The Annals of Statistics
1985, Vol. 13, No. 1, 428–433

ASYMPTOTIC NORMALITY OF THE KERNEL QUANTILE ESTIMATOR

BY MICHAEL FALK

Universität Gesamthochschule-Siegen

Multidimensional asymptotic normality of the kernel quantile estimator is established under fairly general conditions on the underlying distribution function and on the kernel.

Sharpening these assumptions, one can utilize the proof to achieve also a bound for the rate of convergence which entails the comparison of the kernel estimator with the empirical quantile on the basis of their covering probabilities.

1. Introduction and main result. Let P be a probability measure on the real line with distribution function ($\equiv$ df) F. If F is smooth near the q-quantile $F^{-1}(q)$, $q \in (0, 1)$, then one might hope that averaging over order statistics close to the sample q-quantile $F_n^{-1}(q)$ leads to estimators of better performance.

Consequently, it was shown in Falk (1984) that under suitable conditions on F, $\alpha_n > 0$ and $k: \mathbb{R} \to \mathbb{R}$, a kernel type estimator of the form

$$\hat{q}_n(F_n) := \alpha_n^{-1} \int_0^1 F_n^{-1}(x) k((q - x)/\alpha_n)\, dx \tag{1.1}$$

of the q-quantile beats the sample q-quantile on the level of deficiency if (and only if) the number

$$\psi(k) := 2 \int x k(x) K(x)\, dx \tag{1.2}$$

is positive; $K(x) := \int_{-\infty}^{x} k(y)\, dy$.

The kernel quantile estimator for a particular choice of kernels was considered by Harrell and Davis (1982) who also made some empirical studies (see also Kaigh and Lachenbruch, 1982).

Since the above mentioned comparison between $\hat{q}_n(F_n)$ and $F_n^{-1}(q)$ is based on their mean square errors, we had to compute in Falk (1984) moments of these two estimators, and the crucial point for the derivation of these results was the tail behaviour of F.

On the other hand it is well-known that asymptotic normality of the sample q-quantile is achieved under only local assumptions on the underlying df which stimulates the analogous investigation of the kernel estimator.

Therefore, we establish in Theorem 1.3 of the present article multidimensional asymptotic normality of the kernel estimator under fairly general assumptions

Received January 1984; revised September 1984.
AMS 1980 *subject classifications*. Primary 60F05; secondary 62G05, 62G20.
Key words and phrases. Kernel estimator, empirical quantile, central limit theorem, covering probability.

on F and k, i.e. the conditions on F fit completely with the standard assumptions to prove asymptotic normality of the joint distribution of empirical quantiles (see Theorem B on page 80 of Serfling, 1980).

Imposing further conditions on F and k and following the lines of the proof of Theorem 1.3, one can easily derive a bound for the rate at which the distribution of $\hat{q}_n(F_n)$ tends to its limit (see Proposition 1.5 below). Furthermore, in Section 2 of the present article these results will prove useful for the comparison of covering probabilities of the kernel estimator and of the sample q-quantile.

Our main result is the following one. By P^n we denote the n-fold independent product of P, by $P^n * g$ the measure induced by P^n and a measurable function g, and by $\rightarrow_w$ weak convergence.

1.3 THEOREM. *Let $0 < q_1 < \cdots < q_r < 1$. Suppose that F^{-1} has a bounded derivative near q_i which is continuous at q_i, $1 \le i \le r$. Then, if k_i has bounded support, $\int k_i(x)\,dx = 1$ and $0 < \alpha_{in} \rightarrow_{n \in \mathbb{N}} 0$, $1 \le i \le r$, we have*

$$P^n * (n^{1/2}(\hat{q}_{in}(F_n) - \hat{q}_{in}(F))_{i=1}^r) \rightarrow_w N_{(\mathbf{0},\Sigma)}$$

where $\hat{q}_{in}(F_n) = \alpha_{in}^{-1} \int_0^1 F_n^{-1}(x) k_i((q_i - x)/\alpha_{in})\,dx$, $1 \le i \le r$, and $N_{(\mathbf{0},\Sigma)}$ denotes the r-dimensional normal distribution with mean $\mathbf{0}$ and symmetric covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^r$ with $\sigma_{ij} = (F^{-1})'(q_i)(F^{-1})'(q_j)q_i(1 - q_j)$, $i \le j$.

1.4 REMARKS. Notice that the limiting normal distribution for the kernel estimators is the same as for the joint distribution of empirical quantiles. Furthermore, assume that F^{-1} has a bounded $(m + 1)$th derivative near q, $m \ge 0$. Then, if k has bounded support, $\int k(x)\,dx = 1$ and $\int x^i k(x)\,dx = 0$, $i = 1, \cdots, m$, Taylor's formula implies $|\hat{q}_n(F) - F^{-1}(q)| = O(\alpha_n^{m+1})$. Hence, under additional assumptions on α_{in} (which depend on the smoothness of F^{-1}) and on k we may replace $\hat{q}_{in}(F)$ in the preceding result by $F^{-1}(q)$.

PROOF OF THEOREM 1.3. Denote by Q the uniform distribution on (0, 1). Since $Q * F^{-1} = P$, we have for n large

$$\begin{aligned} &P^n * (n^{1/2}(\hat{q}_{in}(F_n) - \hat{q}_{in}(F))_{i=1}^r) \\ &= P^n * \left(n^{1/2}\left(\int_{-c}^{c} k_i(x)\, \langle F_n^{-1}(q_i - \alpha_{in}x) - F^{-1}(q_i - \alpha_{in}x) \rangle\, dx \right)_{i=1}^r \right) \\ &= Q^n * \left(n^{1/2}\left(\int_{-c}^{c} k_i(x) \langle F^{-1}(F_n^{-1}(q_i - \alpha_{in}x)) - F^{-1}(q_i - \alpha_{in}x) \rangle\, dx \right)_{i=1}^r \right) \end{aligned}$$

where the support of k_i is contained in $[-c, c]$. Now, a suitable Bahadur approximation argument (see Kiefer, 1970) together with the continuity of $(F^{-1})'$ at q_i imply that under Q^n the function

$$n^{1/2} \left(\int_{-c}^{c} k_i(x) \langle F^{-1}(F_n^{-1}(q_i - \alpha_{in}x)) - F^{-1}(q_i - \alpha_{in}x) \rangle\, dx \right)_{i=1}^r$$

is asymptotically equivalent to

$$n^{1/2}\left(\int_{-c}^{c} k_i(x)\langle q_i - \alpha_{in}x - F_n(q_i - \alpha_{in}x)\rangle (F^{-1})'(q_i - \alpha_{in}x)\,dx\right)_{i=1}^{r}$$
$$= n^{1/2}\textstyle\sum_{j=1}^{n}\left(\displaystyle\int_{-c}^{c} k_i(x)\langle q_i - \alpha_{in}x - 1_{(0,q-\alpha_{in}x]}(\pi_j)\rangle (F^{-1})'(q_i - \alpha_{in}x)\,dx\right)_{i=1}^{r}$$

where $\pi_j(\mathbf{x}) = x_j$ denotes the jth projection and 1_A the indicator function of a set A. Hence, the assertion is immediate from elementary computations and classical central limit theory.

Imposing further conditions on F, the following result can easily be derived along the lines of the preceding proof. Hence we present it without proof.

1.5 PROPOSITION. *Suppose that F^{-1} has a bounded second derivative near $q \in (0, 1)$ and $(F^{-1})'(q) > 0$. Then, if k has bounded support, $\int k(x)\,dx = 1$ and $0 < \alpha_n \to_{n\in\mathbb{N}} 0$ we have*

$$\sup_{t\in\mathbb{R}} |P^n\{n^{1/2}(\hat{q}_n(F_n) - \hat{q}_n(F)) \le t\sigma_n\} - \Phi(t)| = O(\log(n)n^{-1/4})$$

where

$$\sigma_n^2 := \int_0^1 \left\{\int_{-c}^{c} k(x)[q - \alpha_n x - 1_{(0,q-\alpha_n x]}(y)](F^{-1})'(q - \alpha_n x)\,dx\right\}^2 dy$$
$$\to_{n\in\mathbb{N}} (F^{-1})'^2(q)q(1-q)$$

and Φ denotes the standard normal df.

Notice that bounds for the rate of convergence of $\hat{q}_n(F_n)$ can also be established by utilizing the differential approach to L-statistics due to Boos (1979) (see also Sections 6 and 8.2.4 in Serfling, 1980). However, this leads to rather restrictive differentiability assumptions on the kernel k.

2. Comparison of covering probabilities. In Falk (1984) we evaluated the relative deficiency of the sample q-quantile with respect to the kernel estimator based on their mean square errors and in order to keep these values finite we had to assume that the df F satisfies $\lim_{t\to\infty} t^{\delta}(1 - F(t) + F(-t)) = 0$ for some $\delta > 0$. Since this is a necessary and sufficient growth condition (see Theorem 2.2 of Bickel, 1967) we cannot dispense with it and hence, the tails of the underlying df decide whether at all this measure of performance is finite.

To compare $F_n^{-1}(q)$ and $\hat{q}_n(F_n)$ under only local assumptions on F, we investigate in this section their covering probabilities of intervals symmetric to $F^{-1}(q)$ and to this end, Proposition 1.5 will prove useful.

Therefore, define for any $t > 0$ the interval

$$I_t(n) := [F^{-1}(q) - t\sigma n^{-1/2}, F^{-1}(q) + t\sigma n^{-1/2}] \tag{2.1}$$

where $\sigma^2 := (F^{-1})'^2(q)q(1-q)$ and the number $i_t(n)$ by

$$i_t(n) = \min\{m \in \mathbb{N}: P^m\{F_m^{-1}(q) \in I_t(n)\} \ge P^n\{\hat{q}_n(F_n) \in I_t(n)\}\}. \tag{2.2}$$

The number $i_t(n)$-n denotes the least number of observations which are additionally needed such that the distribution of the sample q-quantile is as concentrated around the true q-quantile as that of the kernel estimator based on n observations.

A similar comparison between the empirical quantile and a linear combination of finitely many order statistics was carried out by Reiss (1980), Section 3.

The following result shows that the sequence $i_t(n) - n$, $n \in \mathbb{N}$, quickly tends to infinity if $\psi(k)$ is positive.

2.3 THEOREM. *Assume that F^{-1} has a bounded $(m+1)$th derivative near q, $m \geq 2$, and $(F^{-1})'(q) > 0$. Let k have finite support and satisfy $\int k(x)\,dx = 1$, $\int x^i k(x)\,dx = 0$, $i = 1, \cdots, m$. Then, if $n\alpha_n^4/\log^4(n) \to_{n\in\mathbb{N}} \infty$ and $n\alpha_n^{2m+1} \to_{n\in\mathbb{N}} 0$ we have for any $t > 0$*

$$\lim_{n\in\mathbb{N}} \alpha_n^{-1}(P^n\{\hat{q}_n(F_n) \in I_t(n)\} - P^n\{F_n^{-1}(q) \in I_t(n)\}) = t\varphi(t)\psi(k)/(q(1-q)) \tag{2.4}$$

where $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Moreover,

$$\liminf_{n\in\mathbb{N}} (i_t(n) - n)/(n\alpha_n) \geq \psi(k)/(q(1-q)). \tag{2.5}$$

This result again suggests the number $\psi(k)$ as a measure of asymptotic performance within the class of kernels and its sign decides whether one does better with the kernel estimator or with the empirical quantile. Thus, Theorem 2.3 fits completely with the results derived in Falk (1983) and (1984), where asymptotically optimal kernels are given also (see also Mammitzsch, 1984, for details).

PROOF. First notice that under the assumptions of Theorem 2.3

$$\sigma_n^2 = (F^{-1})'^2(q)\{q(1-q) - \alpha_n\psi(k)\} + o(\alpha_n) \tag{2.6}$$

where σ_n^2 is defined in Proposition 1.5. Furthermore, Proposition 1.5 implies

$$P^n\{\hat{q}_n(F_n) \in I_t(n)\} = \Phi(t\sigma\sigma_n^{-1} + d_n) - \Phi(-t\sigma\sigma_n^{-1} + d_n) + O(\log(n)n^{-1/4}) \tag{2.7}$$

where $d_n := n^{1/2}(F^{-1}(q) - \hat{q}_n(F))/\sigma_n$. Furthermore, Taylor's formula yields with $t_n := t\sigma\sigma_n^{-1}$

$$\Phi(t_n + d_n) - \Phi(t) - (\Phi(-t_n + d_n) - \Phi(-t)) = 2\varphi(t)(t_n - t) + O(d_n^2 + (t_n - t)^2). \tag{2.8}$$

Now, (2.6) implies

$$t_n - t = t\{\alpha_n(F^{-1})'^2(q)\psi(k) + o(\alpha_n)\}/\{2(F^{-1})'^2(q)q(1-q) + o(1)\}$$

and thus, $\lim_{n\in\mathbb{N}}(t_n - t)/\alpha_n = t\psi(k)/(2q(1-q))$. Moreover, Remark 1.4 implies

$d_n = O(n^{1/2}\alpha_n^{m+1})$ and hence, $d_n^2/\alpha_n = O(n\alpha_n^{2m+1})$. Altogether we have

$$\lim_{n\in\mathbb{N}}\alpha_n^{-1}\{\Phi(t+d_n) - \Phi(-t_n+d_n) - \Phi(t) + \Phi(-t)\} \tag{2.9}$$
$$= t\varphi(t)\psi(k)/(q(1-q)).$$

Now the Berry-Esseen Theorem for sample q-quantiles (see Theorem 2.7 in Reiss, 1976) together with (2.7) and (2.9) imply (2.4).

Furthermore, Theorem 2.7 in Reiss (1976) yields for $t > 0$, $m \in \mathbb{N}$

$$P^m\{F_m^{-1}(q) \in I_t(n)\} = \int_{-t(m/n)^{1/2}}^{t(m/n)^{1/2}} \varphi(x)\,dx + O(m^{-1}) \tag{2.10}$$

which implies $\liminf_{n\in\mathbb{N}} i_t(n)/n \geq 1$. Moreover, by definition of $i_t(n)$

$$\alpha_n^{-1}\langle P^{i_t(n)}\{F_{i_t(n)}^{-1}(q) \in I_t(n)\} - P^n\{\hat{q}_n(F_n) \in I_t(n)\}\rangle \geq 0$$

and hence,

$$\alpha_n^{-1}\left\{\int_{-t(i_t(n)/n)^{1/2}}^{t(i_t(n)/n)^{1/2}} \varphi(x)\,dx - \int_{-t}^{t}\varphi(x)\,dx\right\} \geq t\varphi(t)\psi(k)/(q(1-q)) + o(1)$$

yielding $\liminf_{n\in\mathbb{N}}(i_t(n)^{1/2} - n^{1/2})/(n^{1/2}\alpha_n) \geq \psi(k)/(2q(1-q))$.

Since

$$\begin{aligned}\liminf_{n\in\mathbb{N}}&(i_t(n)^{1/2} - n^{1/2})/(n^{1/2}\alpha_n)\\ &= \liminf_{n\in\mathbb{N}}(i_t(n) - n)/\{n\alpha_n((i_t(n)/n)^{1/2} + 1)\}\\ &\leq \liminf_{n\in\mathbb{N}}(i_t(n) - n)/(2n\alpha_n),\end{aligned}$$

the assertion of Theorem 2.3 is complete.

REFERENCES

BICKEL, P. J. (1967). Some contributions to the theory of order statistics. *Proc. 5th Berkeley Symp.* **I** 575–591.

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959.

FALK, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neerlandica* **37** 73–83.

FALK, M. (1984). Relative deficiency of kernel type estimators of quantiles. *Ann. Statist.* **12** 261–268.

HARREL, F. E. and DAVIS, C. E. (1982). A new distribution-free quantile estimator. *Biometrika* **69** 635–640.

KAIGH, W. D. and LACHENBRUCH, P. A. (1982). A generalized quantile estimator. *Comm. Statist. A–Theory Methods* **11** 2217–2238.

KIEFER, J. (1970). Deviations between the sample quantile process and the sample df. *Proc. Conference Nonparametric Techniques Statist. Infer.* 299–319. Bloomington (ed. by M. L. Puri). Cambridge University Press.

MAMMITZSCH, V. (1984). On the asymptotically optimal solution within a certain class of kernel type estimators. *Statist. Decisions,* **2** 247–255.

REISS, R.-D. (1976). Asymptotic expansions for sample quantiles. *Ann. Probab.* **4** 249–258.

REISS, R.-D. (1980). Estimation of quantiles in certain nonparametric models. *Ann. Statist.* **8** 87–105.

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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SIEGEN
HÖLDERLINSTR. 3
5900 SIEGEN 21
WEST GERMANY