Confidence Bounds for Extreme Quantiles

By

Leo Breiman
Charles J. Stone
and
Charles Kooperberg*

Technical Report No. 167
August 1988

Research supported by ONR under contract N00014-82-K-0054
and NSF Grant DMS-8600409.

Department of Statistics
University of California
Berkeley, California
Confidence Bounds for Extreme Quantiles

LEO BREIMAN, CHARLES J. STONE, and CHARLES KOOPERBERG*

Let $y_p$ be the upper $p$th quantile of the distribution of a random variable $Y$, so that $\Pr(Y \geq y_p) = p$. We consider four related methods for obtaining confidence bounds for $y_p$ when $p$ is very small ($p = 0$) that are obtained by first fitting a parametric model to the $m$ upper order statistics based on a random sample of size $n > m$ from the distribution of $Y$. The exponential-tail (ET) method corresponds to the assumption that the upper tail of the distribution of $Y$ is approximately exponential or, equivalently, that $y_p$ is approximately linear in $\log(1/p)$ for $p = 0$. The quadratic-tail (QT) method corresponds to the assumption that $y_p$ is approximately quadratic in $\log(1/p)$ for $p = 0$. Associated with these two methods are two other methods, ETP and QTP, which involve the use of a preliminary power transformation to make the upper tail more nearly exponential. We also consider the multisample problem, in which the tails of the distributions corresponding to various samples are assumed to have approximately the same shape. When the ETP and QTP methods are applied to the multisample problem, a common power transformation is made to all samples. The confidence bounds we obtain depend on a parameter $t$ that must be adjusted to yield a given nominal coverage probability. We make this adjustment by adaption, via simulation, to the exponential distribution. An extensive simulated study is described, which compares the performance of 90% upper confidence bounds corresponding to the four methods over a wide range of distributions "centered at the exponential"; that is, which are neither too heavy tailed nor too light tailed.

KEY WORDS: Quantile estimation; Exponential-tail model; Quadratic-tail model; Power transformation; Adaption; Tail heaviness.

*This research has been supported by the Office of Naval Research under contract N00014-82-K-0054 and by National Science Foundation Grant DMS-8600409.
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 .01.

As usual, it is important that the estimates we obtain be accompanied by some indication of accuracy. The distributions of estimates of extreme quantiles are far from normal; in particular, they are typically quite skewed. Thus "standard errors" are inappropriate measures of accuracy. Much better are confidence intervals or, equivalently, upper and lower confidence bounds. Since a $100(1-c)\%$ lower confidence bound is a $100c\%$ upper confidence bound (UCB), we can restrict our attention to UCB's.

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

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

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

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

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

$$\Pr(Y_{(1)}^{} \geq y_p^{}) = 1 - (1-p)^n.$$

In particular, $Y_{(1)}^{}$ is a 90% UCB for $y_p$ if and only if

$$p = 1 - (.1)^{1/n} = 1 - \exp(-\log(10)/n) = \frac{\log(10)}{n} = \frac{2.3}{n}, \quad n \geq 1.$$
Thus (for \( n \geq 8 \)) if \( p \leq 2/n \), there is no order statistic that serves as a 90\% UCB for \( y_p \).

When \( p \leq 2/n \), we can obtain a nominal 90\% UCB for \( y_p \) in a standard manner by assuming a Weibull, gamma, lognormal or other classical parametric model for the distribution of \( Y \). But if our assumption is even mildly inaccurate in a given application, the actual coverage probability can differ substantially from .9. In other words, the actual coverage probability of the nominal 90\% UCB for an extreme quantile is very sensitive to model departures.

It is better to fit a more flexible parametric model such as the logspline model discussed in Stone and Koo (1986). But we will not pursue this approach in the present paper.

Another approach is to obtain a nominal 90\% UCB 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 Section 5 and Appendix A. The preliminary power transformation is discussed in Section 2.3; adaption of the parameter \( t \) in a confidence bound to the exponential distribution in Section 2.4; and the multisample problem in Section 2.5. In Section 4 we discuss a reasonably extensive simulation study of 90\% UCB's, in which four specific methods are compared when the actual distribution of \( Y \) is Weibull, generalized gamma, or lognormal. The power parameter of each of these distributions is chosen so that the corresponding tail heaviness, as defined in Section 3, ranges from \(-.2\) to \(.4\) (the tail heaviness of an exponential distribution is zero). The results of the simulation study are presented in graphical form in Section 4.4.

We are unaware of previous work on confidence bounds for extreme quantiles (other than Breiman, et al. (1981)). But there have been many studies of exponential-tail and related methods of estimation for tail probabilities and extreme quantiles. These (mainly
theoretical) studies have focussed 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 exponentially 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; so methods appropriate to approximately exponentially decreasing tails may be applied to data having approximately algebraically decreasing tails by first taking logs. 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 Model

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

\[
\Pr(Y \geq y \mid Y \geq y_{p_0}) = \exp(-(y-y_{p_0})/\alpha)
\]

(2.1)

for \( y \geq y_{p_0} \). For \( y_{p_0} \), 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 \( m/n \leq p_0 \). Then

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

for \( 0 < p \leq p_0 \). It is reasonable to estimate \( y_{m/n} \) by \( Y(m) \) and to estimate \( \alpha \) by

\[
\hat{\alpha} = \frac{1}{m-1} \sum_{i=1}^{m-1} [Y(i) - Y(m)].
\]

This leads to the quantile estimate

\[
\hat{y}_p = Y(m) + \hat{\alpha} \log \left( \frac{m}{np} \right)
\]

(2.3)

for \( 0 < p \leq p_0 \).

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. Then the constant \( t = t_{p,c,n,m} \) can be obtained numerically from the incomplete beta function so that \( U = Y(m) + t\hat{\alpha} \) is an exact 100c\% UCB for \( y_p \). We refer to \( U \) as the exponential-tail 100c\% UCB for \( y_p \).

Under the more general exponential-tail model, the actual coverage probability of the exponential-tail UCB is close to its nominal coverage probability if \( \Pr(Y(m) \geq y_{p_0}) = 1 \). But it is more realistic to consider the exponential-tail model as being a reasonably accurate approximation. Hopefully, the actual coverage probability of the exponential-tail UCB will be close to its nominal coverage probability if \( p \) is not too small. But if \( p \) is extremely small, then the actual and nominal coverage probabilities may well be considerably different.

2.2 Quadratic-tail Model

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 \( m/n \leq p_0 \). Then

\[
y_p = y_{m/n} + \alpha [\log(1/p) - \log(n/m)] + \beta \left[ \log^2(1/p) - \log^2(n/m) \right], \quad 0 < p \leq p_0. \tag{2.4}
\]

Here \( \alpha \) and \( \beta \) are unknown parameters with \( \alpha \geq 0 \). The quadratic 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 \( p \to 0 \). But, even when \( \beta < 0 \), the quadratic-tail model can provide a good approximation to \( y_p \) for values of \( p \) that are not too close to zero.

Given \( m \) and \( p \), set

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

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 \( \tau \) 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 \( \tau \) of the form

\[
\hat{\tau} = \sum_{i=1}^{m-1} \bar{w}_i [Y_{(i)} - Y_{(i+1)}]. \tag{2.6}
\]

It is shown in Section 5 and Appendix A 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 \( w_1, \ldots, w_{m-1} \).

If the exponential-tail model is reasonably accurate and, in particular, if \( \beta = 0 \), 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 5 that this minimization problem has a unique solution, which is given explicitly. (The quadratic model should be thought of as an approximation. In Section 5, the error of approximation is ignored. Thus (2.7) and the solution to the indicated minimization problem should be thought of as informal approximations. The minimization problem itself is reasonable, since it is not possible to choose the weights
to minimize \( \text{Var}(\hat{\tau}) \) for all values of \( \alpha \) and \( \beta \).

In particular, by choosing \( L = 1 \) and \( M = 0 \), we obtain an unbiased estimate of \( \alpha \) having the form
\[
\hat{\alpha} = \frac{m-1}{\sum_{i=1}^{m-1} w_{1i} [Y(i)-Y(i+1)]};
\]
and by choosing \( L = 0 \) and \( M = 1 \), we obtain an unbiased estimate of \( \beta \) having the form
\[
\hat{\beta} = \frac{m-1}{\sum_{i=1}^{m-1} w_{2i} [Y(i)-Y(i+1)]}.
\]
As shown in Section 5, 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 \( \hat{\tau} = L\hat{\alpha} + M\hat{\beta} \); so the corresponding quantile estimate is given by
\[
\hat{y}_p = Y(m) + L\alpha + M\beta
\]
for \( 0 < p \leq p_0 \).

It is shown in Section 5 and Appendix A that
\[
\text{Var}(\hat{y}_p) = C_1\alpha^2 + C_2\alpha\beta + C_3\beta^2,
\]
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
\[
\text{SE}(\hat{y}_p) = (C_1\alpha^2 + C_2\alpha\beta + C_3\beta^2)^{1/2}.
\]
Presumably, under suitable conditions,
\[
\text{Dist} \left( \frac{\hat{y}_p - y_p}{\text{SE}(\hat{y}_p)} \right) = \text{N}(0, 1),
\]
in which case \( \hat{y}_p + z_{1-c} \cdot \text{SE}(\hat{y}_p) \) is an approximate 100c% UCB for \( y_p \); here \( \text{Pr}(Z \geq z_{1-c}) = 1-c \), where \( Z \) has the standard normal distribution.

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. Given a positive constant \( \gamma \), set \( W = Y^\gamma \) and \( W_i = Y_i^\gamma \) for \( 1 \leq i \leq n \). The upper \( p \)th quantile of \( W \) is given by \( w_p = Y_p^\gamma \). 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
\( \hat{w}_p + t \text{SE}(\hat{w}_p) \) be an approximate 100c\% UCB for \( w_p \). Then \( [\hat{w}_p + t \text{SE}(\hat{w}_p)]^{1/\gamma} \) is an approximate 100c\% UCB for \( y_p \).

Let \( 0 < p_0 < 1 \). We would like to choose \( \gamma \) so that the conditional distribution of \( W - w_{p_0} \) given that \( W \geq w_{p_0} \) is approximately exponential. Note that if \( V \) has an exponential distribution, then \( E(V^2) = 2(EV)^2 \). This suggests choosing \( \gamma > 0 \) to satisfy

\[
\frac{E[(Y - \gamma p_0)^2 | Y \geq y_{p_0}]}{E[Y^{\gamma} | Y \geq y_{p_0}]} = 2.
\]

(2.12)

In practice, the power transformation must be determined from the random sample; so 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 \( [\hat{w}_p + t \text{SE}(\hat{w}_p)]^{1/\hat{\gamma}} \) as an approximate 100c\% UCB for \( y_p \). Let \( p_0 = m/n \), where \( 2 \leq m \leq n \). The obvious sample version of (2.12) is to choose \( \hat{\gamma} > 0 \) to satisfy

\[
\frac{1}{m-1} \sum_{i=1}^{m-1} [Y(i)^\hat{\gamma} - Y(m)^\hat{\gamma}]^2
\]

\[
\left[ \frac{1}{m-1} \sum_{i=1}^{m-1} [Y(i) - Y(m)] \right]^2 = 2.
\]

(2.13)

For a refinement of (2.13), let \( Z_1, \ldots, Z_n \) be independent and identically distributed exponential random variables and let \( Z(1), \ldots, Z(n) \) be the corresponding decreasing order statistics. It is shown in Appendix B that

\[
\frac{m}{m-1} E \left[ \frac{1}{m-1} \sum_{i=1}^{m-1} [Z(i) - Z(m)]^2 \right] = 2.
\]

(2.14)

This suggests choosing \( \hat{\gamma} > 0 \) to satisfy

\[
\frac{m}{m-1} \left[ \frac{1}{m-1} \sum_{i=1}^{m-1} [Y(i)^\hat{\gamma} - Y(m)^\hat{\gamma}]^2 \right]
\]

\[
\left[ \frac{1}{m-1} \sum_{i=1}^{m-1} [Y(i) - Y(m)] \right]^2 = 2.
\]

(2.15)
Suppose that \( m \geq 3 \) and that \( Y_{(1)} > \cdots > Y_{(m)} > 0 \). Then the left side of (2.15) is a continuous function of \( \hat{\gamma} \in (0, \infty) \), which has limit

\[
\hat{A}_{\gamma} = \frac{m}{m-1} \left[ \frac{1}{m-1} \sum_{i=1}^{m-1} \frac{1}{\log(Y_{(i)}) - \log(Y_{(m)})} \right]^2 \left[ \frac{1}{m-1} \sum_{i=1}^{m-1} \frac{1}{\log(Y_{(i)}) - \log(Y_{(m)})} \right]^2
\]

as \( \hat{\gamma} \to 0 \) and limit \( m \) as \( \hat{\gamma} \to \infty \); and, as shown in Appendix B, it is a strictly increasing function of \( \hat{\gamma} \). Thus there is at most one value of \( \hat{\gamma} \in (0, \infty) \) that satisfies (2.15); and there is such a value if and only if \( \hat{A}_{\gamma} < 2 \). It is an elementary numerical computation to determine \( \hat{\gamma} \) when it exists. (When \( \hat{\gamma} \) fails to exist, it is reasonable to consider the logarithmic transformation: \( W = \log(Y) \) and \( W_i = \log(Y_i), 1 \leq i \leq n \).)

The use of (2.12)–(2.15) to select the parameter of the power transformation was suggested by the test for exponentiality due to Shapiro and Wilk (1972). The use of a power transformation before applying the ET method was suggested by a similar estimator in Weinstein (1973).

2.4 Adaption to the Exponential Distribution

Consider a confidence bound \( U(t) \) for \( y_p \) that involves a constant \( t \) in its definition, where \( t \) is to be chosen to yield the coverage probability \( c \). Usually this is done by means of an appeal to some central limit theorem to justify normal approximation. We have found that in the context of obtaining confidence bounds for quantiles, the resulting confidence bounds are very unreliable; that is, that the actual coverage probability can differ substantially from \( c \) even when \( Y \) has an exponential distribution.

At least for distributions of \( Y \) similar to those in the simulation study discussed in Section 4, it is more reliable to choose \( t \) by adaption to the exponential distribution; that is, so that \( \Pr(y_p \leq U(t)) = c \) when \( Y \) has an exponential distribution. In practice this must be done by Monte Carlo simulation (as in the implementation of bootstrap methods of obtaining confidence bounds).
Consider, for example, a confidence bound for \( y_p \) of the form
\[
(\hat{w}_p + t \text{SE}(\hat{w}_p))^{1/\hat{\gamma}}.
\]
Let \( t \) be chosen so that, when \( Y \) has an exponential distribution,
\[
\Pr\left[ y_p \leq (\hat{w}_p + t \text{SE}(\hat{w}_p))^{1/\hat{\gamma}} \right] = c
\]
or, equivalently, so that
\[
\Pr\left[ \frac{\hat{w}_p - y_p^{\hat{\gamma}}}{\text{SE}(\hat{w}_p)} \geq -t \right] = c.
\]
Then \(-t\) is the upper \( c \)th quantile of the distribution of
\[
\frac{\hat{w}_p - y_p^{\hat{\gamma}}}{\text{SE}(\hat{w}_p)}
\]
when \( Y \) has an exponential distribution; so \( t \) is easily found by Monte Carlo simulation.

2.5 Multisample Problem

For \( 1 \leq k \leq K \), let \( Y_k \) be a positive random variable having a density \( f_k \) that is continuous and positive on \((0, \infty)\) but otherwise unknown. Let \( y_{kp} \) denote the upper \( p \)th quantile of \( Y_k \). Let \( n_k \) be a positive integer; let \( Y_{k1}, \ldots, Y_{kn_k} \) be a random sample of size \( n_k \) from the distribution of \( Y_k \); and let \( Y_{k(1)}, \ldots, Y_{k(n_k)} \) denote the corresponding decreasing order statistics. It is assumed that the \( n_1 + \cdots + n_K \) random variables obtained by combining the \( K \) random samples are independent. We can obtain separate estimates and confidence intervals for the quantiles \( y_{kp} \), \( 1 \leq k \leq K \).

In many practical applications, the upper tails of the distributions of \( Y_1, \ldots, Y_K \) have approximately the same shape. If so, it is reasonable to use a common power transformation for the \( K \) samples. To this end, we choose positive integers \( m_k \), \( 1 \leq k \leq K \), such that \( 2 \leq m_k \leq n_k \) for \( 1 \leq k \leq K \) and then determine \( \hat{\gamma} \) as the unique positive number such that
\[
\sum_{k=1}^{K} \frac{1}{m_k - 1} \frac{\sum_{i=1}^{m_k-1} (Y_k^{\hat{\gamma}(i)} - Y_k^{\hat{\gamma}(m_k)})^2}{(Y_k^{\hat{\gamma}(i)} - Y_k^{\hat{\gamma}(m_k)})^2} = 2K. \tag{2.16}
\]
Let $G$ denote the tail distribution function of $Y$, which is given by $G(y) = \Pr(Y \geq y)$, $y \in \mathbb{R}$; 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 $0 < p < 1$. In this section, it is assumed, in addition, that $Y$ has a density that is positive and continuously differentiable on the range of $G^{-1}$.

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

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

(3.1)

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

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

An exponential random variable has tail heaviness zero, since its $p$th quantile is a constant multiple of $\log(1/p)$. A random variable is said to be heavy-tailed if its tail heaviness is positive and light-tailed if its tail heaviness is negative.

It follows from (3.1) and elementary calculus that

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

(3.2)

Suppose, for example, that $Y$ has a Weibull distribution with positive power parameter $\beta$; so that $Y$ has the same distribution as $W^\beta$, where $W$ has an exponential distribution. Then $y_p = \alpha \left[ \log(1/p) \right]^{\beta}$ for some positive constant $\alpha$, which is a scale parameter. By (3.1), 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 hence it has tail heaviness zero; when $\beta > 1$, it is heavy-tailed; and when $0 < \beta < 1$, it is light-tailed. For fixed $\beta \neq 1$, the tail heaviness converges very slowly to zero as $p \to 0$.

The tail heaviness corresponding to the quadratic-tail model is given by

$$H(p) = \frac{\beta}{\alpha + \beta \log(1/p)}$$

for $0 < p \leq p_0$. If $\alpha$ and $\beta$ are positive, the tail heaviness is positive; and it converges to zero as $p \to 0$ at the same rate as for heavy-tailed Weibull distributions. If $\alpha$ is positive and $\beta$ is negative, however, the tail heaviness is negative for $p > \exp(\alpha/\beta)$ and it decreases to $-\infty$ as $p \to \exp(\alpha/\beta)$. This is another indication that the quadratic-tail model is unrealistic for sufficiently extreme quantiles of light-tailed distributions.

Suppose next that $Y$ has a lognormal distribution; so 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 $\Phi$ denotes the standard normal distribution function, whose density is denoted by $\varphi$; and the scale parameter $\alpha$ and power parameter $\beta$ are both positive. The random variable $\log(Y)$ has mean $\log(\alpha)$ and standard deviation $\beta$. According to (3.2), the tail heaviness of $Y$ at $y_p$ is given by

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

It follows by straightforward asymptotics that

$$\lim_{p \to 0} H(p) \sqrt{\log(1/p)} = \frac{\beta}{\sqrt{2}}$$

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

Suppose, finally, that $Y$ has a Pareto distribution; so that $y_p = \alpha \exp(\beta \log(1/p))$ for $0 < p < 1$, where the scale parameter $\alpha$ and power parameter $\beta$ are both positive. By (3.1), $H(p) = \beta$ for $0 < p < 1$. In particular, $Y$ is heavy-tailed and its tail heaviness fails to converge to zero as $p \to 0$. 

4. SIMULATION STUDY

4.1 Distributions

The simulation study involves Weibull, generalized gamma(5), and lognormal distributions. In each case, the power parameter $\beta$ takes on values corresponding to seven values, $-2, -1, 0, .1, .2, .3,$ and $.4$, of the tail heaviness, $H(.1)$, at the upper decile. For the Weibull distributions the seven values of $\beta$ are (approximately) $.54$, $.77$, $1.00$, $1.23$, $1.46$, $1.69$, and $1.92$. For the lognormal distributions the seven values of $\beta$ are $.12$, $.30$, $.47$, $.65$, $.82$, $1.00$, and $1.18$.

The generalized gamma(5) distributions considered are distributions of $W^\beta$, where $W$ is distributed as the sum of five independent and identically distributed exponential random variables (so that $W$ has a gamma distribution). The seven values of $\beta$ are $.68$, $1.14$, $1.60$, $2.06$, $2.52$, $2.97$, and $3.43$.

Table 1. Selected Quantiles of Simulated Distributions

<table>
<thead>
<tr>
<th>$p$</th>
<th>$H(.1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$-2$</td>
</tr>
<tr>
<td>Weibull distributions</td>
<td></td>
</tr>
<tr>
<td>$.2$</td>
<td>1.6</td>
</tr>
<tr>
<td>$.02$</td>
<td>2.5</td>
</tr>
<tr>
<td>$.002$</td>
<td>3.3</td>
</tr>
<tr>
<td>$.0002$</td>
<td>3.9</td>
</tr>
<tr>
<td>$.00002$</td>
<td>4.4</td>
</tr>
<tr>
<td>Generalized gamma(5) distributions</td>
<td></td>
</tr>
<tr>
<td>$.2$</td>
<td>1.3</td>
</tr>
<tr>
<td>$.02$</td>
<td>1.7</td>
</tr>
<tr>
<td>$.002$</td>
<td>2.1</td>
</tr>
<tr>
<td>$.0002$</td>
<td>2.4</td>
</tr>
<tr>
<td>$.00002$</td>
<td>2.7</td>
</tr>
<tr>
<td>Lognormal distributions</td>
<td></td>
</tr>
<tr>
<td>$.2$</td>
<td>1.1</td>
</tr>
<tr>
<td>$.02$</td>
<td>1.3</td>
</tr>
<tr>
<td>$.002$</td>
<td>1.4</td>
</tr>
<tr>
<td>$.0002$</td>
<td>1.5</td>
</tr>
<tr>
<td>$.00002$</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 1 shows for the distributions being simulated, the dependence on $H(.1)$ of $y_{.2}$, $y_{.02}$, $y_{.002}$, $y_{.0002}$, and $y_{.00002}$ (selected in light of the discussion later on in this
In this table, the scale parameter is chosen so that the median, \( y_{.5} \), of each distribution is one.

### 4.2 Confidence Bounds

Four methods of obtaining upper confidence bounds for extreme quantiles are evaluated in the simulation study: exponential-tail (ET), quadratic-tail (QT), exponential-tail with power transformation (ETP), and quadratic-tail with power transformation (QTP). In all four methods, the parameter \( t \) is chosen by adaption to the exponential distribution, as described in Section 2.4.

The ET and QT methods, as described in Sections 2.1 and 2.2, respectively, depend on a single integer-valued parameter \( m \). The ETP and QTP methods depend on two integer-valued parameters, \( m_1 \) 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. In the one-sample problem the parameter \( \hat{\gamma} \) of the preliminary power transformation is chosen to satisfy (2.15).

In the simulation study, we also consider the multisample problem with \( K = 10 \). The ET and QT methods treat the ten samples separately. The ETP and QTP methods involve a preliminary power transformation, as described in Section 2.5. In the simulation, the ten sample sizes coincide; the integers \( m_1, \ldots, m_{10} \) introduced in Section 2.5 also coincide, the common value being denoted by \( m_1 \). The exponential-tail method with parameter \( m_2 \) or quadratic-tail method with parameter \( m_2 \) is then separately applied to each of the ten transformed samples to yield upper confidence bounds for the various quantiles of interest. Finally, the inverse power transformation is applied to these upper confidence bounds to yield upper confidence bounds corresponding to the ten original samples.

Consider a confidence bound obtained by using the ETP or QTP method in the context of the one-sample or the multisample problem, with \( t \) 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(5), 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 when the tail heaviness is nonzero.

### 4.3 Parameter Selection

Let \( \text{Med}(U) \) denote the median of a random variable \( U \); so that

\[
\Pr(U \geq \text{Med}(U)) = .5.
\]

The *excess* of an upper confidence bound \( U \) for a quantile \( y_p \) is defined as

\[
\frac{\text{Med}(U) - y_p}{y_p} \times 100\%.
\]

The excesses of the confidence bounds obtained from any of the four methods under investigation depend on the power parameter of the underlying Weibull, generalized gamma(5), or lognormal distribution.

Excesses and coverage probabilities of the nominal 90% upper confidence bounds obtained by using the four methods described in Section 4.2 will be used to compare these methods. Two sample sizes are considered: \( n = 50 \) and \( n = 500 \). Attention is restricted to three quantiles for each sample size: \( y_{1/n}, y_{.1/n}, \) and \( y_{.01/n} \). Thus, for \( n = 50 \), we compare nominal 90% upper confidence bounds for \( y_{.02}, y_{.002}, \) and \( y_{.0002} \); and, for \( n = 500 \), we compare nominal 90% upper confidence bounds for \( y_{.002}, y_{.0002}, \) and \( y_{.00002} \).

In order to determine reasonable values of the pairs \( m_1, m_2 \) of parameters of the ETP and QTP methods, we conducted a preliminary simulation, in which we used 2000 trials to determine values of \( t \) for adaption to the exponential distribution, 1000 trials to determine actual coverage probabilities and excesses for Weibull distributions, 1000 trials to determine those for generalized gamma(5) distributions, and 1000 trials for
lognormal distributions. For each sample size and each of the three families of distributions, the same (pseudo) random numbers were used for both methods, all three quantiles, and all pairs of parameters under investigation.

It is necessary to make tradeoffs between coverage probabilities and excesses. Consider a given sample size and the nominal 90% upper confidence bound for $y_{1/n}$ obtained by either method. We settled on choosing pairs $m_1,m_2$ to meet the objective that the estimated actual coverage probability be at least 88% for each of the three families of distributions. Our second objective is, subject to the constraint of the first objective, to minimize the sum of the excesses of the confidence bounds for the three distributions with zero tail heaviness. For confidence bounds for $y_{.1/n}$ and $y_{.01/n}$ we settled on 85% and 82%, respectively, instead of 88% in the first objective.

We actually made two passes at parameter selection. In the first pass we looked at pairs $m_1,m_2$ broadly spread out and in the second pass, we concentrated on regions that looked most promising in the first pass. Upon reflection, it occurred to us that we could do just about as well by choosing the same pair $m_1,m_2$ for all three quantiles. Since doing this has obvious conceptual and practical advantages, we decided to restrict attention to such quantile-invariant choices. We are satisfied that the pairs we ended up with, shown in Table 2, come close to meeting our two objectives.

<table>
<thead>
<tr>
<th></th>
<th>ETP</th>
<th>QTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td></td>
<td>$n$</td>
</tr>
<tr>
<td>50</td>
<td>500</td>
<td>50</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td>500</td>
</tr>
</tbody>
</table>

Table 2. Values of $m_1,m_2$ for the ETP and QTP Methods

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.
Here, we initially modified our first objective by restricting the power parameter for each of the three families to the seven values indicated in Section 4.1, which correspond to values of the tail heaviness at the upper decile that range from -.2 to .4.

It became apparent, however, that it is not always possible to realize this objective, especially for the lognormal distribution, the highest values of the tail heaviness, and the most extreme quantiles. For the QT method, we ended up by choosing $m$ to give (approximately) the best coverage probabilities for the heavy-tailed distributions for all three quantiles: for $n = 50$ we chose $m = 30$ and for $n = 500$ we chose $m = 45$. These values of $m$ yield reasonably good coverage probabilities and excesses that are not unreasonably large. For the ET method we chose $m = 3$ for both sample sizes, since the coverage probabilities dipped too low for $m > 4$ and the excesses were unreasonably large for $m = 2$.

4.4 Results

In the final simulation, the performance of the four methods for obtaining 90% upper confidence bounds, with $m$ or $m_1, m_2$ as described in Section 4.3, was re-evaluated. Now, 10,000 trials were used to determine values of $t$ for adaption to the exponential distribution, 5000 trials to determine actual coverage probabilities and excesses for Weibull distributions, and 5000 trials each for generalized gamma(5) and lognormal distributions.

Figures 1, 2, and 3 show the results for the one-sample problem with $p = 1/n$, .1/n and .01/n respectively, while Figures 4–6 show those for the ten-sample problem and the same values of $p$. Although all three families were used in the determination of $m$ or $m_1, m_2$, as described in Section 4.3, the final results for the generalized gamma(5) family are omitted from Figures 1–6 in order to save space. These results are intermediate between those for the Weibull family and those for the lognormal family. This is compatible with the quantile values for the three families shown in Table 1 in Section 4.1.

Corresponding to each of the four methods for obtaining an upper confidence bound for $y_p$ is an estimate $\hat{y}_p$ (which does not involve adaption to the exponential distribution). The bias of such an estimate is defined as
In all six figures, the Monte Carlo estimate of the bias of the corresponding estimate is shown along with the coverage probability (expressed as a percentage) and excess of each method. The correspondence between linetypes and methods is as follows:

- - - - - - - - - - - - ET
- - - - - - - - - - - - QT
- - - - - - - - - - - - ETP
- - - - - - - - - - - - - QTP

The ET and QT methods treat separately the ten samples in the ten-sample problem. Thus the results for these methods that are shown in Figures 4–6 coincide with the corresponding results shown in Figures 1–3.

In each plot on each figure, the tail heaviness ranges from −.2 to .4 along the horizontal axis. Note that the coverage probabilities of confidence bounds obtained by the ETP and QTP methods do not depend on tail heaviness, while those obtained by the ET and QT methods do depend on tail heaviness.

According to the coverage and excess plots in Figures 1–3, in the one-sample problem the QT method is best for all three quantiles when \( n = 50 \) and the QTP method is best for all three quantiles when \( n = 500 \). According to Figures 4–6, in the ten-sample problem the ETP method is best for both sample sizes and all three quantiles. (Letting the shapes of the distributions in the ten samples differ somewhat from each other would have favored QTP over ETP.) In particular, the ET method is never the best method.

As of now, someone wanting to apply the QT, ETP or QTP method to real data would need to do a simulation to select \( t \) as described in Section 2.4. Perhaps further simulation studies would be required to select \( m \) or \( m_1, m_2 \) for the sample sizes, quantiles and hypothetical distributions of interest. Although such computer simulations are increasingly feasible because of the ever greater prevalence of powerful workstations, they are still not easy to perform. But nobody has claimed that getting reliable confidence bounds for extreme quantiles would be easy!
Figure 1. One-sample Problem With $p = 1/n$. 

- Weibull
- Lognormal

Coverage

Excess

Bias

n=50
n=500
Figure 2. One-sample Problem With $p = .1/n$. 

Figure 2 shows the comparison of covering, excess, and bias for different distributions (Weibull and Lognormal) and sample sizes (n=50 and n=500). The plots illustrate the performance of the one-sample problem with varying p values, highlighting the effectiveness of the method in terms of coverage, excess, and bias.
Figure 3. One-sample Problem With $p = .01/n$. 

weibull  lognormal  weibull  lognormal 

weibull  lognormal  weibull  lognormal 

weibull  lognormal  weibull  lognormal
Figure 4. Ten-sample Problem With $p = 1/n$. 

- weibull 
- lognormal 

$n=50$ 

$n=500$
Figure 5. Ten-sample Problem With $p = .1/n$. 
Figure 6. Ten-sample Problem With $p = 0.01/n$. 

$n=50$ 

$n=500$
5. QUADRATIC-TAIL MODEL

We now develop the informal properties of the quadratic-tail model that were used in Section 2 to obtain the corresponding upper confidence bounds. To this end, let $U_1, \ldots, U_n$ be the decreasing order statistics based on a random sample of size $n$ from the uniform distribution on $[0, 1]$. Then $G(Y_1), \ldots, G(Y_n)$ have the same joint distribution as $1-U_1, \ldots, 1-U_n$. 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 $1-U_1, \ldots, 1-U_n$ have the same joint distribution as $\exp(-Z_1), \ldots, \exp(-Z_n)$; and $Y_1, \ldots, Y_n$ 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 $\exp(-Z_1), \ldots, \exp(-Z_m)$.

It follows from (2.4) with $p = e^{-y}$ that, for $y \geq \log(n/m)$,

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

for $y \geq \log(1/p_0)$. Thus if $Z_m \geq \log(1/p_0)$, then $G^{-1}(\exp(-Z_i))$, $i = 1, \ldots, m$ coincide respectively with

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

Ignoring the error in the quadratic-tail model and the possibility that $Y_m < \log(1/p_0)$, we conclude that $Y_i$, $i = 1, \ldots, m$ have the same joint distribution as

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

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

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

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

$$\hat{\tau} = \sum_{i=1}^{m-1} iv_i [Y_i - Y_{i+1}]$$

of $\tau$. Observe that $\hat{\tau}$ has the same distribution as

$$\sum_{i=1}^{m-1} iv_i \left[ \alpha [Z^i - Z^i] + \frac{\beta}{2} [Z^2 - Z^2] \right]$$
and hence that
\[
E\hat{\tau} = \alpha \sum_{i=1}^{m-1} iv_i E(Z(i) - Z(i+1)) + \beta \sum_{i=1}^{m-1}iv_i E(Z(i) - Z(i+1)).
\] (5.1)

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
\[
\sum_{j=1}^{n} Z_j, i = 1, \ldots, n.
\]

Consequently, for \(1 \leq i \leq n-1\),
\[
E(Z(i) - Z(i+1)) = E透过分割
\]
and hence
\[
E(Z(i) - Z(i+1)) = \frac{1}{i}.
\] (5.2)

Now
\[
Z^2(i) - Z^2(i+1) = (Z(i) - Z(i+1))^2 + 2Z(i+1)[Z(i) - Z(i+1)]
\]
and hence
\[
E\left[\frac{Z^2(i) - Z^2(i+1)}{i}\right] = E\left[\frac{Z^2}{i}\right] + 2E\left[\frac{Z}{i}\left(\sum_{j=i+1}^{n} \frac{Z_j}{j}\right)\right] = \frac{2}{i} + \frac{2}{i} \sum_{j=i+1}^{n} \frac{1}{j}.
\]

Therefore,
\[
E\left[\frac{Z^2(i) - Z^2(i+1)}{i}\right] = \frac{2u_i}{i},
\] (5.3)

where
\[
u_i = \sum_{j=i}^{n} \frac{1}{j}.
\]

We conclude from (5.1)–(5.3) that
\[
E\hat{\tau} = \alpha \sum_{i=1}^{m-1} \nu_i + \beta \sum_{i=1}^{m-1} \nu_i u_i.
\] (5.4)

Thus \(\hat{\tau}\) is unbiased if and only if
\[
L = \sum_{i=1}^{m-1} \nu_i \quad \text{and} \quad M = \sum_{i=1}^{m-1} \nu_i u_i.
\] (5.5)

The variance of \(\hat{\tau}\) is derived by a simple but lengthy computation given in Appendix A. To state the result, set
\[
u_k^{(2)} = \sum_{j=i}^{n} \frac{1}{j^2} \quad \text{and} \quad \bar{\nu}_i = \frac{1}{i} \sum_{j=1}^{i} \nu_j.
\]
Then

\[
\Var(\hat{\tau}) = \sum_{i=1}^{m-1} \left( \alpha v_i + \beta u_i v_i + \beta u_i v_i \right)^2 + \beta^2 \left( \sum_{i=1}^{m-1} u_i^2 \right) v_i^2 + (m-1)u_m^2 v_{m-1}^2. \tag{5.6}
\]

It follows from (5.6) that (2.7) holds with \( w_i = v_i \) for \( 1 \leq i \leq m-1, \)

\[
c_1 = \sum_{i=1}^{m-1} v_i^2, \tag{5.7}
\]

\[
c_2 = \sum_{i=1}^{m-1} v_i (\tilde{v}_i + u_i v_i),
\]

and

\[
c_3 = \sum_{i=1}^{m-1} \left( (\tilde{v}_i + u_i v_i)^2 + u_i^2 v_i^2 \right) + (m-1)^2 u_m^2 v_{m-1}^2.
\]

Consider the problem of choosing \( v_1, \ldots, v_{m-1} \) to minimize the right side of (5.7) subject to (5.5). It is geometrically clear that there is a unique solution to this minimization problem and that the solution is given by \( v_i = \lambda_1 + \lambda_2 u_i, 1 \leq i \leq m-1, \)

where \( \lambda_1 \) and \( \lambda_2 \) are chosen to satisfy (5.5). It is easily seen that

\[
\lambda_1 = \frac{S_2 L - S_1 M}{D} \quad \text{and} \quad \lambda_2 = \frac{(m-1)M - S_1 L}{D},
\]

where

\[
S_1 = \sum_{i=1}^{m-1} u_i, \quad S_2 = \sum_{i=1}^{m-1} u_i^2,
\]

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

\[
v_i = \frac{L}{D} [S_2 - S_1 u_i] + \frac{M}{D} [(m-1)u_i - S_1]. \tag{5.8}
\]

By choosing \( L = 1 \) and \( M = 0, \) we obtain the unbiased estimate of \( \alpha \) given by

\[
\hat{\alpha} = \sum_{i=1}^{m-1} w_{1i} [Y(i) - Y(i+1)] = \sum_{i=1}^{m-1} i v_{1i} [Y(i) - Y(i+1)],
\]

where

\[
\frac{w_{1i}}{i} = \frac{S_2 - S_1 u_i}{D}
\]

for \( 1 \leq i \leq m-1. \) By choosing \( L = 0 \) and \( M = 1, \) we obtain the unbiased estimate of \( \beta \) given by

\[
\hat{\beta} = \sum_{i=1}^{m-1} w_{2i} [Y(i) - Y(i+1)] = \sum_{i=1}^{m-1} i v_{2i} [Y(i) - Y(i+1)],
\]
where
\[
\frac{w_{2i}}{i} = v_{2i} = \frac{(m-1)u_i - S_1}{D}
\]
for \(1 \leq i \leq m-1\). It now follows from (5.8) that, for arbitrary values of \(L\) and \(M\), the unbiased estimate of \(\tau\) for which \(c_1\) is minimized is given by \(\hat{\tau} = L\hat{\alpha} + M\hat{\beta}\).

The variance of
\[
\hat{Y}_p = Y(m) + \hat{\tau} = Y(m) + \sum_{i=1}^{m-1} iv_i[Y(i) - Y(i+1)]
\]
is the same as the variance of
\[
\alpha Z(m) + \frac{\beta}{2} Z^2(m) + \sum_{i=1}^{m-1} iv_i\left[\alpha [Z(i) - Z(i+1)] + \frac{\beta}{2} [Z^2(i) - Z^2(i+1)]\right].
\]
This variance is clearly a quadratic function of \(\alpha\) and \(\beta\); so that (2.11) holds. It follows from (5.8) 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 Appendix A.
APPENDIX A: QUADRATIC-TAIL MODEL

We now derive (5.6) and determine the constants in (2.11).

Recall that \( Z_1, \cdots, Z_n \) are independent random variables, each having an exponential distribution with mean one. The following facts are easily checked: \( \text{Var}(Z_1) = 1; \) \( \text{Var}(Z_1^2) = 20; \) \( \text{Var}(Z_1 Z_2) = 3; \) \( \text{Cov}(Z_1, Z_1^2) = 4; \) \( \text{Cov}(Z_1, Z_1 Z_2) = 1; \) \( \text{Cov}(Z_1^2, Z_1 Z_2) = 4; \) and \( \text{Cov}(Z_1 Z_2, Z_1 Z_3) = 1. \) It can be assumed that, for \( 1 \leq i \leq n, \)

\[
Z(i) = \sum_{j=i}^{n} \frac{Z_j}{j-i}.
\]

Until further notice, unless otherwise indicated, the variables \( i \) and \( j \) range over \( 1, \cdots, m-1. \) 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 \( i \leq j. \) The following formulas are easily verified:

\[
\text{Cov} \left[ Z_i, Z_{j(j+1)} \right] = \delta_{ij} u_j - \frac{\delta_{ij}}{i},
\]

\[
\text{Cov} \left[ Z_i^2, Z_{j(j+1)} \right] = 4 \left[ \delta_{ij} u_j + \frac{\psi_{ij}}{i} - \frac{\delta_{ij}}{i} \right];
\]

\[
\text{Var} \left[ Z_i Z_{(i+1)} \right] = u_i^2 - \frac{u_i}{i} + 2u_i(2) - \frac{1}{i^2};
\]

and

\[
\text{Cov} \left[ Z_i Z_{(i+1)}', Z_{j(j+1)}' \right] = u_i(2) + \frac{u_i}{i} - \frac{2}{i^2}, \quad i > j.
\]

In verifying (5.6), it can be assumed that

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

\[
= \alpha \sum_i \psi_i Z_i + \beta \sum_i \psi_i \left[ \frac{Z_i^2}{2i} + Z_i Z_{(i+1)} \right].
\]

Thus \( \text{Var}(\hat{\tau}) = c_1 \alpha^2 + c_2 \alpha \beta + c_3 \beta^2, \) where

\[
c_1 = \text{Var} \left[ \sum_i \psi_i Z_i \right] = \sum_i \psi_i^2
\]

and

\[
c_2 = 2 \text{Cov} \left[ \sum_i \psi_i Z_i, \sum_i \psi_i \left[ \frac{Z_i^2}{2i} + Z_i Z_{(i+1)} \right] \right]
\]

\[
= 4 \sum_i \frac{\psi_i^2}{i} + 2 \sum \sum \psi_i \psi_j \left[ \delta_{ij} u_j + \frac{\psi_{ij}}{i} - \frac{\delta_{ij}}{i} \right]
\]

\[
= 2 \sum \psi_i (\bar{\psi}_i + u_i \psi_i).
\]
Also,

\[ c_3 = \text{Var} \left[ \frac{1}{2} \sum_i v_i Z_i^2 + \sum_{i=1}^{N} Y_i Z_i (i+1) \right] = c_4 + c_5 + c_6. \]

Here

\[ c_4 = \frac{1}{4} \text{Var} \left[ \sum_i v_i^2 \right] = 5 \sum_i \frac{v_i^2}{i^2}. \]

Next,

\[ c_5 = \text{Cov} \left[ \sum_i v_i^2 Z_i, \sum_{i=1}^{N} Y_i Z_i (i+1) \right] 
= 4 \sum_i \frac{u_i v_i^2}{i} + 4 \sum_i \frac{\bar{v}_i v_i}{i} - 8 \sum_i \frac{v_i^2}{i^2}. \]

Moreover,

\[ c_6 = \text{Var} \left[ \sum_{i=1}^{N} v_i Z_i (i+1) \right] 
= \sum_i v_i^2 \left[ u_i^2 - \frac{u_i}{i} + 2u_i (2) - \frac{1}{i} \right] + 2 \sum_i v_i \left[ u_i (2) + \frac{u_i}{i} - \frac{1}{i^2} \right] \left[ \bar{v}_i - v_i \right] 
= \sum_i u_i^2 v_i^2 - 4 \sum_i \frac{u_i v_i^2}{i} + 3 \sum_i \frac{v_i^2}{i^2} + 2 \sum_i u_i (2) v_i \bar{v}_i - 4 \sum_i \frac{\bar{v}_i v_i}{i} + 2 \sum_i u_i v_i \bar{v}_i. \]

Consequently,

\[ c_3 = \sum_i u_i^2 v_i^2 + 2 \sum_i \frac{u_i (2) v_i}{i} \bar{v}_i + 2 \sum i u_i v_i \bar{v}_i. \]

Observe that

\[ \sum v_i^2 = \sum \left( u_i (2) - u_{i+1} (2) \right) \left[ \sum v_j \right] = 2 \sum u_i (2) v_i \bar{v}_i - \sum u_i (2) v_i^2 - (m-1)^2 u_m (2) v_m^2 \]

and hence that

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

The last formula for \( c_3 \) and the previous formulas for \( c_1 \) and \( c_2 \) together show that (5.6) is valid.

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

\[ \text{Var} \left[ \hat{y}_p \right] = \text{Var} \left[ Y_{(m)} \right] + 2 \text{Cov} \left[ Y_{(m)}, \hat{\tau} \right] + \text{Var} \left[ \hat{\tau} \right]. \]

Now \( \text{Var}(\hat{\tau}) \) is given explicitly in (5.6). Thus to determine the constants in (2.11), we need to determine explicit formulas for \( \text{Var}(Y_{(m)}) \) and \( \text{Cov}(Y_{(m)}, \hat{\tau}) \).
Now
\[ \text{Var}(Y(m)) = \text{Var}\left[ \alpha Z(m) + \frac{\beta}{2} Z^2(m) \right] \]

and hence
\[ \text{Var}\left[ Y(m) \right] = \alpha^2 \text{Var}\left[ Z(m) \right] + \alpha \beta \text{Cov}\left[ Z(m), Z^2(m) \right] + \frac{\beta^2}{4} \text{Var}\left[ Z(m) \right]. \quad \text{(A.1)} \]

It will be shown below that
\[ \text{Var}\left[ Z(m) \right] = u_m^{(2)}, \quad \text{(A.2)} \]
\[ \text{Cov}\left[ Z(m), Z^2(m) \right] = 2 \left[ u_m^{(3)} + u_m^{(2)} u_m \right], \quad \text{(A.3)} \]

and
\[ \text{Var}\left[ Z^2(m) \right] = 6u_m^{(4)} + 8u_m^{(3)} u_m + 2(u_m^{(2)})^2 + 4u_m^{(2)} u_m, \quad \text{(A.4)} \]

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

Equations (A.1)–(A.4) together yield an explicit formula for \( \text{Var}(Y(m)) \). Also,
\[ \text{Cov}\left[ Y(m)\hat{r}, \hat{t} \right] = \text{Cov}\left[ \alpha Z(m) + \frac{\beta}{2} Z^2(m), \beta \sum_{i} v_i \left[ Z(i+1) - Z(i) \right] \right] \]
\[ = \text{Cov}\left[ \alpha Z(m) + \frac{\beta}{2} Z^2(m), \beta \left[ \sum v_i Z(m) \right] \right] \]

and hence
\[ \text{Cov}\left[ Y(m)\hat{r}, \hat{t} \right] = (m-1)\bar{v}_{m-1} \alpha \beta \text{Var}\left[ Z(m) \right] + \frac{\beta^2}{2} \text{Cov}\left[ Z(m), Z^2(m) \right]. \quad \text{(A.5)} \]

Equations (A.2), (A.3), and (A.5) determine an explicit formula for \( \text{Cov}(Y(m)\hat{r}, \hat{t}) \).

It remains to verify (A.2)–(A.4). To this end, let \( i, j, k, l \) range from \( m \) to \( n \). Then
\[ \text{Var}\left[ Z(m) \right] = \text{Var}\left[ \sum_{i=m}^{n-1} \frac{Z_i}{i} \right] = \sum_{i=m}^{n-1} \frac{1}{i^2} = u_m^{(2)}, \]

so (A.2) holds. Observe next that
\[ \text{Cov}\left[ Z(m), Z^2(m) \right] = \text{Cov}\left[ \sum_{i=m}^{n-1} \frac{Z_i}{i}, \sum_{i=m}^{n-1} \frac{Z_i^2}{i^2} \right] \]
\[ = \sum_{i=m}^{n-1} \sum_{j=1}^{n-1} \text{Cov}\left[ \frac{Z_i}{i}, \frac{Z_j^2}{j} \right] \]
\[ = \sum_{i=m}^{n-1} \text{Cov}\left[ Z_i, Z^2_i \right] + 2 \sum_{i=m}^{n-1} \sum_{j=1}^{n-1} \frac{1}{i j} \text{Cov}(Z_i, Z_j) \]
\[ = \sum_{i=m}^{n-1} \frac{4u_m^{(3)}}{i} + 2 \sum_{i=m}^{n-1} \frac{1}{i^2} \sum_{j=1}^{n-1} \frac{1}{i j} - \frac{1}{i^2} \]
\[ = 2 \left[ u_m^{(3)} + u_m^{(2)} u_m \right], \]
so (A.3) holds.

Finally,

\[ \text{Var} \left[ Z_{(m)} \right] = \text{Var} \left[ \left( \frac{Z_i}{i} \right)^2 \right] = \sum \sum \sum_{ijkl} \frac{1}{ijkl} \text{Cov} \left[ Z_i Z_j, Z_k Z_l \right]. \]

The total contribution of all terms for which \( i = j = k = l \) is

\[ \text{Var} \left[ Z_1^2 \right] \sum_{i} \frac{1}{i^4} = 20u_m^{(4)}. \]

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

\[ 4 \text{Cov} \left[ Z_1^2, Z_1 Z_2, Z_1 Z_3 \right] \sum_{i} \sum_{i \neq j} \frac{1}{ij} = 16 \left[ \frac{u_m^{(3)}}{u_m^{(3)}} - \frac{u_m^{(4)}}{u_m^{(4)}} \right]. \]

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

\[ 4 \text{Cov} \left[ Z_1 Z_2, Z_1 Z_3 \right] \sum_{i \neq j \neq k} \sum_{i \neq j} \frac{1}{ij} = 4 \sum_{i \neq j} \sum_{i \neq j} \frac{1}{ij} \left[ \frac{u_m^{(3)}}{u_m^{(3)}} - \frac{1}{i} - \frac{1}{j} \right] \]

\[ = 8u_m^{(4)} - 8u_m^{(3)} - 4u^{(2)}_m + 4u^{(2)}_m u_m^{(2)}; \]

here \( i \neq j \neq k \) means that \( i, j, k \) 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

\[ 2 \text{Var}(Z_1 Z_2) \sum_{i} \sum_{i \neq j} \frac{1}{ij} = 6 \left[ (u_m^{(2)})^2 - u_m^{(4)} \right]. \]

Equation (A.4) follows by adding up these four totals.
APPENDIX B: POWER TRANSFORMATION

We first verify (2.14). Let $Z_1, \cdots, Z_n$ be independent random variables each having an exponential distribution with (say) mean 1 and let $Z_{(1)}, \cdots, Z_{(n)}$ be the corresponding decreasing order statistics. Let $2 \leq m \leq n$ and let $i$ range over $1, \cdots, m$. Then the conditional distribution of

$$\frac{\sum [Z_{(i)} - Z_{(m)}]^2}{\left[ \sum [Z_{(i)} - Z_{(m)}] \right]^2}$$

given $Z_{(m)}$ is the same as the distribution of $\sum Z_i^2 / (\sum Z_i)^2$. Thus, to verify (2.14), it suffices to verify that the latter random variable has mean $2/m$ or, equivalently, that $Z_1^2((\sum Z_i)^2$ has mean $2/[m(m-1)]$.

Set $V = Z_1$ and $W = Z_2 + \cdots + Z_{m-1}$. We need to verify that

$$E \left[ \left( \frac{V}{V+W} \right)^2 \right] = \frac{2}{m(m-1)} \quad \text{(B.1)}$$

Now $V$ and $W$ are independent random variables; $V$ has the gamma distribution with parameters 1 and 1; $W$ has the gamma distribution with parameters $m-2$ and 1; and $V+W$ has the gamma distribution with parameters $m-1$ and 1. It follows from the change of variables formula involving Jacobians that $W/V$ and $V+W$ are independent and hence that $V/(V+W)$ and $V+W$ are independent. Consequently,

$$2 = E(V^2) = E \left[ \left( \frac{V}{V+W} \right)^2 (V+W)^2 \right] = [(m-1)^2 + m-1]E \left[ \left( \frac{V}{V+W} \right)^2 \right] = m(m-1)E \left[ \left( \frac{V}{V+W} \right)^2 \right]$$

and hence (B.1) holds.

We will now show that the left side of (2.15) is a strictly increasing function of $\hat{\gamma} \in (0, \infty)$. It is enough to verify the following result: Let $W$ be a nonconstant discrete random variable having a finite number of possible values, each of which is greater than one. Define the function $g$ on $(0, \infty)$ by

$$\frac{1}{2} g(\gamma) = E \left[ \left( \frac{W^{\gamma-1}}{E(W^{\gamma-1})} \right)^2 \right]$$

Then $g$ is a strictly increasing function.

To prove this result, we observe that

$$g'(\gamma) = \frac{E[\frac{W^\gamma}{E(W^\gamma)} \log(W)]}{[E(W^{\gamma-1})]^2} - \frac{E[(\log(W))]^2}{E(W^{\gamma-1})} \quad \frac{E[W^{\gamma \log(W)}]}{[E(W^{\gamma-1})]^3}$$
It suffices to verify that \( g' > 0 \) or, equivalently, that

\[
E[W^\gamma - 1] E[(W^\gamma - 1) \log(W)] > E[(W^\gamma - 1)^2] E[W^\gamma \log(W)]
\]  

(B.2)

for \( \gamma > 0 \). Set \( V = W^\gamma \). After noting that \( \log(W^\gamma) = \gamma \log(W) \), we see that (B.2) is equivalent to

\[
E[V - 1] E[(V - 1) \log(V)] > E[(V - 1)^2] E[V \log(V)],
\]

which we can rewrite as

\[
E \left[ \frac{V - 1}{V \log(V)} \right] E \left[ \frac{V \log(V)}{V \log(V)} \right] > E \left[ \frac{(V - 1)^2}{V \log(V)} \right].
\]  

(B.3)

Let \( P \) denote the distribution of \( V \), let \( P^* \) be the distribution on \( \mathbb{R} \) defined by

\[
P^*(dv) = \frac{v \log(v)}{E[V \log(V)]} P(dv),
\]

and let \( U \) be a random variable having distribution \( P^* \). Then (B.3) can be written as

\[
E \left[ \frac{U - 1}{U \log(U)} \right] E(U - 1) > E \left[ \frac{(U - 1)^2}{U \log(U)} \right].
\]  

(B.4)

But (B.4) follows from Schwarz's inequality, provided we can show that \( U - 1 \) is not, with probability one, a constant multiple of

\[
\frac{U - 1}{U \log(U)}.
\]

To this end, it is enough to verify that the function \( h \) on \((1, \infty)\) defined by

\[
h(u) = \frac{u \log(u)}{(u-1)^2}
\]

is strictly decreasing. But

\[
h'(u) = \frac{1 + \log(u)}{(u-1)^2} - \frac{2 u \log(u)}{(u-1)^2},
\]

so it suffices to show that

\[
(u-1)[1 + \log(u)] < 2 u \log(u)
\]

for \( u > 1 \) or, equivalently, that

\[
u \log(u) + \log(u) - u + 1 > 0
\]  

(B.5)

for \( u > 1 \). But (B.5) is clearly valid, since the function defined by the left side of (B.5) equals zero at \( u = 1 \) and its derivative is positive on \((1, \infty)\).

For more general results along these lines, see Breiman, et al. (1979).
REFERENCES


17. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.


32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.


43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?


under information loss.
55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational
Statistics, 12, 101-128.
56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.
60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and
Simulated Annealing.
design for comparing test treatments with a control.
64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.
68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B.
Davies (Fourth edition).
71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.
74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial,
and some continued fractions of Gauss.
77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.
Location based on Partial Information.
two-sample models.


90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.


95. CANCELLED


114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.


118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.


127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.


131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes


166. FAN, JANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.


Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.