J. Statist. Comput. Simul., Vol. 37, pp. 127–149
Reprints available directly from the publisher
Photocopying permitted by license only

ROBUST CONFIDENCE BOUNDS FOR EXTREME UPPER QUANTILES

LEO BREIMAN, CHARLES J. STONE AND CHARLES KOOPERBERG

Department of Statistics, University of California, Berkeley, CA 94720, USA

(Received 4 January 1990; in final form 24 May 1990)

Four related methods are discussed for obtaining robust confidence bounds for extreme upper quantiles of the unknown distribution of a positive random variable. These methods are designed to work when the upper tail of the distribution is neither too heavy nor too light in comparison to the exponential distribution. An extensive simulated study is described, which compares the performance of nominal 90% upper confidence bounds corresponding to the four methods over a wide variety of distributions having light to heavy upper tails, ranging from a half-normal distribution to a heavy-tailed lognormal distribution.

KEY WORDS: Quantile estimation, exponential-tail model, quadratic-tail model, power transformation, adaption, tail heaviness.

1. INTRODUCTION

Often, in applications, we want to estimate extreme quantiles from sample data. For example, we might have 30 years of annual high-water levels on a river and want to estimate the 100-year flood level $y_{0.01}$, defined by the requirement that the probability of annual high-water level exceeding $y_{0.01}$ should be 0.01. For applications to air quality data, see Crager (1982).

Frequently, confidence bounds or confidence intervals are desired in addition to or instead of point estimates. In this paper, we focus on robust 90% upper confidence bounds for the upper pth quantile of a distribution based on a random sample of size n from that distribution when np is so small that no order statistic can serve as such a confidence bound.

Let Y be a random variable (whose distribution function is continuous) and let y_p denote the *upper* pth quantile of Y for $0<p<1$, so that $P(Y \geqq y_p)=p$. Let n be a positive integer and let $Y_1, \ldots, Y_n$ be a random sample of size n from the distribution of Y. Then $Y_1, \ldots, Y_n$ are independent and identically distributed random variables. Let $Y_{(1)}, \ldots, Y_{(n)}$ denote the corresponding upper order statistics, obtained by writing $Y_1, \ldots, Y_n$ in decreasing order: $Y_{(1)} \geqq \cdots \geqq Y_{(n)}$.

Let U be a statistic based on the random sample, which is thought of as an upper confidence bound for y_p. The corresponding coverage probability is $P(y_p \leqq U)$. Let $0<c<1$. If U is derived as a $100c\%$ upper confidence bound for y_p by making various assumptions and approximations, then we refer to c as the nominal coverage probability of U and to $P(y_p \leqq U)$ as its actual coverage probability.

Consider, for example, the maximum value $Y_{(1)}$ in the sample as an upper confidence bound for y_p. Since

$$P(Y_{(1)}<y_p)=P(Y_1<y_p,\ldots,Y_n<y_p)=[P(Y_1<y_p)]^n=(1-p)^n,$$

we see that the actual coverage probability of $Y_{(1)}$ is given by

$$P(Y_{(1)}\geqq y_p)=1-(1-p)^n.$$

In particular, $Y_{(1)}$ is a 90% upper confidence bound for y_p if and only if

$$p=1-(0.1)^{1/n}=1-\exp(-\log(10)/n)\approx\frac{\log(10)}{n}\approx\frac{2.3}{n},\qquad n\gg 1.$$

Thus (for $n\geqq 8$) if $p\leqq 2/n$, there is no order statistic that serves as a 90% upper confidence bound for y_p.

When $p\leqq 2/n$, we can obtain a nominal 90% upper confidence bound for y_p in a standard manner by assuming a Weibull, gamma, lognormal or other classical parametric model for the distribution of Y. However, if our assumption is even mildly inaccurate in a given application, the actual coverage probability can differ substantially from 0.9. In other words, the actual coverage probability of the nominal 90% upper confidence bound for an extreme quantile is very sensitive to model departures. A better approach is to obtain a nominal 90% upper confidence bound for y_p by first fitting a parametric model to the upper tail of the data; that is, to the m upper order statistics $Y_{(1)},\ldots,Y_{(m)}$, where $m<n$. This is the approach that will be followed here.

In Section 2 we describe several such methods for obtaining confidence bounds for extreme quantiles. The well-known exponential-tail method is described in Section 2.1. The quadratic-tail method, briefly described in Section 2.2, was introduced in Breiman *et al.* (1981), which is a precursor to the present work. Further details for this method are given in Sections 6 and 7. A preliminary power transformation is discussed in Section 2.3. Some computer-based refinements of the exponential-tail and quadratic-tail methods are described in Section 3. The results of a reasonably extensive simulation study of four methods of obtaining 90% upper confidence bounds for extreme quantiles (exponential-tail and quadratic-tail, with and without the preliminary power transformation), are presented in graphical form in Section 4 and various conclusions are drawn in Section 5.

When this work was started in the late 1970s in connection with air pollution studies, the prevailing approach in practice was to fit a standard global model, usually Weibull or lognormal, to the data and to obtain confidence bounds and confidence intervals for extreme quantiles by the usual large-sample parametric approach. It was and is our contention that this approach is unrealistic because it ignores modelling errors that can lead to substantial bias in confidence bounds for extreme quantiles. The obvious alternative is "to let the tails of the data speak for themselves" (DuMouchel and Olshen, 1975), but this would seem to suggest using a much smaller proportion of the data than we have found to be desirable.

We are unaware of previous work on confidence bounds for extreme quantiles, other than Breiman *et al.* (1981) and the follow-on work of Crager (1982). On the other hand, there have been many studies of exponential-tail and related methods of estimation for tail probabilities and extreme quantiles. These (mainly theoretical) studies have focused on methods that are appropriate when the tail is (I) in the domain of attraction of some extreme-value distribution; (II) approximately algebraically decreasing; or (III) approximately exponential decreasing. These three conditions are very closely related. For example, the upper tail of Y is approximately algebraically decreasing if and only if that of $\log(Y)$ is approximately exponentially decreasing; thus methods appropriate to approximately exponentially decreasing tails may be applied to data having approximately algebraically decreasing tails by first applying the logarithmic transformation. In category (I) are Maritz and Munro (1967), Pickands (1975), Weissman (1978), Boos (1984), Davis and Resnick (1984), and Smith (1987); in category (II) are Hill (1975), DuMouchel and Olshen (1975), DuMouchel (1983), Hall and Welsh (1985), and Csörgő *et al.* (1985); and in category (III) are Breiman *et al.* (1978, 1979 and 1981), and Crager (1982). See Smith (1987) for a recent and thorough review of this literature.

2. CONFIDENCE BOUNDS

2.1. Exponential-tail Method

Let $0<p_0<1$. Consider the *exponential-tail model*, in which there is an $\alpha>0$ such that

$$P(Y \geqq y \mid Y \geqq y_{p_0}) = \exp(-(y-y_{p_0})/\alpha) \tag{2.1}$$

for $y \geqq y_{p_0}$ or, equivalently, in which y_p is a linear function of $\log(1/p)$ as p ranges over $(0, p_0]$. Let m be a positive integer with $m/n \leqq p_0$. Then

$$y_p = y_{m/n} + \alpha \log\left(\frac{m}{np}\right)$$

for $0<p \leqq p_0$. It is reasonable to estimate $y_{m/n}$ by $Y_{(m)}$. The maximum-likelihood estimate of α based on the data $Y_{(i)}$, $1 \leqq i \leqq m$, is given by

$$\hat{\alpha} = \frac{1}{m-1} \sum [Y_{(i)} - Y_{(m)}]. \tag{2.2}$$

(Summations are over $i \in \{1, \ldots, m-1\}$ unless otherwise noted). The corresponding estimate of the upper pth quantile, for $0<p<p_0$, is given by

$$\hat{y}_p = Y_{(m)} + \hat{\alpha} \log\left(\frac{m}{np}\right). \tag{2.3}$$

Suppose that Y has a (two-parameter) exponential distribution or, equivalently, that y_p is a linear function of $\log(1/p)$ as p ranges over $(0, 1)$. Let $\hat{\alpha}$ and $\hat{y}_p$ be given by (2.2) and (2.3), respectively, and set $SE(\hat{y}_p)=\hat{\alpha}$. Then $t=t_{p,c,n,m}$ can be obtained numerically from the incomplete beta function so that $\hat{y}_p+tSE(\hat{y}_p)$ is an exact $100c\%$ upper confidence bound for y_p.

Consider, in general $U(t)=\hat{y}_p+tSE(\hat{y}_p)$ as a nominal $100c\%$ upper confidence bound for y_p, which we refer to as being obtained by the *exponential-tail* (ET) *method*. Under the more general exponential-tail model, the actual coverage probability of $U(t)$ is close to its nominal coverage probability if $P(Y_{(m)} \geqq y_{p_0}) \approx 1$. It is more realistic, however, to consider the exponential-tail model as being a reasonably accurate approximation. Hopefully, when p is not too small, the actual coverage probability of $U(t)$ will be close to its nominal coverage probability; but when p is extremely small, the nominal and actual coverage probabilities may differ considerably because of the error of approximation.

2.2. *Quadratic-tail Method*

Let $0<p_0<1$. In the corresponding exponential-tail model, y_p is a linear function of $\log(1/p)$ as p ranges over $(0, p_0]$. In order to obtain a more accurate approximation, we consider the *quadratic-tail model*, in which y_p is assumed to be a quadratic function of $\log(1/p)$ as p ranges over $(0, p_0]$. Let m be a positive integer with $m/n \leqq p_0$. Then

$$y_p = y_{m/n} + \alpha[\log(1/p) - \log(n/m)] + \frac{\beta}{2}[\log^2(1/p) - \log^2(n/m)],$$

$$0<p<p_0. \qquad (2.4)$$

Here α and β are unknown parameters with $\alpha \geqq 0$. The quadratic-tail model can be exactly valid if $\beta>0$ or if $\alpha>0$ and $\beta=0$. It cannot be exactly valid when $\beta<0$, for in that case, the quadratic function in (2.4) tends to $-\infty$ as $p \to 0$. Even when $\beta<0$, however, the quadratic-tail model can provide a good approximation to y_p for values of p that are not exceedingly small. The quadratic-tail model was proposed in Breiman *et al.* (1981), but it can be motivated in terms of Proposition 9.1 of Smith (1987).

Given m and p, set

$$L=\log(1/p)-\log(n/m) \quad \text{and} \quad M=\tfrac{1}{2}[\log^2(1/p)-\log^2(n/m)]. \qquad (2.5)$$

It follows from (2.4) that $y_p = y_{m/n} + \tau$, where $\tau = L\alpha + M\beta$. Corresponding to an estimate $\hat{\tau}$ of τ is the estimate $\hat{y}_p = Y_{(m)} + \hat{\tau}$ of y_p.

More generally, let L and M be arbitrary constants and set $\tau = L\alpha + M\beta$. Consider an estimate $\hat{\tau}$ of τ of the form

$$\hat{\tau} = \sum w_i [Y_{(i)} - Y_{(i+1)}]. \qquad (2.6)$$

It is shown in Sections 6 and 7 that

$$\text{var}(\hat{\tau}) = c_1\alpha^2 + c_2\alpha\beta + c_3\beta^2, \tag{2.7}$$

where c_1, c_2 and c_3 are given explicitly in terms of L, M and the weights $w_1, \dots, w_{m-1}$. If the exponential-tail model is reasonably accurate and, in particular, if $\beta \approx 0$, then

$$\text{var}(\hat{\tau}) = c_1\alpha^2. \tag{2.8}$$

In light of (2.8) it is reasonable to choose the weights to minimize c_1 subject to the constraint that $\hat{\tau}$ be unbiased; that is, that, for all values of L and M,

$$E\hat{\tau} = L\alpha + M\beta. \tag{2.9}$$

It is shown in Section 6 that this minimization problem has a unique solution, which is given explicitly. (The quadratic model should be thought of as an approximation. In Section 6 the error of approximation is ignored. Thus (2.7) and the solution to the indicated minimization problem should be thought of as informal approximations. Using the minimization problem stemming from (2.8) to choose the weights is reasonable since it is not possible to choose the weights to minimize $\text{var}(\hat{\tau})$ for all values of α and β.)

In particular, by choosing $L=1$ and $M=0$, we obtain an unbiased estimate of α having the form

$$\hat{\alpha} = \sum w_{1i}[Y_{(i)} - Y_{(i+1)}];$$

by choosing $L=0$ and $M=1$, we obtain an unbiased estimate of β having the form

$$\hat{\beta} = \sum w_{2i}[Y_{(i)} - Y_{(i+1)}].$$

As shown in Section 6, for arbitrary values of L and M the unbiased estimate of $\tau = L\alpha + M\beta$ for which c_1 is minimized is given by $\hat{\tau} = L\hat{\alpha} + M\hat{\beta}$; thus the corresponding quantile estimate is given, for $0 < p \leqq p_0$, by

$$\hat{y}_p = Y_{(m)} + L\hat{\alpha} + M\hat{\beta}. \tag{2.10}$$

It is shown in Sections 6 and 7 that

$$\text{var}(\hat{y}_p) = C_1\alpha^2 + C_2\alpha\beta + C_3\beta^2, \tag{2.11}$$

where C_1, C_2 and C_3 are given explicitly in terms of n, m, L, and M. The corresponding standard error is given by

$$SE(\hat{y}_p) = (C_1\hat{\alpha}^2 + C_2\hat{\alpha}\hat{\beta} + C_3\hat{\beta}^2)^{1/2}.$$

Set $U(t) = \hat{y}_p + tSE(\hat{y}_p)$. Presumably, under suitable conditions, $(\hat{y}_p - y_p)/SE(\hat{y}_p)$

has approximately the standard normal distribution, in which case the actual coverage probability of $U(z_{1-c})$ as a nominal $100c\%$ upper confidence bound for y_p is close to its nominal coverage probability c. (Here $\Phi(z_{1-c})=c$, where Φ is the standard normal distribution function.) We refer to $U(t)$ as having been obtained by the *quadratic-tail* (QT) *method*. Our computer-based experience, however, is that, when n is not extremely large, $p \leqq 2/n$ and $t=z_{0.9} \doteq 1.282$, then the actual coverage probabilities of the QT 90% upper confidence bound for y_p is considerably less than 0.9 even when Y has an exponential distribution. In Section 3.1 we will describe an alternative way of choosing t that leads to actual coverage probabilities that are much closer to their nominal counterparts.

2.3. *Preliminary Power Transformation*

Suppose that Y is a positive random variable. The approximation errors of the exponential-tail and quadratic-tail models can be substantially reduced by a preliminary power transformation (see Weinstein, 1973). Given a positive constant γ, set $W=Y^\gamma$ and $W_i=Y_i^\gamma$ for $1 \leqq i \leqq n$. The upper pth quantile of W is given by $w_p=y_p^\gamma$. We would like to choose γ so that the conditional distribution of $W-w_{p_0}$ given that $W \geqq w_{p_0}$ is approximately exponential. Let $\hat{w}_p$ be an estimator of w_p based on the random sample $W_1,\ldots,W_n$. By applying the inverse power transformation, we obtain the estimate $\hat{y}_p=\hat{w}_p^{1/\gamma}$ of y_p based on the original random sample. Similarly, let $\hat{w}_p+tSE(\hat{w}_p)$ be a nominal $100c\%$ upper confidence bound for w_p. Then $[\hat{w}_p+tSE(\hat{w}_p)]^{1/\gamma}$ is a nominal $100c\%$ upper confidence bound for y_p. In practice, the power transformation must be determined from the sample data; we denote the corresponding parameter by $\hat{\gamma}$. We are led to $\hat{y}_p=\hat{w}_p^{1/\hat{\gamma}}$ as an estimate of y_p and to $[\hat{w}_p+tSE(\hat{w}_p)]^{1/\hat{\gamma}}$ as a nominal $100c\%$ upper confidence bound for y_p.

Let $p_0=m/n$, where $2 \leqq m \leqq n$. A reasonable way to choose $\hat{\gamma}$ is by maximum likelihood based on the two-parameter exponential family with the dependence of $\hat{y}_{p_0}$ on the sample data being ignored. We are lead to choosing $\hat{\gamma}$ to maximize the function

$$g(\gamma)=(m-1)\log(\gamma)+(m-1)\log(\textstyle\sum(Y_{(i)}^\gamma-Y_{(m)}^\gamma)$$
$$+(\gamma-1)\textstyle\sum\log(Y_{(i)})+(m-1)\log(m-1)-(m-1).$$

Observe that

$$g'(\gamma)=\frac{m-1}{\gamma}\left(1-\frac{\sum(Y_{(i)}^\gamma\log(Y_{(i)}^\gamma)-Y_{(m)}^\gamma\log(Y_{(m)}))}{\sum(Y_{(i)}^\gamma-Y_{(m)}^\gamma)}\right)+\sum\log(Y_{(i)}).$$

Under the assumption $0<Y_{(m)}<Y_{(1)}$, it is straightforward to show that $g''(\gamma)<0$ for $\gamma>0$ and hence that $g'(\gamma)$ is a strictly decreasing function of γ. As $\gamma \downarrow 0$, $g'(\gamma)$ has limit $(1-\hat{A}/2)\sum\log(Y_{(i)}/Y_{(m)})$, where

$$\hat{A} = \frac{(m-1)^{-1} \sum \log^2(Y_{(i)}/Y_{(m)})}{[(m-1)^{-1} \sum \log(Y_{(i)}/Y_{(m)})]^2};$$

as $\gamma \to \infty$, $g(\gamma)$ has limit $\sum \log(Y_{(i)}/Y_{(1)})$.

Suppose that $0 < Y_{(m)} \leqq Y_{(m-1)} < Y_{(1)}$. Then $g(\gamma)$ is a strictly decreasing function of γ if $\hat{A} \geqq 2$ and $g(\gamma)$ has a unique maximum $\hat{\gamma}$ if $\hat{A} < 2$. When $\hat{A} < 2$, the numerical value of $\hat{\gamma}$ is easily found by solving the equation $g'(\hat{\gamma}) = 0$ in an iterative manner. (When $\hat{A} \geqq 2$, it is reasonable to consider the logarithmic transformation: $W = \log(Y)$ and $W_i = \log(Y_i)$, $1 \leqq i \leqq n$.) Using maximum-likelihood to determine a preliminary power transformation was suggested in part by Box and Cox (1964).

3. SIMULATION-BASED PARAMETER SELECTION

3.1. Adaption to the Exponential Distribution

Consider a nominal $100c\%$ upper confidence bound $U(t)$ for y_p that involves a constant t in its definition, where t is to be chosen to yield the nominal coverage probability c. Usually this is done by means of an appeal to some central limit theorem to justify normal approximation, but, as noted in Section 2.2, the actual coverage probability of $U(t)$ can be significantly less than c. In the present context, more reliable confidence bounds can be obtained by adapting t to the exponential distribution: choose t such that $P(y_p \leqq U(t)) = c$ when Y is exponentially distributed. In practice this must be done by computationally intense Monte Carlo simulation, as in the implementation of bootstrap methods for obtaining confidence bounds (see Efron, 1981).

Consider, for example, a confidence bound for y_p of the form $(\hat{w}_p + tSE(\hat{w}_p))^{1/\hat{\gamma}}$. Let t be chosen so that, when Y has an exponential distribution,

$$P(y_p \leqq [\hat{w}_p + tSE(\hat{w}_p)]^{1/\hat{\gamma}}) = c$$

or, equivalently, so that

$$P\left(\frac{\hat{w}_p - y_p^{\hat{\gamma}}}{SE(\hat{w}_p)} \geqq -t\right) = c.$$

Then $-t$ is the upper cth quantile of the distribution of $(\hat{w}_p - y_p^{\hat{\gamma}})/SE(\hat{w}_p)$ when Y has an exponential distribution, so t is easily found by Monte Carlo simulation.

3.2. Selection of Tail Sizes

We now consider four specific methods of obtaining upper confidence bounds for extreme quantiles: exponential-tail (ET), quadratic-tail (QT), exponential-tail with preliminary power transformation (ETP), and quadratic-tail with preliminary power transformation (QTP). In the ETP and QTP methods, the parameter $\hat{\gamma}$ of the preliminary power transformation is chosen as described in Section 2.3. In all four methods, the parameter t is chosen by simulation and adaption to the exponential distribution. (For the ET method, t could be obtained without simulation, as described in Section 2.1.)

The ET and QT methods, as described in Sections 2.1 and 2.2, respectively, depend on a *tail size* m. The ETP and QTP methods depend on two tail sizes, m_1 and m_2. Here m_1 is the value of m used in the preliminary power transformation and m_2 is the value of m that is used in the exponential-tail or quadratic-tail method applied to the transformed data. A procedure for choosing the tail size(s) will now be developed.

Let W have the gamma distribution with shape parameter α and let $\beta > 0$. The distribution of $Y = W^{\beta}$ is referred to as a generalized gamma(α) distribution with power parameter β (see the Appendix). A generalized gamma(0.5) distribution with $\beta = 0.5$ is a half-normal distribution. The generalized gamma(1) family is the Weibull family; in particular, a generalized gamma(1) distribution with $\beta = 1$ is an exponential distribution. The lognormal family can be thought of as the generalized gamma(∞) family.

In the Appendix, the "tail-heaviness" of the distribution of Y at (in particular) its upper decile $y_{0.1}$ is defined and formulas for the tail-heaviness of generalized gamma and lognormal distributions are given. The tail-heaviness of an exponential distribution equals zero, that of a "light-tailed" distribution is negative and that of a heavy-tailed distribution is positive.

Consider the generalized gamma(0.5), Weibull, generalized gamma(5) and lognormal families of distributions. For each of these four families, we consider the seven values of the power parameter that corresponds to the values -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4 of the tail-heaviness at the upper decile.

Let $\mathrm{Med}(U)$ denote the median of a random variable U, so that $P(U \geqq \mathrm{Med}(U)) = 0.5$. The *excess* of an upper confidence bound U for a quantile y_p is defined as

$$\frac{\mathrm{Med}(U) - y_p}{y_p} \times 100\%.$$

The excesses of the confidence bounds obtained from any of the four methods under investigation depend on the power parameter of the underlying family.

Excesses and coverage probabilities of the nominal 90% upper confidence bounds based on the four methods will be used to determine the various tail sizes. Attention will be restricted to two sample sizes, $n = 50$ and $n = 500$, and two quantiles, $y_{1/n}$ and $y_{0.1/n}$, for each sample size. Thus, for $n = 50$, we consider nominal 90% upper confidence bounds for $y_{0.02}$ and $y_{0.002}$; for $n = 500$, we consider nominal 90% upper confidence bounds for $y_{0.002}$ and $y_{0.0002}$.

In order to determine reasonable values for the tail sizes of the ETP and QTP methods, we conducted a simulation in which we used 10 000 trials to determine values of t for adaption to the exponential distribution, 5000 trials to determine actual coverage probabilities and excesses for the generalized gamma(0.5) family, and 5000 trials each for the Weibull, generalized gamma(5) and lognormal families.

It is necessary to make tradeoffs between coverage probabilities and excesses.

We settled on choosing pairs m_1, m_2 to meet the objective that the estimated coverage probability be at least 85% for both quantiles and all four families of distributions. (The performance at the lognormal distribution and $y_{0.1/n}$ was critical.) Our second objective was, subject to the constraint of the first objective, to minimize the excesses of the confidence bounds. (Since the various excesses have quite similar behavior, it was not necessary to define the criterion more precisely.)

Upon inspection of the results of the simulation, it became clear that $m_1 = n/2$ was nearly optimal for each sample size; thus, for simplicity, we restricted our attention to this choice of m_1. For ETP these constraints clearly led to the choice of $m_2 = 5$ when $n = 50$ and $m_2 = 7$ when $n = 500$. For QTP, at each sample size, there was a fairly large interval of optimal values of m_2, so we chose the midpoint of this interval: $m_2 = 22$ when $n = 50$ and $m_2 = 130$ when $n = 500$.

The ET and QT methods have a single parameter m. For these methods, the coverage probability depends on the power parameter of the underlying distribution. For the ET method we chose $m = 3$ for both sample sizes, since the coverage probabilities dipped too low for $m \geqq 4$ and the excesses were unreasonably large for $m = 2$.

For the QT method it was not possible to realize the objective of 85% coverage probability for $y_{0.1/n}$ when the underlying distribution was lognormal with tail heaviness 0.4. We ended up by choosing m to give (approximately) the best coverage probability for the heavy tailed distributions: $m = 36$ when $n = 50$ and $m = 45$ when $n = 500$. These values of m yield reasonably good coverage probabilities and excesses that are not unreasonably large.

4. RESULTS

In order to evaluate the performance of the four methods for obtaining 90% upper confidence bounds for extreme upper quantiles, with the values of m or m_1, m_2 that were given explicitly in Section 3.2, we conducted another simulation (in which new pseudorandom numbers were used). Again 10 000 trials were used to determine values of t for adaption to the exponential distribution and 5000 trials were used to determine actual coverage probabilities and excesses for each of the four families under consideration.

Figures 1 and 2 show the results for $n = 50$ and $n = 500$, respectively. The header for generalized gamma(0.5) is "half-normal": when the tail heaviness is -0.20, the underlying distribution is half-normal. The header for generalized gamma(5) is "gamma": when the tail heaviness is -0.13, the underlying distribution is a gamma distribution with shape parameter $\alpha = 5$. The Weibull distribution with tail heaviness zero is an exponential distribution. The lognormal distribution with tail heaviness 0.30 is the distribution of $\alpha \exp(Z)$, where α is a positive scale parameter and Z has the standard normal distribution. In both figures, the Monte Carlo estimates of the coverage probability (expressed as a percentage) and excess of each method are shown. In each plot on each figure, the tail heaviness ranges from -0.2 to 0.4 along the horizontal axis.

Consider a confidence bound obtained by using the ETP or QTP method, with

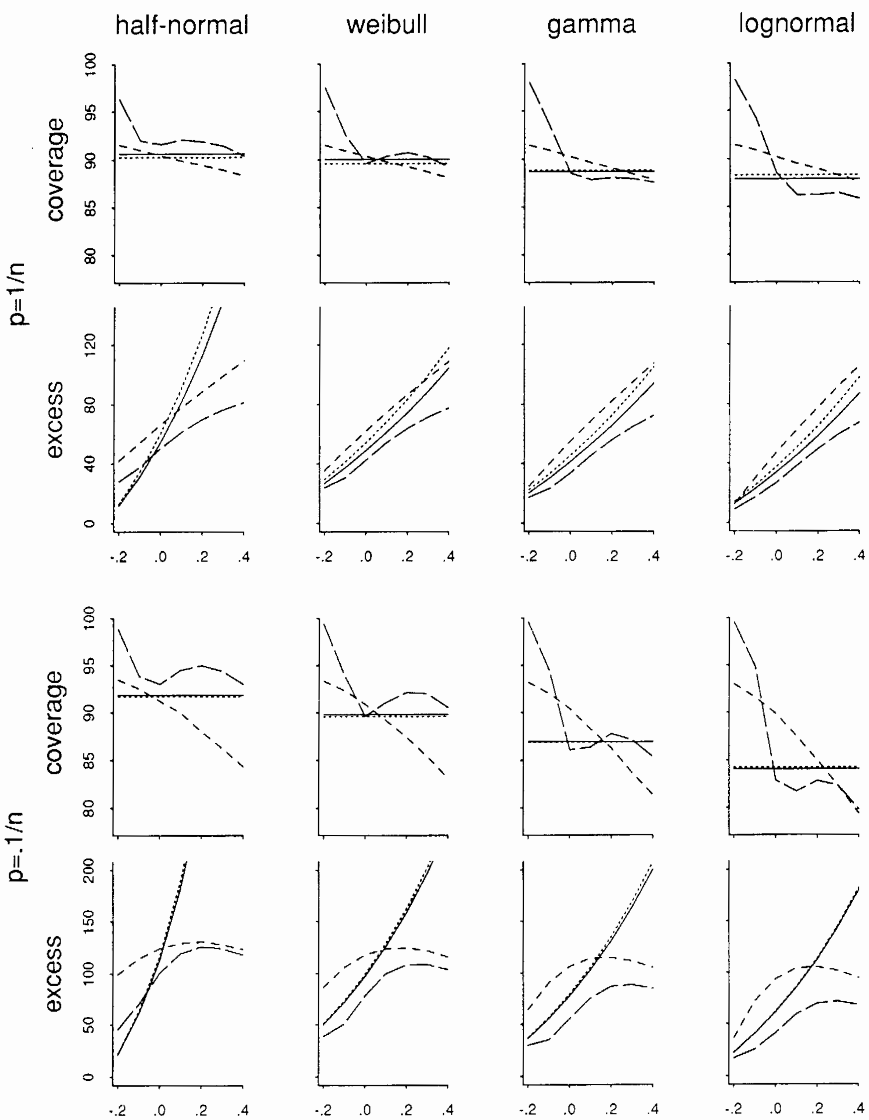

Figure 1 $n = 50$. The correspondence between methods and linetypes is as follows: ET----; QT— — — —; ETP········; QTP————.

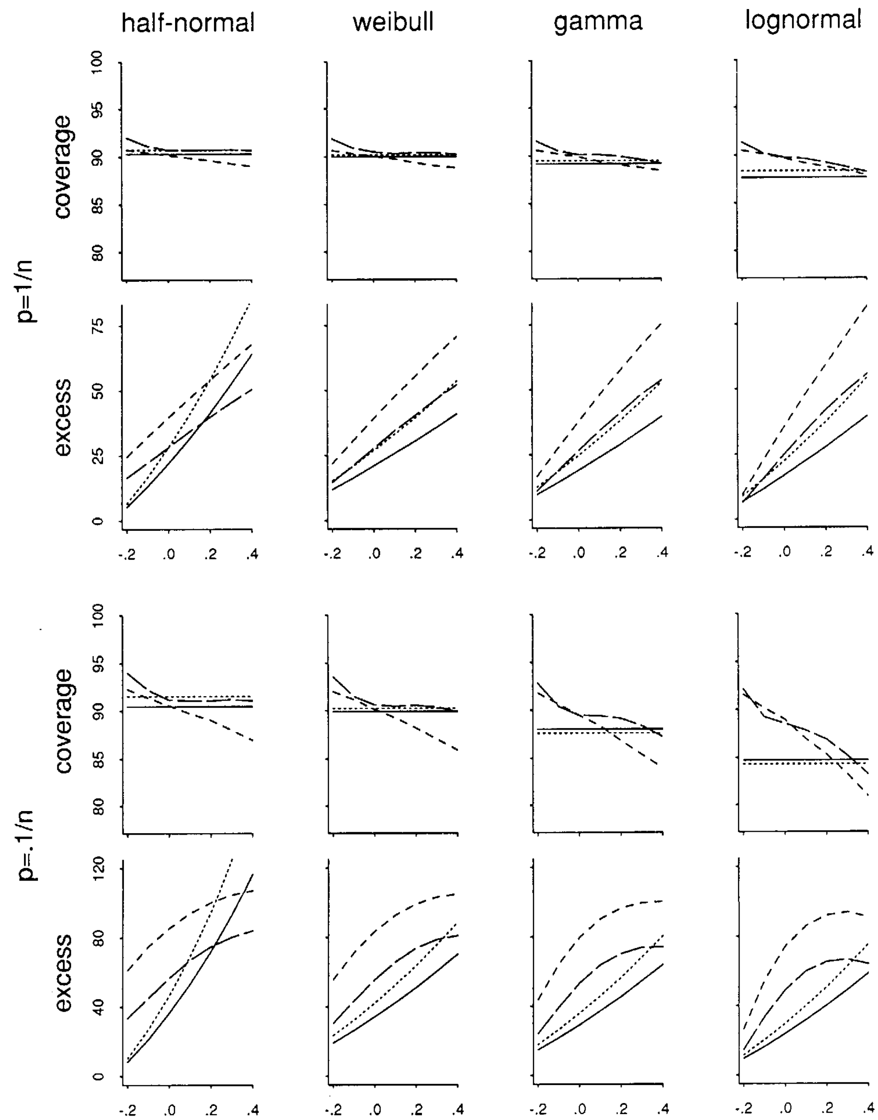

Figure 2 $n = 500$. The correspondence between methods and linetypes is as follows: ET----; QT— — — —; ETP·······; QTP————.

t being chosen by adaption to the exponential distribution (which takes the preliminary power transformation into account). Its actual coverage probability does not depend on the power parameter of the underlying Weibull, generalized gamma, or lognormal distribution. In particular, for underlying Weibull distributions, its actual coverage probability is equal to its nominal coverage probability.

Consider, instead, a confidence bound obtained from the ET or QT method with t being chosen by adaption to the exponential distribution. Its coverage probability does depend on the power parameter of the underlying distribution. In particular, for underlying Weibull distributions, its actual coverage probability is equal to its nominal coverage probability when the tail heaviness is zero, but not otherwise.

5. CONCLUSIONS

Upon examination of Figures 1 and 2, we conclude that, overall, the QT method is best for both quantiles when $n=50$ and the QTP method is best when $n=500$. In particular, the ET method is never the best method. (Other methods that we have tried, not reported here, have performed less well than the better of QT and QTP for various sample sizes.)

Roughly speaking, confidence intervals can be thought of as having bias and variance, the bias being due to errors in modelling approximations and leading to incorrect coverage probabilities and the variance causing the excesses. The ET method can have large bias unless m is very small, in which case it has large variance. For the QT method a substantial proportion of the data ($m=36$) should be used when $n=50$ to avoid large variance. For both the ETP and QTP method, the upper half of the data should be used in estimating the power parameter to avoid unnecessarily large variance. (For values of n substantially larger than 500, less than half of the data should be used to estimate the power parameter.) Even though a substantial proportion of the data is used to estimate the power parameter, the largest observed values have most of the influence on the estimate. Similarly, when the QT method is used, with or without the preliminary power transformation, the largest observed values are the most influential ones.

One important conclusion of this work is that intuition about confidence bounds for extreme quantiles based entirely on asymptotic approximations is very likely to be faulty, for a variety of errors that are asymptotically negligible are actually quite substantial unless the sample size n is unrealistically large. It is only by combining analysis with computer simulation that we can develop sound intuition.

As of now, someone wanting to apply the QT or QTP method to real data would need to do a computer simulation to select t as described in Section 3.1. A further simulation, along the lines of Section 3.2, may be required to select m for the QT method or m_2 for the QTP method for the sample sizes, quantiles and hypothetical distributions of interest. Such computer simulations are increasingly feasible because of the ever greater prevalence and affordability of powerful workstations.

6. QUADRATIC-TAIL MODEL

We now develop the properties of the quadratic-tail model that were used in Section 2 to obtain the corresponding upper confidence bounds. To this end, let $Z_1, \ldots, Z_n$ be a random sample of size n from the exponential distribution with mean one and let $Z_{(1)}, \ldots, Z_{(n)}$ be the corresponding decreasing order statistics. Then $Y_{(1)}, \ldots, Y_{(n)}$ have the same joint distribution as $G^{-1}(\exp(-Z_{(1)})), \ldots, G^{-1}(\exp(-Z_{(n)}))$. In particular, $Y_{(1)}, \ldots, Y_{(m)}$ have the same joint distribution as $G^{-1}(\exp(-Z_{(1)})), \ldots, G^{-1}(\exp(-Z_{(m)}))$.

Since $p_0 \geqq m/n$, it follows from (2.4) with $p = e^{-y}$ that

$$G^{-1}(e^{-y}) = y_{m/n} + \alpha[y - \log(n/m)] + \frac{\beta}{2}[y^2 - \log^2(n/m)], \quad y \geqq \log(n/m).$$

Thus if $Z_{(m)} \geqq \log(1/p_0)$, then $G^{-1}(\exp(-Z_{(i)}))$, $i = 1, \ldots, m$ coincide respectively with

$$y_{m/n} + \alpha[Z_{(i)} - \log(n/m)] + \frac{\beta}{2}[Z_{(i)}^2 - \log^2(n/m), \quad i = 1, \ldots, m.$$

Ignoring the error in the quadratic-tail model and the possibility that $Y_{(m)} < G^{-1}(p_0)$, we conclude that $Y_{(i)}$, $i = 1, \ldots, m$ have the same joint distribution as

$$y_{m/n} + \alpha[Z_{(i)} - \log(n/m)] + \frac{\beta}{2}[Z_{(i)}^2 - \log^2(n/m)], \quad i = 1, \ldots, m.$$

In particular, $Y_{(i)} - Y_{(i+1)}$, $i = 1, \ldots, m-1$, have the same joint distribution as

$$\alpha[Z_{(i)} - Z_{(i+1)}] + \frac{\beta}{2}[Z_{(i)}^2 - Z_{(i+1)}^2], \quad i = 1, \ldots, m.$$

Let L and M be known constants. Consider the parameter $\tau = L\alpha + M\beta$. Let $v_1, \ldots, v_{m-1}$ be known constants and consider the estimate $\hat{\tau} = \sum i v_i [Y_{(i)} - Y_{(i+1)}]$ of τ. Observe that $\hat{\tau}$ has the same distribution as

$$\sum i v_i \left(\alpha[Z_{(i)} - Z_{(i+1)}] + \frac{\beta}{2}[Z_{(i)}^2 - Z_{(i+1)}^2] \right)$$

and hence that

$$E\hat{\tau} = \alpha \sum i v_i E(Z_{(i)} - Z_{(i+1)}) + \frac{\beta}{2} \sum i v_i E(Z_{(i)}^2 - Z_{(i+1)}^2). \tag{6.1}$$

As is well known (see page 37 of Galambos, 1978, or page 21 of David, 1981), $Z_{(i)}$, $i = 1, \ldots, n$, have the same joint distribution as

$$\sum_{j=i}^{n} \frac{Z_j}{j}, \quad i=1,\ldots,n.$$

Consequently, for $1 \leqq i \leqq n-1$,

$$E(Z_{(i)} - Z_{(i+1)}) = E\left(\frac{Z_i}{i}\right)$$

and hence

$$E(Z_{(i)} - Z_{(i+1)}) = \frac{1}{i}. \tag{6.2}$$

Now

$$Z_{(i)}^2 - Z_{(i+1)}^2 = [Z_{(i)} - Z_{(i+1)}]^2 + 2Z_{(i+1)}[Z_{(i)} - Z_{(i+1)}]$$

and hence

$$E(Z_{(i)}^2 - Z_{(i+1)}^2) = E\left(\left[\frac{Z_i}{i}\right]^2\right) + 2E\left(\frac{Z_i}{i} \sum_{j=i+1}^{n} \frac{Z_j}{j}\right) = \frac{2}{i^2} + \frac{2}{i} \sum_{j=i+1}^{n} \frac{1}{j}.$$

Therefore,

$$E(Z_{(i)}^2 - Z_{(i+1)}^2) = \frac{2u_i}{i}, \tag{6.3}$$

where

$$u_i = \sum_{j=1}^{n} \frac{1}{j}.$$

We conclude from (6.1)–(6.3) that

$$E\hat{\tau} = \alpha \sum v_i + \beta \sum u_i v_i. \tag{6.4}$$

Thus $\hat{\tau}$ is unbiased if and only if

$$L = \sum v_i \quad \text{and} \quad M = \sum u_i v_i. \tag{6.5}$$

The variance of $\hat{\tau}$ is derived by a simple but lengthy computation given in Section 7. To state the result, set

$$u_i^{(2)}=\sum_{j=i}^{n}\frac{1}{j^2}\quad\text{and}\quad \bar{v}_i=\frac{1}{i}\sum_{j=1}^{i}v_j.$$

Then

$$\operatorname{var}(\hat{\tau})=\sum(\alpha v_i+\beta\bar{v}_i+\beta u_i v_i)^2+\beta^2\left(\sum u_i^{(2)}v_i^2+(m-1)u_m^{(2)}\bar{v}_{m-1}^2\right). \tag{6.6}$$

It follows from (6.6) that (2.7) holds with $w_i/i=v_i$ for $1\leqq i\leqq m-1$,

$$c_1=\sum v_i^2, \tag{6.7}$$

$$c_2=\sum v_i(\bar{v}_i+u_i v_i),$$

and

$$c_3=\sum\left[(\bar{v}_i+u_i v_i)^2+u_i^{(2)}v_i^2\right]+(m-1)^2u_m^{(2)}\bar{v}_{m-1}^2.$$

Consider the problem of choosing $v_1,\ldots,v_{m-1}$ to minimize the right side of (6.7) subject to (6.5). It is geometrically clear that there is a unique solution to this minimization problem and that the solution is given by $v_i=\lambda_1+\lambda_2 u_i$, $1\leqq i\leqq m-1$, where λ_1 and λ_2 are chosen to satisfy (6.5). It is easily seen that

$$\lambda_1=\frac{S_2L-S_1M}{D}\quad\text{and}\quad\lambda_2=\frac{(m-1)M-S_1L}{D},$$

where

$$S_1=\sum u_i,\quad S_2=\sum u_i^2,$$

and $D=(m-1)S_2-S_1^2$, Thus, for $1\leqq i\leqq m-1$,

$$v_i=\frac{L}{D}\left[S_2-S_1u_i\right]+\frac{M}{D}\left[(m-1)u_i-S_1\right]. \tag{6.8}$$

By choosing $L=1$ and $M=0$, we obtain the unbiased estimate of α given by

$$\hat{\alpha}=\sum w_{1i}\left[Y_{(i)}-Y_{(i+1)}\right]=\sum iv_{1i}\left[Y_{(i)}-Y_{(i+1)}\right],$$

where

$$\frac{w_{1i}}{i}=v_{1i}=\frac{S_2-S_1u_i}{D}$$

for $1\leqq i\leqq m-1$. By choosing $L=0$ and $M=1$, we obtain the unbiased estimate of β given by

$$\hat{\beta}=\sum w_{2i}\left[Y_{(i)}-Y_{(i+1)}\right]=\sum iv_{2i}\left[Y_{(i)}-Y_{(i+1)}\right],$$

where

$$\frac{w_{2i}}{i}=v_{2i}=\frac{(m-1)u_i-S_1}{D}$$

for $1 \leqq i \leqq m-1$. It now follows from (6.8) that, for arbitrary values of L and M, the unbiased estimate of τ for which c_1 is minimized is given by $\hat{\tau}=L\hat{\alpha}+M\hat{\beta}$.

The variance of

$$\hat{y}_p=Y_{(m)}+\hat{\tau}=Y_{(m)}+\sum iv_i[Y_{(i)}-Y_{(i+1)}]$$

is the same as the variance of

$$\alpha Z_{(m)}+\frac{\beta}{2}Z_{(m)}^2+\sum iv_i\left(\alpha[Z_{(i)}-Z_{(i+1)}]+\frac{\beta}{2}[Z_{(i)}^2-Z_{(i+1)}^2]\right).$$

This variance is clearly a quadratic function of α and β, so (2.11) holds. It follows from (6.8) that the constants C_1, C_2 and C_3 in (2.11) depend only on n, m, L, and M. These constants are determined explicitly in Section 7.

7. TECHNICAL DETAILS FOR THE QUADRATIC-TAIL MODEL

We now derive (6.6) and determine the constants in (2.11).

Recall that $Z_1, \ldots, Z_n$ are independent random variables, each having an exponential distribution with mean one. The following facts are easily checked: $\mathrm{var}(Z_1)=1$; $\mathrm{var}(Z_1^2)=20$; $\mathrm{var}(Z_1Z_2)=3$; $\mathrm{cov}(Z_1, Z_1^2)=4$; $\mathrm{cov}(Z_1, Z_1Z_2)=1$; $\mathrm{cov}(Z_1^2, Z_1Z_2)=4$; and $\mathrm{cov}(Z_1Z_2, Z_1Z_3)=1$. It can be assumed that, for $1 \leqq i \leqq n$,

$$Z_{(i)}=\sum_{j=1}^{n}\frac{Z_j}{j}.$$

Until further notice, unless otherwise indicated, the variables i and j range over $1, \ldots, m-1$. Set $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ if $i \neq j$; and set $\psi_{ij}=1$ if $i>j$ and $\psi_{ij}=0$ if $i \leqq j$. The following formulas are easily verified:

$$\mathrm{cov}(Z_i, Z_jZ_{(j+1)})=\delta_{ij}u_j+\frac{\psi_{ij}}{i}-\frac{\delta_{ij}}{i};$$

$$\mathrm{cov}(Z_i^2, Z_jZ_{(j+1)})=4\left(\delta_{ij}u_j+\frac{\psi_{ij}}{i}-\frac{\delta_{ij}}{i}\right);$$

$$\mathrm{var}(Z_iZ_{(i+1)})=u_i^2-2\frac{u_i}{i}+2u_i^{(2)}-\frac{1}{i^2};$$

and

$$\operatorname{cov}(Z_iZ_{(i+1)}, Z_jZ_{(j+1)}) = u_i^{(2)} + \frac{u_i}{i} - \frac{2}{i^2}, \quad i > j.$$

In verifying (6.6), it can be assumed that

$$\hat{\tau} = \sum iv_i\left(\alpha[Z_{(i)} - Z_{(i+1)}] + \frac{\beta}{2}[Z_{(i)}^2 - Z_{(i+1)}^2]\right)$$

$$= \alpha \sum v_i Z_i + \beta \sum v_i \left(\frac{Z_i^2}{2i} + Z_i Z_{(i+1)}\right).$$

Thus $\operatorname{var}(\hat{\tau}) = c_1\alpha^2 + c_2\alpha\beta + c_3\beta^2$, where $c_1 = \operatorname{var}(\sum v_i Z_i) = \sum v_i^2$ and

$$c_2 = 2\operatorname{cov}\left(\sum v_i Z_i, \sum v_i\left(\frac{Z_i^2}{2i} + Z_i Z_{(i+1)}\right)\right)$$

$$= 4\sum \frac{v_i^2}{i} + 2\sum\sum v_i v_j\left(\delta_{ij}u_j + \frac{\psi_{ij}}{i} - \frac{\delta_{ij}}{i}\right)$$

$$= 2\sum v_i(\bar{v}_i + u_i v_i).$$

Also,

$$c_3 = \operatorname{var}\left(\tfrac{1}{2}\sum \frac{v_i}{i} Z_i^2 + \sum v_i Z_i Z_{(i+1)}\right) = c_4 + c_5 + c_6.$$

Here

$$c_4 = \tfrac{1}{4}\operatorname{var}\left(\sum \frac{v_i}{i} Z_i^2\right) = 5\sum \frac{v_i^2}{i^2}.$$

Next,

$$c_5 = \operatorname{cov}\left(\sum \frac{v_i}{i} Z_i^2, \sum v_i Z_i Z_{(i+1)}\right) = 4\sum \frac{u_i v_i^2}{i} + 4\sum \frac{v_i \bar{v}_i}{i} - 8\sum \frac{v_i^2}{i^2}.$$

Moreover,

$$c_6 = \operatorname{var}(\sum v_i Z_i Z_{(i+1)})$$

$$= \sum v_i^2\left(u_i^2 - 2\frac{u_i}{i} + 2u_i^{(2)} - \frac{1}{i^2}\right) + 2\sum v_i\left(u_i^{(2)} + \frac{u_i}{i} - \frac{2}{i^2}\right)(i\bar{v}_i - v_i)$$

$$=\sum u_i^2 v_i^2 - 4\sum \frac{u_i v_i^2}{i} + 3\sum \frac{v_i^2}{i^2} + 2\sum i u_i^{(2)} v_i \bar{v}_i - 4\sum \frac{v_i \bar{v}_i}{i} + 2\sum u_i v_i \bar{v}_i.$$

Consequently,

$$c_3 = \sum u_i^2 v_i^2 + 2\sum i u_i^{(2)} v_i \bar{v}_i + 2\sum u_i v_i \bar{v}_i.$$

Observe that

$$\sum \bar{v}_i^2 = \sum (u_i^{(2)} - u_{i+1}^{(2)}) \left(\sum_{i \geq j} v_j \right)^2 = 2\sum i u_i^{(2)} v_i \bar{v}_i - \sum u_i^{(2)} v_i^2 - (m-1)^2 u_m^{(2)} \bar{v}_{m-1}^2$$

and hence that

$$c_3 = \sum (\bar{v}_i + u_i v_i)^2 + \sum u_i^{(2)} v_i^2 + (m-1)^2 u_m^{(2)} \bar{v}_{m-1}^2.$$

The last formula for c_3 and the previous formulas for c_1 and c_2 together show that (6.6) is valid.

Writing $\hat{y}_p$ as $Y_{(m)} + \hat{\tau}$, we see that

$$\operatorname{var}(\hat{y}_p) = \operatorname{var}(Y_{(m)}) + 2\operatorname{cov}(Y_{(m)}, \hat{\tau}) + \operatorname{var}(\hat{\tau}).$$

The variance of $\hat{\tau}$ is given explicitly in (6.6). Thus to determine the constants in (2.11), we need to determine explicit formulas for $\operatorname{var}(Y_{(m)})$ and $\operatorname{cov}(Y_{(m)}, \hat{\tau})$.

Now

$$\operatorname{var}(Y_{(m)}) = \operatorname{var}\left(\alpha Z_{(m)} + \frac{\beta}{2} Z_{(m)}^2 \right)$$

and hence

$$\operatorname{var}(Y_{(m)}) = \alpha^2 \operatorname{var}(Z_{(m)}) + \alpha\beta \operatorname{cov}(Z_{(m)}, Z_{(m)}^2) + \frac{\beta^2}{4} \operatorname{var}(Z_{(m)}^2). \tag{7.1}$$

It will be shown below that

$$\operatorname{var}(Z_{(m)}) = u_m^{(2)}, \tag{7.2}$$

$$\operatorname{cov}(Z_{(m)}, Z_{(m)}^2) = 2(u_m^{(3)} + u_m^{(2)} u_m), \tag{7.3}$$

and

$$\operatorname{var}(Z_{(m)}^2) = 6u_m^{(4)} + 8u_m^{(3)} u_m + 2(u_m^{(2)})^2 + 4u_m^{(2)} u_m^2, \tag{7.4}$$

where

$$u_m^{(3)} = \sum_{i=m}^{n} \frac{1}{i^3} \quad \text{and} \quad u_m^{(4)} = \sum_{i=m}^{n} \frac{1}{i^4}.$$

Equations (7.1)–(7.4) together yield an explicit formula for $\text{var}(Y_{(m)})$. Also,

$$\text{cov}(Y_{(m)}, \hat{\tau}) = \text{cov}\left(\alpha Z_{(m)} + \frac{\beta}{2} Z_{(m)}^2, \beta \sum i v_i Z_{(i+1)}[Z_{(i)} - Z_{(i+1)}]\right)$$

$$= \text{cov}\left(\alpha Z_{(m)} + \frac{\beta}{2} Z_{(m)}^2, \beta (\sum v_i) Z_{(m)}\right)$$

and hence

$$\text{cov}(Y_{(m)}, \hat{\tau}) = (m-1)\bar{v}_{m-1} \alpha\beta \,\text{var}(Z_{(m)}) + \frac{\beta^2}{2} \text{cov}(Z_{(m)}, Z_{(m)}^2). \tag{7.5}$$

Equations (7.2), (7.3), and (7.5) determine an explicit formula for $\text{cov}(Y_{(m)}, \hat{\tau})$.

It remains to verify (7.2)–(7.4). To this end, let i, j, k, l range from m to n. Then

$$\text{var}(Z_{(m)}) = \text{var}\left(\sum \frac{Z_i}{i}\right) = \sum \frac{1}{i^2} = u_m^{(2)},$$

so (7.2) holds. Observe next that

$$\text{cov}(Z_{(m)}, Z_{(m)}^2) = \text{cov}\left(\sum \frac{Z_i}{i}, \left[\sum \left(\frac{Z_i}{i}\right)\right]^2\right)$$

$$= \sum\sum\sum \frac{1}{ijk} \text{cov}(Z_i, Z_j Z_k)$$

$$= \sum \frac{1}{i^3} \text{cov}(Z_i, Z_i^2) + 2 \sum \frac{1}{i^2} \sum_{i \neq j} \frac{1}{j} \text{cov}(Z_i, Z_i Z_j)$$

$$= \sum 4u_m^{(3)} + 2 \sum \frac{1}{i^2}\left(\sum \frac{1}{j} - \frac{1}{i}\right)$$

$$= 2(u_m^{(3)} + u_m^{(2)} u_m),$$

so (7.3) holds.

Finally,

$$\text{var}(Z_{(m)}) = \text{var}\left(\left(\sum \frac{Z_i}{i}\right)^2\right) = \sum\sum\sum\sum \frac{1}{ijkl} \text{cov}(Z_i Z_j, Z_k Z_l).$$

The total contribution of all terms for which $i = j = k = l$ is

$$\operatorname{var}(Z_1^2)\sum \frac{1}{i^4} = 20u_m^{(4)}.$$

The total contribution of all terms for which exactly three of the four quantities i, j, k, l coincide is

$$4\operatorname{cov}(Z_1^2, Z_1Z_2)\sum \frac{1}{i^3}\sum_{i\neq j}\frac{1}{j} = 16(u_m^{(3)}u_m - u_m^{(4)}).$$

The total contribution of all terms for which i and j are distinct and exactly one of the pair k, l equals i or j is

$$4\operatorname{cov}(Z_1Z_2, Z_1Z_3)\sum\sum_{i\neq j\neq k}\sum \frac{1}{i^2jk} = 4\sum \frac{1}{i^2}\sum_{i\neq j}\frac{1}{j}\left(u_m - \frac{1}{i} - \frac{1}{j}\right)$$

$$= 8u_m^{(4)} - 8u_m^{(3)}u_m - 4(u_m^{(2)})^2 + 4u_m^{(2)}u_m^2;$$

here $i\neq j\neq k$ means that i, j and k are distinct. The total contribution of all terms for which i and j are distinct and (k, l) is either (i, j) or (j, i) is

$$2\operatorname{var}(Z_1Z_2)\sum \frac{1}{i^2}\sum_{i\neq j}\frac{1}{j^2} = 6((u_m^{(2)})^2 - u_m^{(4)}).$$

Equation (7.4) follows by adding up these four totals.

References

Boos, D. D. (1984). Using extreme value theory to estimate large percentiles. *Technometrics* **26,** 33–39.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with Discussion). *J. Roy. Statist. Soc.* Ser. B **26,** 211–252.

Breiman, L., Gins, J. D. and Stone, C. J. (1978). Statistical analysis and interpretation of peak air pollution measurements. *Tech. Report* TSC-PD-A190-10, Technology Service Corporation, Santa Monica.

Breiman, L., Stone, C. J. and Gins, J. D. (1979). New methods for estimating tail probabilities and extreme value distributions. *Tech. Report* TSC-PD-A226-1, Technology Service Corporation, Santa Monica.

Breiman, L., Stone, C. J. and Gins, J. D. (1981). Further development of new methods for estimating tail probabilities and extreme value distributions. *Tech. Report* TSC-PD-A243-1, Technology Service Corporation, Santa Monica.

Crager, M. R. (1982). Exponential tail quantile estimators for air quality data. *Tech. Report* Nos. 4 (Part I) and 5 (Part II), Bay Area Quality Management District, San Francisco.

Csörgő, S., Deheuvels, P. and Mason, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13,** 1050–1077.

David, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

Davis, R. and Resnick, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12,** 1467–1487.

DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *Ann. Statist.* **11,** 1019–1031.

DuMouchel, W. H. and Olshen, R. A. (1975). On the distribution of claims costs: In: *Credibility* (P. M. Kahn, ed.). Academic Press, New York, pp. 23–46.

Efron, B. (1981). *The Jackknife, the Bootstrap and other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.

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

Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13,** 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3,** 1163–1174.

Maritz, J. S. and Munro, A. H. (1967). On the use of the generalized extreme-value distribution in estimating extreme percentiles. *Biometrics* **23,** 79–103.

Pickands, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3,** 119–131.

Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15,** 1174–1207.

Weinstein, S. B. (1973). Theory and application of some classical and generalized asymptotic distributions of extreme values. *IEEE Trans. Information Theory* **IT-19,** 148–154.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73,** 812–815.

APPENDIX. TAIL HEAVINESS

Let G denote the tail distribution function of Y, which is given by $G(y)=P(Y\geqq y)$, $y\in\mathbb{R}$, and let G^{-1} be the inverse function to G, which is assumed to be continuous and strictly decreasing on $(0,1)$. Then $y_p=G^{-1}(p)$ for $0<p<1$. It is also assumed that Y has a density function that is positive and continuously differentiable on the range of G^{-1}.

Let $0<p<1$. The *tail heaviness* at the upper pth quantile y_p of the distribution of Y is defined by

$$H(p)=H_Y(p)=\frac{d^2y_p}{d(\log(1/p))^2}\bigg/\frac{dy_p}{d(\log(1/p))}. \tag{A.1}$$

It follows from (A.1) and elementary calculus that

$$H(p)=-p\left(\frac{d^2y_p}{dp^2}\bigg/\frac{dy_p}{dp}\right)-1=\frac{pG''(y_p)}{[G'(y_p)]^2}-1. \tag{A.2}$$

For an alternative definition of tail-heaviness, suppose that $G'(y)>0$ for $y>0$ and set $\varphi(y)=(d(\log(G(y)/dy)^{-1}$ (see (9.1) in Smith, 1987). Then $H(p)=\varphi'(y_p)$.

The tail heaviness is invariant under location and scale transformations; that is,

$$H_{a+bY}(p)=H_Y(p), \quad a\in\mathbb{R} \text{ and } b>0.$$

The effects of power and logarithmic transformations on the tail heaviness of a positive random variable Y are given by

$$H_{Y^b}(p)=H_Y(p)+\frac{(b-1)}{y_p}\frac{dy_p}{d(\log(1/p))}, \quad b>0,$$

and

$$H_{\log(Y)}(p)=H_Y(p)-\frac{1}{y_p}\frac{dy_p}{d(\log(1/p))}.$$

A random variable is said to be *heavy-tailed* if its tail heaviness is positive and *light-tailed* if its tail heaviness is negative. Suppose that Y has a Weibull distribution with positive power parameter β and hence that Y has the same distribution as W^β, where W has an exponential distribution. The $y_p=\alpha[\log(1/p)]^\beta$ for some positive constant α, which is a scale parameter. By (A.1), the tail heaviness of Y at y_p is given by

$$H(p)=\frac{\beta-1}{\log(1/p)}.$$

When $\beta=1$, Y is exponentially distributed and has tail heaviness zero; when $\beta>1$, it is heavy-tailed; and when $0<\beta<1$, it is light-tailed.

Let W be a positive random variable having tail distribution function G_0 and set $w_p=G_0^{-1}(p)$ for $0<p<1$. Let $\beta>0$ and set $Y=W^\beta$. Then $G(y)=G_0(y^{1/\beta})$ for $y>0$ and $y_p=w_p^\beta$ for $0<p<1$. Moreover, by (A.2) and elementary calculus,

$$H(p)=\frac{p[(1-\beta)G_0'(w_p)+w_pG_0''(w_p)]}{w_p[G_0'(w_p)]^2}-1. \tag{A.3}$$

Suppose, in particular, that W has the gamma distribution with shape parameter α and scale parameter c, whose density function is

$$-G_0'(w)=\frac{w^{\alpha-1}e^{-w/c}}{c^\alpha\Gamma(\alpha)}, \quad w>0.$$

The distribution of $Y=W^\beta$ is referred to as a generalized gamma(α) distribution. It follows from (A.3) that

$$H(p)=\frac{p\Gamma(\alpha)}{(w_p/c)^\alpha\exp(-w_p/c)}\left(\frac{w_p}{c}+\beta-\alpha\right)-1. \tag{A.4}$$

As a special case, suppose that W has the chi-square distribution with one degree of freedom, which is the square of a standard normal random variable Z. Then W has the gamma distribution with parameters $\alpha=0.5$ and $c=2$. Here $w_p=z_{0.5p}^2$, where $P(Z\geqq z_{0.5p})=0.5p$. It follows from (A.4) that

$$H(p)=\frac{p\sqrt{\pi}(2\beta-1+z_{0.5p}^2)}{\sqrt{2}z_{0.5p}\exp(-z_{0.5p}^2/2)}-1.$$

When $\beta=0.5$, $Y=W^{\beta}$ has the same distribution as $|Z|$, which is known as the half-normal distribution.

Suppose now that Y has a lognormal distribution; that is, that $\log(Y)$ is normally distributed. Then $y_p=\alpha\exp(\beta z_p)$ and

$$G(y)=1-\Phi\left(\frac{\log(y/\alpha)}{\beta}\right), \quad y>0.$$

Here Φ denotes the standard normal distribution function, whose density function is denoted by φ; and the scale parameter α and power parameter β are both positive. The random variable $\log(Y)$ has mean $\log(\alpha)$ and standard deviation β. According to (A.2), the tail heaviness of Y at y_p is given by

$$H(p)=\frac{p(z_p+\beta)}{\varphi(z_p)}-1. \tag{A.5}$$

It follows from (A.5) by straightforward asymptotics that

$$\lim_{p\to 0} H(p)z_p=\lim_{p\to 0} H(p)\sqrt{2\log(1/p)}=\beta. \tag{A.6}$$

Thus the tail heaviness is positive for p sufficiently close to zero and it converges exceedingly slowly to zero as $p\to 0$.

If $H(p)\to 0$ as $p\to 0$, then the distribution of Y belongs to the domain of attraction of $\exp(-\exp(-y))$ for the distribution of the maximum (see Theorem 2.7.2 of Galambos, 1978). In particular, generalized gamma and lognormal distributions belong to the indicated domain of attraction. This fact does not appear to be helpful in obtaining confidence bounds for extreme upper quantiles that are reliable for these distributions, however, since the convergence of $H(p)$ to zero as $p\to 0$ can be exceedingly slow. Consider, in particular, the lognormal distribution with $\beta=1$. Suppose we want to treat the upper ten order statistics $Y_{(1)},\ldots,Y_{(10)}$ as coming from a distribution having tail heaviness approximately equal to zero over the relevant portion of the tail. To this end, we might reasonably desire, say, that $H(p)\leqq 0.1$ for $np\geqq 10$, but by (A.5) or (A.6) this would entail that $z_{10/n}\geqq 10$ and hence that $n\geqq 10^{24}$.

Acknowledgment

This research was supported by the Office of Naval Research under contract N00014-82-K-0054 and by National Science Foundation Grants DMS-8600409 and DMS-8902016.