Nonparametric regression with censored survival time data

By

Dorota M. Dabrowska

**Carnegie-Mellon University and
University of California, Berkeley**

Technical Report No. 81
November 1986

Department of Statistics
University of California
Berkeley, California

Nonparametric regression with censored survival time data

Dorota M. Dabrowska

Carnegie-Mellon University and
University of California, Berkeley

Abstract

The paper deals with a class of nonparametric regression estimates introduced by Beran (1981) to estimate conditional survival functions in the presence of right censoring. Weak convergence results are established for kernel and nearest neighbour estimates of the conditional cumulative hazard and survival functions as well as the quantile and L-type regression functionals.

Key words and phrases: Kernel and nearest neighbour regression, right censoring.

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 $Z = (Z_1, \cdots, Z_d)$ be a vector of covariates such as age, blood pressure, cholesterol level. The survival time T is subject to right censoring so that the observable rv's are given by $Y = \min(T, C)$, $\delta = I(T \leq C)$ and Z. Here C is a nonnegative rv representing times to withdrawal from the study. Beran (1981) proposed a class of nonparametric estimates of the conditional survival and cumulative hazard functions, and related quantile and L-type regression functionals. We consider weak convergence results for these estimates.

Since Beran's paper does not seem to be generally available, we briefly summarize his ideas. Denote by $F(t \mid z) = P(T > t \mid Z = z)$, $H_1(t \mid z) = P(Y > t, \delta = 1 \mid Z = z)$ and $H_2(t \mid z) = P(Y > t \mid Z = z)$ the respective conditional survival functions and let

$$\Lambda(t \mid z) = -\int_0^t F(s - \mid z)^{-1} \, dF(s \mid z)$$

be the conditional cumulative hazard function associated with $F(t \mid z)$. It is assumed throughout that T and C are conditionally independent given Z, which is a sufficient condition to ensure identifiability of $\Lambda(t \mid z)$ and $F(t \mid z)$. Specifically, for any t such that $H_2(t \mid z) > 0$ we have

$$\Lambda(t \mid z) = -\int_0^t \frac{dH_1(s \mid z)}{H_2(s - \mid z)},$$

$$F(T \mid z) = \exp\{-\Lambda^c(t \mid z)\} \, \Pi \, \{1 - \Delta\Lambda(s \mid z)\},$$

where $\Lambda^c(s \mid z)$ is the continuous component of $\Lambda(s \mid z)$, the product is taken over the set of discontinuities of $H_1(s \mid z)$ and $s \le t$, and $\Delta\Lambda(s \mid z) = \Lambda(s \mid z) - \Lambda(s - \mid z)$. This is the well known product-integral representation of distribution functions, see for instance Peterson (1977) and Gill (1980).

Let (Y_j, δ_j, Z_j), $j = 1, \cdots, n$, be a sample of i.i.d. rv's each having the same distribution as (Y, δ, Z). The subdistribution functions $H_1(t \mid z)$ and $H_2(t \mid z)$ are estimated by

$$H_{1n}(t \mid z) = \sum_{j=1}^{n} I(Y_j > t, \delta_j = 1) B_{nj}(z),$$

$$H_{2n}(t \mid z) = \sum_{j=1}^{n} I(Y_j > t) B_{nj}(z),$$

where $B_{nj}(z)$ is a random set of nonnegative weights depending on covariates only. Examples of possible weights include kernel type weights, nearest neighbours or local linear weights. Beran's estimates of $\Lambda(t \mid z)$ and $F(t \mid z)$ are provided by

$$\Lambda_n(t \mid z) = -\int_0^t \frac{dH_{1n}(s \mid z)}{H_{2n}(s - \mid z)} \tag{1.1}$$

and

$$F_n(t \mid z) = \prod \{1 - \Delta\Lambda_n(s \mid z)\}, \tag{1.2}$$

where the product is taken over $s \le t$. Both $\Lambda_n(t \mid z)$ and $F_n(t \mid z)$ are right continuous functions of t. jumps occur at discontinuity points of $H_{1n}(t \mid z)$. Note that in the homogeneous case, (1.1) and (1.2) are simply the Aalen - Nelson (Aalen (1978), Nelson (1972)) and Kaplan - Meier (1958) estimates.

Beran studied conditions which entail uniform consistency of $F_n(t \mid z)$. We recall his result in Section 2.1. In this paper, we shall consider asymptotic normality results. Define processes $L_n(t \mid z) = b_n(\Lambda_n(t \mid z) - \Lambda(t \mid z))$ and $S_n(t \mid z) = b_n(F_n(t \mid z) - F(t \mid z))$, where $\{b_n\}$, $b_n \to \infty$ is a sequence of normalizing constants. Let μ denote the distribution of Z. Under suitable conditions, in Section 2.2 we shall show that for μ - almost all z, the processes $S_n(t \mid z)$ and $L_n(t \mid z)$ converge weakly to a mean zero Gaussian process with covariance structure similar to that of the limiting distribution of the Kaplan -Meier and Aalen - Nelson estimates of the unconditional survival and cumulative hazard functions. In particular, the covariance function of the limiting process depends on $C(t \mid z)$, a nondecreasing function of t, given by

$$C(t \mid z) = \int_0^t \frac{\Lambda(ds \mid z)}{H_2(s - \mid z)} \phi(z), \tag{1.3}$$

where $\phi(z)$ is a function depending on the underlying model through the distribution of Z only. The special case of nearest neighbour and kernel estimates is considered in Section 3.

Along with estimation of the conditional survival functions, in practice we would like to deal with some descriptive statistics that estimate parameters of the unknown underlying conditional distributions. In the case of location, mean regression would be the common choice. In the presence of censoring, estimation of the mean regression creates however some problems. Firstly, the mean regression is in general not identifiable. Although this problem can be resolved by assuming mild conditions on the supports of the conditional distributions of survival and censoring variables (see Doksum and Yandell (1981)), in order to ensure

weak convergence results, one needs additional, somewhat cumbersome conditions on the tail behaviour of these distributions. Rather than estimate the mean regression, we can look at estimates of truncated mean, or choose to estimate other location parameters. Quantile regression, for instance and related to it L - type regression functionals provide a broad range of possible estimates of location of the conditional distributions (see Bickel and Lehmann·(1976), Huber (1981)). We briefly discuss asymptotic normality results for these estimates. We refer to Stone (1977), Mallows (1979, 1980), Stützle and Mittal (1979) and Velleman (1977) for a discussion of some recent developments in the area of robust non-parametric regression estimation for uncensored data. For censored data, Doksum and Yandell (1981) considered median regression based on asymmetric nearest neighbour estimates and compared it to the median regression derived from the Cox proportional hazard model.

This fully nonparametric approach towards regression estimation in the presence of censoring was first adopted by Beran (1981). Various alternative approaches exist. Horváth (1981) for instance, proposed to estimate the conditional survival function by integrating an estimate of the conditional density. The latter was constructed as a ratio of two estimates: an estimate of joint density obtained by kernel smoothing the multivariate product limit estimator of Campbell and Földes (1981) and a kernel estimate of the marginal density of covariates. Further, several authors discussed estimation of mean regression. Owing to identifiability problems, these approaches require additional assumptions on the dependence structure among covariates and censoring times. In particular, under assumption that the covariate is independent of the censoring

variable, Doksum and Yandell (1981) developed simultaneous confidence bands for asymmetric nearest neighbour estimates of mean regression. An analogous approach was more recently adopted by Tsai and van Ryzin (1985).

Among other approaches, most popular are methods related to multiplicative intensity models. Define the multivariate counting process $N(t) = (N_1(t), \cdots, N_n(t))$ by $N_i(t) = I(Y_i \leq t, \delta_i = 1)$ and let $\lambda(t) = (\lambda_1(t), \cdots, \lambda_n(t))$ be its random intensity. The Cox (1972) proportional hazard model corresponds to $\lambda_i(t) = \alpha_0(t) \exp\{\beta^T z\} I(Y_i \geq t)$ where α_0 is an unknown hazard function and β is a vector of regression coefficients. In this model, the estimation of the unknown parameters is usually carried out in two steps: first an estimate $\hat{\beta}$ is obtained by maximizing the so-called partial (or rank) likelihood. An estimate of α_0 is next constructed by maximizing a Kiefer - Wolfowitz type nonparametric likelihood function at the point $\beta = \hat{\beta}$. We refer to Cox (1972, 1975), Kalbfleisch and Prentice (1980), Andersen and Gill (1982), Begun et al. (1983) and references therein for a discussion of various theoretical and practical aspects of this approach. Tibshirani (1984) and O'Sullivan (1986) proposed a nonparametric version of this model by replacing $\exp\{\beta^T z\}$ with $\exp\{\eta(z)\}$, where $\eta(z)$ is a smooth function, and used additive approximations and maximum penalized partial likelihood, respectively, to estimate η. Finally, we mention a model introduced by Aalen (1980) who suggested a matrix version of the multiplicative model assuming that $\lambda_i(t) = \Sigma \alpha_j(t) X_{ij}(t)$, where $(\alpha_i, \cdots, \alpha_p)^T$, $p < n$ is a vector of unknown functions and $X_i(t) = \{X_{i1}(t), \cdots, X_{ip}(t) \; i = 1, \cdots, n\}$ is a matrix of predictable

processes.

Methods related to linear models have been generalized to censored data among others by Miller (1976), Buckley and James (1979), Koul et al. (1981), Miller and Halpern (1981) and Leurgans (1984). These methods, although parametric in mind, are usually coupled with some nonparametric techniques in order to reconstruct missing observations. Both parametric and the parametric part of the Cox model may give a systematic bias when estimating for instance the quantile regression.

The nonparametric approach of this paper is useful in the exploratory data analysis and in the analysis of large data sets. Since minimal assumptions on the underlying model are imposed, this method is very flexible and can be used to determine if any of the parametric or semiparametric models give a good fit to the data.

The method has several weak points. A genuine drawback of the nonparametric regression is that it does not work well for $d \geq 3$ and the usual sample sizes. The convergence to the limiting distribution is much slower than in the case of parametric or semiparametric methods. In the case of kernel estimates for instance, we have $(na_n^d)^{1/2}$ - convergence with a_n tending slowly to zero, whereas in the parametric and semiparametric approaches the rate $n^{1/2}$ persists. The amount of data required to avoid an unacceptably large variance increases rapidly with increasing dimensionality. In this respect the parametric and semiparametric methods seem to be advantageous, except when large data sets are available. Projection pursuit (Friedman and Stützle (1981), Huber (1985)) and

generalized additive models (Hastie and Tibshirani (1986)) provide useful alternatives to the nonparametric regression considered in this paper.

Practical user of these methods has to face the problem of the choice of the smoothing parameter (e.g. bandwidth). This parameter determines the degree of smoothing and is a delicate trade-off between the variance and the bias of the estimator. A possible systematic way of choosing it is to determine the value which minimizes the cross-validation mean square error.

2. Main results.

2.1. Uniform consistency.

We recall Beran's (1981) uniform consistency result.

Proposition 2.1. *Let* $\tau(z) < \{s : H_2(s \mid z) > 0\}$. *Suppose that* $H_{in}(t \mid z)$, $i = 1,2$, *are strongly consistent in the sense that for* μ-*almost all* z $\sup_t |H_{in}(t - \mid z) - H_i(t - \mid z)| \to 0$ *a.s. and* $\sup_t |H_{in}(t + \mid z) - H_i(t + \mid z)| \to 0$ *a.s. as* $n \to \infty$. *Then, for* μ-*almost all* z, $\sup |F_n(t \mid z) - F(t \mid z)| \to 0$ *a.s. in the supremum norm on* $[0, \tau(z)]$.

An analogous result holds for the conditional cumulative hazard function.

Proposition 2.2. *Under assumptions of Proposition 2.1, for* μ-*almost all* z,

$\sup |\Lambda_n(t \mid z) - \Lambda(t \mid z)| \to 0$ *a.s. in the supremum norm on* $[0, \tau(z)]$.

Proof. Suppose $0 \le t \le \tau(z)$. Then $H_2(t \mid z) \ge H_2(\tau(z) \mid z) > 0$ and with probability 1, $H_{2n}(t \mid z) \ge H_{2n}(\tau(z) \mid z) > 0$ for μ-almost all z and n sufficiently large. Integration by parts and a little algebra yield

$$|\Lambda_n(t \mid z) - \Lambda(t \mid z)| \le O(1)\sup|H_{2n}(t - \mid z) - H_2(t - \mid z)| +$$

$O(1)\sup|H_{1n}(t\mid z)-H_1(t\mid z)|$, where the suprema are taken over $[0,\tau(z)]$. It follows that $\Lambda_n(t\mid z)$ is pointwise consistent. Since $\Lambda_n(t\mid z)$ and $\Lambda(t\mid z)$ are bounded monotone functions, we also have uniform consistency on $[0,\tau(z)]$.

The following argument, borrowed from Shorack and Wellner (1986, p. 305), is a simple way to see uniform consistency of $F_n(t\mid z)$. From Proposition A.4.1 in Gill (1980, p. 153), we have

$$F_n(t\mid z)-F(t\mid z)=-F(t\mid z)\int_0^t\frac{F_n(s-\mid z)}{F(s-\mid z)}\,d\psi(s\mid z)$$

where

$$\psi(t\mid z)=\int_0^t\left[1-\frac{\Delta F(s\mid z)}{F(s\mid z)}\right]d(\Lambda_n(s\mid z)-\Lambda(s\mid z)).$$

Integration by parts implies $|F_n(t\mid z)-F(t\mid z)|\le O(1)\sup|\psi(t\mid z)|$ a.s., where the supremum is taken over $[0,\tau(z)]$. But

$$\begin{aligned}
\sup|\psi(t\mid z)| &\le \sup|\Lambda_n(t\mid z)-\Lambda(t\mid z)|\\
&\quad+\sup\Sigma\,|\Delta F(s\mid z)|\,F(s\mid z)^{-1}|\,\Delta\Lambda_n(s\mid z)-\Delta\Lambda(s\mid z)|\\
&\le 3\,O(1)\sup|\Lambda_n(t\mid z)-\Lambda(t\mid z)|.
\end{aligned}$$

Here the sum is taken over $s\le t$ such that $\Delta F(s\mid z)>0$. Proposition 2.2 entails pointwise consitency of $F_n(t\mid z)$. Since $F_n(t\mid z)$ and $F(t\mid z)$ are bounded monotone functions, we also have uniform consistency on $[0,\tau(z)]$.

2.2. Weak convergence.

We shall discuss now weak convergence results for the estimates of the conditional survival and hazard function. It is assumed throughout that $H_i(t\mid z)$ is continuous in t for μ-almost all z.

Let $\{b_n\}$ be a sequence such that $b_n \to \infty$ as $n \to \infty$. Further, let $\phi(z)$ be a function depending on the underlying model through the distribution of Z only. For $i = 1,2$, set

$$W_{in}(t \mid z) = b_n(H_{in}(t \mid z) - H_i(t \mid z))$$
$$W_n(t \mid z) = (W_{1n}(t \mid z), W_{2n}(t \mid z)) \tag{2.1}$$

and let $L_n(t \mid z) = b_n(\Lambda_n(t \mid z) - \Lambda(t \mid z))$ and $S_n(t \mid z) = b_n(F_n(t \mid z) - F(t \mid z))$. Finally, given $\tau(z) < \sup\{s : H_2(s \mid z) > 0\}$, let $D[0, \tau(z)]$ be the space of cadlag functions endowed with Skorokhod topology.

The following proposition is an analogue of Theorem 4 in Breslow and Crowley (1974).

Proposition 2.3. *Let* $\tau(z) < \sup\{s : H_2(s \mid z) > 0\}$. *Suppose that, for μ-almost all z, the process $W_n(t \mid z)$ converges weakly in $D[0, \tau(z)] \times D[0, \tau(z)]$ to $W(t \mid z) = (W_1(t \mid z), W_2(t \mid z))$, a two-dimensional mean zero Gaussian process with covariance function*

$$cov\{W_i(s \mid z), W_j(t \mid z)\} \tag{2.2}$$
$$= \{H_{min(i,j)}(t \mid z) - H_i(s \mid z)H_j(t \mid z)\}\phi(z)$$

for $s \leq t$. Then, for μ-almost all z, the processes $L_n(t \mid z)$ and $S_n(t \mid z)$ converge weakly in $D[0, \tau(z)]$ to mean zero Gaussian processes with covariance functions $cov\{L(s \mid z), L(t \mid z)\} = C(s \mid z)$ and $cov\{S(s \mid z), S(t \mid z)\} = F(s \mid z)F(t \mid z)C(s \mid z)$, where $s \leq t$ and $C(t \mid z)$ is given by (1.3).

The proof is postponed to Section 7. The result is not surprising, since $\Lambda(t \mid z)$ and $F(t \mid z)$ are smooth (ie. compactly differentiable) functionals of $H_1(t \mid z)$ and $H_2(t \mid z)$.

For practical purposes we shall need to estimate the covariance function from the data. Let $\phi_n(z)$ be an estimate of $\phi(z)$ and let

$$C_n(t \mid z) = \phi_n(z) \int\limits_0^t H_{2n}^{-1}(s \mid z) d\Lambda_n(s \mid z).$$

Corollary 2.1. *Suppose that the assumptions of Proposition 2.1 are satisfied and that, for μ-almost all z, $\phi_n(z) \to \phi(z)$ a.s. Then, for μ-almost all z,*

$C_n(t \mid z) \to C(t \mid z)$ and $F_n(t \mid z)C_n(t \mid z) \to F(t \mid z)C(t \mid z)$ almost surely in the supremum norm on $[0, \tau(z)]$.

This follows directly from Propositions 2.1 and 2.2. We omit the details.

As an immediate consequence of Proposition 2.1, we obtain asymptotic normality of the truncated mean regression. Let $\tau(z) < \sup\{s : H_2(s \mid z) > 0\}$. Then the truncated mean regression

$$m(z; \tau(z)) = \int\limits_0^{\tau(z)} F(s \mid z) ds \tag{2.3}$$

is identifiable. The sample counterpart $m_n(z; \tau(z))$ of (2.3) can be defined by substituting $F_n(t \mid z)$ for $F(t \mid z)$. Obviously, under assumptions of Propositions 2.1 and 2.3 for μ-almost all z,

$$|m_n(z; \tau(z)) - m(z; \tau(z))| \le \sup |F_n(t \mid z) - F(t \mid z)|\tau(z) \to 0 \qquad \text{a.s.} \qquad \text{and}$$

$b_n(m_n(z; \tau(z)) - m(z; \tau(z)))$ converges weakly to a mean zero normal distribution with variance

$$\sigma^2(z; \tau(z)) = \int\limits_0^{\tau(z)} (\int\limits_s^{\tau(z)} F(t \mid z) dt)^2 ds \; C(s \mid z). \tag{2.4}$$

The asymptotic variance can be estimated consistently by substituting $F_n(t \mid z)$ and $C_n(s \mid z)$ for $F(t \mid z)$ and $C(s \mid z)$ in (2.4).

2.3. Quantile and L-type regression functionals.

As an alternative to the mean regression, Beran (1981) proposed to consider quantile regression and related to it L-type regression functionals. These functionals are attractive because of their robustness properties. Survival time distributions are often skewed to the right. Consequently, the conditional mean and its estimate are pulled to the right by heavy right tails and outlying observations on the right, respectively. For this reason the conditional median or trimmed mean are often more useful.

For any $p \in (0, 1)$ the p-th conditional quantile is taken to be $Q(p \mid z) = \inf\{t : F(t \mid z) \leq 1 - p\}$. Further, we consider L - type regression functionals $l(F(t \mid z)) = l_1(F(t \mid z)) + l_2(F(t \mid z))$ where

$$l_1(F(t \mid z)) = \int Q(p \mid z) J(p) dp$$

$$l_2(F(t \mid z)) = \sum_{j=1}^{m} c_j Q(p_j \mid z)$$

where c_j are some weights and J is a linear combination of densities on $(0, 1)$. Special cases include the a-trimmed mean $(J(p) = I(a < p < 1 - a)/(1 - 2a), m = 0)$, the a-Winsorized mean $(J(p) = I(a < p < 1 - a)/(1 - 2a), m = 2, p_1 = a, p_2 = 1 - a$ and $b_1 = b_2 = a))$, etc. The sample counterparts $Q_n(t \mid z)$ and $l_n(z)$ are defined analogously. In particular, $Q_n(t \mid z) = \inf\{t : F_n(t \mid z) \leq 1 - p\}$ with the exception that if $F_n(t \mid z) > 1 - p$ over its entire domain then $Q_n(p \mid z)$ is left undefined.

To establish weak convergence of $Q_n(p \mid z)$ and $l_n(z)$ we shall require

I. There exists points $0 < a < b < 1$ such that for μ-almost all z

$Q(b \mid z) < \sup\{s : H_2(s \mid z) > 0\}$ and for $t \in [Q(a \mid z), \ Q(b \mid z)]$.

$F(t \mid z)$ has a continuous density $f(t \mid z)$ bounded away from zero.

II. The function J is continuous with compact support $[a, b]$,

$0 < a < b < 1$. For μ-almost all z, $\max\{Q(b \mid z), \ Q(p_j \mid z),$

$j = 1, \cdots, m\} < \sup\{s : H_2(s \mid z) > 0\}$, and the function $F(t \mid z)$ has

positive density in neighbourhoods of its p_j quantiles, $j = 1, \cdots, m$.

With an abuse of notation, we shall write in what follows $f \circ Q(p \mid z)$,

$f \circ Q_n(p \mid z)$, etc. to denote the value of $f(u \mid z)$ at $u = Q(p \mid z)$ and

$u = Q_n(p \mid z)$, respectively. Similar convention applies to $S \circ Q(p \mid z)$,

$C \circ Q(p \mid z)$, etc. The following corollary is an analogue of results of Sander

(1975 a & b), Reid (1981), Aly et al. (1985) and Lo and Singh (1986) for the

unconditional quantiles and L-functionals. See M. Csörgö (1983) for other refer-

ences and theoretical results.

Corollary 2.2. *(i). Under condition I and the assumptions of Propositions 2.1*

and 2.3, for μ-almost all z. $U_n(p \mid z) = b_n\{Q_n(p \mid z) - Q(p \mid z)\}$ converges

weakly in $D[a,b]$ to a mean zero Gaussian process $U(p \mid z)$ given by

$U(p \mid z) = -S \circ Q(p \mid z)/f \circ Q(p \mid z)^{-1}$.

(ii). Under condition II and assumptions of Propositions 2.1 and 2.3, for μ-

almost all z, $b_n(l_n(z) - l(z))$ is asymptotically mean zero normal with variance

$\sigma^2(z) = Var\{-S(t \mid z)J(F(t \mid z))dt + \sum_j c_j U(p_j \mid z)\}$.

As a consequence of this corollary, we can derive approximate confidence intervals for the a-trimmed mean regression. Let $l(F(t \mid z))$ be the a-trimmed mean, i.e. $J(p) = I(a < p < 1 - a)/(1 - 2a)$ and $m = 0$. Then the asymptotic variance of $l_n(z) = l(F_n(t \mid z)$ is equal to

$$\sigma^2(z) = 2(1 - 2a)^{-2} \int\limits_{a}^{1-a} (1 - u)C \circ Q(u \mid z) \left[\int\limits_{u}^{1-a} v dQ(v \mid z) \right] du.$$

The sample analogue $\sigma_n^2(z)$ is a consistent estimate of $\sigma^2(z)$. It follows that an approximate level $(1 - \alpha)$ - confidence interval for the trimmed mean is given by $l_n(z) \pm \omega_{\alpha/2} \sigma_n^2(z)$, where $\omega_{\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the standard normal distribution.

Construction of confidence intervals for the quantile, say median, regression is more complicated since the asymptotic variance depends on the unknown conditional density $f(t \mid z)$. Though estimates of $f(t \mid z)$ could be devised by applying a smoother to $F_n(t \mid z)$, the rate of convergence is very slow. Bootstrap provides an alternative and will be pursued elsewhere.

3. Kernel and nearest neighbour estimates

In this section we consider kernel and nearest neighbour estimates and give conditions under which the assumptions of Propositions 2.1 and 2.3 are satisfied. For the sake of convenience, we assume that the covariate Z is univariate. Extensions to the multidimensional case are briefly mentioned in Section 3.4.

Three estimates are discussed. In the first case the weights $B_{nj}(z)$ are given by

$$B_{nj}(z) = (na_n)^{-1} K\left[\frac{G_n(z) - G_n(Z_j)}{a_n}\right], \qquad (3.1)$$

where K is a nonnegative kernel function, $a_n \to 0$ is a sequence of bandwidths and G_n is the empirical distribution function of Z_j's. The resulting estimators $H_{in}(t \mid z)$ might be considered as smooth nearest neighbour estimates of $H_i(t \mid z)$. In particular, if K is supported on the set $[-1, 1]$ then given z, the kernel K assigns positive weight to $a_n n$ nearest neighbours of z on the right and $a_n n$ nearest neighbours of z on the left. The second estimate is a Nadaraya-Watson kernel type estimate with weights

$$B_{nj}(z) = K\left[\frac{z - Z_j}{a_n}\right] / \sum_{k=1}^{n} K\left[\frac{z - Z_k}{a_n}\right], \qquad (3.2)$$

Finally, we consider asymmetric nearest neighbour estimates with weights

$$B_{nj}(z) = K\left[\frac{z - Z_j}{R_n}\right] / \sum_{k=1}^{n} K\left[\frac{z - Z_k}{R_n}\right], \qquad (3.3)$$

where R_n is the Euclidean distance between z and its m_n-th nearest neighbour among Z_j's. The smooth nearest neighbour estimates were considered in particular by Stute (1984a, 1986a), Yang (1981) and Johnston (1981), kernel estimates by Nadaraya (1964), Watson (1964), Rosenblatt (1969), Schuster (1972) and Stute (1986b), asymmetric nearest neighbour estimates by Stone (1977) and Mack (1981).

Throughout it is assumed that Z has a continuous distribution function G. In the case of kernel and asymmetric nearest neighbour estimates, Z is required to have a density g. The kernel K is a density with compact support, and we write

$$\alpha(K) = \int K^2(u)\, du.$$

3.1 Smooth nearest neighbour estimates.

We shall first consider the problem of uniform consistency. Introduce functions $H_1(t \mid G^{-1}(u)) = P(Y > t \ \delta = 1 \mid G(Z) = u)$ and $H_2(t \mid G^{-1}(u)) = P(Y > t \mid G(Z) = u)$. We need to assume

A1. K is continuously differentiable, $a_n \to 0$, $na_n \to \infty$ and

$$\Sigma \exp\{-\rho na_n^4\} < \infty \tag{3.4}$$

for all $\rho > 0$.

A2. K is a mean zero density. Given z, the functions $H_i(t \mid G^{-1}(u))$, $i = 1,2$ are twice continuously differentiable in a u in a neighbourhood $U(z)$ of $0 < G(z) < 1$ and

$$\sup_{U(z)} \sup_t |H_i''(t \mid G^{-1}(u))| < \infty.$$

Note that condition (3.4) is always satisfied whenever $(n^{-1}\log n)^{1/4} = o(a_n)$. Define

$$\hat{H}_{in}(t \mid z) = a_n^{-1}\int H_i(t \mid u)K\left(\frac{G(z) - G(u)}{a_n}\right)dG(u).$$

Proposition 3.1. *If A1 holds then for μ-almost all z,*

$\sup|H_{in}(t \mid z) - \hat{H}_{in}(t \mid z)| \to 0$ *a.s.* *and* $\sup|H_{in}(t - \mid z) - \hat{H}_{in}(t - \mid z)| \to 0$ *a.s.*

If in addition A2 is satisfied then $\sup|\hat{H}_{in}(t \mid z) - H_i(t \mid z)| \to 0$ *and*

$\sup|\hat{H}_{in}(t - \mid z) - H_i(t - \mid z)| \to 0$ *so that* $H_{in}(t \mid z)$ *and* $H_{in}(t - \mid z)$ *are uniformly consistent.*

The proof of the first part of the proposition is given in Section 4.1. The second part deals with the bias term. More precisely, under assumption A2, for

n sufficiently large and μ-almost all z we have

$$\hat{H}_n(t \mid z) - H(t \mid z) =$$

$$\int\{H(t \mid G^{-1}(G(z) - ua_n)) - H(t \mid G^{-1}(G(z)))\}\, K(u)\,du$$

$$= 1/2\, a_n^2\, H''(t \mid G^{-1}(G(z)))\int u^2 K(u)\,du + o(a_n^2) = O(a_n^2)$$

uniformly in t (see Stute (1984a)).

We shall consider now weak convergence results. We shall need

A3. K is continuously twice differentiable and $na_n^3 \to \infty$.

A4. The sequence a_n satisfies $na_n^5 \to 0$.

Let $W_n(t \mid z)$ be given by (2.1) with $b_n = (na_n)^{1/2}$ and let $\hat{W}_{in}(t \mid z) = (\hat{W}_{in}(t \mid z), \hat{W}_{2n}(t \mid z))$ where $\hat{W}_{in}(t \mid z) = (na_n)^{1/2}(H_{in}(t \mid z) - \hat{H}_{in}(t \mid z))$.

Proposition 3.2. *If A3 holds then for μ-almost all z $\hat{W}_n(t \mid z)$ converges weakly in $D[0, \tau(z)] \times D[0, \tau(z)]$ to a two-dimensional mean zero Gaussian process $W(t \mid z) = (W_1(t \mid z), W_2(t \mid z))$ with covariance function given by (2.2) with $\phi(z) \equiv \alpha(K)$. If in addition A2 and A4 are satisfied, then for μ-almost all z, the process $W_n(t \mid z)$ converges weakly to $W(t \mid z)$.*

The second part of the proposition deals with the bias term and follows from the remark after Proposition 2.1. The proof of the first part is deferred to Section 4.2. For uncensored data but multivariate response variables, Stute (1986a) obtained an analogous result. His proof of tightness rests on some bounds for the oscillation modulus of multivariate empirical processes.

3.2. Kernel estimates.

We turn now to the Nadaraya - Watson kernel estimates with weights (3.2). Beran (1981) considered the question of uniform consistency of $H_{in}(t \mid z)$; the estimates are consistent if $\log n = o(a_n n)$.

To derive weak convergence results, define functions $H_i(t, z) = g(z) H_i(t \mid z)$. The following conditions will be needed.

B1. The sequence a_n satisfies $na_n \to \infty$.

B2. K is a mean zero density and $na_n^5 \to 0$. Given z, the functions $g(u)$ and $H_i(t, u)$ are continuously twice differentiable in u, where u belongs to a neighbourhood $U(z)$ of z. Moreover

$$\sup_{U(z)} \sup_t |H_i''(t, u)| < \infty \quad \sup_{U(z)} |g''(u)| < \infty.$$

Let $W_n(t \mid z)$ be given by (2.1) with $b_n = (na_n)^{1/2}$. Introduce processes $\hat{W}_n(t \mid z) = (\hat{W}_{1n}(t \mid z), \hat{W}_{2n}(t \mid z))$ where $\hat{W}_{in}(t \mid z) = (na_n)^{1/2}(H_{in}(t \mid z) - \hat{H}_{in}(t \mid z))$ and

$$\hat{H}_{in}(t \mid z) = a_n^{-1}\int H_i(t, u)K\left(\frac{z-u}{a_n}\right) du \Big/ \hat{g}_n(u)$$

$$\hat{g}_n(u) = a_n^{-1}\int K\left(\frac{z-u}{a_n}\right) g(u) du. \tag{3.5}$$

Proposition 3.3. *If B1 holds then for μ-almost all z such that $g(z) > 0$, the process $\hat{W}_n(t \mid z)$ converges weakly in $D[0, \tau(z)] \times D[0, \tau(z)]$ to $W(t \mid z) = (W_1(t \mid z), W_2(t \mid z))$, a two dimensional mean zero Gaussian process with covariance function given by (2.2) where $\phi(z) = g^{-1}(z) \alpha(K)$. If in addition B2 is satisfied then. for μ-almost all z, the process $W_n(t \mid z)$ converges weakly to $W(t \mid z)$.*

The proof of the first part of the proposition is given in Section 5. The second part deals with the bias term. We have $EH_n(t \mid z) = \hat{H}_n(t \mid z) + O(n^{-1}a_n^{-1})$ uniformly in t. Under assumption B2, by Taylor expansion

$$\hat{H}_n(t \mid z) - H(t \mid z) = 1/2 a_n^2 g^{-1}(z)H''(t,z) \int u^2 K(u)du$$

$$- 1/2 \, a_n^2 g^{-1}(z)H(t \mid z) = 1/2 \, a_n^2 g^{-1}(z) \int u^2 K(u)du + o(a_n^2) = O(a_n^2)$$

uniformly in t. The condition $na_n^5 \rightarrow 0$ ensures asymptotic unbiasedness of $W_{in}(t \mid z)$.

Since the covariance function depends on the density $g(z)$, in practice we have to estimate it from the data. Parzen's (1962) estimate

$$g_n(z) = (na_n)^{-1} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{a_n}\right) \tag{3.6}$$

is a natural choice, but other strongly consistent estimates can be used as well.

3.3. Asymmetric nearest neighbour estimates.

Our final example is the m_n - nearest neighbour estimate with weights (3.3). Beran (1981) considered the problem of uniform consistency of $H_{in}(t \mid z)$. The estimates are consistent if the sequence m_n satisfies $m_n \rightarrow \infty$, $m_n = o(n)$ and $\log n = o(m_n)$.

To derive weak convergence results, the following conditions will be required.

C1. The sequence m_n satisfies $m_n \rightarrow \infty$ and $m_n = o(n)$. The functions $g(u)$ and $H(t \mid u)$ are continuously differentiable in u in a neighbourhood $U(z)$ of z and

$$\sup_{U(z)} \sup_{t} |H'_{;}(t \mid u)| < \infty \quad \sup_{U(z)} |g'(u)| < \infty.$$

Moreover, $P(|Z - z| > s) = O(s^{-\eta})$ for some $\eta > 0$ as $s \to \infty$.

C2. K is a mean zero density and $m_n^5 n^{-4} \to 0$. The differentiability conditions of C1 are satisfied by the second derivatives.

Let $W_n(t \mid z)$ be given by (2.1) with $b_n = (m_n - 1)^{1/2}$. Define processes $\hat{W}_n(t \mid z) = (\hat{W}_{1n}(t \mid z), \hat{W}_{2n}(t \mid z))$ where $\hat{W}_{in}(t \mid z) =$

$(m_n - 1)^{1/2}(H_{in}(t \mid z) - EH_{in}(t \mid z))$.

Proposition 3.4.

If C1 holds then for z such that $g(z) > 0$, the process $\hat{W}_n(t \mid z)$ converges weakly in $D[0, \tau(z)] \times D[0, \tau(z)]$ to $W(t \mid z) = (W_1(t \mid z), W_2(t \mid z))$, a two dimensional mean zero Gaussian process with covariance function given by (2.2) where $\phi(z) = 2\alpha(K)$. If in addition C2 is satisfied then for z such that $g(z) > 0$, the process $W_n(t \mid z)$ converges weakly to $W(t \mid z)$.

The first part of the proposition is proved in Section 6. The second part deals with the bias term. From Mack (1981), under differentiability conditions of C2, we have

$EH_n(t \mid z) - H(t \mid z) =$
$1/8 \, m_n^2 n^{-2} g(z)^{-3} \{H(t \mid z)g(z)\}'' \int u^2 K(u)du + o(m_n^2 n^{-2}) + o(m_n^{-1})$
uniformly in t. Therefore $m_n^5 n^{-4} \to 0$ implies asymptotic unbiasedness of $W_{in}(t \mid z)$.

3.4. Extensions to the multivariate case.

In practice one usually has to deal with multivariate covariates. Uniform consistency of kernel estimates was studied by Beran (1981) and Stute (1986b), weak

convergence can be established by applying results of Rosenblatt (1969). Uniform consistency and weak convergence of smooth nearest neighbour estimates can be derived from results of Stute (1983). For the asymmetric nearest neighbours, the proof of the weak convergence result is essentially the same as in Section 6. In all cases the required smoothness conditions are analogous to those in assumptions A, B and C, and the covariance structure of the limiting processes has the same form as in Propositions 3.2 - 3.4.

4. Proofs of Propositions 3.1 and 3.2.

4.1. Proof of Proposition 2.1.

We shall verify that for any fixed t and μ-almost all z, $H_{in}(t \pm |z) - \hat{H}_{in}(t \pm |z) \to 0$ a.s. Since $H_{in}(t \pm |z)$ and $\hat{H}_{in}(t \pm |z)$ are bounded monotone functions, we then also have uniform consistency.

We consider $H_{in}(t \mid z)$ only, the proof for $H_{in}(t - \mid z)$ is analogous. Fix t. Let $M_1(y,u) = P(Y \le y, \delta = 1, Z \le u)$, $M_2(y,u) = P(Y \le y, Z \le u)$ and let $M_{1n}(y,u)$ and $M_{2n}(y,u)$ be the respective empiricals. The mean value theorem implies $H_{in}(t \mid z) = I + II$ where

$$I = a_n^{-1} \int I(y > t) K\left[\frac{G(z) - G(u)}{a_n} \right] dM_{in}(y, u)$$

$$II = a_n^{-2} \int I(y > t)(G_n(z) - G_n(u) - G(z) + G(u)) K'(\Delta_n) dM_{in}(y, u).$$

Here Δ_n is a random function assuming values between $a_n^{-1}[G_n(z) - G_n(u)]$ and $a_n^{-1}[G(z) - G(u)]$. Since K has compact support, say $[-1, 1]$, the above expansion is valid for integration restricted to u's for which $|G_n(z) - G_n(u)| \le a_n$.

For μ-almost all z, $I - \hat{H}_{in}(t \mid z) \to 0$ a.s., by Hoeffding's (1963) inequality and A.1. To handle the second term, set $V_n = a_n^{-2}\sup|G_n(z) - G_n(u) - G(z) - G(u)|$ where the supremum is restricted to those u's for which $|G_n(z) - G_n(u)| \le a_n$. We have

$$|II| \le V_n \int|K'(\Delta_n)|dM_{in}(y,u).$$

Since K' is bounded, it is enough to show that $V_n \to 0$ a.s.

Fix $\epsilon > 0$ and let $0 < \eta < 1/8$ be arbitrary. Then for n sufficiently large, $a_n < \eta/3$ and $|G_n(u) - G(u)| < \eta/3$ uniformly in u, with probability 1. It follows that for n large enough, $V_n \le n^{-1/2} a_n^{-2} \omega_n(\eta)$ with probability 1. Here $\omega_n(\eta) = n^{1/2}\sup|G_n(z) - G_n(u) - G(z) - G(u)|$ with sup taken over those u's for which $|G(z) - G(u)| < \eta$. By Lemma 2.4 in Stute (1982), $P(\omega_n(\eta) > n^{1/2} a_n^2\epsilon) \le v_1 \exp\{-v_2 n a_n^4\}$ for some v_1 and $v_2 = v_2(\epsilon)$. Assumption A1 and Borel - Cantelli theorem imply $V_n \to 0$ a.s. μ - a.e. This completes the proof.

4.2. Proof of Proposition 3.2.

Define $W_n^*(t \mid z) = (W_{1n}^*(t \mid z), W_{2n}^*(t \mid z))$ where for $i = 1, 2$

$$W_{in}^*(t \mid z) =$$
$$(na_n^{-1})^{1/2}\int(I(y > t) - H_i(t \mid z)) K\left[\frac{G(z) - G(u)}{a_n}\right] d\,(M_{in}(y,u) - M_i(y,u))$$

Here $M_1(y,u) = P(Y \le y, \delta = 1, Z \le u)$, $M_2(y,u) = P(Y \le y, Z \le u)$ and $M_{1n}(y,u)$ and $M_{2n}(y,u)$ are the respective empiricals. Results of Stute (1984a) imply that for any t, $W_{in}^*(t \mid z) - \dot{W}_{in}(t \mid z) \to 0$ in probability μ - a.e. It can be verified that this holds uniformly in t. Therefore, it is enough to

establish weak convergence of $W_{in}^*(t \mid z)$. We shall verify that for μ-almost all z, the finite dimensional distributions of $W_n^*(t \mid z)$ are asymptotically multivariate normal and the process $W_n^*(t \mid z)$ is tight.

Choose $t_1 < \cdots < t_r < \tau(z)$ and let c_{ip}, $p = 1, \cdots, r$, $i = 1, 2$, satisfy $\sum_i \sum_p c_{ip}^2 \neq 0$. Then $n^{1/2} \sum_i \sum_p c_{ip} W_{in}^*(t_p \mid z)$ is a sum of centered independent rv's whose third absolute moment is of order $O(a_n^{-1/2})$. Further,

$$\sigma_n^2 = Var(\sum_i \sum_p c_{ip} W_{in}^*(t_p \mid z)) =$$

$$\sum_{i,j} \sum_{p,q} c_{ip} c_{jq} \{H_{min(i,j)}(max(t_p, t_q) \mid z) - H_i(t_p \mid z)H_j(t_q \mid z)\}\alpha(K) + o(1)$$

and the variance is bounded away from zero. Berry - Esseen Theorem and Cramer - Wold device complete the proof of the joint asymptotic normality of the finite dimensional distributions.

Further, for μ-almost all z and any $t_1 < t < t_2$ we have

$$E\{W_{in}^*(t \mid z) - W_{in}^*(t_1 \mid z)\}^2 \{W_{in}^*(t_2 \mid z) - W_{in}^*(t \mid z)\}^2 \tag{4.1}$$
$$\leq 3\{H_i(t_1 \mid z) - H_i(t \mid z)\} \{H_i(t \mid z) - H_i(t_2 \mid z)\} \alpha^2(K) + r_n$$

where $r_n \to 0$ Theorem 15.6 in Billingsley (1968), implies that for μ-almost all z the process $W_{in}^*(t \mid z)$ is tight. To verify this inequality, set $W_{in}^*(t_1 \mid z) - W_{in}^*(t \mid z) = (na_n)^{-1/2} \Sigma A_j$ and $W_{in}^*(t \mid z) - W_{in}^*(t_2 \mid z) = (na_n)^{-1/2} \Sigma B_j$. Then the left hand side of (4.1) is equal to $n^{-1}a_n^{-2}\{EA_1^2 B_1^2 + (n-1) EA_1^2 EB_1^2 + 2(n-1)(EA_1B_1)^2\}$. Repeated application of Theorem 3 in Devroye and Györfi (1985, p. 8) yields $n^{-1}a_n^{-2}EA_1^2B_1^2 = O(n^{-1}a_n^{-1})$. $n^{-1}(n-1)a_n^{-2}(EA_1B_1)^2 = p^2q^2\alpha^2(K) + o(1)$ and $n^{-1}(n-1)a_n^{-2}EA_1^2EB_1^2 = p(1-p)q(1-q)\alpha^2(K) + o(1)$. Here

$p = H_i(t_1 \mid z) - H_i(t \mid z)$ and $q = H_i(t \mid z) - H_i(t_2 \mid z)$. The bound on the right hand side of (4.1) follows.

5. Proof of Proposition 3.3.

Define $H_{in}(t, z) = g_n(z) H_{in}(t \mid z)$ and $\hat{H}_{in}(t, z) = \hat{g}_n(z) \hat{H}_{in}(t \mid z)$, where $g_n(z)$ and $\hat{g}_n(z)$ are given by (3.6) and (3.5), respectively. For $i = 1, 2$ set $\hat{W}_{in}(t, z) = (na_n)^{1/2}(H_{in}(t, z) - \hat{H}_{in}(t, z))$ and let $W_{3n}(z) = (na_n)^{1/2}(g_n(z) - \hat{g}_n(z))$.

Let $t_1 < \cdots < t_r$ and let c_{ip}, $i = 1, \cdots, r$, $i = 1, 2$ satisfy $\sum_i \sum_p c_{ip}^2 \neq 0$. Then $n^{1/2} \sum_i \sum_p c_{ip} \hat{W}_{in}(tp \mid z)$ and $n^{1/2} W_{3n}(z)$ are sums of centered independent rv's whose third absolute moment is of order $O(a_n^{-1/2})$. Further, by Theorem 3 in Devroye and Györfi (1985, p.8),

$$Var(\sum_i \sum_p c_{ip} \hat{W}_{ni}(t_p \mid z)) = \sum_{ij} \sum_{pq} H_{min(i,j)}(max(t_p, t_q) \mid z) \alpha(K) + o(1)$$
$$cov(\sum_i \sum_p c_{ip} \hat{W}_{in}(t_p \mid z), W_{3n}(z)) = \sum_i \sum_p c_{ip} H_i(t_p, z)\alpha(K) + o(1)$$
$$Var(W_{3n}(z)) = g(z)\alpha(K) + o(1).$$

Berry - Esseen Theorem and Cramer-Wold device yield joint asymptotic normality of $\hat{W}_{in}(t_p, z)$ and $W_{3n}(z)$, $i = 1, 2$ $p = 1 \cdots r$. It follows that the finite dimensional distributions of $\hat{W}_{in}(t \mid z)$ are asymptotically normal, and a straightforward calculation gives the desired covariance.

It remains to verify tightness. It is enough to consider the processes $\hat{W}_{in}(t, z)$. For μ-almost all z and any $t_1 < t < t_2$ we have

$$E\{\hat{W}_{in}(t_1, z) - \hat{W}_{in}(t, z)\}^2 \{\hat{W}_{in}(t, z) - \hat{W}_{in}(t_2, z)\}^2 \qquad (5.1)$$
$$\leq 3 \{H_i(t_1, z) - H_i(t, z)\} \{H_i(t, z) - H_i(t_2, z)\} \alpha^2(K) + r_n,$$

where $r_n \to 0$. Theorem 15.6 in Billingsley (1968), implies that for μ-almost all z, the process $\hat{W}_{in}(t, z)$ is tight. To verify this inequality, write

$$\hat{W}_{in}(t_1, z) - \hat{W}_{in}(t, z) = (na_n)^{-1/2} \Sigma A_j \text{ and}$$

$$\hat{W}_{in}(t, z) - \hat{W}_{in}(t_2, z) = (na_n)^{-1/2} \Sigma B_j.$$ Then the left-hand side of (5.1) is equal to $n^{-1} a_n^{-2} \{ EA_1^2 B_1^2 + (n-1)EA_1^2 EB_1^2 + 2(n-1)(EA_1 B_1)^2 \}$. Repeated application of Theorem 3 in Devroye and Györfi (1985, p.8) yields

$$n^{-1} a_n^{-2} EA_1^2 B_1^2 = O(n^{-1} a_n^{-1}), \quad n^{-1}(n-1) a_n^{-2} (EA_1 B_1)^2 = p^2 q^2 \alpha^2(K) + o(1)$$

and $n^{-1}(n-1) a_n^{-2} EA_1^2 EB_1^2 = p(1-p)q(1-q)\alpha^2(K) + o(1)$. Here

$p = H_i(t_1, z) - H_i(t, z)$ and $q = H_i(t, z) - H_i(t_2, z)$. The right hand side of (5.1) follows.

6. Proof of Proposition 3.4.

Theorem 3 in Mack (1981) and Cramer - Wold device imply weak convergence of finite dimensional distributions of $\hat{W}_n(t \mid z)$. Furthermore, by Proposition 6 in Mack (1981), for fixed t $(m_n - 1)^{1/2} (E(H_{in}(t \mid z) \mid R_n) - EH_{in}(t \mid z))$ converge in probability to zero μ - a.e. Under assumption C1 this convergence is in fact uniform in t. Therefore to complete the proof, it is enough to verify tightness of $W_n^*(t \mid z) = (W_{1n}^*(t \mid z), W_{2n}^*(t \mid z))$, where $W_{in}^*(t \mid z) = (m_n - 1)^{1/2}(H_{in}(t \mid z) - E(H_{in}(t \mid z) \mid R_n))$. We shall show that for μ - almost all z and $t_1 < t < t_2$

$$E \{ W_{in}^*(t_1 \mid z) - W_{in}^*(t \mid z) \}^2 \{ W_{in}^*(t \mid z) - W_{in}^*(t_2 \mid z) \}^2 \tag{6.1}$$
$$\leq 6\{ H_i(t_1 \mid z) - H_i(t \mid z) \} \{ H_i(t \mid z) - H_i(t_2 \mid z) \} \alpha^2(K) + r_n$$

where $r_n \to 0$. Theorem 15.6 in Billingsley (1968) implies tightness.

We shall check this inequality for $i = 1$. Define $H_1(t_1, t, z) = (H_1(t_1 \mid z) - H_1(t \mid z))g(z)$ and $H_{1n}(t_1, t, z) = (H_{1n}(t_1 \mid z) - H_{1n}(t \mid z))g_n(z)$, where $g_n(z)$ is the nearest neighbour density estimate given by

$$g_n(z) = (nR_n)^{-1} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{R_n}\right).$$

We argue by conditioning on R_n. Since Z has a continuous distribution, almost surely all Z_j's are distinct. Given $R_n = r$, let $I_m(z) = \{j : |Z_j - z| \leq r\}$. Then (Y_j, J_j, Z_j), $j \in I_m(z)$ are conditionally independent given $R_n = r$. Conditional on R_n, $W_{1n}^*(t_1 \mid z) - W_{1n}^*(t \mid z) = \Sigma A_j$ and $W_{1n}^*(t \mid z) - W_{1n}^*(t_2 \mid z) = \Sigma B_j$, where the summation extends over $j \in I_m(z)$ and

$$A_j = \frac{m_n - 1}{nR_n} \left[\frac{I(t_1 < Y_j < t, J_j = 1)}{c_n} - \frac{a_n}{c_n^2}\right] K\left(\frac{z - Z_j}{R_n}\right)$$

$$B_j = \frac{m_n - 1}{nR_n} \left[\frac{I(t < Y_j < t_2, J_j = 1)}{c_n} - \frac{b_n}{c_n^2}\right] K\left(\frac{z - Z_j}{R_n}\right).$$

Here $a_n = E(H_{1n}(t_1, t, z) \mid R_n)$, $b_n = E(H_{1n}(t, t_2, z) \mid R_n)$ and $c_n = E(g_n(z) \mid R_n)$. With this in mind, the left hand side of 6.1 is equal to

$(m_n - 1)^{-1}E\{E(A_j^2 B_j^2 \mid R_n) + 2(m_n - 2)(E(A_j B_j \mid R_n))^2 + (m_n - 2)E(A_j^2 \mid R_n)E(B_j^2 \mid R_n)\}$. It can be easily verfied that $(m_n - 1)^{-1}EA_j^2 B_j^2 = O(m_n^{-1})$. Further, after some algebra,

$$E(A_j B_j \mid R_n) = (m_n - 1)\{a_n b_n f_n c_n^{-4} - a_n e_n c_n^{-3} - b_n d_n c_n^{-3}\}$$

$$E(A_j^2 \mid R_n) = (m_n - 1)\{a_n^2 f_n c_n^{-4} + d_n c_n^{-2} - 2a_n f_n c_n^{-3}\}$$

$$E(B_j{}^2 \mid R_n) = (m_n - 1)\{b_n{}^2 f_n \, c_n{}^{-4} + e_n \, c_n{}^{-2} - 2b_n \, f_n \, c_n{}^{-3}\}$$

where $\quad d_n = E(H_{1n}^2(t_1, t, z) \mid R_n), \quad e_n = E(H_{1n}^2(t, t_2, z) \mid R_n) \quad$ and

$f_n = E(g_n{}^2(z) \mid R_n)$.

Let X be the m_n-th order statistic from an ordered sample of size n from a

uniform distribution on $(0,1)$. If Ψ is the distribution of $|z - Z_j|$, then

$R_n = \Psi^{-1}(X)$. We have

$$a_n = \frac{m_n - 1}{nX} \int K(u) H_1(t_1, t, z - uR_n) du$$

$$b_n = \frac{m_n - 1}{nX} \int K(u) H_1(t, t_2, z - uR_n) du$$

$$c_n = \frac{m_n - 1}{nX} \int K(u) g(z - uR_n) du$$

$$d_n = \frac{m_n - 1}{n^2 R_n X} \int K^2(u) H_1(t_1, t, z - uR_n) du$$

$$e_n = \frac{m_n - 1}{n^2 R_n X} \int K^2(u) H_1(t, t_2, z - uR_n) du$$

$$f_n = \frac{m_n - 1}{n^2 R_n X} \int K^2(u) g(z - uR_n) du$$

Using these expressions, we find after lengthy calculation

$(m_n - 2)(m_n - 1)^{-1} E(E(A_j B_j \mid R_n))^2 = 2p^2 q^2 \alpha^2(K) + O(m_n^{-1}) + O(m_n n^{-1})$

and $(m_n - 1)^{-1}(m_n - 2) E(E(A_j{}^2 \mid R_n) E(B_j{}^2 \mid R_n)) =$

$2pq(1 - p)(1 - q)\alpha^2(K) + O(m_n^{-1}) + O(m_n n^{-1})$, where

$p = H_1(t_1 \mid z) - H_1(t \mid z)$ and $q = H_1(t \mid z) - H_1(t_2 \mid z)$. In computing these

expectations, we use a Taylor expansion of $g(z - uR_n)$ and $H_1(t_1, t, z - uR_n)$

around $z - u \Psi^{-1}(m/(n+1))$ and $s = \Psi^{-1}(t) = g(x)^{-1} t + o(t)$ as $s \downarrow o$, and

follow developments in Mack and Rosenblatt (1979) and Mack (1981). This com-

pletes the proof.

7. Proof of Proposition 2.3 and Corollary 2.2.

7.1. Proof of Proposition 2.3.

We sketch the proof; the details are analogous to Breslow and Crowley (1974). Unless stated otherwise, all suprema are taken over $[0, \tau(z)]$.

Suppose $0 \leq t \leq \tau(z)$. Then $H_2(t \mid z) > 0$ and with probability 1, $H_{2n}(t \mid z) > 0$ for μ-almost all z and n sufficiently large. By appealing to the Skorokhod construction, we can assume without loss of generality that for μ-almost all z, the sample paths of $W(t \mid z)$ are continuous and $W_n(t \mid z)$ converges almost surely to $W(t \mid z)$ in the supremum metric on $[0, \tau(z)] \times [0, \tau(z)]$. We have

$$L_n(t \mid z) = \int_0^t \frac{W_{2n}(s - \mid z) d \Lambda(s \mid z)}{H_2(s - \mid z)} - \int_0^t \frac{dW_{1n}(s \mid z)}{H_2(s - \mid z)} + R_n(z)$$

where $R_n(z)$ is a remainder term. By the assumed weak convergence of $W_n(t \mid z)$, the first two terms converge almost surely in the sup norm on $[0, \tau(z)]$ to

$$L(t \mid z) = \int_0^t \frac{W_2(s \mid z) d \Lambda(s \mid z)}{H_2(s - \mid z)} - \int_0^t \frac{dW_1(s \mid z)}{H_2(s - \mid z)}$$

which is a mean zero Gaussian process. Integration by parts, analogous to Breslow and Crowley (1974), shows that the covariance function of this processes is given by (1.4). The remainder term can be treated in a similar way as in Breslow and Crowley (1974) (see Gill (1983) for a necessary correction). Further, by Taylor expansion, $S_n(t \mid z) \cong -F(t \mid z) L_n(t \mid z) + R_{1n}(t,z) + R_{2n}(t,z)$, where $R_{1n}(t,z) = b_n^{-1} L_n^2(t \mid z) exp(\Lambda_n^*(t \mid z))$ and $R_{2n}(t,z) =$

$b_n \left(F_n \left(t \mid z \right) - exp \left(-\Lambda_n \left(t \mid z \right) \right) \right)$. Here $\Lambda_n^*(t \mid z)$ is a random function assuming values between $\Lambda(t \mid z)$ and $\Lambda_n \left(t \mid z \right)$. By weak convergence of $L_n \left(t \mid z \right)$ and consistency of $\Lambda_n \left(t \mid z \right)$, $\sup |R_{1n} \left(t, z \right)| \to 0$ in probability μ - a.e. Further, for μ-almost all z and n sufficiently large $H_{2n} \left(t \mid z \right) > 0$, $F_n \left(t \mid z \right) > 0$ and $exp\{-\Lambda_n \left(t \mid z \right)\} > 0$ for $t \in [0, \tau(z)]$. After some easy algebra

$$|R_{2n} \left(t, z \right)| \leq b_n \left| \Sigma \left\{ \log \left(1 - \Delta \Lambda_n \left(s \mid z \right) \right) + \Delta \dot{\Lambda}_n \left(s \mid z \right) \right\} \right|$$
$$\leq b_n \; \Sigma \; H_{2n}^{-1}(s \mid z) H_{2n}^{-1}(s - \mid z) \Delta H_{1n}^2(s \mid z)$$
$$\leq b_n \; \sup |\Delta H_{1n} \left(t \mid z \right)| \int_0^t H_{2n}^{-1}(s \mid z) d\Lambda_n \left(s \mid z \right),$$

where the sum is over $s \leq t$. Weak convergence of $H_{1n} \left(t \mid z \right)$ and consistency of $H_{2n} \left(t \mid z \right)$ and $\Lambda_n \left(s \mid z \right)$ entail $\sup |R_{2n} \left(t, z \right)| \to 0$ in probability μ-a.e.

7.2. Proof of Corollary 2.2.

(i). Suppose that z does not lie in the exceptional μ - null set. For $\epsilon(z) > 0$ arbitrary but small enough there exist points $\tau_a(z)$ and $\tau_b(z)$ such that

$$0 < \tau_a(z) < Q \left(a \mid z \right) - \epsilon(z) < Q \left(b \mid z \right) + \epsilon(z) < \tau_b(z) <$$

$\sup\{t : H_2(t \mid z) > 0\}$. Under condition I, $Q(p \mid z)$, $p \in [a, b]$, is the unique solution of $F(t \mid z) \leq 1 - p \leq F(t - \mid z)$ and, by the strong consistency of $F_n \left(t \mid z \right)$, for n sufficiently large $\tau_a(z) < Q_n \left(a \mid z \right) < Q_n \left(b \mid z \right) < \tau_b(z)$ with probability 1.

Further, there exist processes $\hat{S} \left(t \mid z \right)$ and $\hat{S}_n \left(t \mid z \right)$ defined on a common probability space with the same distributions as $S \left(t \mid z \right)$ and $S_n \left(t \mid z \right)$ and such that the sample paths of $\hat{S} \left(t \mid z \right)$ are continuous and $\hat{S}_n \left(t \mid z \right)$ converges to $\hat{S} \left(t \mid z \right)$ almost surely in the supremum norm on $[0, \tau_b(z)]$. Set

$\hat{F}_n(t \mid z) = b_n^{-1}\hat{S}_n(t \mid z) + F(t \mid z)$, $\hat{Q}_n(p \mid z) =$

$\inf\{t : \hat{F}_n(t \mid z) \leq 1 - p\}$ and let $\hat{U}_n(p \mid z)$ and $\hat{U}(p \mid z)$ be defined analogously to $U_n(p \mid z)$ and $U(p \mid z)$. Then, for n sufficiently large $\tau_a(z) < \hat{Q}_n(a \mid z) < \hat{Q}_n(b \mid z) < \tau_b(z)$ with probability 1 and the assertion will follow if we show that $\hat{U}_n(p \mid z) \to \hat{U}(p \mid z)$ a.s. in the sup norm on $D[a,b]$.

We can write

$$\hat{U}_n(p \mid z) = \frac{-\hat{V}_n(p \mid z)}{f \circ Q(p \mid z)} \tag{7.1}$$
$$+ \hat{V}_n(p \mid z)\left[\frac{\hat{U}_n(p \mid z)}{\hat{V}_n(p \mid z)} + \frac{1}{f \circ Q(p \mid z)}\right]$$

where $\hat{V}_n(p \mid z) = -b_n(F \circ \hat{Q}_n(p \mid z) - 1 + p)$. By the uniform continuity of $f \circ Q(p \mid z)^{-1}$, it is enough to show that $\hat{V}_n(p \mid z)$ converges almost surely to $\hat{V}_n(p \mid z) = \hat{S} \circ \hat{Q}(p \mid z)$ and the second term in (7.1) is asymptotically negligible.

Adding and subtracting terms

$$\sup_p |\hat{V}_n(p \mid z) - \hat{V}(p \mid z)| \leq \sup_p |\hat{S}_n \circ \hat{Q}_n(p \mid z) - \hat{S} \circ \hat{Q}_n(p \mid z)|$$
$$+ \sup_p b_n |\hat{F}_n \circ \hat{Q}_n(p \mid z) - 1 + p| + \sup_p |\hat{S} \circ \hat{Q}_n(p \mid z) - \hat{S} \circ \hat{Q}(p \mid z)| \tag{7.2}$$

where the suprema are taken over $p \in [a,b]$. The first term is bounded above by $\sup_t |\hat{S}_n(t \mid z) - \hat{S}(t \mid z)| \to 0$ a.s. with sup taken over $[\tau_a(z), \tau_b(z)]$. Further,

$$\sup_p |F \circ Q_n(p \mid z) - 1 + p| \leq \sup_p |F_n \circ Q_n(p \mid z) - F \circ Q_n(p \mid z)|$$
$$+ \sup_p |F_n \circ Q_n(p \mid z) - 1 + p| \leq \sup_t |F_n(t \mid z) - F(t \mid z)| + R_n(z) \tag{7.3}$$

where $R_n(z) = H_{2n}^{-1}(\tau_b \mid z)\sup_t |\Delta H_{1n}(t \mid z)|$ and the suprema are over $p \in [a,b]$ and $t \in [\tau_a(z), \tau_b(z)]$. The consistency of $H_{in}(t \mid z)$ and $F_n(t \mid z)$

entails that this bound converges in probability to 0. Furthermore, $\sup_p b_n |F_n \circ Q_n(p \mid z) - 1 + p| \leq \sup_t b_n |\Delta F_n(t \mid z)| \leq b_n R_n(z) \to 0$ in probability μ- a.e. This remains valid for the processes $F \circ \hat{Q}_n(p \mid z)$ and $\hat{F}_n \circ \hat{Q}_n(p \mid z)$, except that the convergence is almost sure. It follows that $\hat{F}_n \circ \hat{Q}_n(p \mid z)$ is strongly consistent and the second term in (7.2) converges almost surely to 0 μ-a.e.

Further, for $p \in [a, b]$, we have $\hat{S} \circ \hat{Q}_n(p \mid z) = \hat{S} \circ Q(p' \mid z)$ where $p' = 1 - F \circ \hat{Q}_n(p \mid z)$. The uniform continuity of the sample paths of $S \circ Q(p \mid z)$ and strong consistency of $F \circ \hat{Q}_n(p \mid z)$ implies therefore that the third term in (7.2) converges almost surely to 0 μ-a.e.

Finally, the second term in (7.1) converges almost surely in sup norm on $[a, b]$ to 0 by almost sure boundedness of $\hat{V}_n(p \mid z)$ and (7.2).

(ii). To prove the second part, suppose that z does not lie in the exceptional μ - null set. For $\epsilon(z) > 0$ arbitrary but small enough we can find $0 < \tau_a(z) < Q(a - \epsilon \mid z) < Q(\beta + \epsilon \mid z) < \tau_b(z) < \sup\{t : H_2(t \mid z) > 0\}$. By the strong consistency of $F_n(t \mid z)$, for n sufficiently large,

$$b_n(l_n(z) - l(z)) = -\int S_n(t \mid z) J(F(t \mid z)) dt + \sum_{j=1}^{m} c_j Q(p_j \mid z)$$
$$+ \int S_n(t \mid z) \left[\frac{J(F(t \mid z) - J(F(t \mid z))}{S_n(t \mid z)} - J(F(t \mid z)) \right] dt$$

where the integrals are taken over $[\tau_a(z), \tau_b(z)]$. Replacing processes $S_n(t \mid z)$, $U_n(p \mid z)$ and $F_n(t \mid z)$ by the processes constructed in part (i), it follows that the first two terms converge to a mean zero normal distribution. The third term is asymptotically negligible, by the uniform continuity of $J(u)$, dominated con-

vergence theorem and almost sure convergence of $\hat{S}_n(t \mid z)$ in the sup norm on $[\tau_a(z), \tau_b(z)]$.

Acknowledgment

References

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

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Springer Lect. Notes in Statist.* **2**, 1-25.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Aly, E. E., Csörgo, M. & Horváth, L. (1985). Strong approximations of the quantile process of the product - limit estimator. *J. Multivariate Analysis* **16**, 185-210.

Begun, J. M., Hall, W. J., Huang, W. M. & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric - nonparametric models. *Ann. Statist.* **11**, 432-452.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Tech. Rep., University of California, Berkeley.

Bickel, P. & Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. I - II. *Ann. Statist.* **3**, 1038-1045, 1045-1069.

Billingsley (1968). *Weak convergence of probability measures.* Wiley, New York.

Breslow, N. & Crowley, J. J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437-443.

Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Campbell, G. & Földes, A. (1982). Large sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric statistical inference, Colloquia Mathematica Societatis Janos Bolyai* (Gnedenko, B. V., Puri, M. L. & Vincze, I., Eds.), North Holland, Amsterdam.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-202.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Csörgö, M. (1983). Quantile processes with statistical applications. CBMS - NSF *Conference Series in Appl. Math.* **42**, S.I.A.M., Philadelphia.

Devroye, L. & Györfi, L. (1985). *Nonparametric density estimation: the L_1 view.* Wiley, New York.

Doksum, K. A. & Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. In *Festschrift for Erich L. Lehmann* (Bickel, P. J., Doksum, K. A., Hodges, J. L., Eds.), Wadsworth, 140-156.

Friedman, J. H. and Stützle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 599-608.

Gill, R. D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts 124, Amsterdam.

Gill, R. D. (1983). Large sample behaviour of the product - limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statist. Sc.* **1**, 297-318.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Horváth, L. (1981). On nonparametric regression with randomly censored data. In *Proc. of the Third Pannonian Symp. on Math. Statist., Visegrad, Hungary* (Mogyorodi, J., Vincze, I. & Wertz, W., Eds.) 105-112.

Huber, P. J. (1981). *Robust statistics.* Wiley, New York.

Huber, P. J. (1985). Projection pursuit. *Ann. Statist.* **13**, 435-525.

Johnson, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Analysis* **12**, 402-414.

Kalbfleisch & Prentice (1980). *The statistical analysis of failure time data.* Wiley, New York.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Koul, H., Susarla, V., and van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276-1288.

Leurgans, S. (1984). Linear models, random censoring and synthetic data. Tech. Report, University of Washington, Seattle.

Lo. S. & Singh, K. (1986). The product limit estimator and the bootstrap: some asymptotic representations. *Probability Theor. and Related Fields* **71**, 455-466.

Mack, Y. P. (1981). Local properties of k - NN regression estimates. *SIAM J. Alg. Disc. Meth.* **2**, 311-323.

Mack, Y. P. & Rosenblatt, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Mult. Analysis* **9**, 1-15.

Mallows, C. L. (1979). Some theoretical results on Tukey's 3R smoothers. *Lect. Notes in Math.* **757**, 77-90.

Mallows C. L. (1980). Some theory of nonlinear smoothers. *Ann. Statist.*

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-464.

Miller, R. G. & Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521-532.

Nadaraya, E. A. (1964). On estimating regression. *Theor. Probability Appl.* **9**, 141-142.

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

O'Sullivan, F. (1986). Nonparametric estimation in the Cox proportional hazard model. Tech. Report No. 64, University of California, Berkeley.

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

Peterson, A. V. (1977). Expressing the Kaplan - Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72**, 854-858.

Reid, N. (1981). Influence functions for censored data. *Ann. Statist.* **9**, 78-92.

Rosenblatt, M. (1969). Conditional probability density and regression estimators. *Multivariate Analysis II*, 25-31, Academic Press, New York.

Sander, J. M. (1975 a). The weak convergence of quantiles of the product - limit estimator. Tech. Rep., Stanford University.

Sander, J. M. (1975 b). Asymptotic normality of linear combinations of functions of order statistics with censored data. Tech. Report, Stanford University.

Schuster, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43**, 84-88.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probability* **10**, 86-107.

Stute, W. (1984a). Asymptotic normality of nearest neighbour regression function estimates. *Ann. Statist.* **12**, 917-926.

Stute, W. (1984b). The oscillation behaviour of empirical processes: the multivariate case. *Ann. Probability* **12**, 361-379.

Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.* **14**, 638-647.

Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Probability* **14**, 891-901.

Stützle, W. and Mittal, Y. (1979). Some comments on the asymptotic behavior of robust smoothers. *Lect. Notes in Math.* **757**, 191-195.

Tsai, W. Y. & van Ryzin, J. (1985). On nonparametric regression for censored survival data. Unpublished.

Tibshirani, R. (1984). Local likelihood estimation. Unpublished Ph.D. Thesis, Stanford University.

Velleman, P. (1977). Robust nonlinear data smoothers: definitions and recommendations. *Natl. Acad. Sci. USA* **74**, 434-436.

Watson, G. S. (1984). Smooth regression analysis. *Sankhya Ser. A* **26**, 359-372.

Yang, S. S. (1981). Linear functions of concomitants of order statistics with applications to nonparametric estimation of regression function. *J. Amer.*

Statist. Assoc. **76**, 658-662.

Dorota M. Dabrowska

Department of Statistics

University of California

Berkeley, CA 94720