The Annals of Statistics
1986, Vol. 14, No. 2, 766–773

SOME ASYMPTOTIC PROPERTIES OF KERNEL ESTIMATORS OF A DENSITY FUNCTION IN CASE OF CENSORED DATA

By Jan Mielniczuk

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 $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ of i.i.d. nonnegative random variables such that X_i, Y_i are independent ($i = 1, \dots, n$). 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 R^1. We want to estimate f using the following data:

$$Z_i = \min(X_i, Y_i), \qquad \delta_i = [X_i \le Y_i], \qquad i = 1, \dots, n,$$

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

$$\begin{aligned} 1 - F_n(u) &= \prod_{i:\, Z_{(i)} \le u} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}, \qquad u < Z_{(n)}, \\ &= 0, \qquad u \ge Z_{(n)}, \end{aligned}$$

$\delta_{(i)}$ 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:

$$\hat{f}_n(x) = \frac{1}{h(n)} \int_R K\left(\frac{x-y}{h(n)} \right) dF_n(y), \tag{1.1}$$

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.

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

$$(1.2) \qquad f_n(x) = \frac{1}{R(n)} \int_R K\left(\frac{x-y}{R(n)}\right) dF_n(y),$$

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

$$(1.3) \qquad \hat{f}_n^*(x) = \frac{1}{h(n_1)} \int_R K\left(\frac{x-y}{h(n_1)}\right) dF_n(y),$$

$$(1.4) \qquad f_n^*(x) = \frac{1}{R(n_1)} \int_R K\left(\frac{x-y}{R(n_1)}\right) dF_n(y),$$

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 $\hat{f}_n$ and f_n and in Lemma 2 for $\hat{f}_n^*$ and f_n^*. The link is evident in the second case since then the uncensored observations may be treated as n_1 random variables distributed as $(Z|\delta = 1)$ where $(Z, \delta) \sim (Z_i, \delta_i)$ for any i. In the first case we consider

$$\tilde{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} [Z_i \le y, \delta_i = 1],$$

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 $\hat{f}_n$ and $\hat{f}_n^*$. Finally, we state a (pointwise) mean consistency result for f_n^*.

2. Classical analogues for $\hat{f}_n$, f_n, $\hat{f}_n^*$, and f_n^*. 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

$$\begin{aligned}\sum_{n=1}^{\infty} P(A_n) &= \sum_{n=1}^{\infty} P(Bin(n,p) < k(n)) \\ &\le \sum_{n=1}^{\infty} P(|Bin(n,p) - np| > np - k(n)) \\ &\le n_0 + \sum_{n=n_0}^{\infty} P(|Bin(n,p) - np| > npq) \\ &\le n_0 + \sum_{n=n_0}^{\infty} 2\exp(-2/9npq) < \infty.\end{aligned}$$

(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 $f(x)\overline{G}(x) > 0$ where $\overline{G} = 1 - G$. Let $W(x) = Z \cdot [\delta = 1] + (Z + x + 1) \cdot [\delta = 0]$ and let $W_i(x)$ for $i = 1, \dots, n$ be defined in the same way as $W(x)$ with Z and δ replaced by Z_i and δ_i, respectively. Obviously, $(W_1(x), \dots, W_n(x))$ is an i.i.d. sequence with $W_i(x)$ distributed as $W(x)$ for any i. Let $\overline{R}(n)$ be the distance from x to its $k(n)$th nearest neighbor in the sequence $(W_1(x), \dots, W_n(x))$. 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* $x \in C(f \cdot \overline{G})$ *then*

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

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

$$\overline{R}(n) \to 0 \quad \text{a.s.}$$

Thus the first $k(n)$ neighbors lie in a small neighborhood of x and by the definition of $\overline{R}(n)$ 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). □

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

$$\frac{1}{h(n)} \int_R K\left(\frac{x - y}{h(n)}\right) d\tilde{F}_n(y)$$

or

$$\frac{1}{\overline{R}(n)} \int_R K\left(\frac{x - y}{\overline{R}(n)}\right) d\tilde{F}_n(y),$$

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 $x \in C(f \cdot \bar{G}) \cap C(g)$.*

(i) *If $k(n)/\log\log n \to \infty$ then, with probability* 1,

$$f_n(x) - \frac{1}{\bar{G}(x)\bar{R}(n)} \int_R K\left(\frac{x-y}{\bar{R}(n)}\right) d\tilde{F}_n(y) = O\left((\log\log n/n)^{1/2}\right) + O(k(n)/n).$$

(ii) *If $nh(n)/\log\log n \to \infty$ then, with probability* 1,

$$\hat{f}_n(x) - \frac{1}{\bar{G}(x)h(n)} \int_R K\left(\frac{x-y}{h(n)}\right) d\tilde{F}_n(y) = O\left((\log\log n/n)^{1/2}\right) + O(h(n)).$$

PROOF OF (i). Let $S(x,r) = \{y\colon |y-x| \le r\}$,

$$\left| f_n(x) - \frac{1}{\bar{G}(x)\bar{R}(n)} \int_R K\left(\frac{x-y}{\bar{R}(n)}\right) d\tilde{F}_n(y)\right| < \sup K \frac{k(n)}{\bar{R}(n)} \max_{\substack{z_i \in S(x,\bar{R}(n)) \\ \delta_i = 1}} |1/n\bar{G}(x) - a_n(z_i)|,$$

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

$$\max_{\substack{z_i \in S(x,\bar{R}(n)) \\ \delta_i = 1}} |1/\bar{G}(x) - na_n(z_i)| = O\left((\log\log n/n)^{1/2}\right) + O\left(\frac{k(n)}{n}\right). \tag{2.1}$$

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

$$\sup_{t \le x + \bar{R}(n)} \left| \frac{\bar{F}_n(t)}{\bar{H}_n(t)} - \frac{\bar{F}_n(t)}{\bar{H}(t)} \right| + \sup_{t \le x + \bar{R}(n)} \left| \frac{\bar{F}_n(t)}{\bar{H}(t)} - \frac{\bar{F}(t)}{\bar{H}(t)} \right| + \sup_{t \in S(x,\bar{R}(n))} \left| \frac{1}{\bar{G}(x)} - \frac{1}{\bar{G}(t)} \right|.$$

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 $(G(x + R(n)) - G(x - R(n)))(\bar{G}(x + R(n)))^{-2}$ then (i) follows from

$$(G(x + \bar{R}(n)) - G(x - R(n)))/\bar{R}(n) \to 2g(x) \quad \text{a.s.}$$

and $n\bar{R}(n)/k(n) = O(1)$ a.s. □

PROOF OF (ii). The proof follows the lines of (i) with the equality

$$\lim_{n \to \infty} \sum_{i=1}^{n} [Z_i \in S(x,h(n)), \delta_i = 1]/nh(n) = 2f(x) \cdot \bar{G}(x) \tag{2.2}$$

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 $(W_1(x),\dots,W_n(x))$. □

Let us turn to f_n^* and $\hat{f}_n^*$. Consider $(Z|\delta = 1)$ instead of $W(x)$, and replace $W_1(x),\dots,W_n(x)$ by the sequence of uncensored observations, denoted by $\bar{Z}_1,\dots,\bar{Z}_{n_1}$. The density of $(Z|\delta = 1)$ is equal to $f_1(x) = f(x)\bar{G}(x)/p$.

LEMMA 2. *Let k be a fixed integer $\leq n$. Conditionally on $n_1 = k$, $\bar{Z}_1,\dots,\bar{Z}_{n_1}$ is an i.i.d. sequence with density f_1.*

From Lemma 2, for $T_{n_1}(t) = n_1^{-1}\sum_{i=1}^{n}[Z_i \leq t, \delta_i = 1]$ the estimators

$$\frac{1}{h(n_1)}\int_R K\left(\frac{x-y}{h(n_1)}\right) dT_{n_1}(y), \qquad \frac{1}{R(n_1)}\int_R K\left(\frac{x-y}{R(n_1)}\right) dT_{n_1}(y)$$

can be viewed as the classical kernel estimators for f_1, based on a sample of random size n_1. 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 analogue 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 $\hat{f}_n$ and 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 $\sum_{n=1}^{\infty}\exp(-ck(n)) < +\infty$ 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 $\sum_{n=1}^{\infty}\exp(-cnh(n)) < +\infty$ 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 $\overline{H}(T) > 0$.*

(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,

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

(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 $\tilde{F}_n(y)$ 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

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

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 $h(n)$: $\sum \exp(-cnh^2) < +\infty$ for all positive c and with K of bounded variation, using the result of Nadaraya (1965) and the inequality (Földes and Rejtö (1981a))

$$P\Big(\sup_{0\le x\le T}|F_n(x)-F(x)|>\varepsilon\Big)<d_0\exp\big(-n\varepsilon^2\delta^4 d_1\big),$$

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

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

$$(3.3)\qquad (k(n))^{1/2}(f_n(x)-f(x))\to_{\mathscr{L}} N\Big(0,2f^2(x)\int_R K^2(y)\,dy\Big).$$

(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*

$$(3.4)\quad (nh(n))^{1/2}(\hat{f}_n(x)-f(x))\to_{\mathscr{L}} N\Big(0,(f(x)/\overline{G}(x))\int_R K^2(y)\,dy\Big).$$

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

PROOF OF (i). Observe that for $w_n(x)=(1/\overline{R}(n))\int_R K((x-y)/\overline{R}(n))\,d\tilde{F}_n(y)$ we have

$$(3.5)\quad (k(n))^{1/2}(w_n(x)-f(x)\overline{G}(x))\to_{\mathscr{L}} N\big(0,2(f(x)\overline{G}(x))^2\textstyle\int_R K^2(y)\,dy\big)$$

(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 $h(n)=o(n^{-1/5})$, $w_n(x)=(1/h(n))\int_R K((x-y)/h(n))\,d\tilde{F}_n(y)$ is asymptotically normal with mean $f(x)\overline{G}(x)$ and asymptotic variance $1/(nh(n))f(x)\overline{G}(x)\int_R K^2(y)\,dy$. 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 $\hat{f}_n$ does not depend on censoring and (ii) analogues of Corollaries 1, 2, and 3 for $\hat{f}_n^*$ 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 $(n_1h(n_1))^{1/2}$ 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 $f(x)\overline{G}(x)>0$. 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

BLUM, J. R. and SUSARLA, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. In *Multivariate Analysis* **5** (P. R. Krishnaiah, ed.) 213–222. North-Holland, New York.

DEVROYE, L. P. and WAGNER, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5** 536–540.

DEVROYE, L. P. and WAGNER, T. J. (1979). The L^1 convergence of kernel density estimates. *Ann. Statist.* **7** 1136–1139.

EFRON, B. (1967). The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** 831–853.

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

FÖLDES, A., REJTŐ, L. and WINTER, B. B. (1981b). Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. *Period. Math. Hungar.* **12** 15–29.

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

MACK, Y. P. (1980). Asymptotic normality of multivariate *k-NN* density estimates. *Sankhyā Ser. A* **42** 63.

MACK, Y. P. and ROSENBLATT, M. (1979). Multivariate *k*-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

MCNICHOLS, D. T. and PADGETT, W. J., (1984a). A modified kernel estimator for randomly right censored data. *South African Statist. J.* **18** 13–27.

MCNICHOLS, D. T. and PADGETT, W. J. (1984b). Nonparametric density estimation from censored data. *Comm. Statist. A—Theory Methods* **13** 1581–1611.

MIELNICZUK, J. (1985). Kernel estimators of a density function in case of censored data. ICS PAS Report No. 560, Warsaw.

MOORE, D. S. and YACKEL, J. W. (1976). Large sample properties of nearest neighbor density function estimators. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, eds.) 269–279. Academic, New York.

MOORE, D. S. and YACKEL, J. W. (1977). Consistency properties of nearest neighbor density estimates. *Ann. Statist.* **5** 143–154.

NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probab. Appl.* **10** 186–190.

RÉNYI, A. (1970). *Probability Theory.* Akademiai Kiadó, Budapest.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

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

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimates of a density and its derivatives. *Ann. Statist.* **6** 177–184.

TANNER, M. A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. *Ann. Statist.* **11** 994–997.

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
P.O. BOX 22
PL-00-901 WARSAW, PKIN
POLAND