The Annals of Statistics
1991, Vol. 19, No. 1, 496–503

SOME ASYMPTOTIC RESULTS FOR JACKKNIFING THE SAMPLE QUANTILE

BY XIQUAN SHI

Fudan University

This note studies the asymptotic behavior of the delete-d jackknife in the irregular case of a sample p-quantile, calculated from a sample of n iid r.v.'s.

Two results are obtained: (a) an almost sure rate of convergence of the delete-d jackknife histogram to the normal distribution; (b) almost sure convergence of the delete-d jackknife variance estimate to the asymptotic variance of the sample p-quantile.

Let $\boldsymbol{x} = (X_1, \dots, X_n)$ be iid random observations from $F(t)$ which satisfies the following assumptions $(0 < p < 1)$:

(1) $F(t)$ has a unique p-quantile ξ_p, and its density function $f(t)$ has bounded derivative, and $f(\xi_p) > 0$.

Let $F_n(t)$ be the empirical distribution function based on the samples and $F_n^{-1}(p)$ be the sample p-quantile, we know that

$$\sqrt{n}\left(F_n^{-1}(p) - \xi_p\right) \to_{\mathscr{L}} N\left[0, \frac{p(1-p)}{\left(f(\xi_p)\right)^2}\right]. \tag{2}$$

In order to get the confidence interval of ξ_p, we want to know the distribution of $\sqrt{n}(F_n^{-1}(p) - \xi_p)$ or the variance $p(1-p)/(f(\xi_p))^2$. As is well known, the bootstrap works for the p-quantile, but the Quenouille–Tukey delete-1 jackknife fails. Recently Wu (1987) extended the delete-1 jackknife to the delete-d $(1 < d < n)$ jackknife and studied the weak convergence of the delete-d jackknife histogram to the normal distribution for general nonlinear statistics including the p-quantile. The delete-d jackknife method resamples from $\boldsymbol{x}$ by taking each subset $\boldsymbol{x}_s = (x_{i_1}, \dots, x_{i_r})$, $r = n - d$, of $\boldsymbol{x}$ with equal probability $\binom{n}{r}^{-1}$. Denote this jackknife sampling by $*$. Notation such as P_*, E_* refers to probability calculations under $*$. Let $F_r^*(t)$ and $F_r^{*-1}(p)$ be the empirical distribution and p-quantile based on the jackknife resamples.

In this paper we continue to study the asymptotic properties of the delete-d jackknife p-quantile. We obtain an almost sure rate of convergence of the delete-d jackknife histogram to the normal distribution and almost sure

Received February 1988; revised March 1989.
AMS 1980 *subject classifications*. Primary 62G05, 62E20.
Key words and phrases. Bootstrap, delete-d jackknife, p-quantile, delete-d jackknife variance estimate, a.s. convergence.

convergence of the delete-d jackknife variance estimate to $p(1-p)/(f(\xi_p))^2$. The main results are the following.

THEOREM 1. *Suppose that $F(t)$ satisfies assumption* (1).

(i) *If $d/n \geq \lambda > 0$ and $r(n), d(n) \to \infty$, $\log r(n)/\log\log n \to \infty$, then*

$$\sup_t \left| P_*\left\{\left(\frac{1}{r}-\frac{1}{n}\right)^{-1/2}\left(F_r^{*-1}(p)-F_n^{-1}(p)\right) \leq t\right\} - \Phi\left(\frac{tf(\xi_p)}{\sqrt{p(1-p)}}\right)\right| \to 0 \quad a.s. \tag{3}$$

(ii) *Let $r=[\mu n]$, $0<\mu<1$, where $[x]$ denotes the integer part of x. For any given $0<\delta<1$,*

$$\lim_{n\to\infty}\sup_t n^{1/4}(\log n)^{-3/4-\delta}\left| P_*\left\{\left(\frac{1}{r}-\frac{1}{n}\right)^{-1/2}\left(F_r^{*-1}(p)-F_n^{-1}(p)\right) \leq t\right\} - \Phi\left(\frac{tf(\xi_p)}{\sqrt{p(1-p)}}\right)\right| = 0 \quad a.s. \tag{4}$$

THEOREM 2. *Suppose that $F(t)$ satisfies assumption* (1), *and $r=[\mu n]$, $0<\mu<1$, as $n\to\infty$, we have*

$$\hat{\sigma}^2 = \frac{nr}{d\binom{n}{r}}\sum_s \left(F_r^{*-1}(p)-F_n^{-1}(p)\right)^2 \to \frac{p(1-p)}{f^2(\xi_p)} \quad a.s., \tag{5}$$

where $\sum_s$ denotes summation over all the subsets $s=(i_1,\dots,i_r)\subset(1,2,\dots,n)$.

It is sufficient to consider $F(t)=t$, because we can make the transformation $U=F(X)$ and use Taylor's expansion. Here we only prove part (ii) of Theorem 1, and we need some lemmas.

LEMMA 1. *If $r=[\mu n]$, $0<\mu<1$, for any $0<\delta<1$, as $n\to\infty$,*

$$n^{1/2}(\log n)^{-1-\delta}\sup_t \left| P_*\left\{\left(\frac{1}{r}-\frac{1}{n}\right)^{-1/2}\left(F_r^{*}(p)-F_n(p)\right) \leq t\right\} - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \to 0 \quad a.s. \tag{6}$$

PROOF. Define

$$\hat{\sigma}_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{I(X_i \le p) - F_n(p)\}^2 \to p(1-p) \quad a.s.,$$

where $I(A)$ is the indicator function of set A. By the strong law of large number of iid r.v.'s and the inequality in Wu (1987), it is easy to see that

$$\begin{aligned}
&\sup_t \left| P_*\left\{\left(\frac{1}{r} - \frac{1}{n}\right)^{-1/2}\left(F_r^*(p) - F_n(p)\right) \le t\right\} - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\le \frac{C}{\left(\frac{rd}{n}\right)^{1/2}} \frac{n^{-1}\sum_{i=1}^{n}|I(X_i \le p) - F_n(p)|^3}{\left\{\frac{1}{n-1}\sum_{i=1}^{n}|I(X_i \le p) - F_n(p)|^2\right\}^{3/2}} \\
&\quad + \sup_t \left|\Phi\left(\frac{t}{\hat{\sigma}_1}\right) - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \qquad (7)\\
&\le \frac{C_1}{n^{1/2}} + \sup_{|t| \le C_2 \log n}\left|\Phi\left(\frac{t}{\hat{\sigma}_1}\right) - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\quad + \sup_{|t| \ge C_2 \log n}\left|1 - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| + \sup_{|t| \ge C_2 \log n}\left|1 - \Phi\left(\frac{t}{\hat{\sigma}_1}\right)\right|
\end{aligned}$$

select C_2 such that $C_2/\sqrt{p(1-p)} = \delta + 1$, when n is large enough, we have

$$\begin{aligned}
&\max\left(\sup_{|t| \ge C_2 \log n}\left|1 - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right|, \sup_{|t| \ge C_2 \log n}\left|1 - \Phi\left(\frac{t}{\hat{\sigma}_1}\right)\right|\right) \qquad (8)\\
&\le 1 - \Phi(B \log n) = O(n^{-1/2}) \quad \text{a.s.}
\end{aligned}$$

Using LIL of U-statistics and Taylor's expansion of $t/\hat{\sigma}_1$, it is not difficult to prove

$$\begin{aligned}
&n^{1/2}(\log n)^{-1-\delta} \sup_{|t| \le C_2 \log n}\left|\Phi\left(\frac{t}{\hat{\sigma}_1}\right) - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \qquad (9)\\
&\le n^{1/2}(\log n)^{-1-\delta} C \log n\left\{n^{-1/2}\sqrt{\log\log n}\right. \\
&\qquad\qquad \left. + O(n^{-1}\log\log n)\right\} \to 0 \quad \text{a.s.}
\end{aligned}$$

Thus we complete the proof of Lemma 1. □

LEMMA 2. *For any fixed positive constant C, as $r \to \infty$, we have*

$$P_*\left\{\left|F_r^*(p) - p\right| \ge 4Cr^{-1/2}(\log r)^{1/2}\right\} \to 0 \quad a.s., \qquad (10)$$

$$P_*\left\{\left|F_r^{*-1}(p) - p\right| \ge 4Cr^{-1/2}(\log r)^{1/2}\right\} \to 0 \quad a.s. \qquad (11)$$

PROOF. It is sufficient to prove (10). Define

$$z_j(\alpha,\beta) = I\{\min(\alpha,\beta) \le Y_j \le \max(\alpha,\beta)\} - |F_n(\beta) - F_n(\alpha)| \tag{12}$$

for any given α, β, $0 \le \alpha, \beta \le 1$. Clearly, we have

$$|F_r^*(p) - p| \le \left|\frac{1}{r}\sum_{j=1}^{r} z_j(0,p)\right| + |F_n(p) - p|. \tag{13}$$

From the fact $\sup_t|F_n(t) - t| = O(n^{-1/2}(\log\log n)^{1/2})$ a.s. [Serfling (1980)] and the result of Lemma 1 in Singh (1980) [take $b = 1$, $D = C(r\log r)^{1/2}$],

$$\begin{aligned} P_*\left\{\left|\frac{1}{r}\sum_{j=1}^{r} z_j(0,p)\right| \ge 4Cr^{-1/2}(\log r)^{1/2}\right\} \\ \le 2\exp(-C^2\log r) = 2r^{-C^2} \to 0 \quad \text{a.s.} \end{aligned} \tag{14}$$

Lemma 2 holds. □

LEMMA 3. *Suppose that* $r = r(n) \to \infty$, $\log r(n)/\log\log n \to \infty$, *let* $a_r = r^{-1/2}(\log r)^{1/2}$, *then there exist positive constants* C_1, C_2 *such that*

$$\begin{aligned} P_*\Big\{\left|F_r^{*-1}(p) - p - (p - F_r^*(p))\right| \\ \ge C_1 r^{-3/4}(\log r)^{3/4} + C_2 n^{-1/2}\left(a_r\log\frac{1}{a_r}\right)^{1/2}\Big\} \to 0 \quad a.s. \end{aligned} \tag{15}$$

PROOF. By the proof of Lemma 2, we can select constant h such that

$$P_*\left\{\left|F_r^{*-1}(p) - p\right| \ge hr^{-1/2}(\log r)^{1/2}\right\} \le 2r^{-1} \to 0 \quad \text{a.s.} \tag{16}$$

For fixed h,

$$\begin{aligned} &P_*\Big\{\left|F_r^{*-1}(p) - p - (p - F_r^*(p))\right| \\ &\qquad \ge C_1 r^{-3/4}(\log r)^{3/4} + C_2 n^{-1/2}\left(a_r\log\frac{1}{a_r}\right)^{1/2}\Big\} \\ &\le P_*\Big\{\sup_{|s-p|\le 2hr^{-1/2}(\log r)^{1/2}} \left|F_r^*(s) - s - F_r^*(p) + p\right| \\ &\qquad \ge C_1 r^{-3/4}(\log r)^{3/4} + C_2 n^{-1/2}\left(a_r\log\frac{1}{a_r}\right)^{1/2}\Big\} + Cr^{-1}. \end{aligned} \tag{17}$$

Using the result of Theorem 2 in Stute (1982),

$$\sup_{|s-p|\le 2hr^{-1/2}(\log r)^{1/2}} |F_n(s) - F_n(p) - s + p| = O\left(n^{-1/2}\left(a_r \log\frac{1}{a_r}\right)^{1/2}\right) \quad \text{a.s.} \tag{18}$$

and from the following inequality

$$\begin{aligned}\sup_{|s-p|\le 2hr^{-1/2}(\log r)^{1/2}} &\left|\left(F_r^*(s) - s\right) - \left(F_r^*(P) - p\right)\right| \\ &\le \frac{1}{r} + \left|\max_{1\le|l|\le[2h(r\log r)^{1/2}]+1} \frac{1}{r}\sum_{j=1}^{r} z_j\left(p, p+\frac{l}{r}\right)\right| \\ &\quad + \sup_{|s-p|\le 2hr^{-1/2}(\log r)^{1/2}} |F_n(s) - F_n(p) - s + p|\end{aligned} \tag{19}$$

therefore we only consider the second term of the right side of (19). When r, n are large enough and $1 \le |l| \le [2h(r\log r)^{1/2}] + 1$, clearly

$$\left|F_n\left(p+\frac{l}{r}\right) - F_n(p)\right| \le C_4 r^{-1/2}(\log r)^{1/2} \quad \text{a.s.} \tag{20}$$

Using Lemma 1 in Singh (1980) again [take $b = C_4 r^{-1/2}(\log r)^{1/2}$, $D = (C_2/4)r^{1/4}(\log r)^{3/4}$], then

$$\begin{aligned}P_*&\left\{\left|\max_{1\le|l|\le[2h(r\log r)^{1/2}]+1} \frac{1}{r}\sum_{j=1}^{r} z_j\left(p, p+\frac{l}{r}\right)\right| \ge C_2 r^{-3/4}(\log r)^{3/4}\right\} \\ &\le 2\left(\left[2h(r\log r)^{1/2}\right] + 1\right)\exp\left\{-\frac{C_2^2}{4C_4}\log r\right\} \to 0 \quad \text{a.s.}\end{aligned} \tag{21}$$

□

REMARK. From the proof of Lemma 3, if we select suitable h and C_2 when $r = [\mu n]$, $0 < \mu < 1$, it is not difficult to get

$$\begin{aligned}P_*&\left\{\left|F_r^{*-1}(p) - p - \left(p - F_r^*(p)\right)\right|\right. \\ &\left.\ge C_1 r^{-3/4}(\log r)^{3/4} + C_2 n^{-1/2}\left(a_r \log\frac{1}{a_r}\right)^{1/2}\right\} \\ &= O(n^{-1}) \quad \text{a.s.}\end{aligned} \tag{22}$$

PROOF OF PART (ii) OF THEOREM 2. Note that

$$
\begin{aligned}
F_r^{*-1}(p) - F_n^{-1}(p) &= F_n(p) - F_r^*(p) \\
&\quad + \left[\left(F_r^{*-1}(p) - p\right) - \left(p - F_r^*(p)\right)\right] \\
&\quad - \left[\left(F_n^{-1}(p) - p\right) - \left(p - F_n(p)\right)\right] \\
&= F_n(p) - F_r^*(p) + R_r^* - R_n.
\end{aligned} \tag{23}
$$

From the fact $R_n = O(n^{-3/4}(\log\log n)^{3/4})$ a.s. [Kiefer (1967)], take

$$\delta_{(n)} = \sqrt{\frac{nr}{d}}\left(C_1 r^{-3/4}(\log r)^{3/4} + C_2 n^{-1/2}(a_r \log(1/a_r))^{1/2}\right)$$

and we have

$$
\begin{aligned}
&\sup_t \left| P_*\left\{\left(\frac{1}{r} - \frac{1}{n}\right)^{-1/2}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) \le t\right\} - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\quad \le \sup_t \left| P_*\left\{\left(\frac{1}{r} - \frac{1}{n}\right)^{-1/2}\left(F_r^*(p) - F_n(p)\right) \le t\right\} - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\qquad + \sup_t \left|\Phi\left(\frac{t}{\sqrt{p(1-p)}} \pm \delta_{r(n)}\right) - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\qquad + P_*\left\{\sqrt{\frac{nr}{d}}\,|R_r^*| \ge \delta_{r(n)}\right\}.
\end{aligned} \tag{24}
$$

Since $r = [\mu n]$, $0 < \mu < 1$, it is easy to get

$$
\begin{aligned}
&n^{1/4}(\log n)^{-3/4-\delta} \sup_t \left|\Phi\left(\frac{t}{\sqrt{p(1-p)}} \pm \delta_{r(n)}\right) - \Phi\left(\frac{t}{\sqrt{p(1-p)}}\right)\right| \\
&\quad \le C n^{1/4}(\log n)^{-3/4-\delta}\delta_{r(n)} \to 0.
\end{aligned} \tag{25}
$$

Combining (6), (22) and (25), we complete the proof. □

To prove Theorem 2, in view of Theorem 1, it is sufficient to prove

$$
\begin{aligned}
&E_*\left|\left(\frac{1}{r} - \frac{1}{n}\right)^{-1/2}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right)\right|^{2+\delta} \\
&\quad = (1+\delta)\int_0^\infty t^{1+\delta} P_*\left\{\sqrt{\frac{nr}{d}}\,\left|F_r^{*-1}(p) - F_n^{-1}(p)\right| > t\right\} dt < +\infty
\end{aligned} \tag{26}
$$

for some $\delta > 0$. Following the lines of the arguments in Ghosh, Parr, Singh and Babu (1984) and using the result in Stute (1982), there exists a positive number ε_1 when $t \in [1, C(\delta)/\varepsilon_1(\log n)^{1/2}]$ [where the requirement on the constant $C(\delta)$ is specified later] we have

$$p + \frac{1}{2r} - F_n\left(F_n^{-1}(p) + t\sqrt{\frac{d}{nr}}\right) \le -\varepsilon_1 t\sqrt{\frac{d}{nr}} \quad \text{a.s.} \tag{27}$$

in this interval we can get

$$P_*\left\{\sqrt{\frac{nr}{d}}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) > t\right\} \le Ct^{-4}. \tag{28}$$

Therefore

$$\int_1^{C(\delta)/\varepsilon_1(\log n)^{1/2}} t^{1+\delta} P_*\left\{\sqrt{\frac{nr}{d}}\left|F_r^{*-1}(p) - F_n^{-1}(p)\right| > t\right\} dt \le C < \infty. \tag{29}$$

For $t > C(\delta)/\varepsilon_1(\log n)^{1/2}$, using the same method as in the proof of Lemma 1 in Singh (1980), we can prove

$$P_*\left\{\sqrt{\frac{nr}{d}}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) > t\right\} \le \exp\left(-\frac{1}{16}\frac{d}{n}(C(\delta))^2 \log n\right). \tag{30}$$

Since $r = [\mu n]$, $0 < \mu < 1$, there exists $\lambda > 0$ such that $d/n \doteq \lambda$. Now we take $C^2(\delta) = (16/\lambda)(2 + \delta)(\frac{1}{2} + \delta')$, where δ' is any positive constant. Thus the right-hand side of (30) is $O(n^{-(2+\delta)(1/2+\delta')})$ a.s. So

$$\int_{C(\delta)/\varepsilon_1(\log/n)^{1/2}}^{n^{1/2+\delta'}} t^{1+\delta} P_*\left\{\sqrt{\frac{nr}{d}}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) > t\right\} dt = O(1) \quad \text{a.s.} \tag{31}$$

Since $d/r \doteq C$ (when n is large enough), note that

$$|\xi_{n1}| + |\xi_{n2}| = O(1) \quad \text{a.s.}, \tag{32}$$

where ξ_{n1} is the $[[\mu n]/2]$th-order statistic, and ξ_{n2} is the $\{n - [[\mu n]/2]\}$th order statistic, we have

$$\begin{aligned} &P_*\left\{\sqrt{\frac{nr}{d}}\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) > n^{1/2+\delta'}\right\} \\ &\quad = P_*\left\{\left(F_r^{*-1}(p) - F_n^{-1}(p)\right) > n^{\delta'}\sqrt{\frac{d}{r}}\right\} = 0 \quad \text{a.s.} \end{aligned} \tag{33}$$

Combining (29) and (31) with (33), we can get (26). Thus we complete the proof of Theorem 2. □

Acknowledgment. I am grateful to Professor C. F. Jeff Wu for bringing this problem to my attention and for his encouragement.

REFERENCES

GHOSH, M., PARR, W. C., SINGH, K. and BABU, G. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12** 1130–1135.

KIEFER, J. (1967). On Bahadur's representation of sample quantile. *Ann. Math. Statist.* **38** 1323–1342.

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

SINGH, K. (1980). A note on sample quantile processes of finite population. *Austral. J. Statist.* **22** 358–363.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

WU, C. F. J. (1987). On the asymptotic properties of the jackknife histogram. Technical Report, Dept. Statist., Univ. Wisconsin-Madison.

DEPARTMENT OF STATISTICS
FUDAN UNIVERSITY
SHANGHAI, 200433
PEOPLE'S REPUBLIC OF CHINA