The Annals of Statistics
1989, Vol. 17, No. 3, 1157–1167

UNIFORM CONSISTENCY OF THE KERNEL CONDITIONAL KAPLAN–MEIER ESTIMATE[1]

BY DOROTA M. DABROWSKA

University of California, Los Angeles

We consider a class of nonparametric regression estimates introduced by Beran to estimate conditional survival functions in the presence of right censoring. An exponential probability bound for the tails of distributions of kernel estimates of conditional survival functions is derived. This inequality is next used to prove weak and strong uniform consistency results. The developments rest on sharp exponential bounds for the oscillation modulus of multivariate empirical processes obtained by Stute.

1. Introduction. Let T be a nonnegative random variable (rv) representing the survival time of an individual taking part in a clinical trial or other experimental study and let $\mathbf{Z} = (Z_1, \dots, Z_q)$ be a vector of covariates such as age, blood pressure and cholesterol level. The survival time T is subject to right censoring so that the observable rv's are given by $Y = \min(T, X)$, $D = I(T \le X)$ and $\mathbf{Z}$. Here X is a nonnegative rv representing times to withdrawal from the study. Denote by $F(t|\mathbf{z}) = P(T > t|\mathbf{Z} = \mathbf{z})$, $H_1(t|\mathbf{z}) = P(Y > t, D = 1|\mathbf{Z} = \mathbf{z})$ and $H_2(t|\mathbf{z}) = P(Y > t|\mathbf{Z} = \mathbf{z})$ the respective conditional survival functions and let

$$\Lambda(t|\mathbf{z}) = -\int_0^t F(s-|\mathbf{z})^{-1}\,dF(s|\mathbf{z}) \tag{1.1}$$

be the conditional cumulative hazard function associated with $F(t|\mathbf{z})$. In terms of the cumulative hazard function we have

$$F(t|\mathbf{z}) = \prod_{s \le t} \{1 - d\Lambda(s|\mathbf{z})\}. \tag{1.2}$$

This is the well known product–integral representation of distribution functions; see for instance Gill and Johansen (1987), Gill (1980) and Beran (1981). To avoid trivial cases it is assumed throughout that $p = P(D = 1)$ satisfies $0 < p < 1$. Furthermore, it is assumed that T and X are conditionally independent given $\mathbf{Z}$, which is a sufficient condition to ensure identifiability of $\Lambda(t|\mathbf{z})$ and $F(t|\mathbf{z})$. Specifically, for any t such that $H_2(t|\mathbf{z}) > 0$ we have

$$\Lambda(t|\mathbf{z}) = -\int_0^t (H_2(s-|\mathbf{z}))^{-1}\,dH_1(s|\mathbf{z}).$$

Let $(Y_j, D_j, \mathbf{Z}_j)$, $j = 1, \dots, n$, be a sample of i.i.d. rv's each having the same distribution as $(Y, D, \mathbf{Z})$. To estimate the subsurvival functions $H_i(t|\mathbf{z})$ we use Nadaraya–Watson type kernel estimates of the form $\hat{H}_i(t|\mathbf{z}) = \hat{g}(\mathbf{z})^{-1}\hat{H}_i(t, \mathbf{z})$

Received February 1987; revised November 1988.

[1]Research supported by a University of California Presidential Fellowship and by National Institute of General Medical Sciences Grant SSSY 1RO1 GM35416-02.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G20.

Key words and phrases. Kernel regression, right censoring, oscillation modulus.

where

$$\hat{H}_1(t,\mathbf{z}) = (na_n^q)^{-1} \sum_{j=1}^{n} I(Y_j > t, D_j = 1) K\left(a_n^{-1}(\mathbf{z} - \mathbf{Z}_j)\right),$$

$$\hat{H}_2(t,\mathbf{z}) = (na_n^q)^{-1} \sum_{j=1}^{n} I(Y_j > t) K\left(a_n^{-1}(\mathbf{z} - \mathbf{Z}_j)\right)$$

and $\hat{g}(\mathbf{z})$ is the Parzen–Rosenblatt [Parzen (1962); Rosenblatt (1971)] density estimate

$$\hat{g}(\mathbf{z}) = (na_n^q)^{-1} \sum_{j=1}^{n} K\left(a_n^{-1}(\mathbf{z} - \mathbf{Z}_j)\right).$$

Here K is a kernel function and a_n is a band sequence. These estimates were extensively studied in the literature; Nadaraya (1964), Watson (1964), Rosenblatt (1969), Collomb (1981), Devroye (1981) and Mack and Silverman (1982) are some references. Following Beran (1981), $\Lambda(t|\mathbf{z})$ and $F(t|\mathbf{z})$ are estimated by

$$\hat{\Lambda}(t|\mathbf{z}) = -\int_0^t \frac{d\hat{H}_1(s|\mathbf{z})}{\hat{H}_2(s-|\mathbf{z})} \tag{1.3}$$

and

$$\hat{F}(t|\mathbf{z}) = \prod_{s \le t} \{1 - d\hat{\Lambda}(s|\mathbf{z})\}. \tag{1.4}$$

Both $\hat{\Lambda}(t|\mathbf{z})$ and $\hat{F}(t|\mathbf{z})$ are right continuous functions of t with jumps occurring at discontinuity points of $\hat{H}_1(t|\mathbf{z})$. Note that in the homogeneous case, (1.3) and (1.4) are simply the Aalen–Nelson [Aalen (1978), Nelson (1972)] and Kaplan–Meier (1958) estimates.

We develop an analogue of the Dvoretzky–Kiefer–Wolfowitz (1956) inequality providing an exponential bound for the tails of the distribution of the estimate $\hat{F}(t|\mathbf{z})$. This bound is next used to establish weak and strong uniform consistency of this estimator. The results rest on sharp exponential bounds for the oscillation modulus of univariate and multivariate empirical processes obtained by Stute (1982, 1984a).

For uncensored data, conditional empirical processes were studied among others by Stute (1986a, b) who considered pointwise consistency and weak convergence results for estimates based on kernel and nearest neighbour weights. Horváth and Yandell (1988) obtained functional laws of the iterated logarithm and rates of Gaussian approximation. For censored data, the conditional Aalen–Nelson and Kaplan–Meier estimates were discussed by Beran (1981) and Dabrowska (1987a), who showed their pointwise consistency and, respectively, weak convergence to a time transformed Brownian motion.

2. Main results. In what follows, for any rectangle $I = \prod_{j=1}^q [a_j, b_j]$ in R^q, I_δ denotes a (small) δ-neighbourhood of I of the form $I_\delta = \prod_{j=1}^q [a_j - \delta, b_j + \delta]$.

Further, we write $\|f\|_I^\tau = \sup\{|f(t,\mathbf{z})|: 0 \le t \le \tau, \mathbf{z} \in I\}$ for any real function $f(t,\mathbf{z}) = f(t, z_1,\dots,z_q)$ on $[0,\tau] \times I$, where $\tau < \infty$.

To obtain an exponential bound for the tails of the distribution of $\hat{F}(t|\mathbf{z})$ we shall use a variance–bias type decomposition

$$\hat{F}(t|\mathbf{z}) - F(t|\mathbf{z}) = \left[\hat{F}(t|\mathbf{z}) - F_n(t|\mathbf{z})\right] + \left[F_n(t|\mathbf{z}) - F(t|\mathbf{z})\right],$$

where $F_n(t|\mathbf{z})$ is defined by (1.2) with $\Lambda(t|\mathbf{z})$ replaced by

$$\Lambda_n(t|\mathbf{z}) = -\int_0^t \frac{dH_{1n}(s|\mathbf{z})}{H_{2n}(s-|\mathbf{z})}.$$

Here $H_{in}(t|\mathbf{z}) = g_n(\mathbf{z})^{-1}H_{in}(t,\mathbf{z})$, $i = 1,2$, and

$$H_{in}(t,\mathbf{z}) = a_n^{-1}\int H_i(t|\mathbf{u})K\left(a_n^{-1}(\mathbf{z}-\mathbf{u})\right)dG(\mathbf{u}),$$

$$g_n(\mathbf{z}) = a_n^{-q}\int K\left(a_n^{-1}(\mathbf{z}-\mathbf{u})\right)dG(\mathbf{u}),$$

with G being the joint cdf of $\mathbf{Z}$. The functions $H_{in}(t|\mathbf{z})$ and $F_n(t|\mathbf{z})$ are deterministic and can be thought of as smoothed versions of the original subsurvival functions $H_i(t|\mathbf{z})$ and $F(t|\mathbf{z})$. Theorem 2.1 below provides an exponential bound for distributions of the random term $\hat{F}(t|\mathbf{z}) - F_n(t|\mathbf{z})$.

Throughout we require the joint distribution of $\mathbf{Z}_i$'s to have a density g with respect to Lebesque measure. The marginal cdf's and densities are denoted by G_i and, respectively, g_j, $j = 1,\dots,q$. Further, let $C_0(u_1,\dots,u_q) = P(G_j(Z_j) \le u_j,\ j = 1,\dots,q)$. C_0 is a cdf with uniform marginals satisfying $C_0(G_1(z_1),\dots,G_q(z_q)) = G(z_1,\dots,z_q)$ and its density is given by

$$c_0(\mathbf{u}) = g\left(G_1^{-1}(u_1),\dots,G_q^{-1}(u_q)\right)\Big/\prod_{j=1}^{q} g_j\left(G_j^{-1}(u_j)\right). \tag{2.1}$$

The following assumptions will be needed.

ASSUMPTION A. (i) If $I = \prod_{j=1}^q [a_j, b_j]$ is a rectangle contained in the support of g, then

$$0 < \gamma = \inf\left\{g(\mathbf{z}), g_j(z_j): \mathbf{z} \in I_\delta\right\} < \sup\left\{g(\mathbf{z}), g_j(z_j): \mathbf{z} \in I_\delta\right\} = \Gamma < \infty$$

for some δ-neighbourhood of I. Moreover, $0 < \delta\Gamma < 1$ and $\sup\{c_0(\mathbf{u}): \mathbf{u} \in I_{\delta\Gamma}^G\} = \Gamma_0 < \infty$, where $I_{\delta\Gamma}^G$ is a $\delta\Gamma$-neighbourhood of $I^G = \prod_{j=1}[G_j(a_j), G_j(b_j)]$. For $0 < t < \tau$, $\inf\{H_2(\tau|\mathbf{z}): z \in I_\delta\} > \theta > 0$.

(ii) For $\mathbf{z} \in I_\delta$ the functions $g(\mathbf{z})$ and $H_i(t|\mathbf{z})$, $i = 1,2$, have bounded continuous first partial derivatives with respect to z_j and the derivatives of $g(\mathbf{z})$ are bounded away from zero.

(iii) The same conditions are satisfied by the second derivatives.

Condition A(i) will be used to develop exponential bounds for the tails of the distribution of $\hat{F}(t|\mathbf{z}) - F_n(t|\mathbf{z})$. This condition says that essentially we consider only the central portion of the distribution of $\mathbf{Z}$'s. The assumption that $H_2(\tau|\mathbf{z})$,

$\mathbf{z} \in I_\delta$, stays bounded away from zero is imposed to avoid problems with the tails of the conditional subsurvival functions. Conditions A(ii) and (iii) will be needed to ensure asymptotic unbiasedness of the estimator in consistency results.

As for the kernel K, we assume that K is a density with bounded support and is a function of bounded variation in the sense of Hardy and Krause [see, e.g., Hildebrandt (1963)]. Without loss of generality we assume

ASSUMPTION B. (i) K is a density vanishing outside $(-1,1)^q$ and the total variation of K is less than λ, $\lambda < \infty$.

(ii) The kernel K satisfies $\int u_j K(\mathbf{u})\, d\mathbf{u} = 0$, $j = 1, \dots, q$.

The assumption that K is a probability kernel implies that $\hat{H}_i(t|\mathbf{z})$ and $\hat{F}(t|\mathbf{z})$ are proper (sub)survival functions, i.e., nonnegative and nonincreasing in t. Condition B(ii) will be needed in the strong uniform consistency result to ensure asymptotic unbiasedness of the estimator.

THEOREM 2.1. *Suppose that the conditions* A(i) *and* B(i) *hold. Let* $0 < a_n < 1$ *and* $0 < \varepsilon < 1$ *satisfy* $0 < a_n < \delta \min(1, 1/\Gamma)$, $216 \le \varepsilon\theta^5 n a_n^q \gamma/\lambda$ *and* $e_0 \ge \varepsilon\theta^5\gamma(\Gamma_0\Gamma^q\lambda)^{-1}$ *for some finite* $e_0 > 0$. *There exist constants* $d_1, d_2 > 0$ [*not depending on* n, a_n, ε *or the distribution of* $(Y_i, D_i, \mathbf{Z}_i)$] *such that*

$$P\left(\|\hat{F} - F_n\|_I^\tau > \varepsilon\right) \le d_1 a_n^{-q} \exp\left\{-d_2\varepsilon^2 n a_n^q\right\}.$$

Note that if $\log a_n^{-q}/na_n^q \to 0$, then for $\eta > 0$ such that $d_2 - \eta > 0$ there exist constants $d_1', d_2' > 0$ such that for n sufficiently large

$$d_1 a_n^{-q} \exp\left\{-\eta\varepsilon^2\theta^{25} n a_n^q\right\} \exp\left\{-(d_2 - \eta)\varepsilon^2\theta^{25} n a_n^q\right\} \le d_1' \exp\left\{-d_2'\varepsilon^2\theta^{25} n a_n^q\right\}.$$

This shows the connection with the Dvoretzky–Kiefer–Wolfowitz (1956) inequality even better.

Theorem 2.1 can be used to derive uniform consistency results.

COROLLARY 2.1. *Suppose that the conditions* A(i) *and* B(i) *hold and let* $a_n \to 0$ *and* $na_n^q \to \infty$.

(i) *If* $a_n^{-q}\exp\{-\rho n a_n^q\} \to 0$ *for all* $\rho > 0$, *then* $\|\hat{F} - F_n\|_I^\tau \to_p 0$.

(ii) *If in addition* $\sum a_n^{-q}\exp\{-\rho n a_n^q\} < \infty$ *for all* $\rho > 0$, *then* $\|\hat{F} - F_n\|_I^\tau \to 0$ *a.s.*

(iii) *If* A(ii) *holds, then* $\|F_n - F\|_I^\tau \to 0$.

The first two parts follow directly from Theorem 2.1, whereas part (iii) follows from a one-term Taylor expansion of $H_{in}(t|\mathbf{z}) - H_i(t|\mathbf{z})$, $i = 1, 2$, and integration by parts applied to $F_n(t|\mathbf{z}) - F(t|\mathbf{z})$. Note that condition (i) is satisfied if $na_n^{q+\varepsilon} \to 0$ for some $\varepsilon > 0$. Condition (ii) holds if $\log n/na_n^q \to 0$.

COROLLARY 2.2. *Suppose that the conditions* A(i) *and* B(i) *hold and let* $a_n \to 0$ *and* $na_n^q \to \infty$.

(i) *If* $b_n = \log a_n^{-q}/na_n^q \to 0$, *then* $\|\hat{F} - F_n\|_I^\tau = O_P(b_n^{1/2})$.

(ii) *If in addition* $\sum a_n^\xi < \infty$ *for some* $\xi > 0$, *then* $\|\hat{F} - F_n\|_I^\tau = O(b_n^{1/2})$ *as* $n \to \infty$ *with probability* 1.

(iii) *If* A(iii) *and* B(ii) *hold and* $na_n^{q+4} \to 0$, *then* $b_n^{-1/2}\|F_n - F\|_I^\tau \to 0$.

This follows from Theorem 2.1 applied to $\varepsilon = \varepsilon_n = \{b_n(1 + \xi')/d_2\}^{1/2}$ for arbitrary $\xi' > 0$ in part (i) and $\xi' = \xi$ in part (ii). Part (iii) follows from a two-term Taylor expansion of $\hat{H}_i(t|\mathbf{z}) - H_i(t|\mathbf{z})$. If we let $a_n = n^{-\alpha}$ for some $\alpha > 0$, then the condition for weak and strong uniform consistency with rate $b_n^{1/2}$ is $1/(q + 4) < \alpha < 1/q$.

Corollaries 2.1 and 2.2 can be used to derive consistency results for functions that are supremum norm continuous. In particular, assume Condition A and consider the truncated mean $m(\tau; \mathbf{z}) = \int_0^\tau F(t|\mathbf{z})\,dt$. If $\hat{m}(\tau; \mathbf{z}) = \int_0^\tau \hat{F}(t|\mathbf{z})\,dt$, then, uniformly in $\mathbf{z} \in I$, $\hat{m}(\tau; \mathbf{z}) - m(\tau; \mathbf{z}) = o(1)$ a.s. if $\log n/na_n^q \to 0$ and $\hat{m}(\tau; \mathbf{z}) - m(\tau; \mathbf{z}) = O(b_n^{1/2})$ a.s. provided $na_n^{q+4} \to 0$ and $\log a_n^{-q}/na_n^q \to 0$.

For uncensored data, the results of Theorem 2.1 and Corollaries 2.1 and 2.2 remain valid when the survival time assumes values on the whole line. In particular, the probability bound of Theorem 2.1 is $d_1' a_n^{-q} \exp\{-d_2'\varepsilon^2 na_n^q\}$ for some $d_1', d_2' > 0$ provided $0 < a_n < \delta \min(1, 1/\Gamma) < 1$, $0 < \varepsilon < 1$, $4 \le \varepsilon na_n^q \gamma\lambda^{-1}$ and $e_0 \ge \varepsilon\gamma(\Gamma_0\gamma^q\lambda)^{-1}$ for some $e_0 > 0$. If T has a finite rth moment, $r > 1$, then a truncation argument similar to Mack and Silverman (1982) can be used to establish conditions for uniform consistency of the mean regression.

A drawback of kernel smoothing is that it requires absolute continuity of the distribution of the covariates. If instead we assume only that the joint distribution of **Z**'s is continuous, then we can resort to Stute's (1984b, 1986b) nearest neighbour estimates. Analogues of Theorem 2.1 and Corollaries 2.1 and 2.2 are given in Dabrowska (1987b). In practice we often have to deal with discrete covariates such as sex or type of treatment among m possible treatments. If the covariates assume say l values, then $F(t|\mathbf{z}_j)$ is an ordinary Kaplan–Meier estimate based on the n_j observations for which $\mathbf{Z}_i = \mathbf{z}_j$, $j = 1, \dots, l$. Assuming $\inf_j H_2(\tau|\mathbf{z}_j) > \theta > 0$, the Bonferroni inequality implies that given the numbers n_j,

$$\begin{aligned} P\Big(\max_j \sup_{0 \le t \le \tau} |\hat{F}(t|\mathbf{z}_j) - F(t|\mathbf{z}_j)| > \varepsilon\Big) &\le \sum_{j=1}^{l} P\Big(\sup_{0 \le t \le \tau} |\hat{F}(t|\mathbf{z}_j) - F(t|\mathbf{z}_j)| > \varepsilon/l\Big) \\ &\le \sum_{j=1}^{l} d_1 \exp\{-d_2\varepsilon^2\theta^{25} l^{-2} n_j\} \\ &\le l d_1 \exp\Big\{-d_2\varepsilon^2\theta^{25} l^{-2} \min_j n_j\Big\} \end{aligned}$$

for some constants $d_1, d_2 > 0$. Here the second inequality follows from the Dvoretzky–Kiefer–Wolfowitz (1956) inequality and (4.1) and (4.3).

3. Preliminaries.

3.1. *Local deviations of empirical processes on the unit cube.* Our proofs rest on finite sample tail estimates for the oscillation modulus of univariate and multivariate empirical processes developed by Stute (1982, 1984a).

Let $C(\mathbf{v}, \mathbf{u})$, $\mathbf{v} = (v_1, \dots, v_{q_1})$ and $\mathbf{u} = (u_1, \dots, u_{q_2})$, $q_1 \geq 0$, $q_2 \geq 1$, be a cdf on $[0,1]^{q_1+q_2}$ with uniform marginals and let C_n be the empirical cdf corresponding to a sample of size n from C. For any rectangle $I \subset [0,1]^{q_1+q_2}$, let $C(I)$ and $C_n(I)$ denote the $C-$ and the C_n- measure of I and let $\alpha_n(I) = C_n(I) - C(I)$. Given $0 \leq v_i \leq 1$, $i = 1, \dots, q_1$, and $0 \leq u_{j1} < u_{j2} \leq 1$, $j = 1, \dots, q_2$, let $I_{\mathbf{v},\mathbf{u}} = \prod_{i=1}^{q_1}[v_i, 1] \times \prod_{j=1}^{q_2}[u_{j1}, u_{j2}]$. For $\mathbf{a} = (a_1, \dots, a_{q_2}) \in [0,1]^{q_2}$ define the oscillation modulus

$$\omega_n(\mathbf{a}) = n^{1/2} \sup\left\{|\alpha_n(I_{\mathbf{v},\mathbf{u}})| : 0 \leq v_i \leq 1, |u_{j2} - u_{j1}| \leq a_j\right\},$$

where $i = 1, \dots, q_1$ and $j = 1, \dots, q_2$, $q_1 \geq 0$, $q_2 \geq 1$. Set $\rho(\mathbf{a}) = \sup C(I_{\mathbf{v},\mathbf{u}})$ where supremum extends over the rectangles involved in the definition of $\omega_n(\mathbf{a})$. If no assumptions on C are made, the bound $\rho(\mathbf{a})$ is equal to $\min a_j$. If the last q_2 marginals of C have a bounded joint density $c_0(\mathbf{u})$ with respect to Lebesgue measure on $[0,1]^{q_2}$, then $\rho(\mathbf{a}) = \Gamma_0 \prod_{j=1}^{q_2} a_j$, where $\Gamma_0 = \sup c_0(\mathbf{u})$. The following lemma will be useful.

LEMMA 3.1. *Let s satisfy $2 \leq sn^{1/2}$ and $e_0 n^{1/2} \rho(\mathbf{a}) \geq s$ for some finite $e_0 > 0$. There exist constants $e_1, e_2 > 0$ (not depending on n, a, s or the distribution C) such that $P(\omega_n(\mathbf{a}) > s) \leq e_1(\min a_j)^{-q_2} \exp\{-e_2 s^2/\rho(\mathbf{a})\}$.*

For $q_1 = 0$, the result corresponds to Theorem 1.7 in Stute (1984a). For $q_1 > 0$, a simple modification of the proof of this theorem is needed; see Section 5. Further, note that a variant of the lemma remains valid when the supremum in the definition of $\omega_n(\mathbf{a})$ is restricted to rectangles $I_{\mathbf{v},\mathbf{u}}$ contained in some fixed rectangle $I \subset [0,1]^{q_1+q_2}$.

3.2. *Copula functions.* We return to the framework of Section 2. Let $\tilde{M}_1(x, t, \mathbf{z}) = P(D_i = x, Y_i \leq t, \mathbf{Z}_i \leq \mathbf{z})$, $M_1(t, \mathbf{z}) = P(D_i = 1, Y_i \leq t, \mathbf{Z}_i \leq \mathbf{z})$, $M_2(t, \mathbf{z}) = P(Y_i \leq t, \mathbf{Z}_i \leq \mathbf{z})$ and let $\tilde{M}_{1n}$, M_{1n} and M_{2n} be the corresponding empiricals. (Here $\mathbf{Z}_i \leq \mathbf{z}$ means $Z_{ij} \leq z_j$, $j = 1, \dots, q$.) Under Assumption A(i) the marginals of $\mathbf{Z}_{ij}$'s are continuous. The first marginals of M_1 and M_2 are arbitrary, while the first marginal of $\tilde{M}_1$ is purely discrete assigning mass $1 - p$ to $x = 0$ and p to $x = 1$, where $p = P(D_i = 1)$, $0 < p < 1$. Note also that M_1 is a subdistribution function. We shall find it convenient to replace these cdf's by continuous cdf's. The idea is to spread the jumps uniformly over intervals that will be inserted at each jump point [see, e.g., van Zuijlen (1978) for the discussion of the technique involved].

Let $\{\xi_m : m = 1, 2, \dots\}$ be the set of discontinuities of H_2 and let $\{p_m : m = 1, 2, \dots\}$ be the corresponding heights of jumps, $\sum p_m \leq 1$. Define transformations $\phi_1^*(x) = x + (1-p)I(x > 0) + pI(x > I)$ and $\phi_2^*(t) = t + \sum p_m I(t > \xi_m)$.

Furthermore, let $\{U_{mi}: m = 1, 2, \ldots, i = 1, \ldots, n\}$ and $\{V_{mi}: m = 1, 2, i = 1, \ldots, n\}$ be mutually independent sets of uniform (0, 1) rv's independent of the sample $(D_i, Y_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$. Define new variables $D_i^* = \phi_1^*(D_i) + (1-p)V_{1i}(1 - D_i) + pV_{2i}D_i$ and $Y_i^* = \phi_2^*(Y_i) + \sum p_m U_{mi} I(Y_i = \xi_m)$. Let $\tilde{M}_1^*$ and M_2^* be the joint cdf's of $(D_i^*, Y_i^*, \mathbf{Z}_i)$ and $(Y_i^*, \mathbf{Z}_i)$, respectively, and let $\tilde{M}_{1n}^*$ and M_{2n}^* be the corresponding empiricals. Clearly, $\tilde{M}_1^*$ and M_2^* have continuous marginals and it can be easily verified that $\tilde{M}_1(x, t, \mathbf{z}) = \tilde{M}_1^*(\phi_1^*(x), \phi_2^*(t), \mathbf{z})$ and $M_2(t, \mathbf{z}) = M_2^*(\phi_2^*(t), \mathbf{z})$, and with probability 1, the same relationship holds among the empiricals.

Following Stute (1984a), we shall use a representation of $\tilde{M}_1^*$ and M_2^* in terms of the so-called copula or dependence function. Let H_1^* and H_2^* be the first marginals of $\tilde{M}_1^*$ and M_2^*, respectively. The copula functions pertaining to $\tilde{M}_1^*$ and M_2^* are cdf's C_1 and C_2 defined on $[0,1]^{q+2}$ and $[0,1]^{q+1}$, respectively, satisfying $\tilde{M}_1^*(x^*, t^*, \mathbf{z}) = C_1(H_1^*(x^*), H_2^*(t^*), \mathbf{G}(\mathbf{z}))$ and $M_2^*(t^*, \mathbf{z}) = C_2(H_2^*(t^*), \mathbf{G}(\mathbf{z}))$ where $\mathbf{G}(\mathbf{z}) = (G_1(z_1), \ldots, G_q(z_q))$. Furthermore, if C_{1n} and C_{2n} are empirical cdf's corresponding to C_1 and C_2, then with probability 1, $\tilde{M}_{1n}^*(x^*, t^*, \mathbf{z}) = C_{1n}(H_1^*(x^*), H_2^*(t^*), \mathbf{G}(\mathbf{z}))$ and $M_{2n}^*(t^*, \mathbf{z}) = C_{2n}(H_2^*(t^*), \mathbf{G}(\mathbf{z}))$.

Let us return now to the original processes M_1 and M_2. Let $\beta_{in} = M_{in} - M_i$ and $\alpha_{in} = C_{in} - C_i$, $i = 1, 2$. It follows now that if $I = [t_1, t_2] \times \prod_{j=1}^q [z_{j1}, z_{j2}]$ is a rectangle in R^{q+1}, then with probability 1,

$$\beta_{1n}(I) = \alpha_{1n}(J_0 \times J) \quad \text{and} \quad \beta_{2n}(I) = \alpha_{2n}(J), \tag{3.1}$$

where

$$J_0 = [H_1^*(\phi_1^*(1-)), H_1^*(\phi_1^*(1))]$$

and

$$J = [H_2^*(\phi_2^*(t_1-)), H_2^*(\phi_2^*(t_2))] \times \prod_{j=1}^q [G_j(z_{j1}), G_j(z_{j2})].$$

In conjunction with Lemma 3.1, this representation will be used to compute the oscillation modulus of β_{in}, $i = 1, 2$.

4. Proof of Theorem 2.1. Let Ω_{1n} be the event $\Omega_{1n} = \{\inf \hat{H}_2(t-|\mathbf{z}) > \theta/2, 0 \le t \le \tau, \mathbf{z} \in I\}$. Under Assumption A(i), we have

$$\begin{aligned} P(\Omega_{1n}^c) &\le P\big(\sup\big(\hat{H}_2(t-|\mathbf{z}) - H_{2n}(t-|\mathbf{z})\big) > \inf H_{2n}(t-|\mathbf{z}) - \theta/2\big) \\ &\le P\big(\|\hat{H}_2 - H_{2n}\|_I > \theta/2\big). \end{aligned} \tag{4.1}$$

Further, for $0 < t < \tau$ and $\mathbf{z} \in$ I,

$$\begin{aligned} F_n(t|\mathbf{z}) &= \prod_{s \le t} (1 - d\Lambda_n(s|\mathbf{z})) \\ &\ge \prod_{s \le t} \left(1 + \frac{dH_{2n}(s|\mathbf{z})}{H_{2n}(s-|\mathbf{z})}\right) = H_{2n}(t|\mathbf{z}) > \theta. \end{aligned} \tag{4.2}$$

The second equality follows from the product–integral representation of survival functions given in (1.1) and (1.2). Proposition A.4.1 in Gill (1980), page 153, integration by parts [see Shorack and Wellner (1986), page 305] and (4.2) yield on the event Ω_{1n},

$$\begin{aligned}\|\hat{F} - F_n\|_I^\tau &\le 9\theta^{-3}\|\hat{\Lambda} - \Lambda_n\|_I^\tau \\ &\le 18\theta^{-5}\{\|\hat{H}_2 - H_{2n}\|_i^\tau + 2\|\hat{H}_1 - H_{1n}\|_i^\tau\}.\end{aligned} \tag{4.3}$$

Combining (4.1) and (4.3), we get

$$P(\|\hat{F} - F_n\|_I^\tau > \varepsilon) \le 2P(\|\hat{H}_1 - H_{1n}\|_I^\tau > \varepsilon\theta^5/54) + 2P(\|\hat{H}_2 - H_{2n}\|_I^\tau > \varepsilon\theta^5/54)$$

and clearly it is enough to consider the subsurvival functions only.

For this purpose set $A_{in}(t, \mathbf{z}) = \hat{H}_i(t, \mathbf{z}) - H_{in}(t, \mathbf{z})$, $i = 1, 2$, and $A_{3n}(\mathbf{z}) = A_{2n}(0, \mathbf{z}) = \hat{g}(\mathbf{z}) - g_n(\mathbf{z})$. Under Assumption A(i), $g_n(\mathbf{z}) \ge \gamma$ for $\mathbf{z} \in I$ and all n. Consider the event $\Omega_{2n} = \{\inf \hat{g}(\mathbf{z}) > \gamma/2, \mathbf{z} \in I\}$. We have $P(\Omega_{2n}^c) \le P(\|A_{2n}\|_I^\tau > \gamma/2)$. Further, on the event Ω_{2n},

$$\|\hat{H}_i - H_{in}\|_I^\tau \le \gamma^{-1}(\|A_{in}\|_I^\tau + \|A_{3n}\|_I^\tau) \le \gamma^{-1}(\|A_{in}\|_I^\tau + \|A_{2n}\|_I^\tau).$$

Combining yields

$$P(\|\hat{H}_i - \hat{H}_i\|_I^\tau > \varepsilon\theta^5/54) \le P(\|A_{in}\|_I^\tau > \varepsilon\theta^5\gamma/108) - 2P(\|A_{2n}\|_i^\tau > \varepsilon\theta^5\gamma/108).$$

Define

$$I(t, \mathbf{z}, \mathbf{u}) = [t, \infty] \times \prod_{j=1}^{q} [\min(u_j, z_j), \max(u_j, z_j)].$$

Integration by parts [Hildebrandt (1963)] yields, after some algebra,

$$\begin{aligned}|A_{in}(t, \mathbf{z})| &= a_n^{-q}\left|\int I(y > t) K(a_n^{-1}(\mathbf{z} - \mathbf{u}))\, d\beta_n(y, \mathbf{u})\right| \\ &= a_n^{-q}\left|\int \beta_{in}(I(t, \mathbf{z}; \mathbf{u}))\, dK(a_n^{-1}(\mathbf{z} - \mathbf{u}))\right|,\end{aligned} \tag{4.4}$$

where integration is restricted to those $\mathbf{u}$-values for which $z_j - a_n \le u_j \le z_j + a_n$, $j = 1, \dots, q$. Set

$$J(s, \mathbf{v}; \mathbf{w}) = [s, 1] \times \prod_{j=1}^{q} [v_j + \min(0, w_j), v_j + \max(0, w_j)]$$

and let J_0 be defined as in (3.1). If $s = H_2^*(\phi_2^*(t-))$ and $v_j = G_j(z_j)$, then, after a change of variable in the right-hand side of (4.4), we obtain from (3.1),

$$\begin{aligned}|A_{in}(t, \mathbf{z})| &\le a_n^{-q} \int |\alpha_{1n}(J_0 \times J(s, \mathbf{v}; \mathbf{w}))|\, |d\psi(\mathbf{w})| \quad \text{if } i = 1 \\ &\le a_n^{-q} \int |\alpha_{2n}(J(s, \mathbf{v}; \mathbf{w}))|\, |d\psi(\mathbf{w})| \quad \text{if } i = 2.\end{aligned}$$

Here integration is restricted to those $\mathbf{w}$-values for which

$$G_j(z_j - a_n) - G_j(z_j) \le w_j \le G_j(z_j + a_n) - G_j(z_j), \qquad j = 1, \dots, q.$$

Further, $\psi(\mathbf{w}) = K(\mathbf{w}')$ and $w_j' = (z_j - G_j^{-1}(w_j + G_j(z_j)))/a_n$, $j = 1, \dots, q$. Under Assumption A(i) by the mean value theorem $G_j(z_j + a_n) - G_j(z_j) \le$

Γa_n and $G_j(z_j - a_n) - G_j(Z_j) > -\gamma a_n$, so that $|w_j| \le \Gamma a_n$, $j = 1, \dots, q$. Moreover, the density $c_0(\mathbf{w})$ given by (2.1) is bounded on the rectangle $\Pi_{j=1}^{q}[G_j(a_j) - \delta\Gamma, G_j(b_j) + \delta\Gamma]$ by Γ_0. Lemma 3.1 applied with $q_1 = 2$, $q_2 = q$, $\mathbf{a} = (a_1, \dots, a_q)$, $a_j = \Gamma a_n$ and $s = \varepsilon\theta^5\gamma\lambda^{-1}n^{1/2}a_n^q/108$ entails

$$P\big(\|A_{1n}\|_I^\tau > \varepsilon\theta^5\gamma/108\big) \le e_3\Gamma^{-q}a_n^{-q}\exp\big\{-e_2\gamma^2\lambda^{-2}\Gamma_0^{-1}\Gamma^{-q}\varepsilon^2\theta^{25}na_n^q/11664\big\}$$

for some constants $e_1, e_2 > 0$. The same choice of q_2, $\mathbf{a}$ and s but $q_1 = 1$, yields

$$P\big(\|A_{2n}\|_I^\tau > \varepsilon\theta^5\gamma/108\big) \le e_3\Gamma^{-q}a_n^{-q}\exp\big\{-e_4\gamma^2\lambda^{-2}\Gamma_0^{-1}\Gamma^{-q}\varepsilon^2\theta^{25}na_n^q\big\}$$

for some $e_3, e_4 > 0$. The conclusion of the theorem follows by setting $d_1 = 12\Gamma^{-q}\max(e_1, e_2)$ and $d_2 = \min(e_2, e_4)\gamma^2\lambda^{-2}\Gamma_0^{-1}\Gamma^{-q}/11664$.

5. Proof of Lemma 3.1. The proof rests on a modification of arguments used in the proof of Theorem 1.7 in Stute (1984a). We consider first the case of $q_2 = 1$.

Choose $0 < \delta < \frac{1}{2}$ and let $b_1, \dots, b_{q_1}, a < \frac{1}{2}$ be such that $C(I_{\mathbf{v},\mathbf{u}}) \le \delta/4$, where $\mathbf{v} = (v_1, \dots, v_{q_1})$ and $I_{\mathbf{v},\mathbf{w}} = \Pi_{i=1}^{q_1}[v_i, 1] \times [0, u]$. By Lemma 1.2 in Stute (1984a), there exists $c = c(\delta)$ such that for all $s > 0$, $2 \le sn^{1/2}C(I_{1-b,a})^{1/2}$, $1 - \mathbf{b} = (1 - b_1, \dots, 1 - b_{q_1})$ and $32 \le s^2\delta^2(1 - 2\delta)^2$,

$$\begin{aligned} &P\big(\sup n^{1/2}|\alpha_n(I_{\mathbf{v},u})| > s\big(C(I_{1-\mathbf{b},a})\big)^{1/2}\big)\\ &\quad \le c(\delta)P\big(n^{1/2}|\alpha_n(I_{1-\mathbf{b},a})| \ge s(1 - 2\delta)^{q_1+1}C(I_{1-\mathbf{b},a})^{1/2}\big). \end{aligned}$$

The supremum on the left-hand side extends over $\mathbf{v}$ such that $1 - b_i \le v_i \le 1$, $i = 1, \dots, q_1$ and $0 < u < a$. It can be verified that this inequality remains valid for $b_i = \cdots = b_{q_1} = 1$. Replacing s by $s(C([0,1]^{q_1} \times [0, a])^{-1/2}) = sa^{-1/2}$ we obtain

$$P\Bigg(\sup_{\substack{0 \le v_i \le 1\\ 0 \le u \le a}} n^{1/2}|\alpha_n(I_{\mathbf{v},u})| > s\Bigg) \le c(\delta)P\big(n^{1/2}|\alpha_n(1, \dots, 1, a)| > s(1 - 2\delta)^{q_1+1}\big)$$

provided $2 \le sn^{1/2}$, $32a \le s^2\delta^2(1 - 2\delta)^2$ and $a < \frac{1}{4}$. By Bernstein's inequality, the right-hand side is bounded from above by

$$2c(\delta)\exp\big\{-s^2(1 - 2\delta)^{2q_1+2}/2a\big(1 + e_0(1 - 2\delta)^{q_1+1}/3\big)\big\} \le e_1'\exp\big\{-e_2's^2/a\big\}$$

for some $e_1', e_2' > 0$ depending on δ and e_0.

We proceed next in a fashion similar to the proof of Lemma 2.4 in Stute (1982). Let R be the smallest positive integer satisfying $\delta^2a/4 > 1/R$. Then

$$\begin{aligned} P(\omega_n(a) > s) \le{}& \sum_{i=0}^{R-1} P\Bigg(\sup_{\substack{0 \le v_i \le 1\\ 0 \le u \le a}} n^{1/2}|\alpha_n\big(I_{\mathbf{v},u+i/R}\big) - \alpha_n\big(I_{\mathbf{v},i/R}\big)| > s/(1 + \delta)\Bigg)\\ &+ \sum_{i=0}^{R-1} P\Bigg(\sup_{\substack{0 \le v_j \le 1\\ 0 \le u \le 1/R}} n^{1/2}|\alpha_n\big(I_{\mathbf{v},u+i/R}\big) - \alpha_n\big(I_{\mathbf{v},i/R}\big)|\\ &\qquad > s\delta/(1 + \delta)\Bigg). \end{aligned}$$

Since $1/R < a < \delta/4$, we can apply the exponential bound to both sums to obtain

$$P(\omega_n(a) > s) \le Re_1' \exp\left\{-e_2' s^2/a(1+\delta)^2\right\} + Re_1' \exp\left\{-e_2' s^2\delta^2/\left(4a(1+\delta)^2\right)\right\}$$
$$\le 2e_1' \exp\left\{-e_2' s^2\delta^2/\left(4a(1+\delta)^2\right)\right\}$$

provided $2 \le sn^{1/2}$, $32a \le s^2\delta^2(1-2\delta)^2$ and $a < \delta/4$. Under assumptions of the lemma, since e_1 and e_2 are left unspecified, we can assume without loss of generality that the last two growth conditions are satisfied. The conclusion follows by setting $e_1 = 2e_1'$ and $e_2 = e_2'\delta^2/(4(1+\delta)^2)$.

For $q_2 = 2$, the proof is similar to that of Theorems 1.5 and 1.7 in Stute (1984a). An induction argument shows that the lemma remains valid for arbitrary q_2.

Acknowledgments. I thank the Associate Editor and the referees for their comments.

REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

COLLOMB, G. (1981). Estimation non-paramétrique de la régression: Revue bibliographique. *Internat. Statist. Rev.* **49** 75–93.

DABROWSKA, D. M. (1987a). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14** 181–197.

DABROWSKA, D. M. (1987b). Uniform consistency of nearest neighbour and kernel conditional Kaplan–Meier estimates. Technical Report No. 86, Univ. California, Berkeley.

DEVROYE, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.* **9** 1310–1319.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Mathematical Centre Tracts* **124**. Centre for Mathematics and Computer Science, Amsterdam.

GILL, R. D. and JOHANSEN, S. (1987). Product integrals and counting processes. Technical Report MS-R8707, Centre for Mathematics and Computer Science, Amsterdam.

HILDEBRANDT, T. K. (1963). *Introduction to the Theory of Integration*. Academic, New York.

HORVÁTH, L. and YANDELL, B. S. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Anal.* **26** 184–206.

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

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete* **62** 405–412.

NADARAYA, E. E. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NELSON, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14** 945–966.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 832–837.

ROSENBLATT, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II* (P.R. Krishnaiah, ed.) 25–31. Academic, New York.

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

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

STUTE, W. (1984a). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379.

STUTE, W. (1984b). Asymptotic normality of nearest neighbour regression function estimates. *Ann. Statist.* **12** 917–926.

STUTE, W. (1986a). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.* **14** 891–901.

STUTE, W. (1986b). Conditional empirical processes. *Ann. Statist.* **14** 638–647.

VAN ZUIJLEN, M. C. A. (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann. Probab.* **6** 250–266.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

YOUNG, W. H. (1917). On multiple integration by parts and the second theorem of the mean. *Proc. London Math. Soc.* (*2*) **16** 273–296.

DIVISION OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024-1772