A NOTE ON USING THE JACKKNIFE TO ESTIMATE QUANTILE VARIANCE

BY

MICHAEL A. MARTIN

TECHNICAL REPORT NO. 337
DECEMBER 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A NOTE ON USING THE JACKKNIFE TO ESTIMATE QUANTILE VARIANCE

BY

MICHAEL A. MARTIN

TECHNICAL REPORT NO. 337
DECEMBER 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A NOTE ON USING THE JACKKNIFE TO ESTIMATE QUANTILE VARIANCE

by Michael A. Martin

Stanford University

Abstract. We show that the jackknife technique fails badly when applied to the problem of estimating the variance of a sample quantile. When viewed as a point estimator, the jackknife estimator is known to be inconsistent. We show that the ratio of the jackknife variance estimate to the true variance has an asymptotic Weibull distribution with parameters 1 and $\frac{1}{2}$. We also show that if the jackknife variance estimate is used to Studentize the sample quantile, the asymptotic distribution of the resulting Studentized statistic is markedly non-Normal, having infinite mean. This result is in stark contrast with that obtained in simpler problems, such as that of constructing confidence intervals for a mean, where the jackknife-Studentized statistic has an asymptotic standard Normal distribution.

Key Words. jackknife, quantile, Studentize, Weibull distribution.

AMS 1980 Subject Classification. 62E20.

Abbreviated running head: Jackknife Quantile Variance.

Department of Statistics,
Stanford University,
Stanford, CA 94305,
U.S.A.
1. Introduction. The problem of constructing confidence intervals for quantiles with accurate confidence levels is notoriously difficult. For example, intervals constructed using the sign test have confidence levels confined to a small set of binomial probabilities. An approach often used to obtain accurate intervals in simpler problems (such as inference about the mean of a continuous population) is to base intervals and tests on the distribution of an asymptotically pivotal statistic, such as a Studentized ratio, and to make confidence statements such as "sample quantity plus or minus two standard errors". This raises the critical question of how we should estimate the variance of the sample quantile in order to Studentize. One estimate of sample quantile variance, introduced and discussed by Siddiqui (1960) and Bloch and Gastwirth (1968), is based on a density estimate. If this estimator is used to Studentize, the Studentized quantile has an asymptotic standard Normal distribution, but the resulting intervals have poor coverage accuracy; see Hall and Sheather (1988). Another popular technique, the bootstrap (Maritz and Jarrett, 1978; Efron, 1979), estimates the variance as a functional of the underlying distribution function by the same functional of the empirical distribution function. When this estimate is used to Studentize, the Studentized quantile is asymptotically standard Normal, but again, the resulting confidence intervals suffer from poor coverage accuracy (see McKean and Schrader, 1984; Sheather and McKean, 1987; Hall and Martin, 1989).

In this note we discuss the performance of the jackknife, a major competitor of the bootstrap in simpler problems, in estimating the variance of the sample quantile. This problem has been viewed as a kind of "smoking gun" against the technique, mainly because of the failure of the jackknife to provide a consistent estimate of variance in the case of the median; see Efron (1979), Ghosh, Parr, Singh and Babu (1984), and Sheather (1987). It might be thought, however, that a Studentized ratio, using the jackknife estimate of variance, has a limiting Normal distribution, at least for certain quantiles, such as the median. (For example, Miller (1974, page 8) suggests that in the case of the median, the asymptotic distribution is normal, but with variance 4 when \( n \) is even.) In this note, we show that, in fact, the asymptotic distribution of the jackknife-Studentized quantile is markedly non-Normal, having infinite mean. This result is in contrast to simpler problems, such as Studentizing the sample mean, where the jackknife-Studentized statistic has an exact Student’s-\( t \) distribution with \( n - 1 \) degrees of freedom. The reason for the failure of the jackknife in this problem is the fact that sample quantiles are not sufficiently smooth functionals of the empirical distribution function; see Parr (1985) for a discussion of smoothness conditions required for the jackknife to perform reasonably.

We also prove, under fairly weak conditions on the underlying distribution, that the asymptotic
distribution of the jackknife estimate of variance of the \( p \)'th sample quantile, for \( 0 < p < 1 \), is that of the true variance multiplied by a Weibull random variable with parameters 1 and \( \frac{1}{2} \) (that is, a variable having density \( h(w) = \frac{1}{2} w^{-\frac{3}{2}} \exp(-w^\frac{1}{2}) \) for \( w > 0 \)). Section 2 contains our main results and proofs are given in Section 3.

2. Results. Let \( X, X_1, X_2, \ldots \) denote independent and identically distributed observations from a distribution with distribution function \( F \). Assume that for some \( \epsilon > 0 \), \( \text{E}[|X|^\epsilon] < \infty \). Let \( X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n} \) denote the order statistics of a sample \( X_1, \ldots, X_n \) and, given \( 0 < p < 1 \), put \( r \equiv [np] + 1 \) and \( s = [(n - 1)p] + 1 \), where \([\cdot]\) is the integer part function. Assume that the \( p \)'th population quantile \( \xi_p \), defined by \( F(\xi_p) = p \), is unique, and take the \( p \)'th sample quantile to be \( X_{n,r} \). Denote the variance of the \( p \)'th sample quantile by \( \sigma^2 \). It is well known that under mild conditions on \( F \) (given in Theorem 2.1 below), the variance \( \sigma^2 \) satisfies

\[
\sigma^2 = n^{-1} p(1 - p) f(\xi_p)^{-2} + o(n^{-1}) ,
\]

where \( f \equiv F' \); see David (1981, page 80). Define \( \Phi \) and \( \phi \) to be Standard Normal distribution and density functions respectively.

The delete-one jackknife estimate of \( \xi_p \) is \( \hat{\xi}_p \equiv sn^{-1}X_{n,s+1} + (1 - sn^{-1})X_{n,s} \), and the jackknife estimate of the variance of the sample quantile is thus given by

\[
\hat{\sigma}^2 \equiv (n - 1)(n - s)sn^{-2}(X_{n,s+1} - X_{n,s})^2 ;
\]

see Miller (1974). (When \( p = \frac{1}{2} \) and \( n \) is even, the sample median is often defined by \((X_{n,n/2} + X_{n,n/2+1})/2\). There, \( \hat{\sigma}^2 \) admits a more complicated formula.)

**Theorem 2.1.** Assume \( f \equiv F' \) exists and is continuous at \( \xi_p \) and \( f(\xi_p) > 0 \). Then

\[
\hat{\sigma}^2 \sigma^{-2} = W + o_p(1) ,
\]

as \( n \to \infty \), where \( W \) has a Weibull distribution with parameters 1 and \( \frac{1}{2} \).

**Remark 2.1.** An immediate consequence of Theorem 2.1 is that the jackknife estimate of \( \sigma^2 \) is inconsistent. In fact, \( \hat{\sigma}^2 \) is asymptotically unbiased for twice the true variance. This fact is well-known in the case of the median (Efron, 1979; Ghosh, Parr, Singh and Babu, 1984).

**Theorem 2.2.** Under the conditions of Theorem 2.1,

\[
\sup_{-\infty < z < \infty} |P\{(X_{n,r} - \xi_p)/\hat{\sigma} \leq z\} - G(z)| = o(1),
\]
as \( n \to \infty \), where \( G(z) \equiv \frac{1}{2} + \text{sgn}(z)\Phi(-|z|^{-1})\exp(\frac{1}{2}z^{-2}) \) for \( z \neq 0 \) and \( G(0) = \frac{1}{2} \).

**Remark 2.2.** The distribution function \( G \) is that of the ratio of a standard Normal random variable to an independent, unit-mean, negative exponential random variable. It is easy to show that the \( t \)'th absolute moment of this distribution is \( 2^{t/2}\pi^{-\frac{3}{2}}\Gamma(1-t)\Gamma\left\{\frac{1}{2}(1+t)\right\} \) for \( 0 < t < 1 \), but undefined for \( t \geq 1 \). The asymptotic density of \((X_{n,r} - \xi_p)/\hat{\sigma}\) is \( g(z) \equiv z^{-2}(2\pi)^{-\frac{1}{2}} - \text{sgn}(z)z^{-3}\Phi(-|z|^{-1})\exp(\frac{1}{2}z^{-2}) \) for \( z \neq 0 \) with \( g(0) = \phi(0) \).

**Remark 2.3.** Since the Studentized statistic \((X_{n,r} - \xi_p)/\hat{\sigma}\) is asymptotically pivotal, the quantiles of the distribution function \( G \) can be used to construct confidence intervals for \( \xi_p \) with the correct asymptotic level. However, these intervals are generally very long, because \( G \) is heavy-tailed. In fact, if \( z_\alpha \) denotes the \( 100(1-\alpha) \)'th percentage point of the standard Normal distribution and \( y_\alpha \) is the solution of \( G(y_\alpha) = 1 - \alpha \), i.e., the \( 100(1-\alpha) \)'th percentage point of \( G \), then \( y_{0.1} = 7.344 \), \( y_{0.05} = 15.327 \), and \( y_{0.01} = 31.287 \), which are much larger than the corresponding \( z_\alpha \) values 1.28, 1.645, and 1.96, respectively. Note also that \( G(1.645) = 0.827 \) and \( G(1.96) = 0.847 \), so that if one were to wrongly assume an asymptotic standard Normal distribution for the Studentized statistic when constructing a nominal 95% confidence interval for \( \xi_p \), the asymptotic coverage error would be 0.123 for one-sided intervals and 0.255 for two-sided intervals.

**Remark 2.4.** A different approach to the problem of using the jackknife in constructing confidence intervals for quantiles is presented by Kaigh (1983). Kaigh avoids the problem of inconsistency in the variance estimate by using a U-statistic whose variance estimator is consistent to estimate the quantile. Another approach which overcomes the problem of inconsistency has been suggested by Shao (1987). Shao obtains a consistent estimator of the variance of the sample quantile by using a delete-\( d \) jackknife with \( d = \lambda n \), \( 0 < \lambda < 1 \). The reason the delete-\( d \) jackknife and the bootstrap work in this problem while the delete-one jackknife fails is that the former two variance estimates are "smoother" functionals of the empirical distribution function than the latter, which depends only on the difference between two adjacent order statistics. The delete-\( d \) jackknife variance estimator takes account of \( d + 1 = O(n) \) order statistics and the bootstrap estimate uses all of the data. Shao and Wu (1989) show that consistency of the delete-\( d \) jackknife variance estimator can be achieved by allowing the number of deletions to diverge at a rate determined by a smoothness measure of \( \hat{\xi}_p \). See Parr (1985) and Shao and Wu (1989) for detailed discussions of smoothness criteria for consistency of statistical functionals.
3. Proofs of Theorems

**Proof of Theorem 2.1.** Define \( H(x) \equiv F^{-1}\{\exp(-x)\} \). Using Rényi's representation (David 1981, page 21), let \( Z_1, \ldots, Z_n \) denote independent, unit-mean, negative exponential random variables such that \( \{X_{n,j}; 1 \leq j \leq n\} \) has the same distribution as \( \{H(\sum_{j \leq k \leq n} k^{-1} Z_k); 1 \leq j \leq n\} \). Define \( A_s \equiv \sum_{s \leq k \leq n} k^{-1} Z_k \) and \( a_s \equiv \sum_{s \leq k \leq n} k^{-1} \), and note that, since \( H'(x) = -\exp(-x)/f\{H(x)\} \), it follows that

\[
H(a_s) = \xi_p + O(n^{-1}), \quad H'(a_s) = -pf(\xi_p)^{-1} + O(n^{-1}). \tag{1}
\]

By Taylor expansion, \( X_{n,s+1} - X_{n,s} = D + R \), where \( D \equiv -H'(a_s)s^{-1}Z_s \), and

\[
R \equiv -s^{-1}Z_s \int_0^1 \{H'(A_s - ts^{-1}Z_s) - H'(a_s)\} \, dt.
\]

Now, given \( \zeta > 0 \), with probability tending to one as \( n \to \infty \), \( X_{n,s+1} = H(\sum_{s+1}^n k^{-1} Z_k) \in (\xi_p - \frac{1}{2}\zeta, \xi_p + \frac{1}{2}\zeta) \). Thus, recalling that \( f \) is continuous at \( \xi_p \), we see that \( R = o_p(n^{-1}) \). Therefore,

\[
\hat{\delta}^2 = n^{-2}s^{-1}(n-s)(n-1)H'(a_s)^2Z_s^2 + o_p(n^{-1}) = n^{-1}(1-p)f(\xi_p)^{-2}Z_s^2 + o_p(n^{-1}),
\]

whence \( \hat{\delta}^2 \sigma^{-2} = Z_s^2 + o_p(1) \). The result follows on noting that \( Z_s^2 \) has a Weibull distribution with parameters 1 and \( \frac{1}{2} \).

**Proof of Theorem 2.2.** Define \( H(x) \) as in the previous proof, and let \( \beta_1, \ldots, \beta_n \) denote independent and identically distributed centered exponential random variables. Define \( \mu \equiv \sum_{r \leq k \leq n} k^{-1} \), \( \delta \equiv \sum_{r \leq k \leq n} k^{-1} \beta_k \), and \( \eta_p \equiv H(\mu) \). As for (1), we have

\[
H(\mu) = \xi_p + O(n^{-1}), \quad H'(\mu) = -pf(\xi_p)^{-1} + O(n^{-1}). \tag{2}
\]

Now, \( E(\delta^2) = O(n^{-1}) \), so that \( S \equiv n^{\frac{1}{2}}\delta = O_p(1) \). By Taylor expansion, \( X_{n,r} - \eta_p = D_1 + R_1 \), where \( D_1 \equiv H'(\mu)n^{-\frac{1}{2}}S = -pf(\xi_p)^{-1}n^{-\frac{1}{2}}S + O_p(n^{-1}) \), and

\[
R_1 \equiv \delta \int_0^1 \{H'(\mu + t\delta) - H'(\mu)\} \, dt.
\]

Since \( f \) is continuous at \( \xi_p \), and, given \( \zeta > 0 \), with probability tending to one as \( n \to \infty \), \( X_{n,r} = H(\mu + \delta) \in (\xi_p - \frac{1}{2}\zeta, \xi_p + \frac{1}{2}\zeta) \), we have \( R_1 = o_p(n^{-\frac{1}{2}}) \). Thus, noting (2),

\[
X_{n,r} - \xi_p = X_{n,r} - \eta_p + O(n^{-1}) = -pf(\xi_p)^{-1}n^{-\frac{1}{2}}S + o_p(n^{-\frac{1}{2}}),
\]

and from Theorem 2.1, \( \hat{\sigma} = n^{-\frac{1}{2}}(p(1-p))^{\frac{1}{2}}f(\xi_p)^{-1}Z_s + o_p(n^{-\frac{1}{2}}) \), whence

\[
(X_{n,r} - \xi_p)/\hat{\sigma} = -\{p/(1-p)\}^{\frac{1}{2}}Z_s^{-1}S + o_p(1). \tag{3}
\]
Trivially, $Z_s^{-1} S = Z_s^{-1} n^{\frac{1}{2}} \sum_{k=r}^{n} k^{-1} \beta_k = Z_s^{-1} n^{\frac{1}{2}} \sum_{k=s+1}^{n} k^{-1} \beta_k + o_p(1) = Z_s^{-1} T + o_p(1)$, where 

$T \equiv n^{\frac{1}{2}} \sum_{s+1 \leq k \leq n} k^{-1} \beta_k$. By independence of the $\beta_k$'s, $Z_s$ and $T$ are independent, and the characteristic function of $T$,

$$E\{\exp(i\theta T)\} = \exp\left( -i\theta n^{\frac{1}{2}} \sum_{k=s+1}^{n} k^{-1} \right) \prod_{k=s+1}^{n} (1 - i\theta n^{\frac{1}{2}} k^{-1})^{-1},$$

converges to $\exp[-t^2 p/(2(1-p))]$, the characteristic function of a Normal random variable with mean zero and variance $p^{-1}(1-p)$. The result then follows from (3).

Acknowledgement. I wish to thank Professors William R. Schucany and Peter Hall for helpful discussions and suggestions.

References


