by

$$LINT = (1-\varepsilon)X_{([np])} + \varepsilon X_{([np]+1)} ,$$

where $\varepsilon = np - [np]$. AV performed very well in terms of bias. However, when $np \neq [np]$, LINT outperformed AV in terms of MSE. When $np = [np]$, no clear winner emerged. Other "raw percentile" methods such as k-point inverse interpolation or methods based on nonparametric density estimators (see Azzalini, 1981) may give small improvements over LINT and AV.

--insert Tables 5 and 6 here--

Table 5 lists Monte Carlo estimates of $n^{\frac{1}{2}}$ bias and nMSE for $\hat{\eta}$ in the standard normal. Small amounts of censoring, say # censored/k $\leq$ .1, have only a minor effect on the results. For

example, in the first two rows of Table 5 under $p = .95$, we see that one censored observation caused an inflation of nMSE from 3.31 to 3.57. However, five observations censored (row 3) inflated the nMSE to 6.43. When k gets too large, bias makes a significant contribution to the MSE. At $p = .975$ and $n = 100$ the $n^{\frac{1}{2}}$bias values for $k = 10$, 20, and 50 (none censored) are -.69, -.05, and 3.86, respectively. The associated n·variance values 4.83, 5.33, and 6.40 are relatively stable in comparison.

Table 6 is a basic reduction of Table 5 and lists the ratio of estimates of MSE of $\hat{\eta}_p$ to that of LINT (the censored cases have been deleted). Thus, if an entry in Table 6 is < 1, then $\hat{\eta}_p$ is outperforming LINT in terms of MSE. From Table 6 we see that when $k/c \leq 1$, $\hat{\eta}_p$ is clearly beaten by LINT, and this is also typically true for $k/n \geq .4$. The optimal ratio k/c for $\hat{\eta}_p$ appears to be about 4.

---insert Tables 7 and 8 here---

Table 7 has the same structure as Table 6, except that the data is from a t distribution with 8 degrees of freedom. The pattern of results is very similar to that of Table 6 as one might expect since t_8 is fairly close to a normal distribution. However, Table 8 based on the t distribution with 3 degrees of freedom is markedly different. LINT outperforms $\hat{n}_p$ for most of the entries. It was also discovered, though not listed in Table 8, that censored versions of $\hat{\eta}_p$ often outperformed uncensored versions. It is easy to verify (see Galambos, 1978, Theorem 2.4.3 (iii)) that (1.1) does not hold for any t distribution.

---insert Tables 9-11 here---

Tables 9-11 are for chi-squared distributions with 1, 4, and 8 degrees of freedom, respectively. In these tables we can see that k/c is unimportant, but rather choosing k large tends to optimize the performance of $\hat{\eta}_p$. In Tables 9 and 10 bias begins to play a role in the MSE at $n = 500$, so that $k/n = .2$ is preferred to $k/n = .4$. In Table 11 bias is important even at $n = 100$, so that $k/n = .2$ is preferred for $n \geq 100$ and $p \geq .95$.

In general, the new estimators can be expected to do well in distributions with approximately exponential tails using the rough rule of thumb $k/n = .2$. Of course for n large enough, bias will dominate and raw percentile methods will be optimal. In approximately normal distributions, more attention needs to be given the ratio k/c with the rough rule of thumb $k/c \approx 4$ suggested. In all situations raw percentile methods are preferable to $\hat{\eta}$ for $k/c \leq 1$.

6. NUMERICAL EXAMPLE

The basic situation has been described in Section 2. The specific example considered here is Monte Carlo estimation of the percentiles of $T_n = \min_\mu d(F_n, F_\mu)$, where F_n is the empirical distribution function of $n = 50$ standard logistic random variables, $F_\mu(x) = [1 + \exp(-(x-\mu))]^{-1}$, and $d(\cdot,\cdot)$ is the Anderson-Darling distance (see Boos, 1981, for details). In this particular case the percentiles of T_n have been tabulated in Stephens (1979, Table 1) and appear as the first row of Table 12. The next four rows of Table 12 are the new estimators $\hat{\eta}_p$ using $k = 200$, 150, 100, and 50.

Five observations were outside the reporting gange [0,2] and thus were unavailable. However, I used $\hat{\eta}_p$ based on Type II censoring because the gap between the 995th observation and the upper limit 2 was larger than expected, which suggested that I would have trimmed the five largest observations even if they were available. The last row of Table 12 is the raw percentile estimator LINT which in these five cases is just $X_{(Np)}$ since Np is an integer. At $p = .9$, it would be hard to choose between LINT and $\hat{\eta}$ for $k = 200$ or $k = 150$. At $p = .95$, all three $\hat{\eta}_p$ with $k/c > 1$ are an improvement over LINT. At $p = .975$ and $p = .995$ all four $\hat{\eta}_p$ are better than LINT. And at $p = .99$ all but $k = 200$ are an improvement over LINT. It appears that in the range $p = .95$ to $p = .99$, $k/c \approx 4$-6 is best.

For $k = 50$, $r = 6$, and $c = 10$, $EW = -1.6738$, $Var\ W = .0894$, $\sqrt{\beta_1} = -.5967$, and $\beta_2 = 3.7061$ (see Section 3). From a table of Pearson Curve percentiles (Bouver and Bargmann, 1974) the 5th percentile of W is -1.784 and the 95th percentile is 1.466. Thus, using $\hat{a}_n = .271$, a 90% confidence interval for $\eta_{.99}$ is given by $(1.498 - \hat{a}_n 1.466,\ 1.498 + \hat{a}_n 1.784) = (1.396, 1.660)$. The approximate method of (3.4) yields (1.378, 1.617). The usual distribution-free methods based on order statistics (see Serfling, 1980, p. 103) yields $(X_{(985)}, X_{(995)}) = (1.349, 1.801)$.

7. SUMMARY

The percentile estimators $\hat{\eta}_p$ derived from extreme value theory are biased but often have lower MSE than raw percentile estimators. If the shape of the underlying distribution F is known, then a fairly accurate estimate of the MSE of $\hat{\eta}_p$ can be obtained from L-statistic theory (Theorem 4.2). Simpler but less accurate calculations may be based on Theorem 4.1 or (3.3). Such calculations along with Monte Carlo results indicate that moderate reductions (20-35%) in MSE are possible for normal data using the rule of thumb $k/n(1-p) \approx 4$. Larger reductions may be expected in distributions with approximately exponential tails. For example, in the chi-squared distributions with 1, 4, and 8 degrees of freedom, reductions in MSE of 20-80% were obtained. If a reasonable guess of the shape of F is not possible, then a natural approach would be to let the sample itself select k. This idea will be pursued in a later report.

TABLE 1. Standardized Asymptotic Bias for the Logistic

α_1	$\alpha_2 \backslash p$	.9	.95	.975	.99	.995
.2	0	-.0113	-.0025	.0067	.0133	.0147
.1	0	0	-.0037	-.0009	.0028	.0042
.05	0	.0108	0	-.0013	.00004	.0009

TABLE 2. Standardized Asymptotic Bias for the Weibull*

α_1	$\alpha_2 \backslash p$	.9	.95	.975	.99	.995
.2	0	-.0656	.0326	-.0036	-.0326	-.0417
.1	0	0	.0347	.0185	-.0076	-.0120
.05	0	-.1351	0	.0197	.0069	-.0047

*$F(x) = 1 - \exp(-x^{\frac{1}{2}})$, $x \geq 0$.

TABLE 3. Estimates of Bias, Variance, and MSE in the Standard Normal at p = .99.

			$n^{1/2}$ Bias			n Variance			n MSE		
n	k	α_1=k/n	From Lemma 4.1	From Lemma 4.2	Monte Carlo	From Lemma 4.1	From Lemma 4.2	Monte Carlo	From Lemma 4.1	From Lemma 4.2	Monte Carlo
100	10	.1	-.19	.45	-.50	6.32	10.06	9.29	6.36	10.26	9.54
100	20	.2	-.01	1.87	1.13	5.18	10.05	10.01	5.18	13.57	11.29
100	50	.5	.17	7.95	7.33	3.48	11.40	11.07	3.51	74.60	64.82
500	10	.02	2.67	-.40	-.53	5.97	9.76	9.97	13.10	9.93	10.25
500	20	.04	2.51	-.36	-.73	5.73	9.74	9.67	12.03	9.87	10.20
500	50	.1	2.56	1.01	.68	4.99	10.06	10.19	11.55	11.08	10.65
500	100	.2	2.63	4.19	3.89	3.98	10.05	10.26	10.90	27.62	25.38
500	200	.4	2.69	12.42	12.23	2.93	10.64	10.80	10.16	164.9	160.3

Note: The asymptotic variance of the 99th sample percentile (times n) is 13.94.

TABLE 4. Estimates of Bias, Variance, and MSE in the χ_4^2 at p = .99 .

			$n^{1/2}$ Bias			n Variance			n MSE		
n	k	α_1=k/n	From Lemma 4.1	From Lemma 4.2	Monte Carlo	From Lemma 4.1	From Lemma 4.2	Monte Carlo	From Lemma 4.1	From Lemma 4.2	Monte Carlo
100	10	.1	-4.06	.50	-3.93	298	336	312	314	337	328
100	20	.2	-2.82	2.04	-1.13	244	277	265	252	281	266
100	50	.5	-1.54	8.25	6.18	164	199	193	166	267	231
500	10	.02	.94	-.44	-.97	363	383	378	364	383	379
500	20	.04	-.33	-.39	-2.32	349	378	364	349	379	369
500	50	.1	.08	1.11	-.50	304	336	323	304	338	323
500	100	.2	.62	4.55	3.20	242	277	265	242	298	275
500	200	.4	1.07	13.13	12.34	178	216	210	179	388	362

Note: The asymptotic variance of the 99th sample percentile (times n) is 524.

TABLE 5. Monte Carlo Estimates of Bias and MSE for Percentile Estimators in the Standard Normal

k	No. Censored	p = .9		p = .95		p = .975		p = .99	
		$n^{1/2}$bias	nMSE	$n^{1/2}$bias	nMSE	$n^{1/2}$bias	nMSE	$n^{1/2}$bias	nMSE
				n = 50					
10	0	-.44	2.22	-.56	3.31	-.34	5.08	.31	9.21
10	1	-.39	2.25	-.45	3.57	-.17	5.89	.56	11.3
10	5	-.39	2.93	-.45	6.43	-.17	12.4	.56	25.0
20	0	-.50	2.22	.22	3.44	1.28	7.28	3.04	19.1
20	5	.10	2.87	1.12	6.85	2.48	15.9	4.63	38.8
20	10	.50	5.18	1.71	13.4	3.27	29.4	5.7	65.9
LINT		-.45	2.95	-.69	4.17	-1.14	6.29	-1.92	10.2
AV		-.05	2.77	-.12	4.33	-.73	6.32	-.51	10.7
			(.014)		(.021)		(.021)		(.024)
				n = 100					
10	0	.26	3.05	-.46	3.37	-.69	5.31	-.50	9.54
10	1	.26	3.05	-.41	3.45	-.60	5.70	-.35	10.9
10	5	.26	3.05	-.47	4.35	-.73	9.41	-.56	21.0
20	0	-.58	2.35	-.55	3.41	-.05	5.33	1.13	11.3
20	5	-.34	2.37	-.07	4.29	.68	8.62	2.18	20.8
20	10	-.22	2.91	.16	7.03	1.03	15.5	2.68	36.7
50	0	-.15	2.14	1.62	6.41	3.86	21.3	7.33	64.8
50	5	.49	2.73	2.53	11.0	5.04	33.3	8.88	92.6
50	10	1.07	4.16	3.35	17.1	6.12	47.5	10.3	123
LINT		-.30	2.92	-.54	4.39	-.89	6.56	-1.85	13.0
AV		-.02	2.83	-.07	4.20	-.17	6.96	-.05	11.2
			(.015)		(.019)		(.022)		(.023)
				n = 500					
10	0	5.78	57.4	2.79	22.1	.87	10.1	-.53	10.2
10	1	5.71	59.1	2.75	22.7	.87	10.1	-.50	10.6
10	5	6.25	79.8	3.06	29.1	.94	10.6	-.74	13.4
20	0	2.90	17.3	.72	6.19	-.39	5.58	-.73	10.2
20	5	2.64	17.4	.66	6.23	-.25	5.80	-.33	12.8
20	10	2.66	20.0	.66	6.37	-.26	6.48	-.36	18.7
50	0	.13	2.97	-.80	3.78	-.64	5.46	.68	10.6
50	5	.13	2.97	-.59	3.61	-.23	5.69	1.37	13.8
50	10	.13	2.97	-.44	3.62	.08	6.44	1.88	17.8
100	0	-1.17	3.37	-.74	3.68	.77	6.03	3.89	25.4
100	10	-.86	2.81	-.12	3.55	1.70	9.33	5.23	39.8
100	20	-.60	1.55	.39	4.23	2.46	13.8	6.33	55.2
LINT		-.12	2.92	-.20	4.50	-.38	7.01	-.83	13.4
AV		.00	2.91	.02	4.47	.00	7.21	.00	13.1
			(.017)		(.019)		(.019)		(.020)

Note: In parentheses are the maximum estimated relative standard errors for each column.

TABLE 6. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for Normal Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	.75	4	.79	8	.81	20	.90
50	20	4	.75	8	.83	16	1.16	40	1.87
100	10	1	1.04	2	.77	4	.81	10	.74
100	20	2	.80	4	.78	8	.81	20	.87
100	50	5	.73	10	1.46	20	3.24	50	5.00
500	10	.2	19.69	.4	4.92	.8	1.44	2	.76
500	20	.4	5.94	.8	1.38	1.6	.80	4	.76
500	50	1	1.02	2	.84	4	.78	10	.79
500	100	2	1.15	4	.82	8	.86	20	1.89
500	200	4	.99	8	1.63	16	5.81	40	11.92

TABLE 7. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for t_8 Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	.78	4	.86	8	.81	20	.74
50	20	4	.77	8	.82	16	.77	40	.71
100	10	1	1.08	2	.80	4	.86	10	.75
100	20	2	.78	4	.78	8	.76	20	.57
100	50	5	.75	10	1.02	20	1.30	50	1.17
500	10	.2	20.4	.4	5.23	.8	1.48	2	.83
500	20	.4	4.94	.8	1.35	1.6	.82	4	.80
500	50	1	1.01	2	.76	4	.78	10	.67
500	100	2	.82	4	.75	8	.73	20	.59
500	200	4	.78	8	1.14	16	1.83	40	1.88

TABLE 8. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for t_3 Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	1.29	4	1.38	8	.99	20	.60
50	20	4	1.23	8	1.01	16	.61	40	.36
100	10	1	1.17	2	1.53	4	1.58	10	1.00
100	20	2	1.40	4	1.46	8	1.10	20	.62
100	50	5	1.31	10	.97	20	.56	50	.28
500	10	.2	190	.4	21.6	.8	1.74	2	1.67
500	20	.4	26.7	.8	1.62	1.6	1.56	4	1.58
500	50	1	1.04	2	1.90	4	1.71	10	.87
500	100	2	1.97	4	1.83	8	1.04	20	.60
500	200	4	2.02	8	1.32	16	.59	40	.46

TABLE 9. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for χ_1^2 Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	.79	4	.85	8	.81	20	.78
50	20	4	.74	8	.75	16	.67	40	.62
100	10	1	1.15	2	.79	4	.84	10	.74
100	20	2	.78	4	.76	8	.77	20	.60
100	50	5	.71	10	.67	20	.68	50	.56
500	10	.2	24.39	.4	5.52	.8	1.52	2	.77
500	20	.4	5.77	.8	1.42	1.6	.81	4	.75
500	50	1	1.03	2	.75	4	.76	10	.64
500	100	2	.79	4	.73	8	.68	20	.55
500	200	4	.75	8	.69	16	.79	40	.78

TABLE 10. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for χ_4^2 Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	.76	4	.81	8	.80	20	.77
50	20	4	.74	8	.76	16	.68	40	.61
100	10	1	1.10	2	.75	4	.82	10	.74
100	20	2	.73	4	.74	8	.76	20	.60
100	50	5	.71	10	.67	20	.66	50	.53
500	10	.2	19.80	.4	5.15	.8	1.50	2	.77
500	20	.4	5.07	.8	1.38	1.6	.78	4	.75
500	50	1	1.01	2	.73	4	.75	10	.66
500	100	2	.74	4	.72	8	.69	20	.56
500	200	4	.72	8	.71	16	.76	40	.73

TABLE 11. Ratio of MSE of $\hat{\eta}_p$ to MSE of LINT for χ_8^2 Data

		p = .9		p = .95		p = .975		p = .99	
n	k	k/c	Ratio	k/c	Ratio	k/c	Ratio	k/c	Ratio
50	10	2	.74	4	.83	8	.81	20	.78
50	20	4	.73	8	.78	16	.73	40	.71
100	10	1	1.09	2	.77	4	.82	10	.74
100	20	2	.72	4	.75	8	.78	20	.65
100	50	5	.70	10	.78	20	.94	50	.95
500	10	.2	19.2	.4	5.05	.8	1.48	2	.76
500	20	.4	5.10	.8	1.39	1.6	.79	4	.75
500	50	1	1.02	2	.75	4	.75	10	.67
500	100	2	.79	4	.74	8	.71	20	.68
500	200	4	.76	8	.81	16	1.22	40	1.62

TABLE 12. Estimates of the Percentiles of T_n

p	.9	.95	.975	.99	.995	$\hat{a}_n$
True Percentiles	.854	1.043	1.238	1.502	1.707	
k = 200	.860	1.044	1.227	1.470	1.654	.265
k = 150	.858	1.047	1.235	1.484	1.672	.272
k = 100	.851	1.045	1.240	1.497	1.692	.281
k = 50	.873	1.061	1.249	1.498	1.686	.271
LINT	.849	1.057	1.206	1.471	1.801	

REFERENCES

Azzalini, A. (1981), "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68, 326-328.

Boos, D. D. (1981), "Minimum Anderson-Darling Estimation," preprint.

Bouver, H., and Bargmann, R. E. (1974), "Tables of the Standardized Percentage Points of the Pearson System of Curves in Terms of β_1 and β_2," Technical Report No. 107, Department of Statistics and Computer Science, University of Georgia, Athens.

Chilko, D. M. (1979), "Univariate Procedure," in *SAS User's Guide, 1979 Edition*, eds. J. T. Helwig and K. A. Council, SAS Institute, Inc., 427-432.

Dickey, D. A. (1981), "Histograms, Percentiles, and Moments," *The American Statistician*, 35, 164-165.

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, New York: John Wiley.

Gross, A. M. (1973), "A Monte Carlo Swindle for Estimators of Location," *Applied Statistics*, 22, 347-353.

Marsaglia, G., Ananthanarayanan, K., and Paul, N. J. (1976), "Improvements on Fast Methods for Generating Normal Random Variables," *Information Processing Letters*, 5, 27-30.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Solomon, H., and Stephens, M. A. (1978), "Approximations to Density Functions using Pearson Curves," *Journal of the American Statistical Association*, 73, 153-160.

Stephens, M. A. (1979), "Tests of Fit for the Logistic Distribution Based on the Empirical Distribution Function," *Biometrika*, 66, 591-595.

Weissman, I. (1978), "Estimation of Parameters and Large Quantiles Based on the k Largest Observations, *Journal of the American Statistical Association*, 73, 812-815.