Journal of Mathematical Research & Exposition
Vol. 17, No. 4, 501-507, Nov. 1997

Some Limit Theorems for Kernel-Smooth Quantile Estimators*

Zhou Yong
(Inst. of Appl. Math., Academia Sinica, Beijing 100080)

Abstract Weak convergence and strong consistency of the remainder term in the Bahadur representation of the sample p-quantile are established. From the results we obtain asymptotic normality and the laws of iterated logarithm for smooth quantile estimator.

Keywords Bahadur representation, sample quantile, empirical process and quantile process

Classification AMS(1991) 62E20, 62G30/CCL O212.2, O212.7

1. Introduction

One characteristic of the distribution that is of interest is the quantile function, which is useful in reliability and medical studies.

For the distribution function F, the quantile function is defined by

$$Q(p) = \inf\{x: F(x) \geq p\}, \quad 0 < p < 1.$$

A natural estimator of the quantile function $Q(p)$ is the sample quantile function Q_n defined by

$$Q_n(p) = F_n^{-1}(p) = \inf\{x: F_n(x) \geq p\},$$

where $F_n(\cdot)$ is the empirical distribution function (d.f.).

There are several nonparametric estimators of $Q(p)$ in the literature. For example, the sample quantile function, $F_n^{-1}(p) = \inf\{x: F_n(x) \geq p\}$, $0 \leq p \leq 1$ has been studied, where $F_n(x)$ is the empirical distribution function based on the sample drawing from popular distribution function F. [2] gave many of the known results concerning F_n^{-1}. Another approach has been to solve $\tilde{F}_n(x_p) = p$ for x_p where $\tilde{F}_n(x) = \int_{\infty} f_n(t)dt$ with $f_n(t)$ being a kernel estimator (see [7]). [9] studied a kernel-type estimator which is the smoothed sample quantile function $F_n^{-1}(p)$ based on the kernel method.

The quantile function of the empirical distribution is a step function with jumps corresponding to the observations. The purpose of this paper is to present a smoothed nonpara-

* Received Jan. 14, 1995. This work was supported by the postdoctoral grant of China.

metric estimator for the quantile function based on the kernel method and obtain some asymptotic results for the smoothed quantile estimator. From these results, we can establish Bahadur representation for this smooth quantile estimator with exact convergent order and exact constant in the order. [9] showed that under general conditions this estimator is strongly consistent, and it performs better than the sample quantile function in the sense of smaller mean squared error, particularily when the size of sample is small.

Now, for $0 < p < 1$, define the kernel- type quantile function estimator

$$\tilde{Q}_n(p) = h_n^{-1}\int_0^1 Q_n(t)k\left(\frac{t-p}{h_n}\right)dt = h_n^{-1}\sum_{i=1}^{n} X_{(i)}\int_{(i-1)/n}^{i/n} k\left(\frac{t-p}{h_n}\right)dt, \qquad (1.1)$$

where $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$, $(i = 1, 2, \cdots, n-1)$ are the order statistics of $X_1, X_2, \cdots, X_n$ and $k(t)$ is a probability density function and $h = h_n$ is a sequence of bandwidth.

Let U_i be i. i. d. uniform $(0, 1)$ random variables, and the uniform empirical distribution based on these reduced rv′s is then given by

$$G_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \le y), \qquad y \in (0, 1).$$

where $I(\bullet)$ is the indicator of $(\bullet)$. Let $e_n(t)$ denote the corresponding uniform process

$$e_n(y) = n^{1/2}(G_n(y) - y), \qquad 0 < y < 1.$$

Define q_n, the inverse of G_n, and the uniform empirical process u_n by

$$q_n(y) = \inf\{t: G_n(t) \ge y\},$$

$$u_n(y) = n^{1/2}(y - q_n(y)), \qquad 0 < y < 1.$$

In this paper we shall consider the smoothed quantile process

$$\tilde{\beta}_n(p) = \sqrt{n}\, f(Q(p))(\tilde{Q}_n(p) - Q(p))$$

and the smooth Bahadur- Kiefer process

$$R_n(p) = e_n(Q(p)) + \tilde{\beta}(p), \qquad 0 < p < 1. \qquad (1.2)$$

[1] was the first to investigate the distance between the empirical and quantile processes in the case the sample is coming from the uniform $U(0, 1)$ distribution. The best result concerning this problem, is due to [6], he proved the sharpest order in this distance. In this paper we consider the distance between the empirical and smooth quantile processes for the sample coming from general d. f.

Our main results are the following theorems.

Theorem 1 *Suppose that F is twice differentiable on the neighborhood of $Q(p)$, f is continuous and positive near $Q(p)$ and f' is continuous on $Q(p)$. Let k be a probability density function with finite support $(-c, c)$ for some $c > 0$ and $\int tk(t)dt = 0$. Let $\{h = h_n, n \ge 1\}$ be a sequence of bandwidths satisfying*

$$\sqrt{n}\, h \log h^{-1} \to 0 \quad \text{and} \quad \frac{nh}{\log h^{-1}} \to \infty \tag{1.3}$$

as $n \to \infty$. *Then*

$$\lim P\{n^{\frac{1}{4}} f(Q(p)) \widetilde{R}_n(p) \leq t\} = 2\int_0^\infty \Phi(\frac{t}{u^{1/2}}) N(0, p(1-p)), (du).$$

where $\Phi(\cdot)$ *is Standard normal d. f. and* $N(0, \sigma^2)$ *denotes the normal distribution with expectation zero and variance* σ^2.

Theorem 2 *Suppose that* f *is continuously differentiable and positive on real line. Let* k *be a probability density function with finite support* $(-c, c)$ *for some* $c > 0$ *and* $\int tk(t)dt = 0$. *Let* $\{h = h_n, n \geq 1\}$ *be a sequence of bandwidths satisfying*

$$\frac{\sqrt{n}\, h \log h^{-1}}{(\log\log n)^{3/2}} \to 0 \quad \text{and} \quad \frac{nh}{\log h^{-1}} \to \infty \tag{1.4}$$

as $n \to \infty$. *Then*

$$\limsup_{n\to\infty} n^{1/4} (\log\log n)^{-3/4} |\widetilde{R}_n(p)| = 2^{5/4} 3^{-3/4} (p(1-p))^{\frac{1}{4}} \quad a.s \tag{1.5}$$

for $p \in (0, 1)$ *fixed.*

Theorem 3 *Suppose that* f *is continuously differentiable and positive on real line. Let* k *be a probability density function with finite support* $(-c, c)$ *for some* $c > 0$ *and* $\int tk(t), dt = 0$. *Let* $\{h = h_n, n \geq 1\}$ *be a sequence of bandwidths satisfying*

$$\frac{\sqrt{n}\, h}{(\log h^{-1})^{1/2}} \to 0 \quad \text{and} \quad \frac{nh}{\log h^{-1}} \to \infty \tag{1.6}$$

as $n \to \infty$.

If in addition, that $0 < \inf_{0<x<1} f(Q(x)) < \infty$. Then

$$\limsup_{n\to\infty} n^{1/4} (\log n)^{-1/2} (\log\log n)^{-1/4} \sup_{0\leq p\leq 1} |\widetilde{R}_n(p)| = 2^{1/4} \quad a.s \tag{1.7}$$

From the law of the iterated logarithm for empirical process we immediately get

Corollary 1 *Under all the conditions of Theorem 3 we have*

$$\limsup_{n\to\infty} (\frac{n}{2\log\log n})^{1/2} \sup_{0\leq p\leq 1} f(Q(p)) |\widetilde{Q}_n(p) - Q(p)| = 1 \quad a.s$$

and

$$\liminf_{n\to\infty} n^{\frac{1}{2}} (\log\log n)^{\frac{1}{2}} \sup_{0\leq p\leq 1} f(Q(p)) |\widetilde{Q}_n(p) - Q(p)| = \pi 8^{-1/2} \quad a.s$$

2 Proof of the theorems

For convenience of presentation we shall assume throughout that $q_n(y) = 0$ for $y \le 0$ and $u_n(y) = 0$ for $y \ge 1$. We quote some strong approximation results for the empirical process, which are used in the proofs of our theorems

Let $h = h_n$ for simplity.

Lemma 1 *If $nh/(\log h^{-1}) \to \infty$ as $n \to \infty$, then we have*

$$\sup_{0\le y\le p}\sup_{|t|\le h}\left|K(y+t,n) - K(y,n)\right| = O((nh\log h^{-1})^{1/2}) \quad a.s. \tag{2.1}$$

where define $K(s,t) = 0$ as $s \le 0$ and $s \ge 1$.

Proof See [3] Theorem 1.15.2

Lemma 2 *Assume that the regularity conditions in Theorem 1 are satisfied, then we have*

$$\sup_{0\le y\le 1}\sup_{|s|\le h}\left|u_n(y+s) - u_n(y)\right| = O((h\log h^{-1})^{1/2}) \ a.s. \tag{2.2}$$

Proof From [3] Theorem 4.5.3, we have

$$\begin{aligned} & n^{-1/2}\sup_{0<y\le p}\sup_{|s|\le h}\left|u_n(y+s) - u_n(y)\right| \\ & \le n^{-1/2}\sup_{0<y\le p}\sup_{|s|\le h}\left|u_n(y+s) - n^{-1/2}K(y+s,n)\right| \\ & + n^{-1/2}\sup_{0<y\le p}\left|u_n(y) - n^{-1/2}K(y,n)\right| \\ & + n^{-1}\sup_{0<y\le p}\sup_{|s|\le h}\left|K(y+s,n) - K(y,n)\right| \\ & := A_1 + A_2 + A_3 \end{aligned} \tag{2.3}$$

where

$$A_2 = O(n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}) \quad a.s$$

For sufficiently large n, we obtain

$$A_1 \le n^{-1/2}\sup_{0\le y\le p}\left|u_n(y) - n^{-1/2}K(y,n)\right| = O(n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}) \ a.s$$

By Lemma 1, we obtain

$$A_3 = O((n^{-1}h\log h^{-1})^{1/2}) \ a.s.$$

Thus the result of Lemma 2 follows from the boundedness of A_1, A_2, A_3.

Lemma 3 [[8] **Th. 1A**] *Suppose that $k(t)$ is a bounded integral function on real line and $\lim_{|t|\to\infty}\left|tk(t)\right| = 0$. Define $g_n(x) = \int_{\infty}^{x} g(t)a_n^{-1}k\left(\frac{t-x}{a_n}\right)dt$ Then at every point x of continuity of $g(\cdot)$*

$$\lim_{n\to\infty} g_n(x) = g(x)\int_{\infty}^{\infty} k(t)\,dt,$$

as $g(x)$ is uniform continuous, then the equality above holds true uniformly on x.

Proofs of Theorem 1 and Theorem 2 Let $q_n(p) = \inf\{s: G_n(s) \ge p\} = F(Q_n(p))$, and $\hat{Q}_n(p) = Q(q_n(p))$. For sufficiently large n, we can write

$$\begin{aligned}
|\tilde{R}_n(p) - R_n(p)| &= n^{1/2} f(Q(p)) \left| h^{-1} \int_0^1 Q_n(t) k\left(\frac{p-t}{h}\right) dt - Q_n(p) \right| \\
&= n^{1/2} f(Q(p)) \left| \int_c [Q(q_n(t)) - Q(q_n(p))] \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \right| \\
&= n^{1/2} f(Q(p)) \left| \int_c \frac{1}{f(Q(p))} [q_n(t) - q_n(p)] \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \right. \\
&\quad \left. + \int_c [q_n(t) - q_n(p)]^2 \frac{f'(Q(\xi_t))}{f^3(Q(\xi_t))} \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \right|. \\
&= I_1(p) + I_2(p). \qquad (2.4)
\end{aligned}$$

where ξ_t lies between $q_n(p - ht)$ and $q_n(p)$.

By Lemma 2.1 and the fact that $\int y k(y) dy = 0$, we have

$$\begin{aligned}
I_1(p) &= \left| \int_c (u_n(p - ht) - u_n(p)) k(t) dt \right| \le \sup_{|t| \le c} |u_n(p - ht) - u_n(p)| \int_c k(t) dt \\
&= \sup_{|t| \le ch} |u_n(p + t) - u_n(p)| = O((h \log h^{-1})^{1/2}) \quad a.s. \qquad (2.5)
\end{aligned}$$

By the C_r-inequality, we have

$$\begin{aligned}
I_2(p) \le\ & 2n^{-\frac{1}{2}} \int [u_n(t) - u_n(p)]^2 \frac{|f'(Q(\xi_t))|}{f^3(Q(\xi_t))} \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \\
& + 2n^{\frac{1}{2}} \int [t - p]^2 \frac{|f'(Q(\xi_t))|}{f^3(Q(\xi_t))} \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \qquad (2.6)
\end{aligned}$$

Using Lemma 1 and Lemma 2 we obtain that the first integration of (2.6) is not greater than

$$\begin{aligned}
& n^{-\frac{1}{2}} \sup_{0<t<1} |u_n(t) - u_n(p)|^2 \int \frac{|f'(Q(\xi_t))|}{f^3(Q(\xi_t))} \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \\
&= O(n^{-\frac{1}{2}} h \log h^{-1}) \quad a.s.
\end{aligned}$$

it then follows from Lemma 2 that

$$\int \left(\frac{t-p}{h}\right)^2 \frac{|f'(Q(\xi_t))|}{f^3(Q(\xi_t))} \frac{1}{h} k\left(\frac{p-t}{h}\right) dt \to \frac{f'(Q(p))}{f^3(Q(p))} \int t^2 k(t) dt.$$

Hence the third integration of (2.6) is not greater than $O(n^{\frac{1}{2}} h^2)$. Thus we obtain that

$$I_2(p) = O(n^{-1/2} h \log h^{-1} \vee n^{1/2} h^2) \quad a.s. \qquad (2.7)$$

Since $\sqrt{n}\, h \log h^{-1} \to 0$ $(n \to \infty)$, then we have

$$n^{\frac{1}{4}} I_1(p) \to 0 \quad a.s. \qquad \text{and} \qquad n^{\frac{1}{4}} I_2(p) \to 0 \quad a.s.$$

By Theorem 1 of [4] we immediately obtain Theorem 1.

Since

$$\frac{\sqrt{nh} \log n}{(\log\log n)^{3/2}} \to 0 \quad \text{and} \quad \frac{nh}{\log h^{-1}} \to \infty,$$

as $n\to\infty$, we obtain from (2.5) and (2.7) that

$$n^{1/4}(\log\log n)^{-3/4}I_1\to 0 \quad \text{a.s.} \tag{2.8}$$

and

$$n^{1/4}(\log\log n)^{-3/4}I_2\to 0 \quad \text{a.s.} \tag{2.9}$$

Thus, it follows from (2.4), (2.5) and (2.7) that

$$n^{1/4}(\log\log n)^{-3/4}\left|\widetilde{R}_n(p) - R_n(p)\right|\to 0 \quad \text{a.s.} \tag{2.10}$$

Therefore, Theorem 2 follows from (2.10) and [6] Theorem 1.

Proof of Theorem 3 We can prove Theorem 3 by the same argument used in the proof of Theorem 2, we may write

$$\sup_{0\le p\le 1}\left|\widetilde{R}_n(p) - R_n(p)\right| \le \sup_{0< p\le 1} I_1(p) + \sup_{0< p\le 1} I_2(p). \tag{2.11}$$

By assumption (1.14) and $0 < \inf_{0\le x\le 1} f(Q(x)) < \infty$, we have

$$\sup_{0\le p\le 1}\int\frac{\left|f'(Q(\xi_t))\right|}{f^3(Q(\xi_t))}k(t)\,dt \le \sup_{0\le p\le 1}\frac{\left|f'(Q(p))\right|}{f^3(Q(p))} < \infty.$$

By Lemma 1, it follows that

$$\sup_{0\le p\le 1} I_2(p) \le n^{-\frac{1}{2}}\sup_{0\le p\le 1}\sup_{|t|\le c}\left|q_n(p - ht) - q_n(p)\right|^2 f(Q(p))\int_c^c\frac{\left|f'(Q(\xi_t))\right|}{f^3(Q(\xi_t))}k(t)\,dt$$

$$= O(n^{-\frac{1}{2}}h\log h^{-1}\vee n^{1/2}h^2) \quad \text{a.s.}$$

By Lemma 1, we obtain

$$\sup_{0\le p\le 1} I_1(p) \le \sup_{0\le p\le 1}\sup_{|t|\le ch}\left|u_n(p + t) - u_n(p)\right| = O((h\log(n/h))^{1/2}) \quad \text{a.s.}$$

Since

$$\frac{\sqrt{n}\,h\log h^{-1}}{\log n}\to 0 \quad \text{and} \quad \frac{nh}{\log h^{-1}}\to\infty,$$

we obtain

$$n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4}\sup_{0\le p\le 1} I_1(p)\to 0 \quad \text{a.s.}$$

and

$$n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4}\sup_{0\le p\le 1} I_2(p)\to 0 \quad \text{a.s.}$$

Therefore, we have

$$n^{1/4}(\log n)^{-1/4}(\log\log n)^{-1/4}\sup_{0\le s\le 1}\left|\widetilde{R}_n(p) - R_n(p)\right|\to 0 \quad \text{a.s.} \tag{2.12}$$

[3] Theorem 5.2.2, the result of Theorem 3 follows from (2.12)

References

[1] R. R. Bahadur, *A note on quantiles in large samples*, Ann. Math. Statist., **37** (1966), 577-580.

[2] M. Csörgö, *Quantile Process and their Application in Statistics*, SIAM, Philadelphia PA, 1983.

[3] M. Csörgo and P. Révész, *Strong Approximation in Probability and Statistics*, Academic Press, 1981.

[4] M. Falk, *Weak convergence of the remainder term in the Bahadur representation of extreme quantile*, Statist. Probab. Lett., **8** (1990), 47-50.

[5] J. Kiefer, *On Bahadur's representation of sample quantile*, Ann. Math. Statist., **38** (1967), 1323-1342.

[6] J. Kiefer, *Deviations between the sample quantile and the sample D. F.*, Nonparametric Techniques in Statist. Inference (M. L. Puri, Ed) 299-319, Cambridge Univ. Press London., 1970.

[7] E. A. Nadaraya, *Some new estimates for distribution functions*, Theory Probab. Appl., **9** (1964), 497-500.

[8] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., **33** (1962), 1065-1076.

[9] S. S. Yang, *A smooth nonparametric estimator of aquantile function*, J. Amer. Statist. Assoc., **80** (1985), 1004-1011.