The Annals of Statistics
1995, Vol. 23, No. 6, 2285–2291

UNIFORM CONSISTENCY OF GENERALIZED KERNEL ESTIMATORS OF QUANTILE DENSITY

BY C. CHENG

Johns Hopkins University

Various smoothing methods for quantile density estimation are unified into a generalized kernel smoothing. Based on a stochastic upper bound of the derivatives sequence for a sequence of smoothed Brownian bridges, uniform in-probability consistency of generalized kernel quantile density estimators on any closed subinterval of the open unit interval is derived.

1. Introduction. Let X be an absolutely continuous random variable with cdf F and pdf f. Associated with F is the quantile function (qf) $Q(u) = F^{-1}(u) := \inf\{x\colon F(x) \geq u\}$, $u \in [0,1]$. Once $f(x) > 0$ for all $x \in (x_F, x^F)$ with $x_F = \sup\{x\colon F(x) = 0\}$ and $x^F = \inf\{x\colon F(x) = 1\}$, Q is differentiable on the open unit interval and $q(u) := Q'(u)$, $u \in (0,1)$, is the *quantile density* function (qdf).

Given a random sample $X_1, \dots, X_n$ of X, let $\tilde{F}_n$ be the empirical distribution function (EDF) and define $\tilde{Q}_n(u) := \tilde{F}_n^{-1}(u)$, $u \in [0,1]$, to be the empirical quantile function (EQF). The difference between the uth sample and true quantiles, $\tilde{Q}_n(u) - Q(u)$, is asymptotically normal with zero mean and variance $[q(u)]^2 u(1-u)/n$. The efforts on estimating the qdf q has been motivated by constructing confidence intervals for the population quantiles based on this asymptotic normality. [See Csörgő and Horváth (1989), for an alternative approach.] Histogram estimators of the qdf q were studied by Siddiqui (1960), Bloch and Gastwirth (1968), Bofinger (1975) and Falk (1986). Parzen (1979) introduced convolution kernel estimators, which were subsequently studied by Falk (1986) and Csörgő, Deheuvels and Horváth (1991).

The various smoothing methods applicable to qf and qdf estimation [in addition to the references already mentioned, see Kaigh and Cheng (1991), Vitale (1975), Gawronski (1985), Cheng (1995) and Schoenberg (1965)] can be unified into an integral transform of the EQF $\tilde{Q}_n$ with respect to some kernel:

$$\begin{aligned} \hat{Q}_n(u) &= \int_0^1 \tilde{Q}_n(t) K_n(u,t)\, d\mu_n(t), \\ \hat{q}_n(u) &:= \frac{d}{du}\hat{Q}_n(u) = \frac{d}{du}\int_0^1 \tilde{Q}_n(t) K_n(u,t)\, d\mu_n(t), \qquad u \in (0,1), \end{aligned} \tag{1.1}$$

Received December 1993; revised April 1995.
AMS 1991 *subject classifications*. Primary 62G07, 62G20; secondary 62G05, 62G30.
Key words and phrases. Quantile density function, approximation, smoothing, kernel.

where the measure μ_n and the kernel K_n satisfy appropriate variational properties (cf. Section 2), and K_n is so chosen that $\hat{Q}_n$ is differentiable for (almost) all samples. Then $\hat{q}_n$ is a natural estimator of the qdf $q = Q'$.

This generalized kernel formulation provides a unified treatment of many smoothing methods applicable to qf and qdf estimation. In this paper the uniform in-probability consistency of the qdf estimator $\hat{q}_n(\cdot)$ on any fixed closed subinterval in $(0,1)$ is established. The main result, Theorem 2.1, identifies a particular functional of the smoothing kernel that determines the rate of the stochastic bound of the estimation error. The result is applicable to a wide range of smoothing schemes.

Section 2 contains the main theorem and illustrative examples; Section 3 contains lemmas and proofs required to establish the theorem.

2. The main theorem. Throughout the sequel $U = [a, b]$ is an arbitrarily fixed subinterval of $(0,1)$. The main result, to be proved in the next section, follows from several regularity conditions on the qdf and the smoothing kernel which are given below.

Q_1 (SMOOTHNESS). The qdf $q(\cdot)$ is twice differentiable on $(0,1)$.

Q_2 (CONTROLLED TAIL). There is a $\gamma > 0$ such that $\sup_{u \in (0,1)} u(1-u)|J(u)| \le \gamma$, with $J(u) = d \log q(u)/du$.

Q_3 (TAIL MONOTONICITY). Either $q(0) < \infty$ or $q(u)$ is nonincreasing in some interval $(0, u_*)$, and either $q(1) < \infty$ or $q(u)$ is nondecreasing in some interval $(u^*, 1)$.

K_1. For each n, $0 < \mu_n([0,1]) < \infty$ (but may depend on n), and $\mu_n(\{0,1\}) = 0$.

K_2. For each n and each (u,t), $K_n(u,t) \ge 0$, and, for each $u \in U$, $\int_0^1 K_n(u,t)\, d\mu_n(t) = 1$.

K_3. For each n, $\int_0^1 t K_n(u,t)\, d\mu_n(t) = u$, $u \in U$.

K_4. There is a sequence $\delta_n \downarrow 0$ such that $\sup_{u \in U} |\int_{u-\delta_n}^{u+\delta_n} K_n(u,t)\, d\mu_n(t) - 1| \downarrow 0$, as $n \uparrow \infty$.

The rest of the conditions concern the derivative $K_n'(u,t) := \partial K_n(u,t)/\partial u$. Let S_n be the (unique) closed subset of $(0,1)$ such that $\mu_n((0,1) \setminus S_n) = 0$ and $\mu_n((0,1) \setminus S_n') > 0$ for any $S_n' \subset S_n$. For the sequence δ_n in K_4, let $I_n(u) := [u - \delta_n, u + \delta_n]$, $I_n^c(u) = (0,1) \setminus I_n(u)$, for $u \in U$. Define $\Lambda(u; K_n) := \int_{I_n(u)} |K_n'(u,t)|\, d\mu_n(t)$, $u \in U$; and, for a well-defined function g on $(0,1)$, let $R(g; K_n) := \sup_{u \in U} \int_{I_n^c(u)} |g(t) K_n'(u,t)|\, d\mu_n(t)$.

K_5. For each n, $\sup_{u\in U}\int_0^1 |K_n'(u,t)|\,d\mu_n(t) < \infty$ (but may depend on n).

K_6. (a) For each n and each $u \in U$, $K_n(u,t) \equiv 0$, $t \in I_n^c(u)$; or (b) $S_n \subseteq [\varepsilon, 1-\varepsilon] \subset (0,1)$, with $U \subset [\varepsilon, 1-\varepsilon]$ for some $0 < \varepsilon < \frac{1}{2}$.

K_7. For the δ_n sequence in K_4, $\delta_n^2 \sup_{u\in U} \Lambda(u; K_n) \to 0$ and $R(1; K_n) \to 0$ as $n \uparrow 0$.

For a function g on $(0,1)$, let $M_g := \sup_{u\in U}|g(u)|$. Let $\tilde{q}_n(u) := \int_0^1 Q(t)K_n'(u,t)\,d\mu_n(t)$. Then $d_n := \sup_{u\in U}|\bar{q}_n(u) - q(u)|$ is the deterministic error of the estimator $\hat{q}_n(u)$ in estimating the qdf q (cf. Lemma 3.2). The main result is the following theorem.

THEOREM 2.1. *Under conditions* Q_1–Q_3 *and* K_1–K_7, *the estimator* $\hat{q}_n(u)$ *is uniformly in-probability consistent on* U: $\sup_{u\in U}|\hat{q}_n(u) - q(u)| = O_p(B(q;K_n) + d_n)$, *as* $n \uparrow \infty$, *where*

$$B(q;K_n) = n^{-1/2}\Big[M_q \Lambda_n^* \sqrt{2\delta_n \log \delta_n^{-1}} + M_{q'} + C_0 M_q n^{-1/2} A_\gamma(n)\Lambda_n^*\Big],$$

with $\Lambda_n^* = \sup_{u\in U}\Lambda(u;K_n)$, C_0 *a universal constant and* $n^{-\delta}A_\gamma(n) = o(1)$ *for any* $\delta > 0$.

For illustration, consider first the familiar convolution case

$$K_n(u,t)\,d\mu_n(t) = h_n^{-1}K((t-u)/h_n)\,dt$$

with $K(\cdot)$ a differentiable and symmetric pdf on $[-1,1]$ and $h_n \downarrow 0$. For K_4, $\delta_n = h_n$; for K_7, $\Lambda(u;K_n) = h_n^{-1}\alpha(K)$, with $\alpha(K) = \int_{-1}^1 |K'(x)|\,dx$, and $R(1;K_n) \equiv 0$ for sufficiently large n. Once $n^{-1/2}h_n^{-1} = n^{-\tau} \downarrow 0$, the dominating term in $B(q;K_n)$ is $M_q\alpha(K)h_n^{-1}\sqrt{2h_n^{-1}\log h_n^{-1}}$. It is well-known that, under Q_1, the deterministic error $d_n = O(h_n^2)$ for a second-order kernel K. Hence the best rate of the stochastic bound is $O((n^{-1}\log n)^{2/5})$. For a further illustration, consider the following example.

EXAMPLE 2.1 (Boundary-modified Bernstein polynomial). Let ε be such that $U \subset [\varepsilon, 1-\varepsilon] \subset (0,1)$. The k-degree boundary-modified Bernstein polynomial qdf estimator on U is, for an appropriate kernel $b_k(u,t)$ and μ_k,

$$\begin{aligned}\hat{q}_n^B(u) &:= \frac{d}{du}\int_0^1 \tilde{Q}_n(t)\,b_k(u,t)\,d\mu_k(t)\\ &= \frac{1}{L_\varepsilon^k}\sum_{j=0}^{k-1}\frac{\tilde{Q}_n(t_{j+1}) - \tilde{Q}_n(t_j)}{1/k}\binom{k-1}{j}(u-\varepsilon)^j(1-\varepsilon-u)^{k-1-j},\end{aligned}$$

where $L_\varepsilon = 1 - 2\varepsilon$ and $t_j = \varepsilon + (j/k)L_\varepsilon$, $j = 0,1,\dots,k$. Let $k = k_n \uparrow \infty$ as $n \uparrow \infty$. Conditions K_1–K_3 and K_6 can be easily verified. For condition K_4, Laplace formula [Lorentz (1986), pages 15–18] implies $\delta_n = k^{-1/2+\delta}$, $0 < \delta < \frac{1}{6}$. Next,

$$\begin{aligned}\Lambda(u; K_n) &= \sum_{t_j \in I_n(u)} \left| \frac{db_k(u, t_j)}{du} \right| \\ &= \frac{k-1}{L_\varepsilon} \sum_{t_j \in I_n(u)} \binom{k-1}{j-1} v^{j-1}(1-v)^{k-1-j} \left| \frac{(j/k) - v}{(j/k)(1-j/k)} \right|,\end{aligned}$$

where $v = (u - \varepsilon)/L_\varepsilon$. Thus $\Lambda(u; K_n) \sim (k-1)\delta_n/[L_\varepsilon^2 v(1-v)] = O(k^{1/2+\delta})$ for each $u \in U$ because $|(j/k) - v| \le \delta_n/L_\varepsilon$ for $t_j \in I_n(u)$. Similarly, by Lorentz [(1986), equation (7), page 15], $R(1; K_n) = O(A_s k^{1-2s\delta}) = o(1)$ (choose $s > 1/2\delta$) as $n \uparrow \infty$. So K_5 and K_7 are verified.

Further calculations using Taylor expansion show that the deterministic error $d_n = \sup_{u \in U} |\bar{q}_k(u) - q(u)| \sim a(\varepsilon)k^{-1}$. Assume that $k = k_n \le n$, for each n. Then

$$\begin{aligned}&\sup_{u \in U} \left| \hat{q}_n^B(u) - q(u) \right| \\ &\quad = O_p\Big(C(\varepsilon, M_q) n^{-1/2} k^{(1/4)+(3\delta/2)} \sqrt{\log k^{(1/4)-(\delta/2)}} + a(\varepsilon)k^{-1} \Big).\end{aligned}$$

So the best rate of the stochastic bound is $O(n^{-1}\log n)^{2/(5+6\delta)}$. Because $\delta > 0$ but can be arbitrarily small, this rate is slightly slower than that for the second-order kernels. However, Cheng (1995) shows interesting and desirable oscillation properties of the Bernstein polynomial smoothing in finite samples, which in general are not provided by convolution kernels.

3. Lemmas and proofs. The proof of Theorem 2.1 is divided into several lemmas. The previously defined notation continues to be used below.

Define $\bar{g}_n(u) = (\mathbb{K}_n g)(u) := \int_0^1 g(t) K_n(u, t)\, d\mu_n(t)$ and $\bar{g}_n'(u) := d\bar{g}_n(u)/du$, $u \in U$, for a well-defined function g on $(0, 1)$. Conditions K_5, K_6 and Billingsley [(1986), Theorem 16.8] imply that, for each n, $\bar{g}_n'(u) = \int_0^1 g(t) K_n'(u, t)\, d\mu_n(t)$, $u \in U$.

Arguments using Taylor expansion and Billingsley [(1986), Theorem 16.8] establish the following lemma.

LEMMA 3.1. *Let g be a twice continuously differentiable function on* $(0, 1)$. *Then under conditions* K_1–K_7, $\bar{g}_n$ *and* $\bar{g}_n'$ *approximate g and its derivative g' simultaneously on U in the sense* $\sup_{u \in U} |\bar{g}_n(u) - g(u)| \to 0$ *and* $\sup_{u \in U} |\bar{g}_n'(u) - g'(u)| \to 0$, *as $n \to \infty$.*

Turning to the qdf estimator in (1.1), earlier argument implies that, with probability 1, $\hat{q}_n(u) = \int_0^1 K_n'(u, t)\, d\mu_n(t)$, $u \in U$, for each n. Let $\bar{q}_n(u) = \int_0^1 Q(t) K_n'(u, t)\, d\mu_n(t)$.

LEMMA 3.2. *Under conditions* Q_1–Q_3 *and* K_1–K_7, *there exists a sequence of Brownian bridges* $\{B_n\}_{n=1}^{\infty}$ *such that* $n^{1/2}[\hat{q}_n(u) - \bar{q}_n(u)] = \hat{B}'_n(u) + \hat{e}'_n(u)$, $u \in (0,1)$, *where*

$$\hat{B}'_n(u) = \int_0^1 q(t) B_n(t) K'_n(u,t)\, d\mu_n(t), \qquad u \in U,$$

and

$$\sup_{u \in U} |\hat{e}'_n(u)| \le C_0 n^{-1/2} A_\gamma(n) \Big[\sup_{v \in U} |q(v)| \sup_{u \in U} \Lambda(u; K_n) + R(q; K_n) \Big]$$

with probability 1; C_0 *is a universal constant and* $n^{-1/2} A_\gamma(n) = o(1)$ *as* $n \uparrow \infty$.

PROOF. First note that

$$n^{1/2}[\hat{q}_n(u) - \bar{q}_n(u)] = \int_0^1 n^{1/2}\big[\tilde{Q}_n(t) - Q(t)\big] K'_n(u,t)\, d\mu_n(t).$$

By conditions Q_1–Q_3 and Csörgő and Révész [(1978), Theorem 6], there is a sequence of Brownian bridges $\{B_n\}$ such that, with probability 1,

$$\limsup_{n \to \infty} \left[\frac{n^{1/2}}{A_\gamma(n)} \right] \sup_{u \in (0,1)} \left| \frac{1}{q(t)} n^{1/2} \big[\tilde{Q}_n(t) - Q(t)\big] - B_n(u) \right| \le C_0,$$

where C_0 is a universal constant, $A_\gamma(n)$ depends on γ in Q_2 but $n^{-\tau} A_\gamma(n) = o(1)$ for arbitrary $\tau > 0$. Let $e_n(t) = [1/q(t)] n^{1/2} [\tilde{Q}_n(t) - Q(t)] - B_n(t)$. Then $n^{1/2}[\tilde{Q}_n(t) - Q(t)] = q(t) B_n(t) + q(t) e_n(t)$. So $n^{1/2}[\hat{q}_n(u) - \bar{q}_n(u)] = \hat{B}'_n(u) + \hat{e}'_n(u)$ with $\hat{e}'_n(u) = \int_0^1 q(t) e_n(t) K'_n(u,t)\, d\mu_n(t)$. Moreover, with probability 1,

$$\begin{aligned}
\sup_{u \in U} |\hat{e}'_n(u)| &\le \sup_{u \in U} \int_0^1 q(t) |e_n(t) K'_n(u,t)|\, d\mu_n(t) \\
&\le \sup_{v \in (0,1)} |e_n(v)| \sup_{u \in U} \int_0^1 q(t) |K'_n(u,t)|\, d\mu_n(t) \\
&\le C_0 n^{-1/2} A_\gamma(n) \sup_{u \in U} \int_0^1 q(t) |K'_n(u,t)|\, d\mu_n(t), \qquad n \to \infty,
\end{aligned}$$

and $\sup_{u \in U} \int_0^1 q(t) |K'_n(u,t)|\, d\mu_n(t) \le \sup_{v \in U} q(v) \sup_{u \in U} \Lambda(u; K_n) + R(q; K_n)$. Note further that $R(q; K_n) = o(1)$ by Q_1, K_6 and K_7. □

LEMMA 3.3. *Let* $\hat{B}'_n(u)$ *be as in Lemma* 3.2. *Then* $\sup_{u \in U} |\hat{B}'_n(u)| = O_p(A(q; K_n))$, *as* $n \uparrow \infty$:

$$\begin{aligned}
A(q; K_n) = {}& \sup\nolimits_{v \in U} q(v) \sup\nolimits_{u \in U} \Lambda(u; K_n) \sqrt{2 \delta_n \log \delta_n^{-1}} \\
&+ \sup\nolimits_{u \in U} |q'(u)| + R(q; K_n).
\end{aligned}$$

PROOF. First note that

$$\sup_{u\in U}\left|\hat{B}_n'(u)\right| \le \sup_{u\in U}\left|\int_0^1 q(t)[B_n(t)-B_n(u)]K_n'(u,t)\,d\mu_n(t)\right|$$
$$+\sup_{u\in U}|B_n(u)|\sup_{u\in U}\left|\int_0^1 q(t)K_n'(u,t)\,d\mu_n(t)\right|$$
$$:= T_n + W_n.$$

By Lemma 3.1, $\sup_{u\in U}|\int_0^1 q(t)K_n'(u,t)\,d\mu_n(t)| \sim \sup_{u\in U}|q'(u)|$; $W_n = O_p(\sup_{u\in U}|q'(u)|)$;

$$T_n \le \sup_{u\in U}\int_{I_n(u)}|q(t)[B_n(t)-B_n(u)]K_n'(u,t)|\,d\mu_n(t)$$
$$+\sup_{u\in U}\int_{I_n^c(u)}|g(t)[B_n(t)-B_n(u)]K_n'(u,t)|\,d\mu_n(t)$$
$$:= T_{1,n} + T_{2,n}.$$

Because B_n's are identically distributed as a Brownian bridge,

$$T_{2,n} \le \sup_{u\in U}\sup_{t\in[0,1]}|B_n(t)-B_n(u)|\sup_{u\in U}\int_{I_n^c(u)}q(t)|K_n'(u,t)|\,d\mu_n(t)$$
$$= O_p(1)R(q;K_n),$$

$$T_{1,n} = \sup_{u\in U}\int_{I_n(u)}|q(t)[B_n(t)-B_n(u)]K_n'(u,t)|\,d\mu_n(t)$$
$$\le \sup_{u\in U}\sup_{t\in I_n(u)}q(t)\sup_{v\in I_n(u)}|B_n(v)-B_n(u)|\sup_{u\in U}\int_{I_n(u)}|K_n'(u,t)|\,d\mu_n(t)$$
$$\sim \sup_{v\in U}q(v)\sup_{u\in U}\Lambda(u;K_n)O_p\Big(\sqrt{2\delta_n\log\delta_n^{-1}}\Big);$$

the asymptotic equivalence follows from Csörgő and Révész [(1981), Theorem 1.4.1]. □

REMARK 3.1. Note that, for each n,

$$\hat{B}_n'(u) = \frac{d\int_0^1 q(t)B_n(t)K_n(u,t)\,d\mu_n(t)}{du}$$

with probability 1. The above lemma provides a stochastic upper bound for this derivative process sequence. For related results on smoothed Wiener process and Brownian bridge, see Stadtmüller (1986, 1988) and Xiang [(1994), Lemma 2.1].

Theorem 2.1 follows immediately from the lemmas.

Acknowledgments. The author would like to thank an Associate Editor, the referees and the Editor for their valuable comments and suggestions.

Many thanks are due Professor Emanuel Parzen at Texas A & M University, from whom I received much valuable guidance and encouragement.

REFERENCES

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

BLOCH, D. A. and GASTWIRTH, J. L. (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39** 1083–1085.

BOFINGER, E. (1975). Estimation of a density function using order statistics. *Austral. J. Statist.* **17** 1–7.

CHENG, C. (1995). The Bernstein polynomial estimator of a smooth quantile function. *Statist. Probab. Lett.* **24** 321–330.

CSÖRGŐ, M., DEHEUVELS, P. and HORVÁTH, L. (1991). Estimating the quantile-density function. In *Nonparametric Functional Estimation and Related Topics* (G. Roussas, ed.) 213–223. Kluwer Academic, Boston.

CSÖRGŐ, M. and HORVÁTH, L. (1989). On confidence bands for the quantile function of a continuous distribution function. *Colloq. Math. Soc. János Bolyai* **57** 95–106.

CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882–894.

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.

FALK, M. (1986). On the estimation of the quantile density function. *Statist. Probab. Lett.* **4** 69–73.

GAWRONSKI, W. (1985). Strong laws for density estimators of Bernstein type. *Period. Math. Hungar.* **16** 23–43.

KAIGH, W. D. and CHENG, C. (1991). Subsampling quantile estimators and uniformity criteria. *Comm. Statist. Theory Methods* **20** 539–560.

LORENTZ, G. G. (1986). *Bernstein Polynomials*. Chelsea, New York.

PARZEN, E. (1979). Nonparametric statistical data modeling (with comments). *J. Amer. Statist. Assoc.* **74** 105–131.

SCHOENBERG, I. J. (1965). On spline functions. In *Inequalities* (O. Shisha, ed.) 255–291. Academic Press, New York.

SIDDIQUI, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards*, *Section B* **64** 145–150.

STADTMÜLLER, U. (1986). Asymptotic properties of nonparametric curve estimators. *Period. Math. Hungar.* **12** 83–104.

STADTMÜLLER, U. (1988). Kernel approximations of a Wiener process. *Period. Math. Hungar.* **19** 79–90.

VITALE, R. A. (1975). A Bernstein polynomial approach to density estimation. In *Statistical Inference and Related Topics* (M. L. Puri, ed.) 87–99. Academic Press, New York.

XIANG, X. (1994). A law of the logarithm for kernel quantile density estimators. *Ann. Probab.* **22** 1078–1091.

DEPARTMENT OF MATHEMATICAL SCIENCES
G.W.C. WHITING SCHOOL OF ENGINEERING
JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218