*The Annals of Statistics*
1995, Vol. 23, No. 3, 836–854

# DEFICIENCY OF THE SAMPLE QUANTILE ESTIMATOR WITH RESPECT TO KERNEL QUANTILE ESTIMATORS FOR CENSORED DATA

BY XIAOJING XIANG

*University of Oregon*

Consider a statistical procedure (Method A) which is based on $n$ observations and a less effective procedure (Method B) which requires a larger number $k_n$ of observations to give equal performance under a certain criterion. To compare two different procedures, Hodges and Lehmann suggested that the difference $k_n - n$, called the deficiency of Method B with respect to Method A, is the most natural quantity to examine. In this article, the performance of two kernel quantile estimators is examined versus the sample quantile estimator under the criterion of equal covering probability for randomly right-censored data. We shall show that the deficiency of the sample quantile estimator with respect to the kernel quantile estimators is convergent to infinity with the maximum rate when the bandwidth is chosen to be optimal. A Monte Carlo study is performed, along with an illustration on a real data set.

**1. Introduction.** Comparison of different statistical procedures is an important issue in both practical and theoretical statistics. Consider a statistical procedure (Method A) which is based on $n$ observations and a less effective procedure (Method B) which requires a large number $k_n$ of observations to give equal performance under a certain criterion. To compare two different procedures, Hodges and Lehmann (1970) suggested that the difference $k_n - n$, called the deficiency of Method B with respect to Method A, is the most natural quantity to examine. Let $X_1, X_2, \dots$ be a sequence of independent random variables with common distribution function $F(x)$. For $0 < p < 1$, the $p$th quantile of $F(x)$, which is defined by $Q(p) = \inf\{x: F(x) \geq p\}$, is an important characteristics of $F(x)$ and plays an important role in both statistics and probability. One popular estimator Of $Q(p)$, called the sample quantile estimator, denoted by $\xi_{pn}$, is the $p$th quantile of the empirical distribution function $F_n(x) = (1/n)\sum_{i=1}^n I(X_i \leq x)$. As an alternative, the kernel quantile estimator introduced by Parzen (1979) is of the form

$$q_n(F_n) = \frac{1}{h_n}\int_0^1 F_n^{-1}(x) K\left(\frac{x-p}{h_n}\right) dx \tag{1.1}$$

for an appropriate kernel function $K(x)$ and a bandwidth $h_n \downarrow 0$. Define for any $t > 0$ the interval

$$I_t(n) = \left[Q(p) - t\sigma n^{-1/2}, Q(p) + t\sigma n^{-1/2}\right],$$

Received January 1993; revised August 1994.
*AMS* 1991 *subject classifications.* Primary 62F12, 62G05, 62G20; secondary 62G30.
*Key words and phrases.* Deficiency, kernel estimator, sample quantile, covering probability.

where $\sigma^2 = (p(1-p))/(f^2(Q(p)))$. Then $P(q_n(F_n) \in I_t(n))$ and $P(\xi_{pn} \in I_t(n))$ are the covering probabilities of $q_n(F_n)$ and $\xi_{pn}$ with respect to $I_t(n)$, respectively. Let $i_t(n)$ be defined by

$$i_t(n) = \min\{m \geq 1: P(\xi_{pm} \in I_t(n)) \geq P(q_n(F_n) \in I_t(n))\}.$$

Then $i_t(n)$ is the minimum number of observations required by the sample quantile estimator to give a covering probability equal to that of the kernel quantile estimator and the difference $i_t(n) - n$ is the deficiency of $\xi_{pn}$ with respect to $q_n(F_n)$ in the Hodges and Lehmann sense.

Falk (1985) compared the performance of the sample quantile estimator and the kernel quantile estimator under the criterion of equal covering probability. Falk showed that under certain conditions,

$$\text{(1.2)} \quad \lim_{n \to \infty} h_n^{-1}\big(P(q_n(F_n) \in I_t(n)) - P(\xi_{pn} \in I_t(n))\big) = \frac{t\phi(t)\Psi(K)}{p(1-p)},$$

where

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \qquad \Psi(K) = 2\int_{-c}^{c} xK(x) \int_{-c}^{x} K(u)\, du\, dx$$

and

$$\text{(1.3)} \qquad \liminf_{n \to \infty} \frac{i_t(n) - n}{nh_n} \geq \frac{\Psi(K)}{p(1-p)}.$$

If $\Psi(K) > 0$, it follows that the sample quantile estimator requires more observations to perform as well as the kernel quantile estimator $q_n(F_n)$. More precisely, the approximate number of the additional observations required by the sample quantile is at least equal to $r_n = nh_n(\Psi(K))/(p(1-p))$. Since $nh_n \to \infty$ as $n \to \infty$, $r_n \to \infty$ as $n \to \infty$. In view of (1.3), a lower rate of $h_n$ and a larger value of $\Psi(K)$ will lead to a higher rate of $r_n$. Hence, one would like to choose a kernel function $K$ and bandwidth $h_n$ to make the deficiency $i_t(n) - n$ tend to infinity as quickly as possible. This leads to considerations of the optimal selection of the kernel function and the corresponding bandwidth. The optimal kernel selection has been studied by Falk (1983) and Mammitzsch (1984). However, the optimal bandwidth, based on my knowledge, has not been obtained. The goal of this article is to compare the performance of the sample quantile estimator and two kernel quantile estimators under the criterion of equal covering probability for censored data. The asymptotically optimal bandwidth is obtained.

Let $X_1, \ldots, X_n$ be i.i.d. nonnegative random variables with common distribution function $F(x)$, called the survival time distribution. Our model is that of right random censoring, that is, associated with each $X_i$, there is an independent nonnegative censoring time $Y_i$ and $Y_1, \ldots, Y_n$ are assumed to be i.i.d. random variables with common distribution function $G(x)$. We assume throughout this paper that $G(x)$ is continuous. The observations in this model are the pairs $(T_i, \delta_i)$, where $T_i = \min(X_i, Y_i)$ and $\delta_i = I_{(X_i \leq Y_i)}$, $i =$

$1, 2, \dots, n$. Clearly, $T_i$ are i.i.d. with common distribution function $H(x) = 1 - (1 - F(x))(1 - G(x))$. In the censored model, the sample quantile estimator and the kernel estimator are defined in terms of the product-limit estimator, proposed by Kaplan and Meier (1958). The product-limit estimator is of the form

$$\hat{F}_n(t) = \begin{cases} 1 - \prod_{T_{(i)} \le t} \left( \dfrac{n-i}{n-i+1} \right)^{\delta_{(i)}}, & t < T_{(n)}, \\ 1, & t \ge T_{(n)}, \end{cases}$$

where $T_{(1)} \le T_{(2)} \le \cdots \le T_{(n)}$ are the order statistics of $T_1, T_2, \dots, T_n$ and $\delta_{(1)}, \dots, \delta_{(n)}$ are the corresponding $\delta_i$. The sample quantile estimator $\hat{\xi}_{pn}$ of $Q(p)$ is the $p$th quantile of $\hat{F}_n(t)$ and the kernel quantile estimator first studied by Padgett (1986) is of the form

$$\hat{q}_n(\hat{F}_n) = \frac{1}{h_n} \int_0^1 \hat{F}_n^{-1}(x) K\left( \frac{x-p}{h_n} \right) dx. \tag{1.4}$$

Besides $\hat{q}_n(\hat{F}_n)$, Padgett (1986) also proposed the following kernel quantile estimator as an approximation of $\hat{q}_n(\hat{F}_n)$. Let $s_j$ denote the jump of $\hat{F}_n(t)$ at $T_{(j)}$, that is,

$$s_j = \begin{cases} \hat{F}_n(T_{(1)}), & j = 1, \\ \hat{F}_n(T_{(j)}) - \hat{F}_n(T_{(j-1)}), & j = 2, \dots, n. \end{cases}$$

The approximation of $\hat{q}_n(\hat{F}_n)$, denoted by $\hat{q}_n^*(\hat{F}_n)$, is defined by

$$\hat{q}_n^*(\hat{F}_n) = h_n^{-1} \sum_{i=1}^n T_{(i)} s_i K\left( \frac{\hat{F}_n(T_{(i)}) - p}{h_n} \right). \tag{1.5}$$

In the absence of censoring, $\hat{q}_n^*(\hat{F}_n)$ reduces to the kernel estimator proposed by Yang (1985).

Define, for any $t > 0$, $\bar{I}_t(n) = [Q(p) - t\bar{\sigma} n^{-1/2}, Q(p) + t\bar{\sigma} n^{-1/2}]$, where

$$\bar{\sigma}^2 = \frac{(1-p)^2}{f^2(Q(p))} \int_0^{Q(p)} \frac{dF(y)}{(1 - G(y))(1 - F(y))^2}. \tag{1.6}$$

Let

$$i_{1t}(n) = \min\left\{ m \ge 1 : P\left( \hat{\xi}_{pm} \in \bar{I}_t(n) \right) \ge P\left( \hat{q}_n(F_n) \in \bar{I}_t(n) \right) \right\}$$

and

$$i_{2t}(n) = \min\left\{ m \ge 1 : P\left( \hat{\xi}_{pm} \in \bar{I}_t(n) \right) \ge P\left( \hat{q}_n^*(F_n) \in \bar{I}_t(n) \right) \right\}.$$

We shall compare $P(\hat{\xi}_{pn} \in \bar{I}_t(n))$ with $P(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n))$ and $P(\hat{q}_n^*(\hat{F}_n) \in \bar{I}_t(n))$, respectively, and examine the rates of $i_{1t}(n) - n$ and $i_{2t}(n) - n$. We shall show that $i_{1t}(n) - n$ and $i_{2t}(n) - n$ attain the maximum rates with respect to the asymptotically optimal bandwidth.

This article is organized as follows. The main results are given in Section 2. Section 3 contains the simulation results, along with an illustration on a real data set. The proofs of the theorems are provided in Section 4.

**2. Main results.** Let $T_H = \inf\{t: H(t) = 1\}$. Throughout this paper, we assume that the kernel function $K(x)$ has compact support on $[-1, 1]$. We use the notation $a_n \sim b_n$ if and only if $a_n/b_n \to 1$, as $n \to \infty$. We say the function $f(x)$ is Lipschitz of order 1 if for a universal constant $C$, $|f(x) - f(y)| \le C|x - y|$ for all $x, y \in (-\infty, \infty)$.

THEOREM 2.1. *Assume that $Q(t)$ has a bounded $(m + 1)$th derivative in a neighborhood of $p$ with $0 < p < F(T_H)$, $m \ge 2$, and that $f(Q(p)) > 0$. Let $K(x)$ be Lipschitz of order 1 and let $K(x)$ satisfy $\int_{-1}^{1} K(x)\,dx = 1$, $\int_{-1}^{1} x^i K(x)\,dx = 0$, $i = 1, \dots, m$, $\alpha_m = [(Q^{(m+1)}(p))/(m + 1)!]\int_{-1}^{1} u^{m+1}K(u)\,du \ne 0$. Then if $nh_n^4/\log^4 n \to \infty$,*

$$P\big(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)\big) - P\big(\hat{\xi}_{pn} \in \bar{I}_t(n)\big) = \frac{t\phi(t)\Psi(K)}{f^2(Q(p))(1 - G(Q(p)))\bar{\sigma}^2} h_n - \frac{t\phi(t)\alpha_m^2}{\bar{\sigma}^2} n h_n^{2(m+1)} + o\big(h_n + nh_n^{2(m+1)}\big) \tag{2.1}$$

*and*

$$P\big(\hat{q}_n^*(\hat{F}_n) \in \bar{I}_t(n)\big) - P\big(\hat{\xi}_{pn} \in \bar{I}_t(n)\big) = \frac{t\phi(t)\Psi(K)}{f^2(Q(p))(1 - G(Q(p)))\bar{\sigma}^2} h_n - \frac{t\phi(t)\alpha_m^2}{\bar{\sigma}^2} n h_n^{2(m+1)} + o\big(h_n + nh_n^{2(m+1)}\big). \tag{2.2}$$

*The asymptotically optimal bandwidth is*

$$h_{\text{opt}} = \left\{ n^{-1} \frac{\Psi(K)}{2(m + 1)\alpha_m^2 f^2(Q(p))(1 - G(Q(p)))} \right\}^{1/(2m+1)}, \tag{2.3}$$

*and with $h_{\text{opt}}$,*

$$P\big(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)\big) - P\big(\hat{\xi}_{pn} \in \bar{I}_t(n)\big) \sim \frac{t\phi(t)\Psi(K)}{(1 - G(Q(p)))f^2(Q(p))\bar{\sigma}^2} h_{\text{opt}} \tag{2.4}$$

and

$$P\big(\hat{q}_n^*(\hat{F}_n) \in \bar{I}_t(n)\big) - P\big(\hat{\xi}_{pn} \in \bar{I}_t(n)\big) \sim \frac{t\phi(t)\Psi(K)}{(1 - G(Q(p)))f^2(Q(p))\bar{\sigma}^2} h_{\text{opt}}. \tag{2.5}$$

The following result is about the maximum rate of the deficiency.

THEOREM 2.2. *Under the assumptions of Theorem 2.1, if the bandwidth and the kernel function are chosen such that* $P(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)) - P(\hat{\xi}_{pn} \in \bar{I}_t(n)) \geq 0$ *and* $P(\hat{q}_n^*(\hat{F}_n) \in \bar{I}_t(n)) - P(\hat{\xi}_{pn} \in \bar{I}_t(n)) \geq 0$ *for large n, then*

$$i_{jt}(n) - n = \frac{\Psi(K)}{(1 - G(Q(p)))f^2(Q(p))\bar{\sigma}^2} nh_n - \frac{\alpha_m^2}{\bar{\sigma}^2} n^2 h_n^{2(m+1)} + o\big(nh_n + n^2 h_n^{2(m+1)}\big) \tag{2.6}$$

*for* $j = 1, 2$. *The asymptotically optimal bandwidth is given by* (2.3), *and with* $h_{\text{opt}}$,

$$i_{jt}(n) - n \sim \frac{2m\Psi(K)}{(2m+1)(1 - G(Q(p)))f^2(Q(p))\bar{\sigma}^2} nh_{\text{opt}}, \qquad j = 1, 2. \tag{2.7}$$

REMARK 1. For each fixed $p$, $0 < p < F(T_H)$, using $h_{\text{opt}}$, the deficiencies are of $O(n^{2m/(2m+1)})$. In the absence of censoring, $G(x) = 0$, $\bar{\sigma}^2 = (p(1-p))/(f^2(Q(p))) = \sigma^2$ and the results in the above theorems can be simplified. For example, (2.7) reduces to

$$i_{jt}(n) - n \sim \frac{2m\Psi(K)}{(2m+1)p(1-p)} nh_{\text{opt}} = O(n^{2m/(2m+1)}), \qquad j = 1, 2. \tag{2.8}$$

In contrast, Falk's result gives $i_{1t}(n) - n = o(n^{2m/(2m+1)})$.

REMARK 2. Estimating $h_{\text{opt}}$ may be done by the so-called plug-in method. That is, we substitute the estimators for the unknown quantities involved in $h_{\text{opt}}$ in (2.3). Distribution functions $G(x)$ and $F(x)$ can be estimated directly by using the product-limit estimator. For $l \geq 1$, $Q^{(l)}(p)$ can be estimated by

$$\hat{q}_n^{(l)}(\hat{F}_n) = (-1)^l h_n^{-1-l} \int_0^1 \hat{F}_n^{-1}(x) K^{(l)}\left(\frac{x-p}{h_n}\right) dx \tag{2.9}$$

or

$$\hat{q}_n^{*(l)}(\hat{F}_n) = (-1)^l h_n^{-1-l} \sum_{i=1}^n T_{(i)} s_i K^{(l)}\left(\frac{\hat{F}_n(T_{(i)}) - p}{h_n}\right). \tag{2.10}$$

The asymptotic properties of the estimators in (2.9) and (2.10) are discussed in Xiang (1992, 1994). A Monte Carlo study based on the plug-in method is performed in the following Section 3.

REMARK 3. Lio and Padgett (1992) studied the Nadaraya-type smooth quantile estimator and obtained the asymptotically optimal bandwidth. Though their asymptotically optimal bandwidth is defined by a different criterion, the order of their bandwidth is the same as that of $h_{\text{opt}}$ in (2.3).

REMARK 4. Finally, the type of estimators used in practice will depend on one's particular interests. For instance, the sample quantile estimator for censored data allows one to easily construct bootstrap confidence bands [see Doss and Gill (1992)]. Also the sample quantile estimator performs better than the kernel quantile estimators as $p$ is close to 0 and 1 (see Figure 1 in Section 3).

**3. Monte Carlo study and an example.** In this section, a Monte Carlo study was performed in order to provide some small-sample comparisons of $\hat{q}_n(\hat{F}_n)$ and $\hat{q}_n^*(\hat{F}_n)$ with $\hat{F}_n^{-1}$ and with each other, in the sense of covering probabilities. Consider the interval $\bar{I}_n = [Q(p) - \bar{\sigma} n^{-1/2}, Q(p) + \bar{\sigma} n^{-1/2}]$, where $\bar{\sigma}^2$ is calculated from (1.6). The true covering probabilities of $\hat{q}_n(\hat{F}_n)$, $\hat{q}_n^*(\hat{F}_n)$ and $\hat{F}_n^{-1}$ are $P(\hat{q}_n(\hat{F}_n) \in \bar{I}_n)$, $P(\hat{q}_n^*(\hat{F}_n) \in \bar{I}_n)$ and $P(\hat{F}_n^{-1} \in \bar{I}_n)$, respectively. The estimator with higher covering probability performs better than that with a lower one. In the following Monte Carlo study, the covering probabilities were calculated by finding the proportion of values of the estimators that fall within the interval $\bar{I}_n$. The larger proportion corresponds to higher covering probability. To calculate $\hat{q}_n(\hat{F}_n)$ and $\hat{q}_n^*(\hat{F}_n)$, we proposed a data-based procedure of the optimal bandwidth selection. Besides the Monte Carlo study, a real data set is illustrated in this section. For this data set, we compared the optimal bandwidths calculated from our data-based procedure with those in Padgett (1986) computed based on the bootstrap mean squared errors.

To make things simple, we choose $m = 1$ in the following Monte Carlo study. Notice that the case $m = 1$ is not included in the theorems presented in Section 2, but we suspect that their conclusions hold more generally; the simulation results support these suspicions. For the same reason, we use the triangular kernel function $K(x) = (1 - |x|)I(|x| \le 1)$ for both kernel quantile estimators in (1.4) and (1.5).

Perhaps the most important issue for the kernel quantile estimators is the selection of bandwidth. From the theorems in Section 2, the asymptotically optimal bandwidth is

$$(3.1)\quad h_{\mathrm{opt}} = \left\{ n^{-1} \frac{\Psi(K)}{\left(\int_{-1}^{1} x^2 K(x)\, dx\right)^2 (1 - G(Q(p)))} \left( \frac{Q'(p)}{Q''(p)} \right)^2 \right\}^{1/3},$$

depending on $Q(p)$, $Q'(p)$, $Q''(p)$ and the unknown distribution function $G(x)$. Thus estimators of $Q(p)$, $Q'(p)$, $Q''(p)$ and $G(x)$ are necessary for a data-based choice of $h_{\mathrm{opt}}$. The quantile $Q(p)$ can be estimated by the sample quantile estimator. From (2.9), the estimators of $Q'(p)$ and $Q''(p)$ are

$$\begin{aligned}\hat{q}_n'\left(\hat{F}_n\right) &= -\frac{1}{a_n^2} \int_0^1 \hat{F}_n^{-1}(x) K_*'\left( \frac{x - p}{a_n} \right) dx \\ &= -\frac{1}{a_n} \sum_{i=1}^{n-1} \left[ K_*\left(a_n^{-1}\left(\hat{F}_n(T_{(i+1)}) - p\right)\right) - K_*\left(a_n^{-1}\left(\hat{F}_n(T_{(i)}) - p\right)\right)\right] T_{(i)}\end{aligned}$$

and

$$\hat{q}''_n(\hat{F}_n) = \frac{1}{b_n^3}\int_0^1 \hat{F}_n^{-1}(x) K''_*\left(\frac{x-p}{b_n}\right) dx$$

$$= \frac{1}{b_n^2}\sum_{i=1}^{n-1}\left[K'_*\left(b_n^{-1}\left(\hat{F}_n(T_{(i+1)}) - p\right)\right) - K'_*\left(b_n^{-1}\left(\hat{F}_n(T_{(i)}) - p\right)\right)\right]T_{(i)},$$

where $a_n$ and $b_n$ are bandwidths and $K_*(x)$ is a kernel function which will be given later. The distribution function $G(x)$ can be estimated by the product-limit estimator. However, when the estimator of $G(x)$ is equal to 1, the denominator of the right-hand side of (3.1) is equal to zero. To avoid this problem, we choose the estimator of $G(x)$ of the form

$$\hat{G}_n(t) = \begin{cases} 1 - \prod_{T_{(i)} \le t}\left(\dfrac{n-i}{n-i+1}\right)^{1-\delta_{(i)}}, & t \le T_{(n-1)}, \\ \hat{G}_n(T_{(n-1)}), & t > T_{(n-1)}. \end{cases}$$

As suggested by Sheather and Marron (1990), $K_*$ is chosen to be a higher-order kernel to minimize the problems associated with ratio estimation. We use the following fourth-order kernel, given by Müller (1984),

$$K_*(x) = (315/512)(3 - 20x^2 + 42x^4 - 36x^6 + 11x^8)I(|x| \le 1)$$

to estimate $Q'(p)$ and $Q''(p)$. In the absence of censoring, the asymptotically optimal bandwidths based on the mean squared errors for $\hat{q}'_n(\hat{F}_n)$ and $\hat{q}''_n(\hat{F}_n)$ are given in Sheather and Marron (1990). However, similar results in the censoring case have not been seen in the literature. Here, we choose $a_{\text{opt}}$ and $b_{\text{opt}}$ based on a LIL for estimator $\hat{q}_n^{(l)}(\hat{F}_n)$ [see Xiang (1992) for details]. This LIL gives, corresponding to the fourth-order kernel function, $a_{\text{opt}} = (\log\log n/n)^{1/9}$ and $b_{\text{opt}} = (\log\log n/n)^{1/11}$.

Let $F(x)$ be exponential and let $G(x)$ be exponential with parameter $\alpha$ chosen to give 50% censoring ($\alpha = 1$) or 30% censoring ($\alpha = 3/7$). The covering probabilities of $\hat{q}_n(\hat{F}_n)$, $\hat{q}_n^*(\hat{F}_n)$ and $\hat{F}_n^{-1}$ were computed for various $0 < p < 1$ and sample sizes $n = 50$ and $n = 100$. For each case, 1000 censored samples were generated in the statistical package $S$ on a Sun 2 work station. Tables 1–4 show the covering probabilities for these models. Comparing the covering probabilities, $\hat{q}_n(\hat{F}_n)$ is uniformly better than $\hat{F}_n^{-1}$ for both models and both sample sizes. The estimator $\hat{q}_n(\hat{F}_n)$ also performs equally

TABLE 1
*Covering probabilities*: 50% *censoring. Based on* 1000 *samples of size* $n = 50$

| $p$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{F}_n^{-1}$ | 0.494 | 0.488 | 0.529 | 0.557 | 0.518 | 0.515 | 0.532 | 0.552 | 0709 |
| $\hat{q}_n(\hat{F}_n)$ | 0.567 | 0.566 | 0.568 | 0.573 | 0.546 | 0.538 | 0.571 | 0.659 | 0.742 |
| $\hat{q}_n^*(\hat{F}_n)$ | 0.580 | 0.566 | 0.599 | 0.621 | 0.597 | 0.546 | 0.416 | 0.242 | 0.186 |

TABLE 2
*Covering probabilities*: 30% *censoring. Based on* 1000 *samples of size* $n = 50$

| $p$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{F}_n^{-1}$ | 0.508 | 0.473 | 0.513 | 0.520 | 0.546 | 0.534 | 0.513 | 0.535 | 0.532 |
| $\hat{q}_n(\hat{F}_n)$ | 0.573 | 0.585 | 0.577 | 0.575 | 0.558 | 0.567 | 0.559 | 0.545 | 0.602 |
| $\hat{q}_n^*(\hat{F}_n)$ | 0.570 | 0.574 | 0.572 | 0.596 | 0.604 | 0.588 | 0.580 | 0.507 | 0.337 |

TABLE 3
*Covering probabilities*: 50% *censoring. Based on* 1000 *samples of size* $n = 100$

| $p$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{F}_n^{-1}$ | 0.650 | 0.658 | 0.706 | 0.687 | 0.672 | 0.683 | 0.678 | 0.697 | 0.810 |
| $\hat{q}_n(\hat{F}_n)$ | 0.723 | 0.730 | 0.716 | 0.708 | 0.712 | 0.705 | 0.713 | 0.751 | 0.859 |
| $\hat{q}_n^*(\hat{F}_n)$ | 0.755 | 0.757 | 0.754 | 0.754 | 0.761 | 0.763 | 0.717 | 0.466 | 0.361 |

TABLE 4
*Covering probabilities*: 30% *censoring. Based on* 1000 *samples of size* $n = 100$

| $p$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{F}_n^{-1}$ | 0.660 | 0.701 | 0.693 | 0.709 | 0.667 | 0.680 | 0.691 | 0.702 | 0.686 |
| $\hat{q}_n(\hat{F}_n)$ | 0.728 | 0.755 | 0.758 | 0.729 | 0.714 | 0.726 | 0.719 | 0.703 | 0.746 |
| $\hat{q}_n^*(\hat{F}_n)$ | 0.753 | 0.764 | 0.772 | 0.747 | 0.716 | 0.730 | 0.761 | 0.774 | 0.609 |

well for different values of $p$; $\hat{q}_n^*(\hat{F}_n)$ is better than $\hat{F}_n^{-1}$ except for large values of $p$; $\hat{q}_n^*(\hat{F}_n)$ performs better for larger sample size and lower censoring percentage. For small sample size $n$ and heavy censoring model, $\hat{F}_n(t)$ tends to have big jumps at large $T_{(i)}$. Hence, if $h_n$ is sufficiently small, it may be the case that $|\hat{F}_n(T_{(i)}) - p| > h_n$ for any $T_{(i)}$. From (1.5), this results in $\hat{q}_n^*(\hat{F}_n) = 0$. Moreover, suppose $h_n^{-1}(\hat{F}_n(T_{(i)}) - p) \approx 0$ for some $i$. From (1.5), $\hat{q}_n^*(\hat{F}_n) \approx ((s_i K(0))/h_n)T_{(i)}$. If $h_n$ is sufficiently small and $\hat{F}_n(t)$ has a big jump at $T_{(i)}$, $\hat{q}_n^*(\hat{F}_n)$ may be very large. Hence, $\hat{q}_n^*(\hat{F}_n)$ may perform poorly for large $p$ values. In contrast, from (1.4),
values. In contrast, from (1.4),

$$\hat{q}_n(\hat{F}_n) = \sum_{j=1}^{n} \left( h^{-1} \int_{\hat{F}_n(T_{(j-1)})}^{\hat{F}_n(T_{(j)})} K\left( \frac{x-p}{h_n} \right) dx \right) T_{(j)},$$

where we set $\hat{F}_n(T_{(0)}) = 0$. The sum of the weights in the above $L$-estimator is always equal to 1. This characteristics of $\hat{q}_n(\hat{F}_n)$ overcomes the drawbacks associated with $\hat{q}_n^*(\hat{F}_n)$. For the simulations we did not use the $h_{\text{opt}}$ given in (3.1), but an asymptotically equivalent version $\bar{h}_{\text{opt}} = \min\{0.2, p, 1-p, h_{\text{opt}}\}$, where $h_{\text{opt}}$ is calculated from (3.1) based on the above plug-in method. This

TABLE 5
*The values of $h_{opt}$, calculated based on the life test data*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 |
| $h_{opt}$ | 0.229 | 0.232 | 0.250 | 0.250 | 0.255 | 0.261 | 0.276 | 0.276 | 0.276 |
| $p$ | 0.55 | 0.60 | 0.65 | 0.70 | 0.76 | 0.80 | 0.85 | 0.90 | |
| $h_{opt}$ | 0.289 | 0.289 | 0.331 | 0.331 | 0.331 | 0.331 | 0.331 | 0.331 | |

restriction usually alters $h_{opt}$ only when the sample size is small; this protects against the bias for small sample size associated with the edge effects.

As an example of the quantile estimators, we use the life test data for $n = 40$ mechanical switches reported by Nair (1984) (see Table 5). Two failure modes, $A$ and $B$, were reported assuming the randomly right-censored model. There are 17 uncensored observations, which is 57.5% censoring in this particular sample. The values of $h_{opt}$ are calculated from (3.1) based on the above plug-in method. It is interesting to note that these values of $h_{opt}$ are close to those selected in Padgett (1986) based on the bootstrap mean squared errors. Figure 1 shows the estimators $\hat{q}_n(\hat{F}_n)$ and $\hat{q}_n^*(\hat{F}_n)$ along with $\hat{F}_n^{-1}$. As Figure 1 indicates, as $p$ is close to 1, both $\hat{q}_n(\hat{F}_n)$ and $\hat{q}_n^*(\hat{F}_n)$ are not monotone in $p$ due to the edge effects which result in the bias for the kernel estimators.

**4. Proofs.** Let

$$\hat{q}_n(F) = \frac{1}{h_n}\int_0^1 Q(x)K\left(\frac{x-p}{h_n}\right)dx.$$

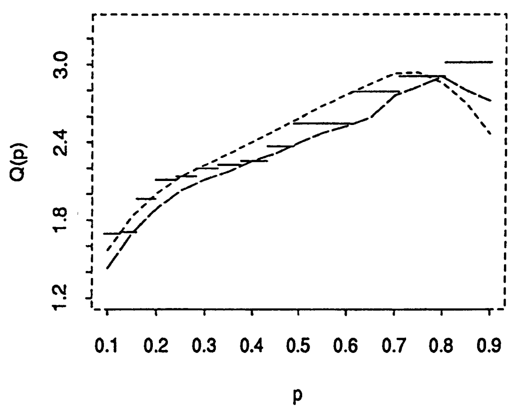


FIG. 1. *Quantile estimators for switch life data. The solid, short dashed and long dashed curves correspond to $\hat{F}_n^{-1}$, $\hat{q}_n(\hat{F}_n)$ and $\hat{q}_n^*(\hat{F}_n)$, respectively.*

The basic idea for the proofs of the theorems is to express $\hat{q}_n(\hat{F}_n) - \hat{q}_n(F)$ as a functional of $\hat{F}_n(t) - F(t)$ and then use a strong embedding result due to Major and Retö (1988). Major and Rejtö (1988) have shown that there exists a Gaussian process $W(u)$, $0 \le u < \infty$, such that

$$(4.1)\qquad P\left(\sup_{u\in[0,T]}\left|n\left(\hat{F}_n(u) - F(u)\right) - \sqrt{n}\,W(u)\right| > \frac{C}{\delta}\log n + x\right) < K\exp(-\lambda\delta^2 x)$$

for all $x > 0$, where $T < T_H$, $0 < \delta < 1 - H(T)$, $C$, $K$ and $\lambda$ are some positive universal constants. The process $W(u)$ has mean zero and the covariance function

$$(4.2)\qquad EW(s)W(t) = (1 - F(s))(1 - F(t))\gamma(\min(s,t)),$$

where

$$(4.3)\qquad \gamma(t) = \int_0^t \frac{dF(y)}{(1 - G(y))(1 - F(y))^2}.$$

Based on Major and Rejtö's result, $\hat{q}_n(\hat{F}_n) - \hat{q}_n(F)$ can be strongly approximated by a Gaussian random variable [see Xiang (1992)]. The variance of this Gaussian random variable is given in the following lemma.

LEMMA 3.1. *Let $W(t)$ be a Gaussian process with mean zero and a covariance function given by* (4.2) *and* (4.3). *Let $Q''(t)$ be continuous in a neighborhood of $p$ and let*

$$Y_n = \int_{-1}^1 Q'(p + uh_n)K(u)W(Q(p + uh_n))\,du.$$

*Assume that $\int_{-1}^1 K(x)\,dx = 1$ and $\int_{-1}^1 xK(x)\,dx = 0$. Then $Y_n$ is a Gaussian random variable with mean zero and variance*

$$(4.4)\qquad \begin{aligned}\sigma_n^2 &= \int_{-1}^1\int_{-1}^1 Q'(p + h_n s)Q'(p + h_n t)(1 - p - h_n s)(1 - p - h_n t)\\ &\quad\times K(s)K(t)\int_0^{Q(p+h_n s)\wedge Q(p+h_n t)} \frac{dF(y)}{(1 - G(y))(1 - F(y))^2}\,ds\,dt\\ &= \frac{(1-p)^2}{f^2(Q(p))}\int_0^{Q(p)} \frac{dF(y)}{(1 - G(y))(1 - F(y))^2}\\ &\quad - \frac{\Psi(K)h_n}{(1 - G(Q(p)))f^2(Q(p))} + o(h_n).\end{aligned}$$

PROOF. Obviously, $Y_n$ is a Gaussian random variable with mean zero. To show (4.4), from the covariance structure of $W(t)$, we have

$$\begin{aligned}\sigma^2(Y_n) &= \int_{-1}^{1}\int_{-1}^{1} Q'(p+h_n s)Q'(p+h_n t)(1-p-h_n s)(1-p-h_n t)\\ &\quad \times K(s)K(t)\int_0^{Q(p+h_n s)\wedge Q(p+h_n t)} \frac{dF(y)}{(1-G(y))(1-F(y))^2}\,ds\,dt\\ &= \int_{-1}^{1}\int_{-1}^{1} Q'(p+h_n s)Q'(p+h_n t)(1-p-h_n s)(1-p-h_n t)\\ &\quad \times K(s)K(t)\,ds\,dt\int_0^{Q(p)} \frac{dF(y)}{(1-G(y))(1-F(y))^2}\\ &\quad + \int_{-1}^{1}\int_{-1}^{1} Q'(p+h_n s)Q'(p+h_n t)(1-p-h_n s)(1-p-h_n t)\\ &\quad \times K(s)K(t)\int_{Q(p)}^{Q(p+h_n s)\wedge Q(p+h_n t)} \frac{dF(y)}{(1-G(y))(1-F(y))^2}\,ds\,dt\\ &= I_1 + I_2.\end{aligned}$$

It is easy to check that

$$I_1 = Q'^2(p)(1-p)^2\int_0^{Q(p)} \frac{dF(y)}{(1-G(y))(1-F(y))^2} + o(h_n). \tag{4.5}$$

For $I_2$, we have

$$\begin{aligned}I_2 &= \iint_{s\le t} Q'(p+h_n s)Q'(p+h_n t)(1-p-h_n s)(1-p-h_n t)\\ &\quad \times K(s)K(t)\int_{Q(p)}^{Q(p+h_n s)} \frac{dF(y)}{(1-G(y))(1-F(y))^2}\,ds\,dt\\ &\quad + \iint_{t\le s} Q'(p+h_n s)Q'(p+h_n t)(1-p-h_n s)(1-p-h_n t)\\ &\quad \times K(s)K(t)\int_{Q(p)}^{Q(p+h_n t)} \frac{dF(y)}{(1-G(y))(1-F(y))^2}\,ds\,dt = S_1 + S_2.\end{aligned}$$

It is easy to verify that

$$S_1 = S_2 = -\frac{Q'^2(p)h_n}{1-G(Q(p))}\int_{-1}^{1} xK(x)\int_{-1}^{x} K(y)\,dy\,dx + o(h_n). \tag{4.6}$$

The lemma follows from combining (4.5) and (4.6). □

To prove the theorems, the following lemma is essential to our approach.

LEMMA 3.2. *let $K(x)$ be Lipschitz of order 1 and let $Q'(t)$ be continuous in a neighborhood of $p$ with $0 < p < F(T_H)$. Assume that $f(Q(p)) > 0$. Then*

$$\sup_{-\infty < x < \infty} \left| P\left(\sqrt{n}\left(\hat{q}_n(\hat{F}_n) - \hat{q}_n(F)\right) \le x\sigma_n\right) - \Phi(x)\right| = O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right) \tag{4.7}$$

*and*

$$\sup_{-\infty < x < \infty} \left| P\left(\sqrt{n}\left(\hat{q}_n^*(\hat{F}_n) - \hat{q}_n(F)\right) \le x\sigma_n\right) - \Phi(x)\right| = O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right), \tag{4.8}$$

*where $\sigma_n$ is given by* (4.4).

The proof of Lemma 3.2 is provided in the Appendix.

PROOF OF THEOREM 2.1. We prove (2.1) only. Let $d_n = \sqrt{n}/\sigma_n(Q(p) - \hat{q}_n(F))$. The smoothness condition on $Q(t)$ implies $d_n^2 \sim \bar{\sigma}^{-2}\alpha_m^2 n h_n^{2(m+1)}$. Write $t_n = t\bar{\sigma}\sigma_n^{-1}$. Using Lemma 3.2, we have

$$\begin{aligned} &P\left(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)\right) \\ &\quad = \Phi(t_n + d_n) - \Phi(-t_n + d_n) + O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right) \\ &\quad = \Phi(t_n) - \Phi(-t_n) + \phi'(t_n)d_n^2 + o(d_n^2) + O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right). \end{aligned}$$

On the other hand, Theorem 2 in Janssen and Veraverbeke (1992) implies

$$P\left(\hat{\xi}_{pn} \in \bar{I}_t(n)\right) = \Phi(t) - \Phi(-t) + O\left(\frac{1}{\sqrt{n}}\right).$$

Hence, using Taylor's theorem,

$$\begin{aligned} &P\left(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)\right) - P\left(\hat{\xi}_{pn} \in \bar{I}_t(n)\right) \\ &\quad = 2\phi(t)(t_n - t) - t\phi(t)\bar{\sigma}^{-2}\alpha_m^2 n h_n^{2(m+1)} \\ &\qquad + o\left(d_n^2 + (t_n - t)^2\right) + O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right). \end{aligned} \tag{4.9}$$

It is easy to check that by Lemma 3.1,

$$t_n - t \sim \frac{t\Psi(K)}{2\sigma^2(1 - G(Q(p)))f^2(Q(p))}h_n.$$

This and (4.9) yield (2.1). □

PROOF OF THEOREM 2.2. We only show (2.6) and (2.7) for $j = 1$. If bandwidths and kernel functions are chosen such that $P(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)) - P(\hat{\xi}_{pn} \in \bar{I}_t(n)) \ge 0$, then from the definition of $i_{1t}(n)$, it is easy to prove that $\liminf_{n\to\infty}(i_{1t}(n))/n \ge 1$. For these bandwidths and kernel functions, we have from (2.1) and again the definition of $i_{1t}(n)$,

$$
\begin{aligned}
&P\Big(\hat{\xi}_{p i_{1t}(n)} \in \bar{I}_t(n)\Big) - P\Big(\hat{\xi}_{pn} \in \bar{I}_t(n)\Big) \\
&\quad = P\big(\hat{q}_n(\hat{F}_n) \in \bar{I}_t(n)\big) - P\Big(\hat{\xi}_{pn} \in \bar{I}_t(n)\Big) + O\left(\frac{1}{\sqrt{i_t(n)}}\right) \\
&\quad = \frac{t\phi(t)\Psi(K)}{f^2(Q(p))(1-G(Q(p)))\bar{\sigma}^2} h_n \\
&\qquad - \frac{t\phi(t)\alpha_m^2}{\bar{\sigma}^2} n h_n^{2(m+1)} + o\big(h_n + n h_n^{2(m+1)}\big).
\end{aligned} \tag{4.10}
$$

This and Theorem 2 in Janssen and Veraverbeke (1992) yield

$$
\begin{aligned}
&2t\phi(t_n^*)\left(\frac{i_{1t}^{1/2}(n)}{n^{1/2}} - 1\right) \\
&\quad = \frac{t\phi(t)\Psi(K)}{f^2(Q(p))(1-G(Q(p)))\bar{\sigma}^2} h_n \\
&\qquad - \frac{t\phi(t)\alpha_m^2}{\bar{\sigma}^2} n h_n^{2(m+1)} + o\big(h_n + n h_n^{2(m+1)}\big),
\end{aligned} \tag{4.11}
$$

where $t_n^* \in (t, t(i_{1t}^{1/2}(n))/n^{1/2})$ or $t_n^* \in (t(i_{1t}^{1/2}(n))/n^{1/2}, t)$ and the right-hand side of (4.11) must tend to 0. Hence, we have $(i_{1f}^{1/2}(n))/n^{1/2} \to 1$ and $t_n^* \to t$ as $n \to \infty$. Therefore,

$$
\begin{aligned}
\frac{i_{1t}^{1/2}(n) - n^{1/2}}{n^{1/2}} &= \frac{\Psi(K)}{2f^2(Q(p))(1-G(Q(p)))\bar{\sigma}^2} h_n \\
&\quad - \frac{\alpha_m^2}{2\bar{\sigma}^2} n h_n^{2(m+1)} + o\big(h_n + n h_n^{2(m+1)}\big)
\end{aligned} \tag{4.12}
$$

and (2.6) follows for $j = 1$. By ignoring the higher-order terms, it is easy to check that $i_{1t}(n) - n$ attains the maximum rate for $h = h_{\text{opt}}$ in (2.3). Equation (2.7) is a consequence of (2.3) and (2.6). □

## APPENDIX

Proof of Lemma 3.2. We prove (4.7) first. Write $G_n(x, o) = \int_0^x K((s-p)/h_n)\, ds$. Then for large $n$, using the change of variable theorem [Billingsley (1986), page 219],

$$
\begin{aligned}
\hat{q}_n(\hat{F}_n) - \hat{q}_n(F) &= \frac{1}{h_n}\int_{-\infty}^{\infty} x\, d\big(G_n(\hat{F}_n(x), p) - G_n(F(x), p)\big) \\
&= -\frac{1}{h_n}\int_{-\infty}^{\infty}\big(G_n(\hat{F}_n(x), p) - G_n(F(x), p)\big)\, dx \\
&= -\int_{-\infty}^{\infty}\left(\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n} K(u)\, du\right) dx = I_{1n} + I_{2n},
\end{aligned}
$$

where

$$(A.1)\qquad I_{1n} = -\int_{-\infty}^{\infty} K\left(\frac{F(x)-p}{h_n}\right)\left(\frac{\hat{F}_n(x)-F(x)}{h_n}\right)dx$$

and

$$(A.2)\qquad I_{2n} = -\int_{-\infty}^{\infty}\left(\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}\left(K(u)-K\left(\frac{F(x)-p}{h_n}\right)\right)du\right)dx.$$

From Major and Rejtö's result, write $I_{1n} = Z_n + r_n$ with

$$Z_n = -\frac{1}{\sqrt{n}\,h_n}\int_{-\infty}^{\infty} K\left(\frac{F(x)-p}{h_n}\right)W(x)\,dx$$

and

$$r_n = -\frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{F(x)-p}{h_n}\right)\tau_n(x)\,dx,$$

where $W(x)$ is a Gaussian process with mean zero and satisfies (4.2) and (4.3) and $\tau_n(x)$ is the difference between $\hat{F}_n(x) - F(x)$ and $(1/\sqrt{n})W(x)$. After a simple transformation, it is easy to see that, using Lemma 3.1, $\sqrt{n}\,Z_n$ is a Gaussian random variable with mean zero and variance $\sigma_n^2$ given by (4.4). It follows that, with probability 1,

$$(A.3)\qquad \frac{\sqrt{n}}{\sigma_n}\left(\hat{q}_n(\hat{F}_n) - \hat{q}_n(F)\right) = N(0,1) + \frac{\sqrt{n}}{\sigma_n}(r_n + I_{2n}),$$

where $N(0,1)$ stands for a standard normal random variable. Hence, using a well known lemma [cf. Serfling (1980), page 228], to prove (4.7), we only need to show

$$(A.4)\qquad p\left(\sqrt{n}\,|r_n + I_{2n}| \geq \frac{(\log n)^2}{h_n\sqrt{n}}\right) = O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right).$$

Let $\delta_n = (\log n)^2/h_n n$. First of all, for $n$ sufficiently large, there is a constant $T$ with $T < T_H$ and $0 < \delta < 1 - H(T)$ such that $Q(p + h_n y) \leq T < T_H$ for all $|y| \leq 1$. Hence for large $n$, from (4.1),

$$\begin{aligned} P(|r_n| \geq \delta_n) &= P\left(\left|\int_{-1}^{1} Q'(p + h_n y)K(y)\tau_n(Q(p + h_n y))\,dy\right| \geq \delta_n\right) \\ (A.5)\qquad &\leq P\left(\sup_{0\leq u\leq T} |n\tau_n(u)| \geq C_0 n\delta_n\right) \\ &\leq K\exp\left\{-\lambda\delta^2\left(C_0\frac{(\log n)^2}{h_n} - \frac{C}{\delta}\log n\right)\right\} = O\left(\frac{1}{n}\right), \end{aligned}$$

where $C_0$ is a constant. On the other hand, for an $\varepsilon > 0$ with $1 + \varepsilon \le (1 - p)/h_n$ (this holds if $n$ is large), we write

$$\begin{aligned} I_{2n} &= h_n \int_{-1-\varepsilon}^{1+\varepsilon} Q'(p + h_n x) \int_x^{(\hat{F}_n(Q(p+xh_n))-p)/h_n} (K(u) - K(x))\,du\,dx \\ &\quad + h_n \int_{1+\varepsilon}^{(1-p)/h_n} Q'(p + h_n x) \int_x^{(\hat{F}_n(Q(p+xh_n))-p)/h_n} (K(u) - K(x))\,du\,dx \\ &\quad + h_n \int_{-p/h_n}^{-1-\varepsilon} Q'(p + h_n x) \int_x^{(\hat{F}_n(Q(p+xh_n))-p)/h_n} (K(u) - K(x))\,du\,dx \\ &= S_{1n} + S_{2n} + S_{3n}. \end{aligned}$$

By

$$\begin{aligned} |S_{2n}| &\le h_n \int_{1+\varepsilon}^{(1-p)/h_n} |Q'(p + h_n x)| I\left( \frac{\hat{F}_n(Q(p + xh_n)) - p}{h_n} < 1 \right) \\ &\qquad \times \left| \int_x^{(\hat{F}_n(Q(p+xh_n))-p)/h_n} (K(u) - K(x))\,du \right| dx \\ &\le h_n \int_{1+\varepsilon}^{(1-p)/h_n} |Q'(p + h_n x)| \\ &\qquad \times I\left( \frac{\hat{F}_n(Q(p + (1+\varepsilon)h_n)) - (p + (1+\varepsilon)h_n)}{h_n} < -\varepsilon \right) \\ &\qquad \times \left| \int_x^{(\hat{F}_n(Q(p+xh_n))-p)/h_n} (K(u) - K(x))\,du \right| dx, \end{aligned}$$

Theorem 2 in Földes and Rejtö (1981) implies

$$\begin{aligned} P(|S_{2n}| \ge \delta_n) &\le P\big(|\hat{F}_n(Q(p + (1+\varepsilon)h_n)) \\ &\qquad - (p + (1+\varepsilon)h_n)| \ge \varepsilon h_n\big) = O\left(\frac{1}{n}\right). \end{aligned} \tag{A.6}$$

Similarly,

$$P(|S_{3n}| \ge \delta_n) = O\left(\frac{1}{n}\right). \tag{A.7}$$

Since $K(x)$ is Lipschitz of order 1, for a constant $C_1$,

$$|S_{1n}| \le C_1 h_n^{-1} \sup_{|u| \le 1+\varepsilon} \left| \hat{F}_n(Q(p + uh_n)) - (p + uh_n) \right|^2.$$

Again, using Theorem 2 in Földes and Rejtö (1981),

$$P(|S_{1n}| \geq \delta_n) \leq P\left( C_1 h_n^{-1} \sup_{|u| \leq 1+\varepsilon} \left| \hat{F}_n(Q(p+uh_n)) - (p+uh_n) \right|^2 \geq \delta_n \right)$$

$$\leq P\left( \sup_{|u| \leq 1+\varepsilon} \left| \hat{F}_n(Q(p+uh_n)) - (p+uh_n) \right| \geq C_1^{-1/2} h_n^{1/2} \delta_n^{1/2} \right) \tag{A.8}$$

$$\leq d_0 \exp\{-d_1 C_1^{-1} n \delta \delta_n h_n\} = O\left(\frac{1}{n}\right).$$

Combining (A.5)–(A.8), (A.4) follows.

Now we prove (4.8). Write

$$\hat{q}_n^*(\hat{F}_n) = \frac{1}{h_n} \int_0^1 \hat{F}_n^{-1}(x) R_n\left(\frac{x-p}{h_n}\right) dx$$

and introduce

$$\tilde{q}_n(F) = \frac{1}{h_n} \int_0^1 Q(x) K_n\left(\frac{x-p}{h_n}\right) dx,$$

where $K_n$ is defined by

$$K_n\left(\frac{x-p}{h_n}\right) = K\left(\frac{i/n - p}{h_n}\right), \qquad \frac{i-1}{n} < x \leq \frac{i}{n}, \qquad i = 0, \pm 1, \pm 2, \dots .$$

It is easy to check that

$$K_n(u) = 0 \quad \text{if } |u| \geq 1 + \frac{1}{nh_n}$$

and

$$\sup_{-\infty < u < \infty} |K_n(u) - K(u)| = O\left(\frac{1}{nh_n}\right).$$

The difference between $\hat{q}_n^*(\hat{F}_n)$ and $\hat{q}_n(F)$ is

$$\hat{q}_n^*(\hat{F}_n) - \hat{q}_n(F) = \hat{q}_n(\hat{F}_n) - \hat{q}_n(F) + \tilde{q}_n(F) - \hat{q}_n(F)$$
$$- \int_{-\infty}^{\infty} \int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n} (K_n(u) - K(u))\, du\, dx.$$

Hence, to complete the proof, we need to show

$$\tilde{q}_n(F) - \hat{q}_n(F) = O\left(\frac{(\log n)^2}{nh_n}\right) \tag{A.9}$$

and

$$P\left( \left| \int_{-\infty}^{\infty} \int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n} (K_n(u) - K(u))\, du\, dx \right| \geq \delta_n \right) = O\left(\frac{(\log n)^2}{h_n \sqrt{n}}\right), \tag{A.10}$$

where $\delta_n = (\log n)^2/h_n n$. Equation (A.9) follows easily from

$$|\tilde{q}_n(F) - \hat{q}_n(F)| \le \int_{|u|\le 1+1/nh_n} |Q(t+uh_n)||K_n(u) - K(u)|\,du$$

(A.11)

$$= O\left(\frac{1}{nh_n}\right) = o\left(\frac{(\log n)^2}{nh_n}\right).$$

To prove (A.10), we write, for a given $\varepsilon > 0$ with $\varepsilon > 1/(nh_n)$,

$$\int_{-\infty}^{\infty}\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\,dx$$

$$= \int_{Q(p+h_n+2h_n\varepsilon)}^{\infty}\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\,dx$$

$$+ \int_{Q(p-h_n-2h_n\varepsilon)}^{Q(p+h_n+2h_n\varepsilon)}\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\,dx$$

$$+ \int_{-\infty}^{Q(p-h_n-2h_n\varepsilon)}\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\,dx = T_1 + T_2 + T_3.$$

From

$$|T_1| \le \int_{Q(p+h_n+2h_n\epsilon)}^{\infty} I\left(\frac{\hat{F}_n(x) - p}{h_n} < 1 + \epsilon\right)$$

$$\times\left|\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\right|dx$$

$$\le \int_{Q(p+h_n+2h_n\varepsilon)}^{\infty} I\left(\frac{\hat{F}_n(Q(p+h_n+2h_n\varepsilon)) - (p+h_n+2h_n\varepsilon)}{h_n} < -\varepsilon\right)$$

$$\times\left|\int_{(F(x)-p)/h_n}^{(\hat{F}_n(x)-p)/h_n}(K_n(u) - K(u))\,du\right|dx$$

and an argument similar to that in the proof of (A.6),

(A.12)
$$P(|T_1| \ge \delta_n) = O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right).$$

Similarly,

(A.13)
$$P(|T_3| \ge \delta_n) = O\left(\frac{(\log n)^2}{h_n\sqrt{n}}\right).$$

Finally, there exists a constant $C_2$, such that

$$P(|T_2| \geq \delta_n)$$

$$\text{(A.14)} \quad \leq P\left( \frac{C_2}{nh_n} \sup_{Q(t-h_n-2h_n\varepsilon) \leq x \leq Q(t+h_n+2h_n\varepsilon)} \left| \hat{F}_n(x) - F(x) \right| \geq \delta_n \right)$$

$$= O\left( \frac{(\log n)^2}{h_n\sqrt{n}} \right).$$

Combining (A.11)–(A.14), (A.10) follows. □

**Acknowledgments.** The author is very grateful to the referees and especially to an Associate Editor for their valuable and constructive comments and helpful criticisms. The author would like to thank Professor Paul Vos for his help.

## REFERENCES

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

DOSS, H. and GILL, R. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Amer. Statist. Assoc.* **87** 869–877.

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

FALK, M. (1985). Asymptotic normality of the kernel quantile estimator. *Ann. Statist.* **13** 428–433.

FÖLDES, A. and REJTŐ, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete* **56** 75–86.

HODGES, J. L. and LEHMANN, E. L. (1970). Deficiency. *Ann. Math. Statist.* **41** 783–801.

JANSSEN, P. and VERAVERBEKE, N. (1992). The accuracy of normal approximations in censoring models. *Journal of Nonparametric Statistics* **1** 205–217.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

LIO, Y. L. and PADGETT, W. J. (1992). Asymptotically optimal bandwidth for a smooth nonparametric quantile estimator under censoring. *Nonparametric Statist.* **1** 219–229.

MAJOR, P. and REJTŐ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.

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

MÜLLER, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.* **12** 766–774.

NAIR, V. N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26** 265–275.

PADGETT, W. J. (1986). A kernel-type estimator of a quantile function from right-censored data. *J. Amer. Statist. Assoc.* **81** 215–222.

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

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

SHEATHER, S. J. and MARRON, J. S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.* **85** 410–416.

XIANG, X. (1992). Asymptotic theory for linear functions of ordered observations. Ph.D. dissertation, Dept. Statistics, Univ. Chicago.

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

YANG, S. S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.* **80** 1004–1011.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OREGON 97403