Some Asymptotic Properties of Kernel Estimators of a Density Function in Case of Censored Data

Author(s): Jan Mielniczuk


Published by: Institute of Mathematical Statistics

Stable URL: https://www.jstor.org/stable/2241251

Accessed: 19-10-2019 10:25 UTC
SOME ASYMPTOTIC PROPERTIES OF KERNEL ESTIMATORS OF A DENSITY FUNCTION IN CASE OF CENSORED DATA

BY JAN MIELNICKUZ

Polish Academy of Sciences

The kernel estimator is a widely used tool for the estimation of a density function. In this paper its adaptation to censored data using the Kaplan–Meier estimator is considered. Asymptotic properties of four estimators, arising naturally as a result of considering various types of bandwidths, are investigated. In particular we show that (i) both proposed estimators stemming from the nearest neighbor estimator have censoring-free variances and (ii) one of them is pointwise mean consistent.

1. Introduction. Consider the random censorship model with two sequences $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of i.i.d. nonnegative random variables such that $X_i, Y_i$ are independent ($i = 1, \ldots, n$). Let $F$ and $G$ be the unknown right continuous distribution functions of the $X$'s and the $Y$'s, respectively. It is assumed that $X_i$ and $Y_i$ have densities $f$ and $g$ with respect to Lebesgue measure on $\mathbb{R}^1$. We want to estimate $f$ using the following data:

$$Z_i = \min(X_i, Y_i), \quad \delta_i = \lfloor X_i \leq Y_i \rfloor, \quad i = 1, \ldots, n,$$

where $[A]$ for any event $A$ denotes the indicator function of $A$. Let $H$ be the distribution function of the $Z$'s and let $Z^{(1)}, \ldots, Z^{(n)}$ denote the ordered sample. The well-known KM estimator (Kaplan and Meier (1958)) is defined by

$$1 - F_n(u) = \prod_{i: Z^{(i)} \leq u} \left( \frac{n - i}{n - i + 1} \right)^{\delta^{(i)}}, \quad u < Z^{(n)},$$

$$= 0, \quad u \geq Z^{(n)},$$

$\delta^{(i)}$ being the concomitant of $Z^{(i)}$. The KM empirical survival function $1 - F_n$ will be denoted by $\bar{F}_n$.

Blum and Susarla (1980) introduced a kernel-type estimator of $f$, considered then by Földes, Rejtö, and Winter (1981b). The estimator is based on the KM estimator:

$$\hat{f}_n(x) = \frac{1}{h(n)} \int_{\mathbb{R}} K\left( \frac{x - y}{h(n)} \right) dF_n(y),$$

Received June 1984; revised July 1985.

AMS 1980 subject classifications. Primary 62G05; secondary 60F15.

Key words and phrases. Censored data, density estimator, $k$ nearest neighbor estimator, Kaplan–Meier estimator, kernel, random censorship model.

766
where \( h(n) \) is a sequence of positive numbers such that \( h(n) \to 0, \, nh(n) \to \infty, \) and \( K \) is a density function. Analogously, we define a \( k(n) \)th nearest uncensored neighbor estimator

\[
 f_n(x) = \frac{1}{R(n)} \int_R K \left( \frac{x-y}{R(n)} \right) dF_n(y),
\]

where \( R(n) \) is the distance from \( x \) to its \( k(n) \)th nearest uncensored neighbor and \( k(n) \) is a given sequence of integers such that \( k(n) \to \infty \) and \( k(n)/n \to 0 \). Tanner (1983) used this type of bandwidth in hazard rate estimation from censored data. Classical nearest neighbor estimators were studied by Moore and Yackel (1976, 1977), Mack and Rosenblatt (1980), and Mack (1980). Moreover, we introduce

\[
 f_n^*(x) = \frac{1}{h(n)} \int_R K \left( \frac{x-y}{h(n)} \right) dF_n(y),
\]

\[
 f_n^x(x) = \frac{1}{R(n)} \int_R K \left( \frac{x-y}{R(n)} \right) dF_n(y),
\]

where \( n_1 = \sum_{i=1}^n \delta_i \) is the number of uncensored (\( \delta_i = 1 \)) observations. Two other estimators were studied by Blum and Susarla (1980) and McNichols and Padgett (1984a). A thorough survey of recent results in density estimation for censored data is given in McNichols and Padgett (1984b).

We shall show that some properties of estimators (1.1)-(1.4) may be deduced from the properties of classic kernel estimators when the observations are not censored. The connections with the classical case are stated in Lemma 1 for \( f_n \) and \( f_n \) and in Lemma 2 for \( f_n^* \) and \( f_n^x \). The link is evident in the second case since then the uncensored observations may be treated as \( n_1 \) random variables distributed as \( (Z, \delta) \sim (Z, \delta_i) \) for any \( i \). In the first case we consider

\[
 F_n(y) = \frac{1}{n} \sum_{i=1}^n [Z_i \leq y, \delta_i = 1],
\]

which, on any compact interval \([0, d]\) may be interpreted as the empirical distribution function of some random variable \( W(d) \). In Section 3 some results on the consistency and weak convergence of the introduced estimators are proved. In particular it is shown that the asymptotic variances of \( f_n \) and \( f_n^* \) do not depend on censoring, as opposed to the asymptotic variances of \( f_n \) and \( f_n^* \). Finally, we state a (pointwise) mean consistency result for \( f_n^* \).

2. Classical analogues for \( f_n^*, f_n, f_n^*, \) and \( f_n^x \). Put \( p = P(\delta = 1) \) and \( q = 1 - p \). Observe first that defining \( R(n) \) as the distance to the \( k(n) \)th uncensored observation leaves \( R(n) \) undefined on the set \( A_n = \{ k(n) > n_1 \} \). However this has no influence on the asymptotic properties of \( R(n) \): If \( n_0 \) is such
that \( n > n_0 \) implies \( np - k(n) > npq \) then

\[
\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(\text{Bin}(n, p) < k(n)) \\
\leq \sum_{n=1}^{\infty} P(|\text{Bin}(n, p) - np| > np - k(n)) \\
\leq n_0 + \sum_{n=n_0}^{\infty} P(|\text{Bin}(n, p) - np| > npq) \\
\leq n_0 + \sum_{n=n_0}^{\infty} 2\exp(-2/9npq) < \infty.
\]

(At the end of this argument Bernstein's inequality was used; cf. Rényi (1970).)

From now on we denote by \( x \) a fixed point of \( R^+ \) such that \( f(x)G(x) > 0 \) where \( G = 1 - G \). Let \( W(x) = Z \cdot [\delta = 1] + (Z + x + 1) \cdot [\delta = 0] \) and let \( W_i(x) \) for \( i = 1, \ldots, n \) be defined in the same way as \( W(x) \) with \( Z \) and \( \delta \) replaced by \( Z_i \) and \( \delta_i \), respectively. Obviously, \((W_1(x), \ldots, W_n(x))\) is an i.i.d. sequence with \( W_i(x) \) distributed as \( W(x) \) for any \( i \). Let \( \bar{R}(n) \) be the distance from \( x \) to its \( k(n) \)th nearest neighbor in the sequence \((W_1(x), \ldots, W_n(x))\). For any function \( f \) denote by \( C(f) \) the set of continuity points of \( f \).

**Lemma 1.** If \( k(n)/\log \log n \to \infty \) and \( x \in C(f \cdot G) \) then

(i) \( \bar{R}(n) = R(n) \) for all but finitely many \( n \) a.s.
(ii) \( k(n)/2nR(n) \to f(x)G(x) \) a.s.

**Proof.** Using the results of Moore and Yackel (1976) we have

\[
\bar{R}(n) \to 0 \quad \text{a.s.}
\]

Thus the first \( k(n) \) neighbors lie in a small neighborhood of \( x \) and by the definition of \( \bar{R}(n) \) they must be uncensored. Since on \([0, x + 1)\) the random variable \( W \) has a density which is continuous at the point \( x \), assertion (ii) follows from (i) and Theorem 1 of Moore and Yackel (1976). \( \square \)

Note that \( \bar{F}_n(y) \) is the e.d.f. of \( W(x) \) on \([0, x + 1)\) and thus we may estimate \( f(x)G(x) \) by means of

\[
\frac{1}{h(n)} \int_R K\left(\frac{x-y}{h(n)}\right) d\bar{F}_n(y)
\]

or

\[
\frac{1}{\bar{R}(n)} \int_R K\left(\frac{x-y}{\bar{R}(n)}\right) d\bar{F}_n(y),
\]

using a kernel \( K \) which has a compact support.
THEOREM 1. Let $K$ be a bounded density function with support in $[-1,1]$. Assume that $x \in C(\mathbf{f} \cdot G) \cap C(g)$.

(i) If $k(n)/\log \log n \to \infty$ then, with probability 1,
$$f_{n}(x) - \frac{1}{G(x)R(n)} \int_{R} K\left(\frac{x - y}{R(n)}\right) d\tilde{F}_{n}(y) = O\left(\frac{\log \log n}{n}\right) + O\left(\frac{k(n)}{n}\right).$$

(ii) If $nh(n)/\log \log n \to \infty$ then, with probability 1,
$$f_{n}(x) - \frac{1}{G(x)h(n)} \int_{R} K\left(\frac{x - y}{h(n)}\right) d\tilde{F}_{n}(y) = O\left(\frac{\log \log n}{n}\right) + O(h(n)).$$

PROOF OF (i). Let $S(x, r) = \{y: |y - x| \leq r\}$,
$$\left|f_{n}(x) - \frac{1}{G(x)R(n)} \int_{R} K\left(\frac{x - y}{R(n)}\right) d\tilde{F}_{n}(y)\right| < \sup_{z \in R(n)} \max_{z_{i} \in S(x, R(n))} \frac{1}{nG(x)} - a_{n}(z_{i}),$$
where $a_{n}(z_{i})$ is the value of the jump of the KM estimator in $z_{i}$. Using (ii) of Lemma 1 it is enough to show that
$$\max_{z_{i} \in S(x, R(n))} \frac{1}{nG(x)} - a_{n}(z_{i}) = O\left(\frac{\log \log n}{n}\right) + O\left(\frac{k(n)}{n}\right).$$

Since $na_{n}(z_{i}) = \bar{F}_{n}(z_{i} - 0)/\bar{H}_{n}(z_{i} - 0)$ (Efron (1967)), where $\bar{H}_{n}$ is the empirical survival function for all observations, it follows that (2.1) is majorized by
$$\sup_{t \leq x + R(n)} \left|\frac{\bar{F}_{n}(t)}{\bar{H}_{n}(t)} - \bar{F}_{n}(t)\right| + \sup_{t \leq x + R(n)} \left|\frac{\bar{F}_{n}(t)}{\bar{H}(t)} - \bar{F}(t)\right|$$
$$+ \sup_{t \in S(x, R(n))} \left|\frac{1}{G(x)} - \frac{1}{G(t)}\right|.$$

The two first terms are $O(\log \log n/n)^{1/2}$ a.s. in view of the LIL for the Kolmogorov-Smirnov distance (cf. Serfling (1980)) and the result of Földes and Rejtö (1981a), respectively. Since the last term is majorized by $(G(x + R(n)) - G(x - R(n)))/(\bar{G}(x + R(n)) - \bar{G}(x - R(n)))^{-2}$ then (i) follows from
$$\frac{(G(x + R(n)) - G(x - R(n)))/R(n)}{2g(x) a.s.}$$
and $nR(n)/k(n) = O(1)$ a.s. $\Box$

PROOF OF (ii). The proof follows the lines of (i) with the equality
$$\lim_{n \to \infty} \sum_{i=1}^{n} [Z_{i} \in S(x, h(n)), \delta_{i} = 1]/nh(n) = 2f(x) \cdot \bar{G}(x)$$
used instead of (ii) of Lemma 1. Formula (2.2) is obvious in view of the strong
consistency result for the classical kernel estimator with the kernel uniform on 
[−1, 1] applied to \((W_i(x), \ldots, W_n(x))\).

Let us turn to \(f_n^*\) and \(\hat{f}_n^*\). Consider \((Z|\delta = 1)\) instead of \(W(x)\), and replace
\(W_1(x), \ldots, W_n(x)\) by the sequence of uncensored observations, denoted by
\(\bar{Z}_1, \ldots, \bar{Z}_n\). The density of \((Z|\delta = 1)\) is equal to \(f_i(x) = f(x)\bar{G}(x)/p\).

**Lemma 2.** Let \(k\) be a fixed integer \(\leq n\). Conditionally on \(n_i = k\), \(\bar{Z}_1, \ldots, \bar{Z}_n\)
is an i.i.d. sequence with density \(f_1\).

From Lemma 2, for \(T_{n_i}(t) = n_i^{-1}\sum_{i=1}^{n_i} [Z_i \leq t, \delta_i = 1]\) the estimators
\[\frac{1}{h(n_i)} \int_R K\left(\frac{x-y}{h(n_i)}\right) dT_{n_i}(y), \quad \frac{1}{R(n_i)} \int_R K\left(\frac{x-y}{R(n_i)}\right) dT_{n_i}(y)\]
can be viewed as the classical kernel estimators for \(f_1\), based on a sample of
random size \(n_i\). From Lemma 2 it is also easy to see that asymptotic properties of
these estimators such as convergence in probability and weak convergence, are
the same as the respective asymptotic properties of analogous estimators based
on \(n\) observations from the distribution of \((Z|\delta = 1)\). Moreover, an exact ana-
logue of Theorem 1 is true for \(f_n^*\) and \(\hat{f}_n^*\) and its proof is similar to that of
Theorem 1. Theorem 1 and its analogue allows us to study the asymptotic
properties of the proposed estimators.

**3. Asymptotic properties of the proposed estimators.** Below we list
some properties of the estimators \(f_n^*\) and \(\hat{f}_n^*\). They rely upon analogous properties
of their classical counterparts and, as for the consistency results, on the following
theorem of Moore and Yackel (1977). Let \(K\) satisfy the assumptions of Theorem
1 and the condition \(K(cu) \geq K(u)\) for any \(0 \leq c \leq 1\). For any fixed sequence
\(k(n)\) consider an arbitrary consistency result holding for the estimator with
kernel \(K\) and bandwidth \(h(n) = k(n)/n\) and also for the estimator with the
uniform kernel and bandwidth \(h(n)\). Then this result holds for the nearest
neighbor estimator with kernel \(K\) and the bandwidth based on \(k(n)\). The only
qualification to this argument is that the conditions on \(h(n)\) must also be
satisfied by \(\alpha h(n)\) for any \(\alpha > 0\). We assume that the conditions of Theorem 1
are satisfied.

**Corollary 1.** (i) If \(\sum_{i=1}^{\infty} \exp(-ck(n)) < +\infty\) for every \(c > 0\), \(K(cu) \geq K(u)\) for \(0 \leq c \leq 1\), then
\[f_n(x) - f(x) \to 0 \quad a.s.\]

(ii) If \(\sum_{i=1}^{\infty} \exp(-c\alpha h(n)) < +\infty\) for every \(c > 0\), then
\[\hat{f}_n(x) - f(x) \to 0 \quad a.s.\]

Corollary 1(ii) is an analogue of Theorem 1 in Devroye and Wagner (1979) in
the case of censored data.
If $K$ is the uniform kernel on $[-1,1]$ then $f_n(x)$ and $\hat{f}_n(x)$ are strongly consistent under the assumptions of Theorem 1.

**Corollary 2.** Let $g$ be continuous and let $f \cdot \overline{G}$ be continuous and positive. Assume that $H(T) > 0$.

(i) Suppose that $K$ is continuous and $K(cu) \geq K(u)$ for $0 \leq c \leq 1$. If $k(n)$ is a sequence of integers such that $k(n)/\log n \to +\infty$ then, with probability 1,

$$\lim_{n \to \infty} \sup_{0 \leq x \leq T} |f_n(x) - f(x)| = 0.$$  \hfill (3.1)

(ii) If $K$ is a continuous kernel then (3.1) holds with $f_n$ replaced by $\hat{f}_n$.

Corollary 2(ii) is a censored data version of Theorem A in Silverman (1978). The proof of Corollary 2 relies on the fact that $\overline{F}_n(y)$ is the e.d.f. of $W(T)$ on $[0, T + 1)$.

**Proof of (i)** By the aforementioned theorem of Silverman (1978) it is enough to show that strong convergence in Theorem 1 can be replaced by uniform strong convergence on $[0, T]$. To see this observe that since $k(n)/\log n \to \infty$ and $f \cdot \overline{G}$ is uniformly continuous on $[0, T]$, in view of Theorem 1 in Devroye and Wagner (1977) we have

$$\sup_{0 \leq x \leq T} |k(n)/n\overline{R}(n, x) - 2f(x)\overline{G}(x)| \to 0 \quad \text{a.s.}$$  \hfill (3.2)

It remains to consider the last term of the majorant occurring in the proof of Theorem 1 and to show that

$$\sup_{0 \leq x \leq T} G(x + \overline{R}(n)) - G(x - \overline{R}(n)) = O(k(n)/n) \quad \text{a.s.}$$

We have

$$\sup_{0 < x < T} G(x + \overline{R}(n)) - G(x - \overline{R}(n)) = \sup_{0 < x < T} \frac{k(n)}{n} \frac{n\overline{R}(n)}{k(n)} \frac{(G(x + \overline{R}(n)) - G(x - \overline{R}(n)))}{\overline{R}(n)}.$$  

Since $\sup \overline{R}(n)$ on $[0, T]$ tends to 0 a.s. and $g$ is uniformly continuous we have

$$\sup_{0 \leq x \leq T} |(G(x + \overline{R}(n)) - G(x - \overline{R}(n)))/\overline{R}(n) - 2g(x)| \to 0 \quad \text{a.s.}$$

Thus the proof of (i) is completed in view of (3.2) and the fact that $f \cdot \overline{G}$ is positive on $[0, T]$.

The proof of (ii) is similar.

**Remark.** Observe that the uniform strong convergence of $\hat{f}_n$ on $[0, T]$ is obtained, with the stronger condition on $h(n)$: $\Sigma \exp(-cnh^2) < +\infty$ for all positive $c$ and with $K$ of bounded variation, using the result of Nadaraya (1965)
and the inequality (Földes and Rejtö (1981a))

\[ P \left( \sup_{0 \leq x \leq T} |F_n(x) - F(x)| > \epsilon \right) \leq d_0 \exp\left( -n \epsilon^2 \delta^4 d_1 \right), \]

where \( \bar{H}(T) > \delta > 0 \) and \( \epsilon > 2^7/n \delta^2 \), \( d_0, d_1 \) being universal constants.

**Corollary 3.** (i) Assume that \( f \cdot \bar{G} \) has a bounded derivative in a neighborhood of \( x \). If \( k(n) = o(n^{2/3}) \) then

\[ (k(n))^{1/2}(f_n(x) - f(x)) \to \mathcal{N}\left(0, 2 f^2(x) \int K^2(y) \, dy\right). \tag{3.3} \]

(ii) Assume that \( K \) is an even function, \( f \) has a second derivative which is bounded in a neighborhood of \( x \), and \( h(n) = O(n^{-1/3}) \). Then

\[ (nh(n))^{1/2}(\hat{f}_n(x) - f(x)) \to \mathcal{N}\left(0, (\int f(x)/\bar{G}(x)) \int K^2(y) \, dy\right). \tag{3.4} \]

The corresponding uncensored data theorems are given in Moore and Yackel (1976) and in Rosenblatt (1971).

**Proof of (i).** Observe that for \( w_n(x) = (1/\bar{R}(n)) f_R K((x - y)/\bar{R}(n)) \, d\bar{R}(y) \)
we have

\[ (k(n))^{1/2}(w_n(x) - f(x)\bar{G}(x)) \to \mathcal{N}\left(0, 2 (f(x)\bar{G}(x))^2 \int R K^2(y) \, dy\right). \tag{3.5} \]

(Moore and Yackel (1976)). Since \( k(n) = o(n^{2/3}) \) implies \( [k(n)]^{1/2} = o(n/k(n)) \), (i) follows from (3.5) and Theorem 1.

**Proof of (ii).** Rosenblatt (1971) proved that under the conditions imposed on \( K \) in (ii) and \( h(n) = o(n^{-1/3}) \), \( w_n(x) = (1/h(n)) f_R K((x - y)/h(n)) \, d\bar{R}(y) \)
is asymptotically normal with mean \( f(x)\bar{G}(x) \) and asymptotic variance \( 1/(nh(n)) f(x)\bar{G}(x) \int R K^2(y) \, dy \). The result follows from the fact that \( (nh(n))^{1/2} = o(1/h(n)) \) for \( h(n) = o(n^{-1/3}) \).

Observe that (i) asymptotic variance of \( f_n \) does not depend on censoring and (ii) analogues of Corollaries 1, 2, and 3 for \( f_n^* \) and \( f_n^* \) are also true. The only difference is that in Corollary 3 the scaling sequences \( (nh(n))^{1/2} \) and \( (k(n))^{1/2} \) are replaced by \( (n_1 h(n_1))^{1/2} \) and \( (k(n_1))^{1/2} \), respectively.

We also state (Mielniczuk (1985)):

**Theorem 2.** Assume that conditions of Corollary 1(i) are satisfied. Suppose that \( \log n \cdot k(n)/n \to 0 \), \( f_1 \) is a bounded density function in a neighborhood of \( x \), and \( x \) satisfies \( f(x)\bar{G}(x) > 0 \). Then

\[ \int |f_n^*(x) - f(x)| \, dP \to 0. \]

This theorem is a censored data version of Theorem 4 of Moore and Yackel (1976). Basically, the proof of Theorem 2 is parallel to the proof of the corresponding theorem.
Acknowledgment. I thank J. Koronacki for his comments.

REFERENCES


