CONSISTENCY OF JACKKNIFE ESTIMATORS OF THE VARIANCES OF SAMPLE QUANTILES

by

Jun Shao

Purdue University
Technical Report #87-47

Department of Statistics
Purdue University

October 1987

CONSISTENCY OF JACKKNIFE ESTIMATORS OF THE VARIANCES OF SAMPLE QUANTILES

Jun Shao

Department of Statistics
Purdue University
West Lafayette, IN 47907

Key words and phrases: jackknife; delete-d jackknife; sample quantile; consistency; Helly's theorem.

ABSTRACT

Let σ^2 be the asymptotic variance of the sample p-quantile ($0<p\leq 1$). Consistency of the delete-d jackknife estimators of σ^2 with d being a fraction of n is proved under very weak conditions. Some other results, such as the asymptotic orders of the moments of the jackknife histograms and an analog of the generalized Helly's theorem, are also established.

1. INTRODUCTION

Let F be a distribution function defined on $\mathbf{R}$. The p-quantile of F is defined to be $\theta=F^{-1}(p)=inf\{t: F(t)\geq p\}$, $0<p<1$. Throughout the paper, we assume that

F is continuously differentiable in a neighborhood of θ and $F'(\theta)>0$.

$$ \tag{1.1} $$

Let $\{X_n\}$ be i.i.d. samples from F and $F_n(x)=n^{-1}\sum_{i=1}^{n} I(X_i\leq x)$ be the empirical distribution function, where $I(A)$ is the indicator function of the set A. An estimator of θ is the sample p-quantile $\hat{\theta}=F_n^{-1}(p)$. Under (1.1), Ghosh (1971) proved that

$$\hat{\theta} = \theta+\frac{1}{n}\sum_{i=1}^{n}\phi_F(X_i)+R_n, \tag{1.2}$$

where $\phi(x)=[p-I(x\leq\theta)]/F'(\theta)$ and $R_n=o_p(n^{-\frac{1}{2}})$. From (1.2) and the central limit theorem,

$$n^{\frac{1}{2}}(\hat{\theta}-\theta)\rightarrow N(0,\sigma^2) \textit{ in distribution},$$

where

$$\sigma^2 = p(1-p)/[F'(\theta)]^2. \tag{1.3}$$

As σ^2 is not known, making further inferences, such as setting confidence interval for θ, requires a (weak) consistent estimator of σ^2.

The jackknife method introduced by Quenouille (1956) and Tukey (1958) provides a very convenient way of estimating the accuracy of a point estimator. For the sample p-quantile, however, the traditional delete-1 jackknife estimator of σ^2 is known to be inconsistent. Shao and Wu (1986) studied the general delete-d jackknife estimator $\hat{\sigma}^2_{J(d)}$ (defined in (2.1)) and proved its consistency.

For the sample p-quantile $\hat{\theta}$, an application of their general result shows that if d diverges to infinity at the same rate as n,

$$\hat{\sigma}^2_{J(d)} \to_p \sigma^2 \tag{1.4}$$

under the conditions (1.1) and

$$\lim_{n\to\infty} nVar\hat{\theta} = \sigma^2, \tag{1.5}$$

where $\to_p$ denotes convergence in probability.

However, the justification of (1.5) may not be trivial in many situations. Usually it requires some moment conditions and some further smoothness conditions on F (e.g., $EX_1^2<\infty$ and F'' exists in a neighborhood of θ, see Duttweiler (1973)). Without any moment condition, $Var\hat{\theta}$ may not even exist.

In the present paper we show that under (1.1), (1.4) holds for certain choices of d, without any further condition (Theorem 1). For the proof of this result, we state in Section 2 some other results and give their proofs in Section 3. These results are stated as theorems since they are of interest in their own right.

2. THE MAIN RESULTS

We first define the delete-d jackknife estimator of σ^2. For a given n, let $d=d(n)\leq n$ be an integer and $r=n-d$. Let $\mathbf{S}_r$ be the collection of subsets of $\{1,...,n\}$ which have size r. For any $s\in\mathbf{S}_r$, let $F_{n,s}(x)=r^{-1}\sum_{i\in s} I(X_i\leq x)$ and $\hat{\theta}_s=F^{-1}_{n,s}(p)$. The delete-$d$ jackknife estimator of σ^2 is

$$\hat{\sigma}^2_{J(d)} = \frac{nr}{dN}\sum_s(\hat{\theta}_s-\bar{\theta})^2, \quad \bar{\theta}=\frac{1}{N}\sum_s\hat{\theta}_s, \tag{2.1}$$

where $N=\binom{n}{d}$ and $\sum_s$ is the summation over all subsets in $\mathbf{S}_r$.

Let s^* be a random element satisfying

$$P_*(s^*=s) = \frac{1}{N}, \quad s \in \mathbf{S}_r,$$

and E_* and Var_* be the expectation and variance taken under P_*. Then $\hat{\sigma}^2_{J(d)}$ defined in (2.1) is actually equal to

$$\hat{\sigma}^2_{J(d)} = \frac{nr}{d} Var_*(\hat{\theta}_{s^*}) = \frac{nr}{d} Var_*(\hat{\theta}_{s^*} - \hat{\theta})$$

$$= \frac{nr}{d}\{\textstyle\sum_{j=k}^{m} \binom{n}{d}^{-1}\binom{j-1}{k-1}\binom{n-j}{r-k}(X_{(j)})^2 - [\sum_{j=k}^{m} \binom{n}{d}^{-1}\binom{j-1}{k-1}\binom{n-j}{r-k} X_{(j)}]^2\}, \qquad (2.2)$$

since

$$P_*(\hat{\theta}_{s^*} = X_{(j)}) = \binom{n}{d}^{-1}\binom{j-1}{k-1}\binom{n-j}{r-k}, \quad k \le j \le m, \qquad (2.3)$$

where $X_{(j)}$ is the jth order statistic, $m=k+d$, $k=[rp]$ and $[x]$ is the largest integer $\le x$. (2.2) provides a convenient way of evaluating $\hat{\sigma}^2_{J(d)}$ without involving computations of $\hat{\theta}_s$ for $\binom{n}{d}$ subsets $s \in \mathbf{S}_r$.

Given $X=(X_1, \ldots, X_n)'$, the jackknife histogram is defined to be

$$J(t\,|X) = P_*[\,(\tfrac{nr}{d})^{1/2}(\hat{\theta}_{s^*} - \hat{\theta}) \le t\,], \quad t \in \mathbf{R}, \qquad (2.4)$$

which can be used as an estimator of $P[\,n^{1/2}(\hat{\theta} - \theta) \le t\,]$. Note that $\hat{\sigma}^2_{J(d)}$ is the variance of $J(t\,|X)$ for given X.

In this paper we assume that d is chosen so that

$$d = [\lambda n]+1, \qquad \text{for } a \text{ fixed } \lambda,\ 0<\lambda<1. \tag{2.5}$$

Theorem 1: Assume (1.1). Then for any d satisfying (2.5),

$$\hat{\sigma}^2_{J(d)} \to_p \sigma^2,$$

where σ^2 and $\hat{\sigma}^2_{J(d)}$ are defined in (1.3) and (2.1), respectively.

The following results are needed for the proof of Theorem 1. Their proofs are in the next section. Theorem 2 is a general result (i.e., $\hat{\theta}$ is not necessarily the sample p-quantile) for the consistency of the jackknife histogram $J(t\,|X)$.

Theorem 2: Let $\hat{\theta}$ be an estimator of θ and admit an expansion (1.2) with $\phi(x)$ satisfying $E\phi(X_1)=0$ and $0<E\phi^2(X_1)<\infty$ and $R_n=o_p(n^{-\frac{1}{2}})$. Then,

$$sup_{t\in\mathbf{R}}\,|J(t\,|X)-P[n^{\frac{1}{2}}(\hat{\theta}-\theta)\leq t]\,| \to_p 0,$$

where $J(t\,|X)$ is defined in (2.4).

The following result gives asymptotic orders of the moments of $J(t\,|X)$.

Theorem 3: Let $\hat{\theta}=F_n^{-1}(p)$ and $\hat{\theta}_s=F_{n,s}^{-1}(p)$. Assume (1.1). Then for any d satisfying (2.5) and $\delta\geq 0$,

$$E_*[\,(\frac{nr}{d})^{\frac{1}{2}}|\hat{\theta}_{s^*}-\hat{\theta}|\,]^{2+\delta} = O_p(1).$$

The next result is an analog of the generalized Helly's theorem (Serfling, 1980, Appendix).

Theorem 4: Let Y_n be random n-vectors, $n=1,2,\ldots$, and Z be a random variable. Suppose that for any fixed n and given $Y_n=y$, $Z_n(y)$ is a random variable and as $n\to\infty$,

$$P[\,Z_n(Y_n)\le t\ |Y_n]-P(\,Z\le t\,)\to_p 0,$$

Then as $n\to\infty$,

$$E[h(Z_n(Y_n))|Y_n]\to_p E[h(Z)]$$

for any real-valued bounded continuous function h.

Proof of Theorem 1: Let $Z_{n,s}=(\frac{nr}{d})^{1/2}(\hat{\theta}_s-\hat{\theta})$ and Z be a random variable distributed as $N(0,\sigma^2)$. For any $\varepsilon>0$ and $\tau>0$, from Theorem 3, there is an $A>0$ such that

$$P(\,E_*|Z_{n,s^*}|^{2+\delta}>A\,)<\varepsilon, \tag{2.6}$$

where $\delta>0$ is a fixed number. Choose an $M>0$ such that

$$E[Z^2I(Z^2>M)]<\tau/2 \tag{2.7}$$

and

$$4^{-1}\tau M^{\delta/2}>A. \tag{2.8}$$

For this M, from Theorems 2 and 4

$$\Delta_n = |E_*[Z^2_{n,s^*}I(Z^2_{n,s^*}\le M)]-E[Z^2I(Z^2\le M)]|\to_p 0. \tag{2.9}$$

From (2.6)-(2.8) and

$$|E_* Z^2_{n,s^*} - EZ^2| \leq \Delta_n + E_*[Z^2_{n,s^*} I(Z_{n,s^*} > M)] + E[Z^2 I(Z^2 > M)],$$

we have

$$P[\ |E_* Z^2_{n,s^*} - EZ^2| \geq \tau\ ] \leq P(\Delta_n \geq \tau/4) + P(E_* |Z_{n,s^*}|^{2+\delta} \geq A)$$

$$\leq \varepsilon + P(\Delta_n \geq \tau/4).$$

Hence from (2.9),

$$E_* Z^2_{n,s^*} \to_p EZ^2 = \sigma^2.$$

A similar argument shows $E_* Z_{n,s^*} \to_p EZ=0$. Thus the result follows. □

3. SOME TECHNICAL PROOFS

Proof of Theorem 2: Without loss of generality, we assume that $E\phi^2(X_1)=1$. It suffices to show that

$$sup_{t \in \mathbf{R}} |J(t\,|X) - \Phi(t)| \to_p 0,$$

where Φ is the standard normal distribution function. For any fixed $s \in \mathbf{S}_r$, from (1.2),

$$\hat{\theta}_s = \theta + \frac{1}{r} \sum_{i \in s} \phi_F(X_i) + R_{n,s}$$

with $R_{n,s} = o_p(r^{-1/2})$. Let

$$\xi_{n,s^*} = \left(\frac{nr}{d}\right)^{1/2} \left[\frac{1}{r} \sum_{i \in s^*} \phi_F(X_i) - \frac{1}{n} \sum_{i=1}^{n} \phi_F(X_i)\right]$$

and

$$\alpha_{n,s^*} = \left(\frac{nr}{d}\right)^{1/2}(R_{n,s^*} - R_n).$$

Then $\left(\frac{nr}{d}\right)^{1/2}(\hat{\theta}_{s^*} - \hat{\theta}) = \xi_{n,s^*} + \alpha_{n,s^*}$. For any $\varepsilon>0$, from

$$P_*(\xi_{n,s^*} \le t-\varepsilon) - P_*(|\alpha_{n,s^*}| > \varepsilon) \le P_*(\xi_{n,s^*} \le t-\varepsilon, |\alpha_{n,s^*}| \le \varepsilon)$$

$$\le P_*(\xi_{n,s^*} + \alpha_{n,s^*} \le t, |\alpha_{n,s^*}| \le \varepsilon) \le P_*(\xi_{n,s^*} \le t+\varepsilon),$$

we have

$$sup_{t\in\mathbf{R}} |P_*(\xi_{n,s^*} + \alpha_{n,s^*} \le t) - \Phi(t)| \le 2P_*(|\alpha_{n,s^*}| > \varepsilon)$$

$$+ sup_{t\in\mathbf{R}}[\max(|P_*(\xi_{n,s^*} \le t+\varepsilon) - \Phi(t)|, |P_*(\xi_{n,s^*} \le t-\varepsilon) - \Phi(t)|)]$$

$$\le sup_{t\in\mathbf{R}} |P_*(\xi_{n,s^*} \le t) - \Phi(t)| + 2P_*(|\alpha_{n,s^*}| > \varepsilon)$$

$$+ sup_{t\in\mathbf{R}}[\max(|\Phi(t+\varepsilon) - \Phi(t)|, |\Phi(t-\varepsilon) - \Phi(t)|)]$$

$$\le sup_{t\in\mathbf{R}} |P_*(\xi_{n,s^*} \le t) - \Phi(t)| + (2\pi)^{-1/2}\varepsilon + 2P_*(|\alpha_{n,s^*}| > \varepsilon).$$

From Theorem 1 of Wu (1987),

$$sup_{t\in\mathbf{R}} |P_*(\xi_{n,s^*} \le t) - \Phi(t)| \to_p 0.$$

It remains to be shown that

$$P_*(|\alpha_{n,s^*}| > \varepsilon) \to_p 0, \quad \text{for } any \ \varepsilon>0.$$

Let $s_0 = \{1,\ldots,n\}$. Since $X_1, \ldots, X_n$ are i.i.d.,

$$E[P_*(|\alpha_{n,s^*}| > \varepsilon)] = P\left[\left(\frac{nr}{d}\right)^{1/2} |R_{n,s_o} - R_n| > \varepsilon\right] \to 0$$

by (2.5) and $R_{n,s_o} = o(r^{-1/2})$. This completes the proof. □

Note that under (1.1), there is an interval $[\theta-\tau, \theta+\tau]$ on which $F'(x)\geq l$, where $\tau>0$ and $l>0$ are some constants. Hence we have the following result.

Lemma 1: Assume (1.1). Let $s_0=\{1,\ldots,r\}$. If $0<t<\tau r^{1/2}$,

$$P[\, (\frac{nr}{d})^{1/2} |\hat{\theta}_{s_o}-\theta| > \frac{t}{2}\,] \leq 2e^{-l^2\lambda t^2/2},$$

and

$$P[\, (\frac{nr}{d})^{1/2} |\hat{\theta}-\theta| > \frac{t}{2}\,] \leq 2e^{-l^2\lambda t^2/2}.$$

Also,

$$\max[\, P(\, |\hat{\theta}-\theta| > \tau\lambda^{1/2}/2\,),\, P(\, |\hat{\theta}_{s_o}-\theta| > \tau\lambda^{1/2}/2\,)\,] \leq 2e^{-2r\rho},$$

where $\rho=\min[\, F(\theta+\tau\lambda^{1/2}/2)-F(\theta),\, F(\theta)-F(\theta-\tau\lambda^{1/2}/2)\,] > 0$.

Proof: From Theorem 2.3.2 of Serfling (1980),

$$P[\, (\frac{nr}{d})^{1/2} |\hat{\theta}_{s_o}-\theta| > \frac{t}{2}\,] \leq 2e^{-2r\delta_n^2} \quad \text{for } all\ n,$$

where $\delta_n=\min[\, F(\theta+\frac{t}{2}(\frac{d}{nr})^{1/2})-F(\theta),\, F(\theta)-F(\theta-\frac{t}{2}(\frac{d}{nr})^{1/2})\,]$. Since $0<t<\tau r^{1/2}$, $\delta_n \geq \frac{t}{2}(\frac{d}{nr})^{1/2} l$. Hence the first and the second inequalities follow from $d/n \geq \lambda$. The last inequality follows directly from Theorem 2.3.2 of Serfling (1980). □

Proof of Theorem 3: Let

$$A_n = \int_0^{\tau r^{1/2}} t^{1+\delta} P_*[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s^*} - \hat{\theta}| > t\,]dt,$$

$$B_n = \int_{\tau r^{1/2}}^{r} t^{1+\delta} P_*[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s^*} - \hat{\theta}| > t\,]dt,$$

and

$$C_n = \int_r^{\infty} t^{1+\delta} P_*[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s^*} - \hat{\theta}| > t\,]dt.$$

Then

$$E_*[(\frac{nr}{d})^{1/2} |\hat{\theta}_{s^*} - \hat{\theta}|]^{2+\delta} = (2+\delta)(A_n + B_n + C_n).$$

From $P[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s_o} - \hat{\theta}| > t\,] \le P[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s_o} - \theta| > \frac{t}{2}\,] + P[\,(\frac{nr}{d})^{1/2} |\hat{\theta} - \theta| > \frac{t}{2}\,]$

and Lemma 1,

$$E(A_n) = \int_0^{\tau r^{1/2}} t^{1+\delta} P[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s_o} - \hat{\theta}| > t\,]dt$$

$$\le 4\int_0^{\infty} t^{1+\delta} e^{-l^2 \lambda t^2/2} dt < \infty.$$

Hence $E(A_n) = O(1)$ and therefore $A_n = O_p(1)$. Also,

$$E(B_n) = \int_{\tau r^{1/2}}^{r} t^{1+\delta} P[\,(\frac{nr}{d})^{1/2} |\hat{\theta}_{s^*} - \hat{\theta}| > t\,]dt$$

$$\le P(\,|\hat{\theta}_{s_o}-\hat{\theta}|>\tau\lambda^{½}\,)\int_{\tau r^{½}}^{r} t^{1+\delta}dt$$

$$\le (2+\delta)^{-1} r^{2+\delta}[P(\,|\hat{\theta}_{s_o}-\theta|>\tau\lambda^{½}/2\,)+P(\,|\hat{\theta}-\theta|>\tau\lambda^{½}/2\,)]$$

$$\le 4(2+\delta)^{-1} r^{2+\delta} e^{-2r\rho},$$

where the last inequality follows from Lemma 1. Hence $E(B_n)=o(1)$ and therefore $B_n=o_p(1)$. From (2.3),

$$\max_{s\in S_r}|\hat{\theta}_s-\hat{\theta}| \le |X_{([rp])}-X_{([rp]+d)}| = W_n,$$

and it is easy to see that $W_n/r^{½}\to 0$ *a.s.* Then

$$|W_n| \le (\lambda r)^{½}\ a.s.,$$

and therefore $C_n=o(1)$ *a.s.* Thus the result follows. □

For the proof of Theorem 4, we need the following lemma.

Lemma 2: Let G_{ny} be a distribution function for given $Y_n=y$. If for any continuity point x of a distribution function G,

$$G_{nY_n}(x)-G(x)\to_p 0, \tag{3.1}$$

then there is a countable set $\Omega\subset[0,\,1]$ such that for $t\in[0,\,1]-\Omega$,

$$P(\,|G_{nY_n}^{-1}(t)-G^{-1}(t)|>\varepsilon\,)\to 0 \quad \text{for } any\ \varepsilon>0. \tag{3.2}$$

Proof: Suppose that $t\in[0,\,1]$ such that (3.2) does not hold for an $\varepsilon_o>0$. Then there are $\delta>0$ and $\varepsilon_1>0$ such that $G^{-1}(t)+\varepsilon_1$ and $G^{-1}(t)-\varepsilon_1$ are continuity points of G, and

$$P(\ |G_{nY_n}^{-1}(t)-G^{-1}(t)|>\varepsilon_1\ ) > \delta \quad \text{for } \textit{infinitely many } n.$$

This implies

$$P(\ G_{nY_n}^{-1}(t)\geq G^{-1}(t)+\varepsilon_1\ ) > \delta/2 \quad \text{for } \textit{infinitely many } n, \tag{3.3}$$

since from (3.1),

$$P(\ G_{nY_n}^{-1}(t)\leq G^{-1}(t)-\varepsilon_1\ ) > \delta/2 \quad \text{for } \textit{infinitely many } n$$

implies $t\leq G(G^{-1}(t)-\varepsilon_1)$, which implies $G^{-1}(t)\leq G^{-1}(t)-\varepsilon_1<G^{-1}(t)$. Hence from (3.1) and (3.3), $t\geq G(G^{-1}(t)+\varepsilon_1)\geq G(G^{-1}(t))$. Thus $t=G(x)$ for $x\in[G^{-1}(t), G^{-1}(t)+\varepsilon_1]$. This shows that there are at most countablely many t for which (3.2) does not hold. □

Proof of Theorem 4: Let $M=sup|h(x)|$ and G_{ny} and G be the distribution functions of $Z_n(y)$ and Z, respectively. Let $W_n(t,y)=G_{ny}^{-1}(t)$ and $W(t)=G^{-1}(t)$. From Lemma 2, for any $\varepsilon>0$ and $t\in[0, 1]-\Omega$,

$$P(\ |W_n(t, Y_n)-W(t)|>\varepsilon\ ) \rightarrow 0,$$

where Ω is a countable set. Let T be a uniform random variable and independent of Y_n. Then

$$E[P(\ |W_n(t, Y_n)-W(t)|>\varepsilon|T=t\ )] = \int_0^1 P(\ |W_n(t, Y_n)-W(t)|>\varepsilon\ )dt \rightarrow 0.$$

Hence

$$P(\ |h(W_n(T, y))-h(W(T))|>\varepsilon|Y_n=y\ ) \rightarrow_p 0.$$

Note that

$$E[|h(W_n(T, y))-h(W(T))|\ |Y_n=y] \geq \delta$$

implies

$$P(\ |h(W_n(T,y)) - h(W(T))| > \delta/2 \mid Y_n = y\ ) \geq \delta/4M.$$

Hence

$$E[h(W_n(T,y) \mid Y_n = y] - E[h(W(T))] \to_p 0.$$

The result follows since for given $Y_n = y$,

$$E[h(W_n(T,y)) \mid Y_n = y] = E[h(Z_n(y)) \mid Y_n = y]$$

and

$$E[h(W(T))] = E[h(Z)]. \ \square$$

REFERENCES

Duttweiler, D. L. (1973). "The mean-square error of Bahadur's order-statistic approximation," *Ann. Statist.* **1**, 446-453.

Ghosh, J. K. (1971). "A new proof of the Bahadur representation of quantiles and an application," *Ann. Math. Statist.* **42**, 1957-1961.

Quenouille, M. (1956). "Notes on bias in estimation," *Biometrika* **43**, 353-360.

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

Shao, J. and Wu, C. F. J. (1986). "Some general theory for the jackknife," Tech. Report, Dept. of Statistics, University of Wisconsin, Madison.

Tukey, J. (1958). "Bias and confidence in not quite large samples," *Ann. Math. Statist.* **29**, 614.

Wu, C. F. J. (1987). "On the asymptotic properties of the jackknife histograms," Tech. Report, Dept. of Statistics, University of Wisconsin, Madison.