

NONPARAMETRIC ESTIMATION OF HAZARD FUNCTIONS AND THEIR DERIVATIVES UNDER TRUNCATION MODEL*

ÜLKÜ GÜRLER¹ AND JANE-LING WANG²

¹*Bilkent University, Faculty of Industrial Engineering, Ankara, Turkey*

²*Division of Statistics, University of California, Davis, CA 95616-8705, U.S.A.*

(Received August 26, 1991; revised September 16, 1992)

Abstract. Nonparametric kernel estimators for hazard functions and their derivatives are considered under the random left truncation model. The estimator is of the form of sum of identically distributed but dependent random variables. Exact and asymptotic expressions for the biases and variances of the estimators are derived. Mean square consistency and local asymptotic normality of the estimators are established. Adaptive local bandwidths are obtained by estimating the optimal bandwidths consistently.

Key words and phrases: Adaptive bandwidth choice, consistency, Hájek projection, kernel estimate, mean square error, tightness.

1. Introduction

Let X be a random variable (r.v.) of interest, referred to as the lifetime. In practice the observation of X may be prevented by another independent random variable Y called the truncation variable. Suppose (X_i, Y_i) , $i = 1, \dots, N$ is a random sample of (X, Y) . Then, under the random left truncation model, one observes only those i.i.d. pairs (X_i, Y_i) for which $Y_i \leq X_i$. We index those observed pairs by $i = 1, \dots, n$. There is a similarity of the left truncation model to the left censoring model studied by Csörgö and Horváth (1980), but the number of observations n in the former is a random variable.

Much of the literature has been devoted to the censoring model and the statistical interest in the truncation model spurred only more recently, partly due to its applicability to AIDS data (Lui *et al.* (1986), Lagakos *et al.* (1988), Kalbfleisch and Lawless (1989)). More applications of the random left truncation model can be found in Allredge and Gates (1985), among others.

Let F and G be the (right continuous) distribution functions (d.f.) of X and Y respectively. The nonparametric maximum likelihood estimator (MLE) of F was

* Research supported by Air Force Grant AFOSR 89-0386. Part of the work of Ülkü Gürler was done while she was a Ph.D. student at the Department of Statistics, the Wharton School of the University of Pennsylvania.

first suggested by Lynden-Bell (1971) and studied by Woodroffe (1985), Wang *et al.* (1986), Chao and Lo (1988), Gu and Lai (1990) and Keiding and Gill (1990). In this paper, our interest focuses on the hazard function λ of F defined as

$$(1.1) \quad \lambda(x) = f(x)/[1 - F(x)], \quad \text{for } F(x) < 1,$$

where f is the probability density function of F .

The hazard function is important for the assessment of risks and has been studied extensively for randomly censored data, e.g., hazard estimates of the type (2.10) were studied by Ramlau-Hansen (1983), Tanner and Wong (1983), Yandell (1983), Diehl and Stute (1988) and Müller and Wang (1990) among others. However, little is known about hazard estimation for truncated data although this problem is of applied interest. For example, in the Channing House data in Hyde (1977), F is the lifetime distribution for males, and $\lambda(t)$ is their hazard at age t which is of demographic interest. However the lifetime X is subject to left truncation since the data consists of only those males who were alive at the time the study started; thus the truncation variable Y is the age at entry into the study. A detailed analysis of these data can be found in Gürler and Wang (1992).

Although our main interest is the hazard function itself, we consider the more general problem of estimating its r -th derivative, $\lambda^{(r)}$, for $r \geq 0$. One motivation being that they are involved in the choice of data-dependent optimal bandwidths. We consider the kernel hazard estimator $\hat{\lambda}^{(r)}$ in (2.10) of $\lambda^{(r)}$ by convolving the kernel with a cumulative hazard estimator. Exact and asymptotic expansions for the mean and variance of $\hat{\lambda}^{(r)}(z)$ are given in Theorems 3.1 and 3.2 which then imply the mean square consistency of $\hat{\lambda}^{(r)}(z)$. The computation of the variance term of $\hat{\lambda}^{(r)}(z)$ (cf. (3.2) and Appendix A) is much more complicated under the present truncation model than the censoring model. Asymptotic normality of $\hat{\lambda}^{(r)}(z)$ is obtained in Theorem 3.3 via the Hájek projection method (Hájek (1968)).

It is well known that the choice of bandwidths is crucial for the quality of the resulting kernel estimate and that the optimal local bandwidth depends on the curvature at a point. This was first noticed, for kernel density estimators by Parzen ((1962), equation (4.15)). This effect is magnified, even for i.i.d. observations, for kernel estimated hazard functions since the variance of $\hat{\lambda}(z)$ tends to infinity as z tends to the right end of the support of F . The left truncation scheme further complicates the situation and the variance of $\hat{\lambda}^{(r)}(z)$ also blows up as z tends to zero (cf. (3.5)). Local bandwidth choice is therefore considered here instead of a global one. The optimal local bandwidth b^* depends on unknown quantities (cf. (3.8)). We show in Theorem 4.1 that any consistent estimator of it will give rise to a kernel hazard estimator which possesses the same limiting distribution as the kernel hazard estimator employing the optimal bandwidth. Such procedures provide efficient methods for hazard estimation and the resulting bandwidths are called locally adaptive bandwidths. Some choices of locally adaptive bandwidths are given in Section 4. Lengthy proofs are relegated to the Appendices.

2. Kernel hazard estimates

We shall assume without loss of generality that both X and Y are nonnegative random variables. We adopt Woodroffe's (1985) notation throughout the presentation. The cumulative hazard function of F (or X) is:

$$(2.1) \quad \Lambda(x) = \int_0^x \lambda(t)dt = -\log(1 - F(x)).$$

For any d.f. W , define

$$a_W = \inf\{t : W(t) > 0\} \quad \text{and} \quad b_W = \sup\{t : W(t) < 1\},$$

as the left and right endpoints of the support of W . As Woodroffe (1985) points out, in random truncation models, F can be estimated completely only if $a_G \leq a_F$. We shall assume this and put

$$(2.2) \quad \alpha \equiv \alpha(F, G) = P(Y \leq X) = \int_0^\infty G(x)dF(x) > 0.$$

Let H^* denote the joint distribution of the observed (X, Y) pair, and F^* and G^* denote the corresponding marginal d.f.'s. Then

$$(2.3) \quad H^*(x, y) = P(X \leq x, Y \leq y \mid Y \leq X) = \alpha^{-1} \int_0^x G(\min(y, t))dF(t),$$

$$(2.4) \quad F^*(x) = H^*(x, \infty) = \alpha^{-1} \int_0^x G(t)dF(t),$$

$$(2.5) \quad G^*(y) = H^*(\infty, y) = \alpha^{-1} \int_0^\infty G(\min(y, t))dF(t).$$

Theorem 1 of Woodroffe (1985) gives the following representation for the cumulative hazard Λ :

$$(2.6) \quad \Lambda(x) = \int_0^x dF^*(z)/C(z),$$

where

$$(2.7) \quad C(z) = P(Y \leq z \leq X) = G^*(z) - F^*(z-) = \alpha^{-1}G(z)[1 - F(z-)].$$

Note that C is not monotone and $C(z)$ tends to zero as z tends to either a_G or b_F . The representation (2.6) then suggests estimating $\Lambda(z)$ by

$$(2.8) \quad \hat{\Lambda}_n(z) = \int_0^z [C_n(x)]^{-1} dF_n^*(x) = \sum_{i: X_i \leq z} [nC_n(X_i)]^{-1},$$

where F_n^* and C_n are the empirical functions of F^* and C , e.g.,

$$(2.9) \quad C_n(z) = \#\{i : Y_i \leq z \leq X_i\}/n.$$

Notice that $C_n(X_i) \geq 1/n$, however, it is not a monotone function.

We consider the following kernel estimator for $\lambda^{(r)}(z)$ by convolving a kernel K_r with $\hat{\Lambda}_n$ in (2.8):

$$(2.10) \quad \hat{\lambda}^{(r)}(z) = \frac{1}{b^{r+1}} \int K_r \left(\frac{z-x}{b} \right) d\hat{\Lambda}_n(x) \equiv \sum_{i=1}^n K_{r,b}(z - X_i) \frac{1}{nC_n(X_i)},$$

where $K_{r,b}(x) = b^{-(r+1)}K_r(x/b)$, and $b = b_n$ is the bandwidth sequence.

To obtain the properties of $\hat{\lambda}^{(r)}$ we need to assume that:

(A1) for some $p \geq r$, λ is p times continuously differentiable at z .

As for the bandwidth sequence we require that:

(B1) $b_n \rightarrow 0$,

(B2) $nb_n^{2r+1} \rightarrow \infty$.

For the kernel function it is assumed that: K_r is a function of bounded variation with support $[-1, 1]$ and it is a function of order (r, p) , i.e., K_r satisfies

$$(2.11) \quad K_r \in M_{r,p} = \left\{ q \in L^2[-1, 1] : \int q(x)x^j dx = \begin{cases} (-1)^r r! & j = r \\ 0 & 0 \leq j < p, j \neq r \\ \neq 0 \text{ but finite} & j = p \end{cases} \right\}.$$

Note that under (2.11) K_r and $K_{r,b}(x)$ implicitly involve p ; however, for brevity of notation this is suppressed.

3. Mean square consistency and asymptotic normality

We will derive in this section the properties of $\hat{\lambda}^{(r)}(z)$ for $a_G < z < b_F$. The notations in Sections 1 and 2 are used. All expectations hereafter are with respect to conditioning on n , the number of observations.

THEOREM 3.1. (Mean and variance)

$$(3.1) \quad E(\hat{\lambda}^{(r)}(z)) = \int K_{r,b}(z-x)[1 - (1 - C(x))^n]d\Lambda(x)$$

and

$$(3.2) \quad \text{Var}(\hat{\lambda}^{(r)}(z)) = \int K_{r,b}^2(z-x)I_n(C(x))d\Lambda(x) \\ + 2 \int_{t < s} K_{r,b}(z-t)K_{r,b}(z-s) \\ \cdot \left\{ [1 - C(s)]^n - \frac{1 - F(t)}{F(s) - F(t)} [[1 - C(s)]^n - P^n(s, t)] \right. \\ \left. - [1 - C(t)]^n [1 - C(s)]^n \right\} d\Lambda(t)d\Lambda(s),$$

where for $0 \leq y \leq 1$

$$I_n(y) = \sum_{k=1}^n \frac{1}{k} \binom{n}{k} y^k (1-y)^{n-k},$$

$$\begin{aligned} P(s, t) &= P(\text{neither } s \text{ nor } t \text{ is in } [Y, X] \mid Y \leq X) \\ &= 1 - \alpha^{-1} [G(s)[1 - F(s)] + G(t)[F(s) - F(t)]]. \end{aligned}$$

PROOF. The proof is given in Appendix A. \square

Remark. The I_n function is also used in Watson and Leadbetter ((1964), formula (2.3)). Note that $nyI_n(y) \leq 2$ and $nI_n(y)$ converges uniformly to y^{-1} on any interval $[a, b]$ with $a > 0$ and $b < 1$.

Asymptotic behavior of the bias term and the variance is given in the following theorem. The proof is in Appendix B.

THEOREM 3.2. (a) For $p > r$ and under (A1),

$$(3.3) \quad \text{bias}(\hat{\lambda}^{(r)}(z)) = b^{p-r} \lambda^{(p)}(z) B_{r,p} + o(b^{p-r}),$$

where

$$(3.4) \quad B_{r,p} = \frac{(-1)^p}{p!} \int K_r(y) y^p dy.$$

(b) If z is a continuity point of G , then

$$(3.5) \quad \text{Var}(\hat{\lambda}^{(r)}(z)) = \frac{1}{nb^{2r+1}} \left\{ \frac{\lambda(z)}{C(z)} V_{r,p} + o(1) \right\},$$

where

$$(3.6) \quad V_{r,p} = \int_{-1}^1 K_r^2(y) dy.$$

COROLLARY 3.1. Under the conditions of Theorem 3.2,

(a) If (B1) holds, then $\hat{\lambda}^{(r)}(z)$ is asymptotically unbiased for $\lambda^{(r)}(z)$.

(b) If (B1) and (B2) hold, then $\hat{\lambda}^{(r)}(z)$ is a mean square consistent and hence consistent estimator of $\lambda^{(r)}(z)$.

(c)

$$(3.7) \quad \begin{aligned} \text{MSE}(\hat{\lambda}^{(r)}(z)) &= \frac{1}{nb^{2r+1}} \frac{\lambda(z)}{C(z)} V_{r,p} + (b^{p-r} \lambda^{(p)}(z) B_{r,p})^2 \\ &\quad + o\left(\frac{1}{nb^{2r+1}} + b^{2(p-r)}\right). \end{aligned}$$

The optimum bandwidth which minimizes the leading term in (3.7) is given by

$$(3.8) \quad b^*(z) = n^{-1/(2p+1)} \left[\frac{2r+1}{2(p-r)} \frac{\lambda(z)}{C(z)} \frac{V_{r,p}}{(\lambda^{(p)}(z)B_{r,p})^2} \right]^{1/(2p+1)} \\ \equiv n^{-1/(2p+1)} \omega^*(z).$$

Note that the optimum rate, $n^{-1/(2p+1)}$ for bandwidth and the optimum rate, $n^{-2(p-r)/(2p+1)}$ for MSE, are analogous to the i.i.d. case (i.e. without truncation). The optimal bandwidth $b^*(z)$ depends on the unknown quantities $\lambda(z)$, $C(z)$ and $\lambda^{(p)}(z)$. Data dependent adaptive bandwidth choices will be addressed in Section 4.

Next, we will derive the local limiting distribution of $\hat{\lambda}^{(r)}(z)$. Notice that $\hat{\lambda}^{(r)}(z)$ is a sum of identically distributed but not independent terms since $C_n(X_i)$ depends on the entire sample. As mentioned earlier, we will utilize the Hájek projection method. This method was also used by Tanner and Wong (1983) for kernel hazard estimates based on randomly censored data.

Let W be a function of i.i.d. random variables Z_1, Z_2, \dots, Z_n . Hájek (1968) defines the projection W^* of W to the space \mathcal{S} of the sum of i.i.d. variables as follows:

$$(3.9) \quad W^* - E(W^*) = \sum_{i=1}^n [E(W | Z_i) - E(W)],$$

$$(3.10) \quad E(W^*) = E(W), \quad E(W^* - W)^2 = \text{Var}(W) - \text{Var}(W^*).$$

In the truncation setting, $Z_i = (X_i, Y_i)$, $W = \hat{\lambda}^{(r)}(z)$, and $\lambda^{*(r)}(z)$ denotes the Hájek projection W^* of $\hat{\lambda}^{(r)}(z)$. The following lemma gives the form of $\lambda^{*(r)}(z)$. The derivation is in Appendix C.

LEMMA 3.1. (Hájek projection) (a)

$$(3.11) \quad \lambda^{*(r)}(z) - E(\lambda^{*(r)}(z)) \\ = n^{-1} \sum_{i=1}^n \{ K_{r,b}(z - X_i) [C(X_i)]^{-1} [1 - [1 - C(X_i)]^n] \\ - \int I(Y_i \leq s \leq X_i) K_{r,b}(z - s) \\ \cdot [C(s)]^{-1} [1 - (1 - C(s))^n] d\Lambda(s) \} \\ + \sum_{i=1}^n \int [I(Y_i \leq s \leq X_i) - C(s)] \\ \cdot K_{r,b}(z - s) [1 - C(s)]^{n-1} d\Lambda(s) \\ \equiv n^{-1} \sum_{i=1}^n \xi_i(z) + \sum_{i=1}^n \eta_i(z),$$

where $E(\xi_i(z)) = E(\eta_i(z)) = 0$.

(b) If z is a continuity point of G , then for $V_{r,p}$ defined in (3.6),

$$(3.12) \quad \text{Var}(\lambda^{*(r)}(z)) = \frac{1}{nb^{2r+1}} \left\{ \frac{\lambda(z)}{C(z)} V_{r,p} + o(1) \right\}.$$

THEOREM 3.3. (Asymptotic normality) *Assume G is continuous at z and (B1)–(B2) are satisfied. We then have*

- (a) $(nb^{2r+1})^{1/2}([\lambda(z)/C(z)]V_{r,p})^{-1/2}[\hat{\lambda}^{(r)}(z) - E(\hat{\lambda}^{(r)}(z))] \xrightarrow{\mathcal{L}} N(0, 1)$,
- (b) If $d = \lim_{n \rightarrow \infty} nb^{2p+1} < \infty$, then

$$(nb^{2r+1})^{1/2}[\hat{\lambda}^{(r)}(z) - \lambda^{(r)}(z)] \xrightarrow{\mathcal{L}} N(d^{1/2}\lambda^{(p)}(z)B_{r,p}, [\lambda(z)/C(z)]V_{r,p}).$$

PROOF. (a) It follows from (3.5), (3.10) and (3.12) that

$$\text{Var}(\hat{\lambda}^{(r)}(z))/\text{Var}(\lambda^{*(r)}(z)) \rightarrow 1 \quad \text{and} \quad \frac{E(\hat{\lambda}^{(r)}(z) - \lambda^{*(r)}(z))^2}{\text{Var}(\hat{\lambda}^{(r)}(z))} \rightarrow 0.$$

Therefore, $[\text{Var}(\hat{\lambda}^{(r)}(z))]^{-1/2}[\hat{\lambda}^{(r)}(z) - E(\hat{\lambda}^{(r)}(z))]$ has the same asymptotic distribution as $Z_n(z) = [\text{Var}(\lambda^{*(r)}(z))]^{-1/2}[\lambda^{*(r)}(z) - E(\lambda^{*(r)}(z))]$ and it suffices to show that $Z_n \rightarrow N(0, 1)$. This is accomplished by verifying Lindeberg’s condition for a triangular array.

(b) Follows immediately from (3.3) and (a). \square

4. Adaptive bandwidth choice

Consider the estimator (2.10) with local bandwidth $b(z) = n^{-1/(2p+1)}\omega(z) \equiv n^{-1/(2p+1)}\omega$, which attains the optimal rate of convergence by (3.8) and denote it as

$$(4.1) \quad \hat{\lambda}^{(r)}(z, \omega) = \frac{1}{[n^{-1/(2p+1)}\omega]^{r+1}} \sum_{i=1}^n K_r \left(\frac{z - X_i}{n^{-1/(2p+1)}\omega} \right) \frac{1}{nC_n(X_i)}.$$

Thus $\lambda^{(r)}(z, \omega^*(z))$ is optimal in terms of minimizing the asymptotic MSE. In this section we show that locally adaptive bandwidth choices are indeed feasible. More precisely, it is shown that the estimator $\hat{\lambda}^{(r)}(z, \hat{\omega}^*(z))$, where $\hat{\omega}^*(z)$ is a consistent estimator of $\omega^*(z)$, has the same asymptotic distribution as the hypothetical optimal estimator $\hat{\lambda}^{(r)}(z, \omega^*(z))$. To obtain this result, it will be convenient to deal with a suitably normalized form of (4.1), given as

$$(4.2) \quad U_n(z, \omega) = n^{(p-r)/(2p+1)}[\hat{\lambda}^{(r)}(z, \omega) - \lambda^{(r)}(z)].$$

For fixed z choose ω_a, ω_b such that $0 < \omega_a < \omega^*(z) < \omega_b < \infty$.

Let $Lip_\alpha(A)$ denote the class of real functions on the set A which satisfy Lipschitz continuity of order $\alpha > 0$. The next lemma provides the key to the main result, Theorem 4.1, of this paper. The proof of Lemma 4.1 is in Appendix D.

LEMMA 4.1. Assume (B1), (B2), and G is continuous at z . If $K_r \in Lip_\alpha(-\infty, \infty)$ where $\alpha > 0.5$ and $p > r$, then for $0 < \omega_a \leq \omega \leq \omega_b < \infty$, and $\omega_b - \omega_a \leq 1$ the process $U_n(z, \omega)$ given by (4.2) converges weakly in $C[\omega_a, \omega_b]$ to a Gaussian process $U(z, \omega)$ with

$$(4.3) \quad E(U(z, \omega)) = [\omega(z)]^{p-r} \lambda^{(p)}(z) B_{r,p},$$

and

$$(4.4) \quad \begin{aligned} \text{Cov}(U(z, \omega_1), U(z, \omega_2)) \\ = (\omega_1, \omega_2)^{-(r+1)} [\lambda(z)/C(z)] \int K_r \left(\frac{u}{\omega_1} \right) K_r \left(\frac{u}{\omega_2} \right) du, \end{aligned}$$

where $B_{r,p}$ is given by (3.4).

THEOREM 4.1. (Locally adaptive bandwidth choice) Under the conditions of Lemma 4.1 and for any estimator $\hat{\omega}(z)$ satisfying $\hat{\omega}(z) \xrightarrow{P} \omega^*(z)$, both $U_n(z, \hat{\omega}(z))$ and $U_n(z, \omega^*(z))$ converge weakly to a normal distribution $N([\omega^*(z)]^{p-r} \lambda^{(p)}(z) \cdot B_{r,p}; \omega^*(z)^{-(2r+1)} [\lambda(z)/C(z)] V_{r,p})$, where $B_{r,p}$ and $V_{r,p}$ are given by (3.4) and (3.6).

PROOF. Lemma 4.1 implies that $U_n(z, \hat{\omega}) - U_n(z, \omega^*) \rightarrow 0$ in probability. The result then follows from Lemma 4.1 and application of Slutsky's Theorem. \square

Remarks 1. Note that Lemma 4.1 is only a tool to show the adaptive bandwidth choice result in Theorem 4.1. In practice one doesn't need to locate ω_a and ω_b .

2. Theorem 4.1 requires construction of consistent estimators for $\omega^*(z)$ which reduces to estimating the quantity $\lambda(z)/[C(z)\lambda^{(p)}(z)]$ consistently. By Corollary 3.1(b), consistent estimators for $\lambda(z)$ and $\lambda^{(p)}(z)$, denoted by $\hat{\lambda}_0(z)$ and $\hat{\lambda}_0^{(p)}(z)$ respectively, can be obtained via selecting proper K_0 , K_p and initial bandwidths b_0 and b_p . The initial bandwidth for $\hat{\lambda}_0^{(p)}(z)$ should be larger than the initial bandwidth for $\hat{\lambda}_0(z)$ ($nb_0 \rightarrow \infty$ but $nb_p^{2p+1} \rightarrow \infty$). As for estimating $C(z)$, the $C_n(z)$ given by (2.9) is not appropriate for the present purpose since it may assume zero value. Let \tilde{C}_n be any modified version of C_n which is nonzero and consistent for C , e.g., let $\tilde{C}_n(z) = 1/(n+1)$, whenever $C_n(z) = 0$. Then, a candidate for adaptive bandwidth choice can be given as:

$$(4.5) \quad \hat{b}^*(z) = n^{-1/(2p+1)} \left[\frac{2r+1}{2(p-r)} \frac{\hat{\lambda}_0(z)}{\tilde{C}_n(z)} \frac{V_{r,p}}{[\hat{\lambda}_0^{(p)}(z) B_{r,p}]^2} \right]^{1/(2p+1)}$$

3. Another choice of adaptive bandwidth can be obtained using the fact that $\lambda = dF^*/C$ which follows from (2.6), and that $dF^* = f^*$ can be estimated using the ordinary kernel estimate

$$f_n^*(z) = \int K_{0,b}(z-x) dF_n^*(x) = \frac{1}{nb_n} \sum_{i=1}^n K_0 \left(\frac{z-X_i}{b_n} \right).$$

An alternative candidate for adaptive bandwidth choice is then:

$$(4.6) \quad \tilde{b}^*(z) = n^{-1/(2p+1)} \left[\frac{2r+1}{2(p-r)} \frac{f_n^*(z)V_{r,p}}{[\tilde{C}_n(z)\hat{\lambda}_0^{(p)}(z)B_{r,p}]^2} \right]^{1/(2p+1)}.$$

Acknowledgement

We would like to thank Abba Krieger for a useful suggestion in the computation of $\text{Var}(\hat{\lambda}^{(r)}(z))$ in Theorem 3.1.

Appendix A

PROOF OF THEOREM 3.1. The mean of $\hat{\lambda}^{(r)}(x)$ follows directly from Lemma 2 of Woodroffe (1985). To find the variance, consider

$$(a.1) \quad \begin{aligned} E(\hat{\lambda}^{(r)}(z))^2 &= E \left(\sum_{i=1}^n K_{r,b}^2(z - X_i) [nC_n(X_i)]^{-2} \right) \\ &\quad + 2E \left(\sum_{i < j} K_{r,b}(z - X_i) K_{r,b}(z - X_j) \right. \\ &\quad \left. \cdot [n^2 C_n(X_i) C_n(X_j)]^{-1} \right) \\ &= \text{I} + \text{II}. \end{aligned}$$

Now, observe that given X_i , $nC_n(X_i) - 1 \sim \text{Binomial}(n - 1, C(X_i))$, and $E([n^2 C_n^2(X_i)]^{-1} | X_i) = I_n(C(X_i))/[nC(X_i)]$. Hence by (2.6)

$$(a.2) \quad \text{I} = E\{K_{r,b}^2(z - X_1) I_n(C(X_1))/C(X_1)\},$$

which is the first term in (3.2). To evaluate II, first consider the following conditional expectation:

$$\text{II} = 2E \left\{ \sum_{i < j} K_{r,b}(z - X_i) K_{r,b}(z - X_j) \cdot E([n^2 C_n(X_i) C_n(X_j)]^{-1} | X_i, X_j, Y_i, Y_j) \right\}.$$

For $X_i < X_j$, $nC_n(X_j) = 1 + M_2 + M_3$, and

$$nC_n(X_i) = \begin{cases} 1 + M_1 + M_3, & \text{if } X_i < Y_j \\ 2 + M_1 + M_3, & \text{if } X_i \geq Y_j, \end{cases}$$

where (M_1, M_2, M_3, M_4) have a multinomial distribution with parameters $n - 2$ and $P_k = P_k(X_j, X_i)$, $k = 1, 2, 3, 4$. The cell probabilities P_k 's are defined as follows:

$$\begin{aligned} P_1(s, t) &= P(Y \leq t, t < X \leq s \mid Y \leq X) = \alpha^{-1}G(t)[F(s) - F(t)], \\ P_2(s, t) &= P(t < Y \leq s, X > s \mid Y \leq X) = \alpha^{-1}[G(s) - G(t)][1 - F(s)], \\ P_3(s, t) &= P(Y \leq t, X > s \mid Y \leq X) = \alpha^{-1}G(t)[1 - F(s)], \\ P_4(s, t) &= P(\text{neither } t \text{ nor } s \text{ is in } [Y, X] \mid Y \leq X) \\ &= 1 - P_1(s, t) - P_2(s, t) - P_3(s, t). \end{aligned}$$

Hence,

$$\begin{aligned} \text{(a.3)} \quad I(X_i < X_j)E([n^2 C_n(X_i)C_n(X_j)]^{-1} \mid X_i, X_j, Y_i, Y_j) \\ = I(X_i < Y_j)E[(1 + M_1 + M_3)(1 + M_2 + M_3)]^{-1} \\ + I(Y_j \leq X_i < X_j)E[(2 + M_1 + M_3)(1 + M_2 + M_3)]^{-1}. \end{aligned}$$

Similarly, one can replace $I(X_i < X_j)$, $I(X_i < Y_j)$ and $I(Y_j \leq X_i < X_j)$ in (a.3) by $I(X_i \geq X_j)$, $I(X_j < Y_i)$ and $I(Y_i \leq X_j \leq X_i)$ respectively. Now using the facts that:

$$(1) \quad (ab)^{-1} = \int_0^1 \int_0^1 x^{a-1}y^{b-1}dxdy, \text{ for } a, b > 1.$$

$$(2) \quad (a + b + c + d)^n = \sum_{k_1+k_2+k_3+k_4=n} \binom{n}{k_1 k_2 k_3 k_4} a^{k_1} b^{k_2} c^{k_3} d^{k_4}, \text{ for integer } n \geq 1,$$

it can be shown that (details are available in Appendix A of Uzunogullari and Wang (1990)),

$$\begin{aligned} \text{(a.4)} \quad \Pi &= 2E \left\{ K_{r,b}(z - X_i)K_{r,b}(z - X_j) \right. \\ &\quad \cdot \left[I(X_i < X_j) \left(\frac{1}{C(X_i)C(X_j)} - \frac{[1 - C(X_j)]^n}{P_1 C(X_j)} \right) \right. \\ &\quad \left. \left. - I(X_i < Y_j < X_j) \left(\frac{[1 - C(X_i)]^n}{P_2 C(X_i)} - \frac{P_4^n}{P_1 P_2} \right) \right] \right\}. \\ &= 2 \int_{t < s} K_{r,b}(z - s)K_{r,b}(z - t) \\ &\quad \cdot \left\{ 1 - \frac{1 - F(t)}{F(t) - F(s)} [[1 - C(s)]^n - P_4^n(s, t)] - [1 - C(t)]^n \right\} \\ &\quad \cdot d\Lambda(t)d\Lambda(s). \end{aligned}$$

Theorem 3.1 now follows from (a.1), (a.2), (a.4), and noting that $P_4(s, t) = P(s, t)$, and

$$\begin{aligned} E^2(\hat{\lambda}^{(r)}(z)) \\ = 2 \int_{t < s} K_{r,b}(z - t)K_{r,b}(z - s) \\ \cdot [1 - [1 - C(s)]^n - [1 - C(t)]^n + [1 - C(t)]^n [1 - C(s)]^n] d\Lambda(t)d\Lambda(s). \quad \square \end{aligned}$$

Appendix B

PROOF OF THEOREM 3.2. (a) Using integration by parts and the moment conditions in (2.11), it follows that if one defined $K_{r-1}(x) = \int_{-1}^x K_r(y)dy$, for $r \geq 1$, then

$$K_{r-1} \in M_{r-1,p-1}, \quad K_{r-j} \in M_{r-j,p-j} \quad \text{and} \quad K \equiv K_0 \in M_{0,p-r}.$$

Hence

$$\int K_{r,b}(z-x)d\Lambda(x) = \int_{-1}^1 K(y)\lambda^{(r)}(z-yb)dy.$$

By (3.1) this leads to the following bias expansion:

$$\begin{aligned} \text{(b.1)} \quad \text{bias}(\hat{\lambda}^{(r)}(z)) &= \left[\int_{-1}^1 K(y)\lambda^{(r)}(z-yb)dy - \lambda^{(r)}(z) \right] \\ &\quad - \int K_{r,b}(z-x)[1-C(x)]^n d\Lambda(x) \\ &= \text{I} + \text{II}. \end{aligned}$$

Now utilizing the assumptions (A1) and Taylor expansion, it follows that

$$\text{(b.2)} \quad \text{I} = b^{p-r}\lambda^{(p)}(z)\frac{(-1)^p}{p!} \int K_r(y)y^p dy + o(b^{p-r}).$$

On the other hand, for n large enough and $\alpha_G < x < b_F$, there exists $\delta > 1$ such that $1 - C(x) \leq \delta$. Therefore,

$$\text{(b.3)} \quad |\text{II}| \leq \delta^n b^{-r} \int_{-1}^1 |K_r(y)|\lambda(z-yb)dy = O(\delta^n b^{-r}) = o(b^{p-r})$$

since $K_r \in L^2[-1, 1]$ and λ is continuous at z . Part (a) now follows from (b.1) to (b.3).

(b) Consider the first term in (3.2). Using the assumptions on the kernel and bandwidth, the continuity of λ/C at z and the uniform convergence of $nI_n(C(x))$ to $1/C(x)$ on $[z - b_n, z + b_n]$, one can show that

$$nb^{2r+1} \int K_{r,b}^2(z-x)I_n(C(x))d\Lambda(x) \rightarrow \lambda(z)V_{r,p}/C(z).$$

It remains to show that the second term in (3.2) is of the order $o((nb^{2r+1})^{-1})$. To see this observe that:

$$\begin{aligned} \text{(b.4)} \quad & \left| [1 - C(s)]^n - \frac{1 - F(t)}{F(t) - F(s)} [[1 - C(s)]^n - P^n(s, t)] \right. \\ & \quad \left. - [1 - C(s)]^n [1 - C(t)]^n \right| \\ & \leq [1 - C(s)]^n + \frac{1 - F(t)}{F(t) - F(s)} [[1 - C(s)]^n - P^n(s, t)] \\ & \leq (n + 1)[1 - C(s)]^{n-1}, \end{aligned}$$

where the last inequality follows from the fact that $1 - C(s) - P(s, t) = \alpha^{-1}G(t) \cdot [F(s) - F(t)]$, $P(s, t) \leq \min\{1 - C(s), 1 - C(t)\}$ and the following polynomial expansion:

$$\begin{aligned} & \frac{1 - F(t)}{F(t) - F(s)} [[1 - C(s)]^n - P^n(s, t)] \\ &= C(t) [[1 - C(s)]^{n-1} + [1 - C(s)]^{n-2}P(s, t) + \dots + P^{n-1}(s, t)] \\ &\leq n[1 - C(s)]^{n-1}. \end{aligned}$$

For large n , (b.4) implies that for some $\delta > 0$,

$$\begin{aligned} & (nb^{2r+1}) \cdot [\text{second term in (3.2)}] \\ &\leq (nb^{2r+1}) \int_{t < s} K_{r,b}(z - s)K_{r,b}(z - t)(n + 1)[1 - C(s)]^{n-1} d\Lambda(t)d\Lambda(s) \\ &\leq n(n + 1)b^{2r+1}\delta^{n-1} \left[\int K_{r,b}(z - t)d\Lambda(t) \right]^2 \rightarrow 0. \quad \square \end{aligned}$$

Appendix C

PROOF OF LEMMA 3.1. (a) Application of (3.9) to $\hat{\lambda}^{(r)}(z)$ yields

$$\begin{aligned} \text{(c.1)} \quad & \lambda^{*(r)}(z) - E(\lambda^{*(r)}(z)) \\ &= \sum_{j=1}^n \{E(W_j | X_j, Y_j) + (n - 1)E(W_i | X_j, Y_j) - E(\hat{\lambda}^{(r)}(z))\}, \end{aligned}$$

where $W_k = K_{r,b}(z - X_k)[nC_n(X_k)]^{-1}$, and

$$\text{(c.2)} \quad E(W_j | X_j, Y_j) = [nC(X_j)]^{-1}K_{r,b}(z - X_j)[1 - (1 - C(X_j))^n],$$

by Lemma 2 of Woodroffe (1985). Also,

$$\text{(c.3)} \quad E(W_i | X_j, Y_j) = E\{K_{r,b}(z - X_i)E[(nC_n(X_i))^{-1} | X_i, Y_i, X_j, Y_j] | X_j, Y_j\}.$$

Let $p = C(X_i)$ and observe that, given X_i, Y_i, X_j, Y_j , and n , the conditional distribution of $nC_n(X_i)$ is

$$nC_n(X_i) \sim \begin{cases} 2 + \text{Binomial}(n - 2, p), & \text{if } Y_j \leq X_i \leq X_j \\ 1 + \text{Binomial}(n - 2, p), & \text{otherwise.} \end{cases}$$

Hence for $Y_j \leq X_i \leq X_j$,

$$\begin{aligned} \text{(c.4)} \quad E([nC_n(X_i)]^{-1} | X_i, Y_i, X_j, Y_j) &= \sum_{k=2}^n k^{-1} \binom{n-2}{k-2} p^{k-2}(1-p)^{n-k} \\ &= [n(n-1)p^2]^{-1}[np - 1 + (1-p)^n]. \end{aligned}$$

Similarly, for $X_i < Y_j$ or $X_j \leq X_i$,

$$(c.5) \quad E([nC_n(X_i)]^{-1} \mid X_i, Y_i, X_j, Y_j) = [(n-1)p]^{-1}[1 - (1-p)^{n-1}].$$

Combining (c.4) and (c.5) we have

$$(c.6) \quad \begin{aligned} E([nC_n(X_i)]^{-1} \mid X_i, Y_i, X_j, Y_j) &= [(n-1)p]^{-1}[1 - (1-p)^{n-1}] \\ &\quad + [n(n-1)p^2]^{-1}[(1-p)^n + np(1-p)^{n-1} - 1]I(Y_j \leq X_i \leq X_j). \end{aligned}$$

Replacing p back by $C(X_i)$, and plugging (c.6) into (c.3), we obtain

$$(3.11) \quad \begin{aligned} E(W_i \mid X_j, Y_j) &= (n-1)^{-1} \left\{ \int K_b(z-s)[1 - [1 - C(s)]^{n-1}]d\Lambda(s) \right. \\ &\quad - \int I(Y_j \leq s \leq X_j)K_b(z-s)[nC(s)]^{-1} \\ &\quad \left. \cdot [1 - [1 - C(s)] - nC(s)[1 - C(s)]^{n-1}]d\Lambda(s) \right\}. \end{aligned}$$

(3.11) now follows from (c.1), (c.2), (c.6) and (3.1). The fact that ξ_i and η_i have mean zero follows from (2.6), (2.7), and the fact that the first and second term in ξ_i have the same expectation.

(b) For this part we utilize the following result whose proof is given in Appendix C of Uzunogullari and Wang (1990):

$$(c.7) \quad \text{Var}(\xi_i(z)) = \int K_{r,b}^2(z-x)[1 - [1 - C(x)]^n]^2[C(s)]^{-1}d\Lambda(s).$$

Using the continuity of λ/C at z , the fact that $K \in L^2[-1, 1]$ and the dominated convergence theorem, (c.7) can be written as

$$(c.8) \quad \begin{aligned} \text{Var}(\xi_i(z)) &= \frac{1}{b^{2r+1}} \int K_r^2(y) \frac{\lambda(z-by)}{C(z-by)} [1 - [1 - C(z-by)]^n]^2 dy \\ &= \frac{1}{b^{2r+1}} \left[\frac{\lambda(z)}{C(z)} V_{r,p} + o(1) \right]. \end{aligned}$$

Next, consider $\eta_i(z)$. For some $\delta > 0$,

$$(c.9) \quad \begin{aligned} |\eta_i(z)| &= \left| \int K_{r,b}(z-s)[I(Y_i \leq s \leq X_i) - C(s)][1 - C(s)]^{n-1}d\Lambda(s) \right| \\ &\leq \left| \int K_{r,b}(z-s)[1 - C(s)]^{n-1}d\Lambda(s) \right| \\ &= b^{-r}\delta^n \left| \int K_r(y)\lambda(z-by)dy \right|. \end{aligned}$$

Formula (3.12) now follows from (c.8), (c.9) and application of the Cauchy-Schwarz inequality for the covariance term. \square

Appendix D

PROOF OF LEMMA 4.1. The course of the proof is to show that

- (a) the finite dimensional distributions of $U_n(z, \omega)$ converge to a multivariate normal distribution, with the covariance structure given by (4.4), and
- (b) the process $U_n(z, \omega)$ is tight.

Part (a). By the hypothesis of the lemma, Theorem 3.2(a) and Theorem 3.3(a), it follows that

$$(d.1) \quad U_n(z, \omega) \rightarrow N(\omega(z)^{p-r} \lambda^{(p)}(z) B_{r,p}, \omega^{-(2r+1)} [\lambda(z)/C(z)] V_{r,p}).$$

The Cramér-Wold device then implies the weak convergence of the finite dimensional distributions of $U_n(z, \omega)$ to a multivariate normal distribution with mean given by (4.3). It remains to verify the covariance structure of the limiting multivariate normal distribution.

Let $a_n = n^{-1/(2p+1)}$. Then following the proof of Theorem 3.1(b) and Theorem 3.2(b) for the variance computations, we arrive at

$$(d.2) \quad \begin{aligned} & \text{Cov}(U_n(z, \omega_1), U_n(z, \omega_2)) \\ &= n^{2(p+1)/(2p+1)} (\omega_1 \omega_2)^{-(r+1)} \\ & \cdot \left\{ \int K_r \left(\frac{z-x}{a_n \omega_1} \right) K_r \left(\frac{z-x}{a_n \omega_2} \right) I_n[C(x)] d\Lambda(x) \right. \\ & \quad + \int_{t < s} \left[K_r \left(\frac{z-t}{a_n \omega_1} \right) K_r \left(\frac{z-s}{a_n \omega_2} \right) + K_r \left(\frac{z-t}{a_n \omega_2} \right) K_r \left(\frac{z-s}{a_n \omega_1} \right) \right] \\ & \quad \cdot \left[[1 - C(s)]^n [1 - (1 - C(t))^n] \right. \\ & \quad \left. \left. - \frac{1 - F(t)}{F(s) - F(t)} ([1 - C(s)]^n - p^n(s, t)) \right] \right\} d\Lambda(t) d\Lambda(s) \\ &= (\omega_1 \omega_2)^{-(r+1)} \frac{\lambda(z)}{C(z)} \int K_r \left(\frac{t}{\omega_1} \right) K_r \left(\frac{t}{\omega_2} \right) dt + o(1). \end{aligned}$$

Part (a) is now completed by (d.2).

Part (b). To see that the process $U_n(z, \omega)$ is tight, consider

$$(d.3) \quad \begin{aligned} & E(U_n(z, \omega_1) - U_n(z, \omega_2))^2 \\ &= E[n^{(p-r)/(2p+1)} (\hat{\lambda}^{(r)}(z, \omega_1) - \hat{\lambda}^{(r)}(z, \omega_2))]^2 \\ &\equiv n^{2(p+1)/(2p+1)} E \left\{ \sum_{i=1}^n H(z, X_i, \omega_1, \omega_2) / [nC_n(X_i)] \right\}^2, \end{aligned}$$

where $H(z, X_i, \omega_1, \omega_2) = (1/\omega_1^{r+1}) K_r((z-X_i)/a_n \omega_1) - (1/\omega_2^{r+1}) K_r((z-X_i)/a_n \omega_1)$. Observe that the last expression in curled brackets is similar in form to $\hat{\lambda}^{(r)}(z)$ in (2.10) and therefore one can obtain this expectation from the proof of $E[\hat{\lambda}^{(r)}(z)]^2$

in Appendix A since H is a fixed function of X_i . Hence we have

$$\begin{aligned} E[U_n(z, \omega_1) - U_n(z, \omega_2)]^2 &= n^{2(p+1)/(2p+1)} \left[\int H^2(z, s, \omega_1, \omega_2) I_n[C(x)] d\Lambda(s) \right. \\ &\quad + 2 \int_{t < s} H(z, s, \omega_1, \omega_2) H(z, t, \omega_1, \omega_2) \\ &\quad \cdot \left\{ 1 - \frac{1 - F(t)}{F(s) - F(t)} [[1 - C(s)]^n - P_4^n] \right. \\ &\quad \left. \left. - [1 - C(s)]^n [1 - C(t)]^n \right\} d\Lambda(t) d\Lambda(s) \right] \\ &= \text{I} + \text{II}. \end{aligned}$$

Now consider term I, with $a_n = n^{-1/(2p+1)}$:

$$\begin{aligned} \text{(d.4)} \quad \text{I} &= \int n I_n[C(z - a_n t)] (\omega_1^{-(r+1)} K_r(t/\omega_1) - \omega_2^{-(r+1)} K_r(t/\omega_2))^2 \\ &\quad \cdot \lambda(z - a_n t) dt \\ &= \int n I_n[C(z - a_n t)] \{ \omega_1^{-(r+1)} [K_r(t/\omega_1) - K_r(t/\omega_2)] \\ &\quad + [\omega_1^{-(r+1)} - \omega_2^{-(r+1)}] K_r(t/\omega_2) \}^2 \lambda(z - a_n t) dt \\ &\leq 2 \int n I_n[C(z - a_n t)] \{ \omega_1^{-2(r+1)} [K_r(t/\omega_1) - K_r(t/\omega_2)]^2 \\ &\quad + [\omega_1^{-(r+1)} - \omega_2^{-(r+1)}]^2 K_r^2(t/\omega_2) \} \lambda(z - a_n t) dt \\ &\leq \text{constant} \int n I_n[C(z - a_n t)] \cdot |\omega_1 - \omega_2|^{\min(2\alpha, 2)} \lambda(z - a_n t) dt \\ &\leq \text{constant} |\omega_1 - \omega_2|^{\min(2\alpha, 2)}, \end{aligned}$$

where the second last step follows from the Lipschitz condition on K_r and that $|\omega_1 - \omega_2| \leq 1$; the last step follows from the continuity of λ/C at z and the fact that $nyI_n(y) \leq 2$ for $0 \leq y \leq 1$.

Term II can similarly be bounded by $L(\omega_1 - \omega_2)^{\min(2\alpha, 2)}$ for some $L > 0$. Therefore

$$E[U_n(z, \omega_1) - U_n(z, \omega_2)]^2 \leq \text{constant} (\omega_1 - \omega_2)^{\min(2\alpha, 2)},$$

for all $(\omega_1, \omega_2) \in [\omega_a; \omega_b]$. This implies the tightness of $U_n(z, \omega)$ by Theorem 12.3 of Billingsley (1968). \square

REFERENCES

Allredge, J. R. and Gates, C. E. (1985). Line transect estimators for left truncated distributions, *Biometrics*, **41**, 273-280.
 Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

- Chao, M. T. and Lo, S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data, *Ann. Statist.*, **16**, 661–668.
- Csörgö, S. and Horváth, L. (1980). Random censorship from the left, *Studia Sci. Math. Hungar.*, **15**, 397–401.
- Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring, *J. Multivariate Anal.*, **25**, 299–310.
- Gu, M. G. and Lai, T. L. (1990). Functional law of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, *Ann. Probab.*, **18**, 160–189.
- Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternative, *Ann. Math. Statist.*, **39**, 325–346.
- Hyde, J. (1977). Testing survival under right censoring and left truncation, *Biometrika*, **64**, 225–230.
- Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on transfusion-relation AIDS, *J. Amer. Statist. Assoc.*, **84**, 360–372.
- Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes, *Ann. Statist.*, **18**, 582–602.
- Lagakos, S. W., Barraj, L. M. and DeGruttola, V. (1988). Nonparametric analysis of truncated survival data with application to AIDS, *Biometrika*, **75**, 515–523.
- Lui, K. J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. W. and Bergman, D. J. (1986). A model based approach for estimating the mean incubation period of transfusion associated acquired immunodeficiency syndrome, *Proc. Nat. Acad. Sci. U.S.A.*, **88**, 3051–3055.
- Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices of the Royal Astronomical Society*, **155**, 95–118.
- Müller, H. G. and Wang, J. L. (1990). Locally adaptive hazard smoothing, *Probab. Theory Related Fields*, **85**, 523–538.
- Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065–1076.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions, *Ann. Statist.*, **11**, 453–466.
- Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method, *Ann. Statist.*, **11**, 989–993.
- Uzunogullari, U. and Wang, J. L. (1990). Nonparametric estimation of hazard functions and their derivatives under truncation model, Tech. Report, #156, University of California, Davis.
- Uzunogullari, U. and Wang, J. L. (1992). A comparison of hazard rate estimators for left truncated and right censored data, *Biometrika*, **79**, 297–310.
- Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986). Asymptotic properties of the product limit estimate under random truncation, *Ann. Statist.*, **14**, 1597–1605.
- Watson, G. S. and Leadbetter, M. R. (1964). Hazard Analysis II, *Sankhyā Ser. A*, **26**, 101–116.
- Woodroffe, M. (1985). Estimating a distribution function with truncated data, *Ann. Statist.*, **13**, 163–177.
- Yandell, S. B. (1983). Nonparametric inference for rates with censored survival data, *Ann. Statist.*, **11**, 1119–1135.