Probab. Th. Rel. Fields 80, 461–473 (1989)

Probability Theory and Related Fields

Density and Hazard Rate Estimation for Censored Data via Strong Representation of the Kaplan-Meier Estimator

S.H. Lo[1], Y.P. Mack[2] and J.L. Wang[2]

[1] Harvard University, Department of Statistics, Cambridge, MA 02138, USA

[2] University of California, Division of Statistics, Davis, CA 95616, USA

Summary. We study the estimation of a density and a hazard rate function based on censored data by the kernel smoothing method. Our technique is facilitated by a recent result of Lo and Singh (1986) which establishes a strong uniform approximation of the Kaplan-Meier estimator by an average of independent random variables. (Note that the approximation is carried out on the original probability space, which should be distinguished from the Hungarian embedding approach.) Pointwise strong consistency and a law of iterated logarithm are derived, as well as the mean squared error expression and asymptotic normality, which is obtain using a more traditional method, as compared with the Hajek projection employed by Tanner and Wong (1983).

1. Introduction

Suppose $T_1, \ldots, T_n$ are i.i.d. nonnegative random variables ("lifetimes") with common continuous distribution function (d.f.) F and suppose $C_1, \ldots, C_n$ are i.i.d. nonnegative random variables ("censoring sequence") with common d.f. G. Assume also that the lifetimes and censoring sequence are independent. In the setting of survival analysis data with random right censorship, one observes the bivariate sample $(X_1, \delta_1), \ldots (X_n, \delta_n)$, where

$$X_i = T_i \wedge C_i, \qquad \delta_i = I\{T_i \leqq C_i\}$$

with $\wedge$ denoting minimum and $I\{\cdot\}$ denoting the indicator function on a set. One question of interest in survival analysis is the estimation of the hazard the function h defined as follows when it is further assumed that F has a density f:

$$h(x) = \frac{d}{dx}[-\log \bar{F}(x)] = f(x)/\bar{F}(x), \qquad \text{for } F(x) < 1,$$

with $\bar{F} = 1 - F$. (The quantity $H(x) = -\log \bar{F}(x)$ is called the cumulative hazard function.) In the setting without censoring, parametric models of monotone failure rate have been extensively studied (see Ch. 3 of Barlow and Proschan (1975)). The nonparametric estimation of $h(x)$ was initiated by Watson and Leadbetter (1964a, 1964b). Subsequent research works include Rice and Rosenblatt (1976) and Singpurwalla and Wong (1983). There are essentially three variants based on the delta-sequence smoothing introduced by Watson and Leadbetter (1964a, 1964b) and Rice and Rosenblatt (1976):

$$h_n^{(1)}(x)=\int k_n(x-u)\,dF_n(u)/\bar{F}_n(x), \qquad F_n(x)<1; \tag{1.1}$$

$$h_n^{(2)}(x)=\int k_n(x-u)\,dF_n(u)/\bar{F}_n(u)=\sum_n k_n(x-X_{(j)})/(n-j+1); \tag{1.2}$$

and

$$h_n^{(3)}(x)=\int k_n(x-u)\,dH_n(x)=\sum_n k_n(x-X_{(j)})\log[1-(n-j+1)^{-1}], \tag{1.3}$$

where F_n is the empirical d.f., $H_n(x)=-\log \bar{F}_n(x)$, $X_{(j)}$ is the j^{th} order statistic from the sample $\{X_i,\ i=1,\dots,n\}$ (note that since there is no consoring, $X_i=T_i$, $i=1,\dots,n$); and $\{k_n\}$ is a delta-sequence (see Walter and Blum (1979)), which in the kernel case is specialized by taking

$$k_n(v)=(1/b_n)\,k(v/b_n), \tag{1.4}$$

where k is usually a bounded, symmetric, density function, and $\{b_n\}$ is a so-called bandwidth sequence such that $b_n\to 0$, $nb_n\to\infty$ as $n\to\infty$. It was shown in Rice and Rosenblatt (1976) that $h_n^{(2)}(x)-h_n^{(3)}(x)=O_p(n^{-1})$, $h_n^{(i)}(x)$, $i=1,2,3$, all have the same asymptotic variance (i.e., variance of the asymptotic distribution), but $h_n^{(1)}(x)$ has a different asymptotic bias. No expressions for $Eh_n^{(1)}(x)$ or Var $h_n^{(1)}(x)$ were obtained.

When the data are subjected to random right censoring, the problem becomes more complex, primarily because the estimate of $F(\cdot)$, due to Kaplan and Meier (1958), now take on a product form:

$$KM_n(x)=\begin{cases}1-\prod\limits_{X_{(i)}\leqq x}^{n}\left(\dfrac{n-i}{n-i+1}\right)^{\delta(i)} & \text{if } x\leqq X_{(n)};\\ 1 \quad \text{if } x>X_{(n)} \text{ and the largest observation is uncensored.}\end{cases}$$

Here $\delta_{(i)}$ is the induced order statistic corresponding to $X_{(i)}$.

Since many well-studied properties of the empirical d.f. cannot be readily transferred to the Kaplan-Meier estimator, several researchers circumvented this technical difficulty by considering an equivalent problem on the uncensored observations (for example, Blum and Susarla (1980), Burke (1983)). Some researchers (for instance, Ramlau-Hansen (1983)) employed the method of counting processes. Still others (Földes, Rejtö and Winter (1981), Burke and Horvath (1984)) used a Chung Smirnov type result on the Kaplan-Meier estimator. To the credit of Tanner and Wong (1983), expressions for the bias and variance in the kernel case (essentially the form $h_n^{(2)}$ given in (1.2)) were obtained by direct calculations and asymptotic normality was proved by appealing to Hajek's projection. Padgett and McNichols (1984) gave a review of density and failure rate estimators for censored data.

Our present research is motivated by a recent result of Lo and Singh (1986) which establishes a strong uniform approximation of the Kaplan-Meier estimator by an average of i.i.d. random variables with a sufficiently small error. This allows for a more traditional approach to the hazard estimation problem for censored data. As contrasted with approaches mentioned in the paragraph above, our method will be a direct one. Although it will become apparent that we could equally well have considered the variants $h_n^{(2)}(x)$ or $h_n^{(3)}$, since there have been fewer investigations carried out for $h_n^{(1)}(x)$ (see (1.1)) with $F_n(x)$ replaced by a modified version $\Gamma_n(x)$ of the Kaplan-Meier estimator defined as follows to avoid the possibility that $KM_n(x)=1$:

$$\Gamma_n(x)=\begin{cases} 1-\prod\limits_{X_{(i)}\leqq x}^{n}\left(\dfrac{n-i+1}{n-i+2}\right)^{\delta_{(i)}}, & \text{if } x\leqq X_{(n)}; \\ \Gamma_n(X_{(n)}) & \text{if } x>X_{(n)} \text{ and the largest observation is uncensored.} \end{cases}$$

It is easily checked that $\bar{\Gamma}_n(x)\geqq(n+1)^{-1}$ for all x, and that

$$\sup_{0\leqq x\leqq T}|KM_n(x)-\Gamma_n(x)|=O(n^{-1}) \quad \text{a.s.}$$

for any $0<T<\inf\{t\geqq 0\colon\ L(t)=1\}$, where $\bar{L}(x)=\bar{F}(x)\cdot\bar{G}(x)=P(T_i>x,\ C_i>x)$. (Hereafter, a.s. will be an abbreviation for "almost surely".)

The main contribution of this paper is to derive an expression for the mean squared error (MSE) of $h_n^{(1)}(x)$ (Theorem 4.1) which to the best of our knowledge has not been hitherto obtained rigorously. In addition, a law of iterated logarithm and asymptotic normality are obtained for both the density and hazard rate estimators. Our arguments are based on Lemma 2.1 which is an improved version of Theorem 1 of Lo and Singh (1986).

In Section 2, we state the preliminaries needed for our presentation. In Section 3, we focus our attention on kernel density estimation under censoring via strong approximation. In Section 4, we present the law of iterated logarithm, asymptotic normality and mean squared error expression of our hazard rate estimate. Finally, in the last section, we conclude with relevant comments and some comparison with the nearest-neighbor method. For more details of the proofs we refer the reader to Lo, Mack and Wang (1985).

2. Preliminaries

We shall concentrate our analysis on the kernel method. *We assume throughout our discussion that f is continuous at x, $f(x)>0$, G is continuous at x and $L(x)<1$ for a given point x under consideration.* Let

$$f_n(x)=b_n^{-1}\int k(x-u)/b_n)\,d\Gamma_n(u) \tag{2.1}$$

be the kernel density estimate of $f(x)$ with kernel k and bandwith b_n. The assumptions we made on the kernel k are as follows:

(k1) k is a symmetric density function;

(k2) k is compactly supported with support $[-c, c]$;

(k3) k is continuous;

(k4) k is of bounded variation.

These assumptions are the "usual" ones encountered in the kernel method of curve estimation. We shall comment on the use of kernels with vanishing moments in the last section.

The assumptions we make on the bandwidth sequence $\{b_n\}$ are:

(b1) $b_n \to 0 \quad as \quad n \to \infty$;

(b2) $(\log n)^2 (n b_n)^{-1} \to 0 \quad \text{as} \quad n \to \infty$;

(b3) $b_n/b_m \to 1$ as $n, m \to \infty$ with $n/m \to 1$, and $(\log n)^4 = o[n b_n (\log\log n)]$;

(b4) $b_n = o[(\log\log n/n)^{1/5}]$.

These assumptions are not mutually exclusive. For example, (b3) implies (b2). Assumptions (b1) and (b2) are similar to the usual requirements one imposes in kernel estimation for the uncensored case. The extra factor $(\log n)^2$ in (b2) is a very small price one pays for using the Lo-Singh representation. (b3) (the same as condition (11) in Hall (1981)) is needed to derive the law of iterative logarithm for kernel density estimates (Theorem 3.4 (ii)).

The estimate that we consider is modelled after $h_n^{(1)}(x)$ (we continue to label this as $h_n^{(1)}(x)$ for convenience):

$$h_n^{(1)}(x) = f_n(x)/\bar{\Gamma}_n(x). \tag{2.2}$$

To analyze the asymptotic behavior of $h_n^{(1)}(x)$, we need to first analyze that of $f_n(x)$. As mentioned earlier our technique is motivated by the strong representation result (Theorem 1) of Lo and Singh (1986). In Lemma 2.1 we shall give an improved modified version of their result. Note that the rate of remainder $r_n(x)$ in (2.5) below is of order (log n/n), as compared to $(\log n/n)^{3/4}$ in their Theorem 1. We begin with some notations. Let $L_1(t) = P(X_i \leqq t,\ \delta_i = 1)$. For positive real z and x, and δ taking values 0 or 1, let

$$\zeta(z, \delta, x) = -g(z \wedge x) + [\bar{L}(z)]^{-1} \cdot I\{z \leqq x, \delta = 1\}, \tag{2.3}$$

where

$$g(y) = \int_0^y [\bar{L}(s)]^{-2}\, dL_1(s).$$

Let $\zeta_i(x) = \zeta(X_i, \delta_i, x)$, also let T be such that $L(T) < 1$. Note that the random variables $\zeta_i(x)$ are bounded, uniformly in $0 \leqq x \leqq T$, $E\zeta_i(x) = 0$, and

$$\mathrm{Cov}(\zeta_i(x), \zeta_i(y)) = g(x \wedge y), \tag{2.4}$$

(see Lo and Singh (1986)).

The following lemma provides the key ideas involved in the results in Sect. 3 and 4, the proof of which is given in Appendix A.1.

Lemma 2.1. *Let* $\xi_i(x)=\bar{F}(x)\cdot\zeta_i(x)$. *We have*

$$\Gamma_n(x)=F(x)+n^{-1}\sum_n \xi_i(x)+r_n(x), \tag{2.5}$$

where

$$\sup_{0\leqq x\leqq T} |r_n(x)|=O(\log n/n) \quad \text{a.s.}, \tag{2.6}$$

and for any $\alpha\geqq 1$,

$$\sup_{0\leqq x\leqq T} E|r_n(x)|^\alpha=O([\log n/n]^\alpha). \tag{2.7}$$

Finally, we state a lemma which by now is a standard device in the kernel estimation literature:

Lemma 2.2. *Assume the kernel* k *is a bounded density and* $(b1)$ *holds. Let* q *be an integrable function.*

(a) *For every continuity point* x *of* q, *we have*

$$\lim_n b_n^{-1}\int k((x-u)/b_n)\,q(u)\,du=q(x). \tag{2.8}$$

(b) *If in addition* k *is symmetric with finite second moment, and* q *is twice continuously differentiable at* x, *then*

$$b_n^{-1}\int k((x-u)/b_n)\,q(u)\,du=q(x)+(q''(x)/2)\int v^2\,k(v)\,dv\cdot b_n^2+o(b_n^2).$$

3. Strong Approximation of $f_n(x)$

We shall obtain, in Proposition 3.1, a strong approximation of $f_n(x)$ and use it to derive asymptotic properties of the kernel density estimate. Let ξ_i and r_n be defined as in Sect. 2. Also define

$$\beta_n^{(x)}=\int f(x-v\,b_n)\,k(v)\,dv-f(x), \tag{3.1}$$

$$\sigma_n(x)=(n\,b_n)^{-1}\sum_n\int \xi_i(x-v\,b_n)\,d\,k(v), \tag{3.2}$$

and

$$e_n(x)=b_n^{-1}\int r_n(x-v\,b_n)\,d\,k(v). \tag{3.3}$$

We note that the integrals are well-defined for large n since k is compactly supported. $\beta_n(x)$ and $\sigma_n(x)$ are essentially the bias and random fluctuation components of $f_n(x)$, respectively, and $e_n(x)$ is the error of approximation in Proposition 3.1 below.

Proposition 3.1. *Suppose k satisfies $(k2)-(k4)$. Then $f_n(x)$ admits the strong approximation on the interval* $[0, T]$:

$$f_n(x) = f(x) + \beta_n(x) + \sigma_n(x) + e_n(x), \tag{3.4}$$

where

$$\sup_{0 \leqq x \leqq T} |e_n(x)| = O([\log n/(nb_n)) \quad \text{a.s.} \tag{3.5}$$

and for any $\alpha \geqq 1$,

$$\sup_{0 \leqq x \leqq T} E|e_n(x)|^\alpha = O([\log n/(nb_n)]^\alpha). \tag{3.6}$$

Proof. Using the integration by parts lemma of Földes, Rejtö and Winter (1981) and Lemma 2.1, under assumptions $(k2)$ and $(k3)$, we have that if $x < T$, where $L(T) < 1$, then

$$\begin{aligned}
f_n(x) &= b_n^{-1} \int k((x-u)/b_n)\, d\Gamma_n(u) \\
&= b_n^{-1} \int \Gamma_n(x - vb_n)\, dk(v) \\
&= b_n^{-1} \int [F(x - vb_n) + n^{-1} \sum_n \xi_i(x - vb_n) + r_n(x - vb_n)]\, dk(v) \\
&= \int f(x - vb_n)\, k(v)\, dv + (nb_n)^{-1} \sum_n \int \xi_i(x - vb_n)\, dk(v) \\
&\quad + b_n^{-1} \int r_n(x - vb_n)\, dk(v) \\
&= f(x) + \beta_n(x) + \sigma_n(x) + e_n(x).
\end{aligned}$$

(3.5) and (3.6) follow from (2.6), (2.7) and $(k4)$. □

Theorem 3.2. (Bias and variance.) *Suppose k satisfies $(k1)-(k4)$, $\{b_n\}$ satisfies* $(b1)$, $(b2)$, $f(x) > 0$, *and that f is twice continuously differentiable at x, then*

$$E f_n(x) = f(x) + (f''(x)/2) \int v^2 k(v)\, dv \cdot b_n^2 + o(b_n^2) + o((nb_n)^{-1/2}). \tag{3.7}$$

$$\operatorname{Var} f_n(x) = (nb_n)^{-1} [f(x)/\bar{G}(x)] \int k^2(v)\, dv + o((nb_n)^{-1}). \tag{3.8}$$

Proof. See Appendix A.2. □

Note. The MSE of $f_n(x)$ can be obtained from the above theorem, and the usual balancing between bias and variance will give the optimal rate $n^{-1/5}$ for the bandwidth b_n as in the uncensored case.

The asymptotic normality of $f_n(x)$ now follows from Proposition 3.1 and Theorem 3.2:

Corollary 3.3 *(Asymptotic normality).* (i) *Suppose k satisfies $(k1)-(k4)$, $\{b_n\}$ satisfies* $(b1)$, $(b2)$, *we have as $n \to \infty$,*

$$(nb_n)^{1/2} [f_n(x) - E f_n(x)] \xrightarrow{d} N(0, \sigma^2).$$

(ii) *If in addition f is twice continuously differentiable at x, and $\{b_n\}$ satisfies $b_n = o(n^{-1/5})$ also, then as $n \to \infty$,*

$$(n b_n)^{1/2} [f_n(x) - f(x)] \xrightarrow{d} N(0, \sigma^2).$$

Here $\xrightarrow{d}$ *means convergence in distribution, and* $\sigma^2 = [f(x)/\bar{G}(x)] \int k^2(u)\, du$.

The next result can be obtained by verifying conditions (z) and (3) in Theorem 1 of Hall (1981) using the strong embedding results for the bivariate empirical process.

Theorem 3.4. *(Law of iterated logarithm.) Suppose k satisfies $(k1)-(k4)$.*

(i) *If $\{b_n\}$ satisfies $(b1)$ and $(b3)$, then*

$$\limsup_n \, [2 \log\log n/(n b_n)]^{-1/2} |\sigma_n(x)| = \{[f(x)/\bar{G}(x)] \int k^2(v)\, dv\}^{1/2} \quad \text{a.s.}$$

(ii) *If in addition f is twice continuously differentiable at x and $\{b_n\}$ satisfies also $(b4)$, then*

$$\limsup_n \, [2 \log\log n/(n b_n)]^{-1/2} |f_n(x) - f(x)| = \{[f(x)/\bar{G}(x)] \int k^2(v)\, dv\}^{1/2} \quad \text{a.s.}$$

4. Kernel Estimation of the Hazard Rate

Let $h_n^{(1)}(x)$ be the hazard rate estimate defined in (2.2). Using results in the previous section, we shall establish its asymptotic properties.

Theorem 4.1. *(Mean squared error). Under the assumptions of Theorem 3.2,*

$$MSE[h_n^{(1)}(x)] = \{[f''(x)/[2\bar{F}(x))] \int v^2 k(v)\, dv\}^2 \cdot b_n^4 + \{(h(x)/\bar{L}(x)) \int k^2(v)\, dv\} \cdot (n b_n)^{-1} + o(b_n^4 + (n b_n)^{-1}).$$

Proof. See Appendix A.3. □

Theorem 4.2. *(Asymptotic normality.) Let $\tau^2 = [h(x)/\bar{L}(x)] \int k^2(v)\, dv$.*

(i) *Under the assumptions of Corollary 3.3 (i), as $n \to \infty$,*

$$(n b_n)^{1/2} [h_n^{(1)}(x) - E\, h_n^{(1)}(x)] \xrightarrow{d} N(0, \tau^2).$$

((ii) *Under the assumptions of Corollary 3.3* (ii), *as $n \to \infty$,*

$$(n b_n)^{1/2} [h_n^{(1)}(x) - h(x)] \xrightarrow{d} N(0, \tau^2).$$

Proof. (i) is immediate from Corollary 3.3 (i) and Slutsky's Theorem. (ii) follows from Corollary 3.3 (ii) and the fact that

$$(n b_n)^{1/2}[h_n^{(1)}(x)-h(x)]=(n b_n)^{1/2}\{f_n(x)[(\bar{F}_n(x))^{-1}-(F(x))^{-1}] \\ +(\bar{F}(x))^{-1}[f_n(x)-f(x)]\},$$

since the first term on the right converges to zero in probability as $n\to\infty$. □

Corollary 4.3. *(Law of iterated logarithm.) Under the assumptions in Corollary 3.4* (ii),

$$\limsup_n [2\log\log n/(n b_n)]^{-1/2}|h_n^{(1)}(x)-h(x)|=[(h(x)/\bar{L}(x))\int k^2(v)\,dv]^{1/2} \quad \text{a.s.}$$

5. Concluding Comments

(a) We have seen in the above discussion the use of Lo and Singh's (1986) strong representation of the Kaplan-Meier estimator in analyzing kernel estimation of hazard rate functions. Although our technique can also be applied to $h_n^{(2)}(x)$ and $h_n^{(3)}(x)$, we have chosen to consider the estimates given by $h_n^{(1)}(x)$ as it is less explored. Our variance expression and asymptotic normality results are similar to those of $h_n^{(2)}(x)$ studied by Yandell (1983), although we have employed a more traditional approach. The bias for the three variants appear to be different in the scale constant but not the rate.

(b) Tanner (1983) mentioned that a nearest-neighbor approach may be preferable to the fixed bandwidth sequence approach from an extensive simulation experiment. This observation appears to have some theoretical support judging from the recent work of Liu and Van Ryzin (1985) which essentially used an asymmetric nearest-neighbor window. Both their findings (Theorems 4.3 and 4.4) and the findings of some other researchers on nearest-neighbor density estimation with censored data (for instance, Mielniczuk (1986)) suggest that the censoring mechanism may have no effect on the asymptotic variance for nearest-neighbor estimates. This may be an advantage in terms of constructing a confidence interval at a fixed point or a simultaneous confidence band if one wants to test for goodness-of-fit. Nevertheless, one cautions that the bias behavior of the Liu and van Ryzin (1985) variable histogram estimator suffers essentially the same drawback as nearest-neighbor density estimators in that it may be quite large at the tail regions of F (see Mack and Rosenblatt (1979)).

(c) A number of researchers in kernel estimation have studied the effects of kernels which may have vanishing moments. Its use, coupled with the assumption of a higher degree of smoothness of $h(x)$, can make the convergence of the bias to zero faster. This point of view was taken in Singpurwalla and Wong (1983). Of course one pays the price that the estimator so constructed may take on negative values if the sample size is not "large enough". For this reason we have kept the non-negativity of the kernel in this paper.

Acknowledgment. We are grateful to an anonymous referee for constructive comments. The research of J.L. Wang is supported in part by U.S. Air Force grant AFOSR-85-0268.

Appendix

A.1. Proof of Lemma 2.1

The following lemma is needed in the proof of Lemma 2.1.

Lemma A.1.

$$\log \bar{\Gamma}_n(x) - \log \bar{F}(x) = n^{-1} \sum_n \zeta_i(x) + R_n(x), \tag{A.1.1}$$

where

$$P\left(\sup_{0 \leqq x \leqq T} |R_n(x)| > a_n \right) = O(n^{-\beta}), \tag{A.1.2}$$

for any $\beta > 0$ *with* $a_n = \theta \cdot [\log n/n]$ *for some constant* θ *depending on* β.

Proof. Let $L_n(t) = n^{-1} \sum_n I\{X_i \leqq t\}$ and $L_{ln}(t) = n^{-1} \sum_n I\{X_i \leqq t,\ \delta_i = 1\}$ be the empirical distribution and subdistribution function respectively. If one checks the proof of Theorem 1 of Lo and Singh (1986) carefully, one will find that $R_n(x)$ is composed of three terms:

$$R_n(x) = R_{n1}(x) + R_{n2}(x) + R_{n3}(x),$$

where

$$R_{n1}(x) = \log \bar{\Gamma}_n(x) + \int_0^x [\bar{L}_n(s)]^{-1}\, dL_{1n}(s),$$

$$R_{n2}(x) = \int_0^x ([\bar{L}(s)]^{-1} - [\bar{L}_{1n}(s)]^{-1})\, dL_1(s) + \int_0^x ([\bar{L}(s)]^{-2} [\bar{L}(s) - \bar{L}_n(s)]\, dL_1(s)$$

$$= -\int_0^x \frac{[\bar{L}_n(x - \bar{L}(s)]^2}{[\bar{L}(s)]^2\, \bar{L}_n(s)}\, dL_1(s),$$

$$R_{n3}(x) = \int_0^x ([\bar{L}(s)]^{-1} - [\bar{L}_n(s)]^{-1})\, d(L_{1n}(s) - L_1(s)).$$

Tracing Lo and Singh's proof carefully and using their Lemma 1, it can be shown that

$$P\{ \sup_{0 \leqq x \leqq T} |R_{ni}(x)| > a_n \} = O(n^{-\beta}) \qquad \text{for } i = 1, 2. \tag{A.1.3}$$

A recent result of Burke, Csörgő and Horváth (1988) (Lemma) then implies that (A.1.3) also holds for $R_{n3}(x)$. Lemma A.1. is thus verified. $\square$

Proof of Lemma 2.1. By Taylor's expansion and Lemma A.1.,

$$-[\bar{\Gamma}_n(x) - F(x)] = [\log \bar{\Gamma}_n(x) - \log \bar{F}(x)] \cdot \bar{F}(x) + \Delta_n \cdot [\log \bar{\Gamma}_n(x) - \log \bar{F}(x)]^2$$

$$= -n^{-1} \sum_n \xi_i(x) + \bar{F}(x) \cdot R_n(x) + \Delta_n \cdot [\log \bar{\Gamma}_n(x) - \log \bar{F}(x)]^2, \tag{A.1.4}$$

where Δ_n is between $\bar{\Gamma}_n(x)$ and $\bar{F}(x)$ and is therefore bounded by one.

The last term on the right of the expansion can be shown to be $O(a_n)$ a.s. by (A.1.1). The second term on the right is $O(a_n)$ a.s. by (A.1.2) and the Borel-Cantelli lemma. Hence (2.5) and (2.6) follow.

For (2.7), we shall only demonstrate the case $\alpha=1$. Since $\xi_i(x)$'s are uniformly bounded and $(n+1)^{-1} \leqq \bar{\Gamma}_n(x) \leqq 1$ for all x in $[0, T]$, we have $\sup_{0 \leqq x \leqq T} |R_n(x)| = O(\log(n+1))$. Lemma A.1 then implies (choosing $\beta > 1$)

$$\sup_{0 \leqq x \leqq T} E|R_n(x)| = O(a_n) \tag{A.1.5}$$

and

$$\sup_{0 \leqq x \leqq T} E(R_n(x)^2) = O(a_n^2). \tag{A.1.6}$$

Using (A.1.5), (A.1.6) and the Taylor expansion (A.1.4) above, we have

$$\begin{aligned} \sup_{0 \leqq x \leqq T} E|\log \bar{\Gamma}_n(x) - \log \bar{F}(x)|^2 &\leqq \sup_{0 \leqq x \leqq T} 2[E(n^{-1} \sum_n \xi_i(x))^2 + E(R_n(x)^2)| \\ &= O(n^{-1}). \end{aligned}$$

Hence

$$\begin{aligned} \sup_{0 \leqq x \leqq T} E|r_n(x)| &\leqq \sup_{0 \leqq x \leqq T} \bar{F}(x) \cdot E|R_n(x)| + \sup_{0 \leqq x \leqq T} E|\log \bar{\Gamma}_n(x) - \log \bar{F}(x)|^2 \\ &= O(a_n). \end{aligned}$$

Lemma 2.1 is thus proved. □

A.2. Proof of Theorem 3.2.

Since $(b2)$ implies that $a_n/b_n = o((nb_n)^{-1/2})$, (3.7) follows from Proposition 3.1, (2.9) and (3.6).

To verify (3.8), consider first

$$\operatorname{Var} \sigma_n(x) = (nb_n^2)^{-1} \iint \bar{F}(x-ub_n) \cdot \bar{F}(x-vb_n)\, g[(x-ub_n) \wedge (x-vb_n)]\, dk(u)\, dk(v),$$

where we recall $g(y) = \int_0^y [\bar{L}(t)]^{-2}\, dL_1(t)$, with derivative

$$g'(t) = [\bar{L}(t)]^{-2}\, dL_1(t)/dt = f(t)/[\bar{G}(t) \cdot \bar{F}(t)^2]. \tag{A.2.1}$$

There exists a positive constant $\rho = \rho(x, c) < \infty$ such that for n sufficiently large,

$$\bar{F}(x-ub_n) \cdot \bar{F}(x-vb_n) = \bar{F}(x)^2 + \varepsilon_n(x, u, v) \cdot b_n \tag{A.2.2}$$

with $\sup_{u,v} |\varepsilon_n(x, u, v)| \geqq \rho$, since $|u| \leqq c$, $|v| \leqq$c. For n sufficiently large, write

$$\operatorname{Var} \sigma_n(x) = \sigma_n^*(x) + \varepsilon_n^*(x),$$

where

$$\sigma_n^*(x) = [\bar{F}(x)^2/nb_n^2)] \iint g[(x-ub_n)\wedge(x-vb_n)]\,dk(u)\,dk(v)$$

and

$$\varepsilon_n^*(x) = (nb_n)^{-1} \iint \varepsilon_n(x,u,v)\,g[(x-ub_n)\wedge(x-vb_n)]\,dk(u)\,dk(v).$$

Using integration by parts, we have

$$\int g[(x-ub_n)\wedge(x-vb_n)]\,dk(u) = (-b_n)\int_{-c}^{v} g'(x-ub_n)\,k(u)\,du.$$

Thus by Fubini Theorem and a change of variable, we have

$$\iint g[(x-ub_n)\wedge(x-vb_n)]\,dk(u)\,dk(v) = b_n \int k^2(u)\,g'(x-ub_n)\,du. \quad \text{(A.2.3)}$$

Hence by continuity of F and G at x and the dominated convergence theorem

$$(nb_n)\,\sigma_n^*(x) \to \bar{F}(x)^2\,g'(x)\int k^2(u)\,du. \quad \text{(A.2.4)}$$

Similarly, using (A.2.2), it can be shown that $\varepsilon_n^*(x) = O(n^{-1})$. Thus (3.8) follows by applying (3.6), (A.2.1) to (A.2.4), and Schwarz inequality to an expansion of Var $f_n(x)$ via (3.4). □

A.3. Proof of Theorem 4.1.

Consider

$$E[h_n^{(1)}(x) - h(x)]^2 = E[\mathrm{I} + \mathrm{II} + \mathrm{III}]^2, \quad \text{(A.3.1)}$$

where

$$\mathrm{I} = f_n(x)[(1/\bar{\Gamma}_n(x) - (1/\bar{F}(x))]$$

$$\mathrm{II} = [f_n(x) - E\,f_n(x)]/\bar{F}(x), \quad \text{(A.3.2)}$$

$$\mathrm{III} = [E\,f_n(x) - f(x)]/\bar{F}(x). \quad \text{(A.3.3)}$$

Note that III is deterministic and $E(\mathrm{II}\cdot\mathrm{III}) = 0$.

We shall show below that the main contribution comes from $E(\mathrm{II}^2)$ and $E(\mathrm{III}^2)$, all other terms in the quadratic expansion being of smaller order.

First observe that

$$E|\mathrm{I}| = E\left[f_n(x)\cdot\left|\frac{\bar{F}(x) - \bar{\Gamma}_n(x)}{\bar{F}(x)\,\bar{\Gamma}_n(x)}\right|\right]$$
$$\leq M(\bar{F}(x))^{-1}\{E[\bar{F}(x) - \bar{\Gamma}_n(x)]^2\}^{1/2}\cdot\{E[\bar{\Gamma}_n(x)]^{-2}\}^{1/2}$$

by Schwartz inequality, where $M=\sup_x |f_n(x)| < \infty$ by ($k2$) and ($k3$). Lemma 2.1 implies that $E[\bar{F}(x)-\bar{\Gamma}_n(x)]^2 = O(n^{-1})$ and

$$\begin{aligned} E[\bar{\Gamma}_n(x)]^{-2} &= [\bar{F}(x)-d_n]^{-2} \cdot P\{|\bar{\Gamma}_n(x)-\bar{F}(x)| \leqq d_n\} \\ &\quad + (n+1)^2 \cdot P\{|\bar{\Gamma}_n(x)-\bar{F}(x)| > d_n\} \\ &= O(1), \end{aligned}$$

where $d_n = d \cdot (\log n/n)^{1/2}$ for some $d>0$. Hence $E|\mathrm{I}| = O(n^{-1/2})$. Similarly, one can show that $E[\mathrm{I}^2] = O(n^{-1})$.

These facts together with Theorem 3.2 now imply that

$$\begin{aligned} E|\mathrm{I}\cdot\mathrm{III}| &= |\mathrm{III}| \cdot E|\mathrm{I}| = O(n^{-1/2}\, b_n^2) + o(n^{-1/2}(n b_n)^{-1/2}) \\ E|\mathrm{I}\cdot\mathrm{II}| &= O(n^{-1/2}(n b_n)^{-1/2} + O(n^{-1/2} \cdot (a_n/b_n)^{1/2} \cdot (n b_n)^{-1/4}). \\ &= o((n b_n)^{-1}). \end{aligned}$$

Let b_n be of the form cn^{-p}, where c, p are both positive constants. Since b_n^4 dominates $n^{-1/2}\, b_n^2$ for $p<1/4$, and $(n b_n)^{-1}$ dominates $n^{-1/2}\, b_n^2$ for $p>1/6$, the term $O(n^{-1/2}\, b_n^2)$ is always dominated by either b_n^4 or $(n b_n)^{-1}$ for any $p>0$. Also, $(n b_n)^{-1}$ always dominates $n^{-1/2}(n b_n)^{-1/2}$, for any $p>0$. Hence $E(\mathrm{II}^2)$ and $E(\mathrm{III}^2)$ will be the main contribution to the MSE of $h_n^{(1)}(x)$. The theorem now follows from (A.3.2) (A.3.3) and Theorem 3.2. $\square$

References

1. Barlow, R.E., Proschan, F.: Statistical theory of reliability and life testing: probability models. Holt, Rinehart and Winston, New York 1975
2. Blum, J.R., Susarla, V.: Maximal deviation theory of density and failure rate function estimates based on censored data. In: Krishnaiah, P.R. (ed.) Mult. Analysis vol. V, pp. 213–222. New York: North Holland, 1980
3. Breslow, N., Crowley, J.: A large sample study of the life table and product-limit estimates under random censorship. Ann. Stat. **2**, 437–443 (1974)
4. Burke, M.D.: Approximations of some hazard rate estimators in a competing risks model. Stoch. Processes Appl. **14**, 157–174 (1983)
5. Burke, M.D., Csörgő, S., Horváth, L.: A correction to and an improvement of strong approximation of some biometric estimates under random censorship. Probab. Th. Rel. Fields **79**, 51–57 (1988)
6. Burke, M.D., Horváth, L.: Density and failure rate estimation in a competing risks model. Sankhyā Ser. A. **46**, 135–154 (1984)
7. Földes, A., Rejtö, L., Winter, B.B.: Strong consistency properties of nonparametric estimators for randomly censored data. II: Estimation of density and failure rate. Period. Math. Hung. **12**, 15–29 (1981)
8. Hajek, J.: Asymptotic normality of simple linear statistics under alternatives. Ann. Math. Stat. **39**, 325–346 (1968)
9. Hall, P.: Laws of the iterated logarithm for nonparametric density estimators. Z. Wahrscheinlichkeitstheor. Verw. Geb. **56**, 47–61 (1981)
10. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**, 457–481 (1958)
11. Liu, R.Y.C., Van Ryzin, J.: A histogram estimator of the hazard rate with censored data. Ann. Stat. **13**, 592–605 (1985)

12. Lo, S.H., Mack, Y.P., Wang, J.L.: Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. Tech. Rep. No. 64, University of California, Davis. (1985)
13. Lo, S.-H., Singh, K.: The product-limit estimator and the bootstrap: Some asymptotic representations. Probab. Th. Rel. Fields **71**, 455–465 (1986)
14. Mack, Y.P., Rosenblatt, M.: Multivariate k-nearest neighbor density estimates. J. Multivariate. Anal. **9**, 1–15 (1979)
15. Mileniczuk, J.: Some asymptotic properties of kernel estimators of a density function in case of censored data. Ann. Stat. **13**, 766–773 (1986)
16. Padgett, W.J., McNichols, D.T.: Nonparametric density estimation from censored data. Commun. Stat. Theory Methods **13**, 1581–1613 (1984)
17. Ramlau-Hansen, H.: Smoothing counting process intensities by means of kernel functions. Ann. Stat. **11**, 453–466 (1983)
18. Rice, J., Rosenblatt, M.: Estimation of the log survivor function and hazard function. Sankhya Ser. A **38**, 60–78 (1976)
19. Singpurwalla, N.D., Wong, M.-Y.: Kernel estimators of the failure rate function and density function: An analogy. J. Am. Stat. Assoc. **78**, 478–481 (1983)
20. Tanner, M.A.: A note of the variable kernel estimator of the hazard function from randomly censored data. Ann. Stat. **11**, 994–998 (1983)
21. Tanner, M.A., Wong, W.H.: The estimation of the hazard function from randomly censored data by the kernel method. Ann. Stat. **11**, 989–993 (1983)
22. Walter, G., Blum, J.R.: Probability density estimation using delta sequences. Ann. Stat. **7**, 328–340 (1979)
23. Watson, G.S., Leadbetter, M.R.: Hazard analysis. I. Biometrika **51**, 175–184 (1964a)
24. Watson, G.S., Leadbetter, M.R.: Hazard analysis. II. Sankhyā Ser. A **26**, 101–116 (1964b)
25. Yandell, B.S.: Nonparametric inference for rates and densities with censored serial data. Ph.D. dissertation, Biostatistics Program, University of California, Berkeley (1981)
26. Yandell, B.S.: Nonparametric inference for rates and densities with censored survival data. Ann. Stat. **11**, 1119–1135 (1983)

Received November 1, 1986; in revised form May 10, 1988