JOURNAL OF MULTIVARIATE ANALYSIS 49, 278–286 (1994)

Law of the Logarithm for Density and Hazard Rate Estimation for Censored Data*

XIAOJING XIANG

University of Oregon and University of Chicago

In this note, we establish law of the logarithm for kernel-type density and hazard rate estimators based on censored data. These results are applied to get optimal bandwidths with respect to strong uniform consistency. © 1994 Academic Press, Inc.

1. INTRODUCTION

Arbitrarily right-censored data arise naturally in industrial life testing and medical follow-up studies. Let $X_1, ..., X_n$ be independent and identically distributed (i.i.d.) nonnegative random variables with common distribution function $F(t)$, 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 $Y_1, ..., Y_n$ are assumed to be i.i.d. random variables with common distribution function $G(t)$. The observations in this model are the pairs (T_i, δ_i), where $T_i = \min(X_i, Y_i)$ and $\delta_i = I_{(X_i \leq Y_i)}$, $i = 1, 2, ..., n$. Clearly, T_i are i.i.d. with common distribution function $D(t) = 1 - (1 - F(t))(1 - G(t))$. Throughout this paper we assume that $F(t)$ and $G(t)$ are continuous. Let $f(t) = F'(t)$ be the density function of X_1. The hazard rate function is defined by

$$h(t) = \frac{f(t)}{1 - F(t)}, \qquad \text{for} \quad F(t) < 1. \tag{1.1}$$

Received June 29, 1992; revised September 4, 1993.

AMS 1980 subject classifications: primary 60F15, secondary 62G05.

Key words and phrases: random censorship, Kaplan–Meier estimator, kernel density estimator, hazard function, optimal bandwidth, strong Gaussian approximation, oscillation modulus.

* This manuscript was prepared using computer facilities supported in part by the National Science Foundation Grants DMS 89-05292, DMS 87-03942, DMS 86-01732, and DMS 84-04941 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

0047-259X/94 $6.00

Based on such right-censored data, one would like to estimate $f(t)$ and $h(t)$ uniformly on an interval. A very popular estimator of $f(t)$ is the kernel estimator defined by

$$f_n(t) = \frac{1}{a_n} \int K\left(\frac{t-x}{a_n}\right) d\hat{F}_n(x), \tag{1.2}$$

where $K(t)$ is an appropriate kernel function, $\{a_n\}$ is a sequence of bandwidths with $a_n \downarrow 0$, and $\hat{F}_n$ is the Kaplan–Meier estimator. The Kaplan–Meier estimator is of the form

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

where $T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(n)}$ are the order statistics of T_i and $\delta_{(1)}, ..., \delta_{(n)}$ are the corresponding δ_i. From (1.1) and (1.2), a natural estimator of $h(t)$ is

$$h_n(t) = \frac{f_n(t)}{1 - \hat{F}_n(t)}, \qquad \text{for} \quad t < T_{(n)}. \tag{1.3}$$

Estimation of $f(t)$ or $h(t)$ in the presence of censoring has been widely studied by Blum and Susarla [1], Földes *et al.* [4], Tanner and Wong [14], Padgett and McNichols [10], Mielniczuk [9], Marron and Padgett [8], Diehl and Stute [2], Lo *et al.* [6], and Karunamuni and Yang [5], among others. The strong uniform consistency rate of the estimator (1.2) can be found in [2, 5]. Let

$$\bar{f}_n(t) = \frac{1}{a_n} \int K\left(\frac{t-x}{a_n}\right) dF(x).$$

Under certain conditions, Diehl and Stute [2] showed

$$\lim_{n \to \infty} \sqrt{\frac{na_n}{\log a_n^{-1}}} \sup_{t \in J} \sqrt{\frac{1-G(t)}{f(t)}}\, |f_n(t) - \bar{f}_n(t)| = \left(2 \int K^2(x)\, dx\right)^{1/2} \quad \text{a.s.}, \tag{1.4}$$

where $J = [c, d]$ is an interval. Furthermore, if $f^{(r)}(t)$ is continuous on $J_\varepsilon = [c - \varepsilon, d + \varepsilon]$ for some $\varepsilon > 0$, Diehl and Stute's result implies

$$\sup_{t \in J} |f_n(t) - f(t)| = O\left(\left(\frac{\log n}{n}\right)^{r/(2r+1)}\right) \quad \text{a.s.} \tag{1.5}$$

with bandwidth $a_n = (\log n/n)^{1/(2r+1)}$. The rate given by (1.5) is optimal. In contrast, the strong uniform consistency rate obtained by Karunamuni and Yang [5, Thm. 2.2] is incorrect. In the absence of censoring, (1.4) reduces to Theorem 1.3 of Stute [13].

The first aim of this note is to prove (1.4) under the weaker condition. In [2], Diehl and Stute has required $g(t) = G'(t)$ is bounded. This condition has been relaxed based on a different approach. The second aim of this paper is to show a similar result to (1.4) for estimator (1.3). This result is applied to get the strong uniform consistency rate of $h_n(t)$ to $h(t)$.

In this paper, we require that $K(t)$ is symmetric and for some $l \geqslant 1$,

$$K(t) \in C^l(-\infty, \infty), \qquad K(t) \text{ has compact support } [-1, 1], \tag{1.6}$$

where

$$C^l(-\infty, \infty) = \{g\colon g^{(l)} \text{ is continuous on } (-\infty, \infty)\}.$$

We also require that for some integer $r \geqslant 2$,

$$\begin{gathered}\int_{-1}^{1} K(x)\, dx = 1; \qquad \int_{-1}^{1} x^j K(x)\, dx = 0,\ j = 1, ..., r-1; \\ \int_{-1}^{1} x^r K(x)\, dx = \alpha_r \neq 0.\end{gathered} \tag{1.7}$$

2. Main Results

In our approach, a strong embedding result due to Major and Rejtö [7] plays an important role. Let $H(t) = P(T_1 \leqslant t)$, $H^u(t) = P(T_1 \leqslant t,\ \delta_1 = 1)$, $H^c(t) = P(T_1 \leqslant t,\ \delta_1 = 0)$, and $T_H = \inf\{t\colon H(t) = 1\}$. Major and Rejtö [7] have shown that, for $t < T_H$,

$$\hat{F}_n(t) - F(t) = \frac{1}{\sqrt{n}} W_n(t) + r_n(t), \tag{2.1}$$

where

$$\begin{aligned} W_n(t) = (1 - F(t)) \Bigg\{ &\int_0^t \frac{B_n(H^u(y)) - B_n(1 - H^c(y))}{(1 - H(y))^2}\, dH^u(y) \\ &+ \frac{B_n(H^u(t))}{1 - H(t)} - \int_0^t \frac{B_n(H^u(y))}{(1 - H(y))^2}\, dH(y) \Bigg\}, \end{aligned} \tag{2.2}$$

and $B_n(t)$, $0 \leqslant t \leqslant 1$, is a Brownian bridge. Moreover,

$$\sup_{t \in [0, T]} |r_n(t)| = O\left(\frac{(\log n)^2}{n}\right), \qquad \text{a.s. for } T < T_H. \tag{2.3}$$

In order to use the result of modulus of continuity of Brownian bridge, we require hereafter that for the sequence of bandwidth $\{a_n\}$

$$\text{(i)}\ \ na_n \uparrow \infty; \qquad \text{(ii)}\ \ \frac{\log a_n^{-1}}{na_n} \to 0; \qquad \text{(iii)}\ \ \frac{\log a_n^{-1}}{\log \log n} \to \infty. \tag{2.4}$$

Let $B_n(t)$ be a Brownian bridge and $A(t)$ be a function defined on $[0, \infty)$ with $0 \leqslant A(t) \leqslant 1$. Assume that $A(t)$ has a uniformly continuous derivative $a(t)$ with $0 < \delta \leqslant a(t) \leqslant M < \infty$ for all $t \in J_\varepsilon \subset [0, \infty)$. We claim that the results of Stute [12] for $\alpha_n(t)$ and $\beta_n(t)$ also hold for $B_n(t)$ and $B_n(A(t))$, respectively. For example, from Shorack and Wellner [11, p. 559], we have

$$\lim_{n \to \infty} \sup_{\substack{\underline{c}a_n \leqslant t-u \leqslant \bar{c}a_n \\ t, u \in J}} \frac{|B_n(t) - B_n(u)|}{\sqrt{2(t-u) \log a_n^{-1}}} = 1 \qquad \text{a.s.}, \tag{2.5}$$

where $0 < \underline{c} \leqslant \bar{c} < \infty$ are fixed numbers. (2.5) is similar to Theorem 2.10 of Stute [12] and the analogue of Theorem 2.13 of Stute [12] is

$$\lim_{n \to \infty} \sup_{\substack{\underline{c}a_n \leqslant t-u \leqslant \bar{c}a_n \\ t, u \in J}} \frac{|B_n(A(t)) - B_n(A(u))|}{\sqrt{2(t-u)\, a(x_{u,t}) \log a_n^{-1}}} = 1 \qquad \text{a.s.}, \tag{2.6}$$

where $x_{u,t}$ is any point between u and t. Let

$$L_n(t) = \frac{1}{a_n} \int K\left(\frac{t-x}{a_n}\right) dB_n(A(x)). \tag{2.7}$$

With an argument similar to that of Stute [13], we have the following lemma.

LEMMA 2.1. *Suppose that* $a(t) = A'(t)$ *is continuous on* J_ε *with* $0 < \delta \leqslant a(t) \leqslant M < \infty$ *for all* $t \in J_\varepsilon$. *Assume that* (1.6) *holds for* $l = 1$. *Then with probability* 1,

$$\lim_{n \to \infty} \sqrt{\frac{a_n}{\log a_n^{-1}}} \sup_{t \in J} \frac{|L_n(t)|}{\sqrt{a(t)}} = \left(2 \int_{-1}^{1} K^2(x)\, dx\right)^{1/2}. \tag{2.8}$$

Let $J_\varepsilon \subset (0, T_H)$. The following result is a refined version of Corollary 2 of Diehl and Stute [2].

THEOREM 2.2. *Assume that $f(t)$ is continuous on J_ε with $0<\delta\leqslant f(t)\leqslant M<\infty$ for all $t\in J_\varepsilon$ and (1.6) holds for $l=1$. Then if $na_n=n^\alpha$ for some $\alpha>0$,*

$$\lim_{n\to\infty}\sqrt{\frac{na_n}{\log a_n^{-1}}}\sup_{t\in J}\frac{\sqrt{1-G(t)}\,|f_n(t)-\bar{f}_n(t)|}{\sqrt{f(t)}}=\left(2\int_{-1}^{1}K^2(x)\,dx\right)^{1/2}. \tag{2.9}$$

Remark. Assume that $f^{(r)}(t)$ is continuous on J_ε for some $r\geqslant 2$ and that (1.7) holds: Then from Theorem 2.2 and

$$\bar{f}_n(t)-f(t)=\frac{a_n^r}{r!}\alpha_r f^{(r)}(t)+o(a_n^r)$$

with the same argument as that of Stute [13], the optimal bandwidth is obtained by minimizing the term

$$\frac{a_n^r}{r!}\sup_{t\in J}\frac{\sqrt{1-G(t)}\,|f^{(r)}(t)|}{\sqrt{f(t)}}\int_{-1}^{1}|K(u)u^r|\,du+\left(\frac{2\log a_n^{-1}}{na_n}\int_{-1}^{1}K^2(u)\,du\right)^{1/2}.$$

The optimal bandwidth is of the order $O((\log n)/n)^{1/(2r+1)})$. The bandwidth with order $O((\log n)/n)^{1/(2r+1)})$ yields (1.5).

Let

$$\bar{h}_n(t)=\frac{\bar{f}_n(t)}{1-F(t)},\qquad \text{for}\quad F(t)<1.$$

The following result is a law of logarithm for the estimator $h_n(t)$.

THEOREM 2.3. *Under the assumptions of Theorem 2.2.,*

$$\lim_{n\to\infty}\sqrt{\frac{na_n}{\log a_n^{-1}}}\sup_{t\in J}\sqrt{\frac{1-D(t)}{h(t)}}\,|h_n(t)-\bar{h}_n(t)|=\left(2\int_{-1}^{1}K^2(x)\,dx\right)^{1/2}. \tag{2.10}$$

Remark. From Theorem 2.3, under the assumptions in the remark of Theorem 2.2, for $a_n=((\log n)/n)^{1/(2r+1)})$,

$$\sup_{t\in J}|h_n(t)-h(t)|=O\left(\left(\frac{\log n}{n}\right)^{1/(2r+1)}\right)\qquad\text{a.s.} \tag{2.11}$$

The rate given by (2.11) is optimal.

In the following proof, we use the notation $b_n\sim c_n$ if and only if $b_n/c_n\to 1$, as $n\to\infty$.

Proof of Theorem 2.2. From (2.1) and (2.3), we can write

$$f_n(t)-\bar{f}_n(t)\sim\frac{1}{\sqrt{n}\,a_n(1-G(t))}\int K\left(\frac{t-x}{a_n}\right)dW_{1n}(x)$$
$$+\sum_{i=2}^{3}\frac{1}{\sqrt{n}\,a_n}\int K\left(\frac{t-x}{a_n}\right)dW_{in}(x)+O\left(\frac{(\log n)^2}{na_n}\right), \qquad (2.12)$$

where

$$W_{1n}(t)=B_n(H^u(t))$$
$$W_{2n}(t)=(1-F(t))\int_0^t\frac{B_n(H^u(y))-B_n(1-H^c(y))}{(1-H(y))^2}\,dH^u(y)$$

and

$$W_{3n}(t)=-(1-F(t))\int_0^t\frac{B_n(H^u(y))}{(1-H(y))^2}\,dH(y).$$

Let

$$\omega_i(h)=\sup_{|u-t|\leqslant h,\,u,\,t\in J}|W_{in}(t)-W_{in}(u)|, \qquad i=1,2,3,$$

be the oscillation modulus of $W_{in}(t)$. Thus as $h\downarrow 0$, Lévy's theorem (cf. Shorack and Wellner [11, p. 534]) and the smoothness conditions imposed on $G(t)$ and $F(t)$ imply that with probability 1,

$$\omega_1(h)=O(h^{1/2}(\log h^{-1})^{1/2}) \qquad\text{and}\qquad \omega_i(h)=O(h), \qquad =2,3. \qquad (2.13)$$

Hence, if we write

$$L_n(t)=\frac{1}{a_n}\int K\left(\frac{t-x}{a_n}\right)dB_n(A(x)) \qquad (2.14)$$

with $A(t)=H^u(t)$, (2.13) implies that with probability 1,

$$f_n(t)-\bar{f}_n(t)\sim\frac{1}{\sqrt{n}\,(1-G(t))}L_n(t). \qquad (2.15)$$

On the other hand, from

$$H^u(t)=\int_0^t(1-G(y))\,dF(y),$$

we have

$$a(t)=A'(t)=(1-G(t))\,f(t).$$

Hence, Lemma 2.1 and (2.14) imply

$$\lim_{n\to\infty} \sqrt{\frac{a_n}{\log a_n^{-1}}} \sup_{t\in J} \frac{|L_n(t)|}{\sqrt{f(t)(1-G(t))}} = \left(2\int_{-1}^{1} K^2(x)\,dx\right)^{1/2}. \tag{2.16}$$

The conclusion follows from (2.15) and (2.16).

Proof of Theorem 2.3. Write

$$\begin{aligned} h_n(t)-\bar{h}_n(t) &= \frac{f_n(t)-\bar{f}_n(t)}{1-F(t)} + [f_n(t)-\bar{f}_n(t)]\left\{\frac{\hat{F}_n(t)-F(t)}{[1-\hat{F}_n(t)][1-F(t)]}\right\} \\ &\quad + [\bar{f}_n(t)-f(t)]\left\{\frac{\hat{F}_n(t)-F(t)}{[1-\hat{F}_n(t)][1-F(t)]}\right\} \\ &\quad + f(t)\left\{\frac{\hat{F}_n(t)-F(t)}{[1-\hat{F}_n(t)][1-F(t)]}\right\} \\ &= \frac{f_n(t)-\bar{f}_n(t)}{1-F(t)} + \sum_{i=1}^{3} I_{in}(t). \end{aligned} \tag{2.17}$$

The conclusion follows if we can show

$$\sup_{t\in J} |I_{in}(t)| = o\left(\left(\frac{\log a_n^{-1}}{na_n}\right)^{1/2}\right) \quad \text{a.s.} \quad \text{for } i=1,2,3. \tag{2.18}$$

We only show (2.18) for $i=3$. Other cases follow in a similar manner. Choose $0<\delta\leqslant(1/2)[1-F(d)]$ and write

$$\begin{aligned} I_{3n}(t) &= f(t)\left\{\frac{\hat{F}_n(t)-F(t)}{[1-F(t)-(\hat{F}_n(t)-F(t))][1-F(t)]}\right\} \\ &\quad \times I(\sup_{0\leqslant t\leqslant d} |\hat{F}_n(t)-F(t)|\leqslant\delta) \\ &\quad + f(t)\left\{\frac{\hat{F}_n(t)-F(t)}{[1-F(t)-(\hat{F}_n(t)-F(t))][1-F(t)]}\right\} \\ &\quad \times I(\sup_{0\leqslant t\leqslant d} |\hat{F}_n(t)-F(t)|>\delta) \\ &= S_{1n}(t)+S_{2n}(t). \end{aligned}$$

Let $M=\max_{t\in J} f(t)$. Since

$$\sup_{t\in J} |S_{1n}(t)| \leqslant \frac{2M}{(1-F(d))^2} \sup_{t\in J} |\hat{F}_n(t)-F(t)|,$$

Corollary 1 of Földes and Rejtö [3] implies

$$\sup_{t \in J} |S_{1n}(t)| = o\left(\left(\frac{\log a_n^{-1}}{na_n}\right)^{1/2}\right) \qquad \text{a.s.}$$

Moreover, for each $\varepsilon > 0$,

$$P\left(\sqrt{\frac{na_n}{\log a_n^{-1}}} \sup_{t \in J} |S_{2n}(t)| \geqslant \varepsilon\right) \leqslant P(\sup_{t \in J} |\hat{F}_n(t) - F(t)| > \delta).$$

Hence, Theorem 2 of Földes and Rejtö [3] and the Borel–Cantelli lemma imply

$$\sup_{t \in J} |S_{2n}(t)| = o\left(\left(\frac{\log a_n^{-1}}{na_n}\right)^{1/2}\right) \qquad \text{a.s.}$$

(2.18) holds for $i = 3$.

We now consider the strong uniform consistency rate of the kernel estimators of higher derivatives. Assume that (1.6) holds for $l = m+1$ for some $m \geqslant 1$. Consider the kernel estimator

$$f_n^{(m)}(t) = \frac{1}{a_n^{m+1}} \int K^{(m)}\left(\frac{t-x}{a_n}\right) d\hat{F}_n(x). \tag{2.19}$$

Define

$$\bar{f}_n^{(m)}(t) = \frac{1}{a_n^{m+1}} \int K^{(m)}\left(\frac{t-x}{a_n}\right) dF_n(x). \tag{2.20}$$

We only give the following result without proof.

THEOREM 2.4. *Assume that* (1.6) *holds for* $l = m+1$ *for some* $m \geqslant 1$ *and* f *is continuous on* J_ε *with* $0 < \delta \leqslant f(t) \leqslant M < \infty$ *for all* $t \in J_\varepsilon$. *Then if* $na_n = n^\alpha$ *for some* $\alpha > 0$,

$$\lim_{n \to \infty} \sqrt{\frac{na_n^{2m+1}}{\log a_n^{-1}}} \sup_{t \in J} \frac{\sqrt{1 - G(t)}\, |f_n^{(m)}(t) - \bar{f}_n^{(m)}(t)|}{\sqrt{f(t)}}$$
$$= \left(2 \int_{-1}^{1} [K^{(m)}(x)]^2\, dx\right)^{1/2}. \tag{2.21}$$

In addition if for some $r \geqslant 2$, $f^{(m+r)}(t)$ is continuous on J_ε, and assume that (1.7) holds. Then with $a_n = ((\log n)/n)^{1/(2(r+m)+1)}$,

$$\sup_{t \in J} |f_n^{(m)}(t) - f^{(m)}(t)| = O\left(\left(\frac{\log n}{n}\right)^{r/(2(r+m)+1)}\right) \qquad \text{a.s.} \tag{2.22}$$

Acknowledgments

I thank the referee and the Editor for their valuable comments which led to this greatly improved version of the manuscript.

References

[1] Blum, J. R., and Susarla, V. (1980). Maximal derivation theory of density and failure rate function estimates based on censored data. In *Multivariate Analysis* (P. R. Krishniah, Ed.), pp. 213–222. North-Holland, New York.

[2] Diehl, S., and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *J. Multivariate Anal.* **25** 299–310.

[3] Földes, A., and Rejtö, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete* **56** 75–86.

[4] Földes, A., Rejtö, L., and Winter, B. B. (1981). Strong consistency properties of non-parametric estimators for randomly censored data, part II: Estimation of density and failure rate. *Period. Math. Hungar.* **12** 15–29.

[5] Karunamuni, R. J., and Yang, S. (1991). Weak and strong uniform consistency rates of kernel density estimates for randomly censored data. *Canad. J. Statist.* **19** 349–359.

[6] Lo, S. H., Mack, Y. P., and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probab. Theory Related Fields* **80** 461–473.

[7] Major, P., and Rejtö, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.

[8] Marron, J. S., and Padgett, W. J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Ann. Statist.* **15** 1520–1535.

[9] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *Ann. Statist.* **14** 766–773.

[10] Padgett, W. J., and McNichols, D. T. (1984). Nonparametric density estimation from censored data. *Comm. Statist. Theory Methods* **13**, 1581–1611.

[11] Shorack, G. R., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[12] Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

[13] Stute, W. (1982). A law of the logarithm for kernel density estimators. *Ann. Probab.* **10** 414–422.

[14] Tanner, M. A., and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by kernel method. *Ann. Statist.* **11** 989–993.