Sankhyā : The Indian Journal of Statistics
1993, Volume 55, Series A, Pt. 1, pp. 66-79.

ON THE HAJEK PROJECTION FOR TRUNCATED AND CENSORED DATA

By ÜLKÜ GÜRLER
University of Pennsylvania and Bilkent University
and
JANE-LING WANG*
University of California

SUMMARY. Large sample properties of the product-limit estimators for truncated or censored data are usually achieved via the empirical cumulative hazard function estimators. Hajek projection of the empirical cumulative hazard function estimator is derived for truncated data and expressed for censored data. It turns out that both projections are asymptotically $n^{\frac{1}{2}}$-equivalent but not equal to the respective influence curves. Weak convergences of the empirical cumulative hazard processes are deduced accordingly.

1. Introduction

In the collection of scientific data it often happens that one cannot observe completely the data of interest. Incomplete data may occur in various forms and we restrict our attention to two particular forms, censoring and truncation, in this paper.

Let X denote the time of occurrence of an event of interest, called the lifetime in standard survival analysis, with distribution function F. The observation of X is sometimes prevented by another independent variable Y, called censoring time or truncation time depending on the situation, with distribution function G. In the random (right) censoring model, the total number of items, n, under study is known in advance, and for the i-th item under study, one observes only the minimum of the failure time and the censoring time, along with an indicator of the censoring status. In the (left) truncation model the total number of items, N, under study is unknown and one observes only those pairs (X_i, Y_i) such that $Y_i \leqslant X_i$. The total number

Paper received. April 1989 ; revised December 1990.

AMS (1980) *subject classification.* 62G05, 62E20.

Key words and phrases. Censored data, truncated data, product-limit estimate, cummulative hazard function, Hajek projection.

*Research supported in part by Air Force grant AFOSR-85-0268. Part of this work was done while the author was at the Department of Statistics, University of Pennsylvania.

of observations n is a random quantity in cntrast to the censoring case where it is fixed in advance. Also, whenever $Y_i \leqslant X_i$, one observes both X_i and Y_i instead of just the minimum Z_i.

Let $\tilde{F}_n, \hat{F}_n$ denote respectively the product-limit estimator under the censoring and truncation model, described in (3.3) and (2.6), and derived by Kaplan and Meier (1958) and Lynden-Bell (1971). Due to the product form, the finite sample properties of $\tilde{F}_n$ and $\hat{F}_n$ are hard to grasp and large sample properties are usually achieved via the cumulative hazard function. The cumulative hazard function of a distribution function F (taken to be right continuous with $F(0^- = 0)$ is defined by

$$\Lambda(x) = \int_0^x dF(t)/[1-F(t-)],\ 0 \leqslant t < \infty, \qquad \dots \quad (1.1)$$

Note that (1.1) can be inverted so that the cumulative hazard function Λ uniquely determines the distribution F (cf. formula (3) of Woodroofe (1985) p. 166 noting that $\Lambda(t) = -\log(1-F(t))$ for continuous F).

The cumulative hazard function Λ can be estimated empirically using the representations (2.4) and (3.1). For censored data this estimator $\tilde{\Lambda}_n$ was first proposed by Nelson (1972) and is given in (3.2). For truncated data a derivation of the estimator $\hat{\Lambda}_n$ was illustrated in Woodroofe (1985) and given in (2.5). Although (2.5) and (3.2) imply that both $\hat{\Lambda}_n$ and $\tilde{\Lambda}_n$ take the form of a sum of identically distributed random variables, these variables are not independent. For continuous F and G and t in a compact interval, Breslow and Crowley (1974), Woodroofe (1985) and Wang, Jewell and Tsai (1986) decomposed $\tilde{\Lambda}_n(t)-\Lambda(t)$ and $\hat{\Lambda}_n(t)-\Lambda(t)$ into a mean of i.i.d. random variables plus a remainder of the order $o(n^{-1/2})$. The weak convergence of the cumulative hazard processes $n^{1/2}[\tilde{\Lambda}_n(t)-\Lambda(t)]$ and $n^{1/2}[\hat{\Lambda}_n(t)-\Lambda(t)]$ were then obtained from that of the mean processes, and the weak convergence of the product-limit processes $n^{1/2}[\tilde{F}_n(t)-\Lambda(t)]$ and $n^{1/2}[\hat{F}_n(t)-\Lambda(t)]$ follows from the inversion algorithm of (1.1).

The orders of the aforementioned remainder terms were further improved by Lo and Singh (1986), Burke *et al.* (1988), Major and Rajtö (1988) for the censored case, and by Chao and Lo (1988) for the truncation case. More precisely, let log denote the natural logarithm, then

$$\tilde{\Lambda}_n(t)-\Lambda(t) = n^{-1}\Sigma\, \eta(X_i, \delta_i, t)+R_n'(t) = \bar{\eta}(t)+R_n'(t), \qquad \dots \quad (1.2)$$

$$\hat{\Lambda}_n(t)-\Lambda(t) = n^{-1}\Sigma\, \xi(X_i, Y_i, t)+R_n(t) = \bar{\xi}(t)+R_n(t), \qquad \dots \quad (1.3)$$

where the means of both $\eta(X_i, \delta_i, t)$ and $\xi(X_i, Y_i, t)$ are zero and the supremum of $|R_n'(t)|$ and $|R_n(t)|$ on compact intervals are $o((\log n/n))$ a.s. and $o(n^{-1/2})$ a.s., respectively. The functions ξ and η are also the influence curves of $\tilde{\Lambda}_n$ and $\hat{\Lambda}_n$, respectively, as shown by Ried (1981) and Chao (1987).

A classical method to obtain asymptotic properties of a statistic is the projection method of Hejek (1968). An interesting question is whether the Hajek projection of the centered variables $\tilde{\Lambda}_n(t)-E(\tilde{\Lambda}_n(t))$ and $\hat{\Lambda}_n-E(\hat{\Lambda}_n(t))$ is $\bar{\eta}(t)$ and $\bar{\xi}(t)$, respectively, or not. If not, the question arises, what are the Hajek projections ?

The Hajek projection W' of $\tilde{\Lambda}_n$ can be derived following the approach of Tanner and Wang (1983) who calculated the Hajek projection of a kernel hazard rate estimate based on $\tilde{\Lambda}_n$. This was demonstrated in Gaenssler and Stute (1987). We show in Section 2 how to obtain the Hajeck projection V' of $\hat{\Lambda}_n$ for truncated data. It turns out that $\bar{\xi}(t)$ is not the Hejek projection of $\hat{\Lambda}_n-E(\hat{\Lambda}_n(t))$. However, as shown in Theorem 2, it is $n^{1/2}$-equivalent to the Hajek projection. Similar results are also available in Section 3 for $\tilde{\Lambda}_n$. As applications of the Hajek projection principle, we show in Section 4 how to derive the asymptotic normality and weak convergence of $\hat{\Lambda}_n(t)$ and $\tilde{\Lambda}_n(t)$.

2. Hajek projection of $\hat{\Lambda}_n(t)$ under truncation

Assume the truncation model in the previous section and adopt the notation in Woodroofe (1985). That is, $(X_1, Y_1), \ldots, (X_N, Y_N)$ are independent copies of (X, Y) with X_i and Y_i independent for each i, and F, G are the distribution functions of X and Y respectively. The observations are those pairs (X_i, Y_i) for which $i \leqslant N$ and $Y_i \leqslant X_i$. Assume that there is at least one such pair, and let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote these pairs. Then given n, the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are conditionally i.i.d. with joint distribution

$$H_*(x, y) = P(X \leqslant x, Y \leqslant y \mid Y \leqslant X) = \alpha^{-1} \int_0^x G(y \wedge z)\, dF(z) \quad \ldots \quad (2.1)$$

and marginal distributions $F_*(x) = H_*(x, \infty)$ and $G_*(y) = H_*(\infty, y)$, where $\alpha = \alpha(F, G) = P(Y \leqslant X) = \int G dF$ is assumed to be positive, $y \wedge z$ denotes the minimum of y and z.

Remark 1. All the probability statements in this section are conditional on $n = n_N = \#\{i \leqslant N : Y_i \leqslant X_i\}$. However, the conditional distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$ given n is the same as their unconditional distribution and n has a binomial $B(N, \alpha)$ distribution (cf. first paragraph on p. 170 of Woodroofe (1985)). Therefore, the large sample results for $n \to \infty$ hold for $N \to \infty$ as well.

For any distribution function $K(t)$ on $[0, \infty)$, let

$$a_K = \inf\{t : K(t) > 0\},\ b_K = \sup\{t : K(t) < 1\} \quad \ldots \quad (2.2)$$

be the endpoints of the support of K. As Woodroofe (1985) pointed out, one can estimate F and G only if they satisfy the identifiability condition that $(F, G) \in \mathcal{K}_0$, where $\mathcal{K}_0 = \{(F, G) : a_G \leqslant a_F,\ b_G \leqslant b_F,\ \alpha(F, G) > 0\}$. We will therefore assume that $(F, G) \in \mathcal{K}_0$. Let

$$C(t) = P(Y \leqslant t \leqslant X \mid Y \leqslant X) = G_*(t) - F_*(t-) = \alpha^{-1}G(t)\,[1-F(t-)]. \quad \ldots \quad (2.3)$$

Theorem 1 of Woodroofe (1985) gives the following representation of the cumulative hazard function Λ :

$$\Lambda(t) = \int_0^t dF_*(x)/C(x), \quad \ldots \quad (2.4)$$

Since F_* and G_* can be estimated empirically from $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, denoting their empirical distribution functions by F_n^* and G_n^*, the representation (2.4) suggests estimating Λ by

$$\hat{\Lambda}_n(t) = \int_0^t dF_n^*(x)/C_n(x) = \sum_{x_i < t} [nC_n(x_i)]^{-1}, \quad \ldots \quad (2.5)$$

where $C_n(t) = G_n^*(t) - F_n^*(t-)$. Using the algorithm (3) in Woodroofe (1985) and letting $r(x_i) = \#\{k \leqslant n : x_k = x_i\}$ for $1 \leqslant i \leqslant n$, the corresponding distribution function for $\hat{\Lambda}_n$ is thus

$$\hat{F}_n(t) = 1 - \prod_{x_i \leqslant t} \left[1 - \frac{r(x_i)}{nC_n(x_i)}\right] \quad \ldots \quad (2.6)$$

which is the Lynden-Bell (1971) estimate of $F(t)$.

We will show in this section how to derive the Hajek projection of $\hat{\Lambda}_n$. Let V be a statistic based on a sequence of i.i.d. random variables $U_1, \ldots, U_n$. Hajek (1968) showed that the projection V' of V onto the subspace $\mathcal{S} = \{\Sigma\phi(U_i)$, where ϕ is any real-valued function$\}$, is given by :

$$V' - E(V) = \Sigma\,[E(V \mid U_i) - E(V)] \quad \ldots \quad (2.7)$$

where $EV' = EV$, and for any S in $\mathcal{S}$,

$$E((V' - V)^2) = \text{var}\,(V) - \text{var}\,(V') \leqslant E((S - V)^2) \quad \ldots \quad (2.8)$$

Here and hereafter Σ means sum from 1 to n. Note that $V'-E(V)$ is the Hajek projection of $V-E(V)$. Recall from (2.5) that $\hat{\Lambda}_n(t) \equiv V(t) = \Sigma V_i(t)$ where $V_i(t) = 1(X_i \leqslant t)[nC_n(X_i)]^{-1}$, $i = 1, \ldots, n$, are identically distributed but not independent random variables. To compute the Hajek projection of $\hat{\Lambda}_n$ we first need the following lemma. For the rest of this section we will assume that F is a continuous distribution function and the expectations are the conditional expectations given n.

Lemma 1. *For $j \neq i$ and $t < b_F$,*

$$E(V_j(t) \mid X_i, Y_i) = (n-1)^{-1}\Big[\int_0^t 1-[1-C(x)]^{n-1}\, d\Lambda(x)$$

$$+ \int_0^t 1(Y_i \leqslant x \leqslant X_i)\, \{[1-C(x)]^{n-1}-[nC(x)]^{-1}[1-(1-C(x))^n]\} d\Lambda(x)\Big].$$

Proof. The proof is given in the Appendix.

Next consider,

$$\xi(x, y, t) = 1(x \leqslant t)\,[C(x)]^{-1} - \int_0^t 1(y \leqslant s \leqslant x)\,[C(x)]^{-1} d\Lambda(s). \quad \ldots \quad (2.9)$$

For any $b < b_F$, Chao and Lo (1988) showed that

$$\hat{\Lambda}_n(t)-\Lambda(t) = n^{-1}\Sigma\, \xi(X_i, Y_i, t)+R_n(t) = \bar{\xi}(t)+R_n(t),$$

where

$$\sup_{0 \leqslant t \leqslant b} |R_n(t)| = O((\log n/n)^{3/4}) \text{ a.s., if } a_G < a_F\,;$$

and

$$\sup_{0 \leqslant t \leqslant b} |R_n(t)| = o(n^{-1/2}) \text{ a.s. if } a_G = a_F$$

and

$$\lim_{x \to 0+} F(x)/G(x) = 0.$$

The function ξ is also the influence curve of $\hat{\Lambda}_n$ as shown by Chao (1987). We now express the Hajek projection V' of $\hat{\Lambda}_n$ in terms of the function ξ. Note from Lemma 2 of Woodroofe (1985) that

$$E(\hat{\Lambda}_n(t)) = \Lambda(t) - \int_0^t [1-C(x)]^n d\Lambda(x). \quad \ldots \quad (2.10)$$

Theorem 1. *The Hajek projection $V'(t)$ of $\hat{\Lambda}_n(t)$ is given as* :

$$V'(t)-E(\hat{\Lambda}_n(t)) = \bar{\xi}(t) - n\int_0^t C(x)\,[1-C(x)]^{n-1} d\Lambda(x)$$

$$-n^{-1}\Sigma\, e_{1i}(t)+n^{-1}\Sigma\, e_{2i}(t),$$

where

$$e_{1i}(t) = 1(X_i \leqslant t)\,[1-C(X_i)]^n\,[C(X_i)]^{-1}. \quad \ldots \quad (2.11)$$

$$e_{2i}(t) = \int_0^t 1(Y_i \leqslant x \leqslant X_i)\{[1-C(x)]^n [C(x)]^{-1} + n[1-C(x)]^{n-1}\}\, d\Lambda(x). \quad \ldots \quad (2.12)$$

Proof. Notice that

$$\begin{aligned} E(V_i(t) \mid X_i, Y_i) &= 1(X_i \leqslant t) \,.\, E([nC_n(X_i)]^{-1} \mid X_i, Y_i) \\ &= 1(X_i \leqslant t) \,.\, E([nC_n(X_i)]^{-1} \mid X_i) \\ &= 1(X_i \leqslant t) \,.\, [nC(X_i)]^{-1} [1-(1-C(X_i))^n], \end{aligned}$$

where the last step follows from formula (11) of Woodroofe (1985). Using this (2.9), (2.10) and Lemma 1, we thus have

$$\begin{aligned} & V'(t) - E(\hat{\Lambda}_n(t)) \\ &= \Sigma[E(\hat{\Lambda}_n(t) \mid X_i, Y_i) - E(\hat{\Lambda}_n(t))] \\ &= \Sigma[E(V_i(t) \mid X_i, Y_i) + (n-1)E(V_j(t) \mid X_i, Y_i) - E(\hat{\Lambda}_n(t))] \\ &= \Sigma 1(X_i \leqslant t)\, [nC(X_i)]^{-1} [1-(1-C(X_i))^n] + n \int_0^t 1-[1-C(x)]^{n-1} d\Lambda(x) \\ &+ \Sigma \int_0^t 1(Y_i \leqslant x < X_i) \{[1-C(x)]^{n-1} - [nC(x)]^{-1} [1-(1-C(x))^n]\} d\Lambda(x) \\ &- n\Lambda(t) + n \int_0^t [1-C(x)]^n d\Lambda(x) \\ &= n^{-1}\Sigma\, \xi(X_i, Y_i, t) - n^{-1}\, \Sigma e_{1i} + n^{-1}\Sigma e_{2i} - n \int_0^t C(x)\, [1-C(x)]^{n-1} d\Lambda(x). \end{aligned}$$

Remark 2. Since the Hajek projection of $\hat{\Lambda}_n(t) - E(\hat{\Lambda}_n(t))$ is $V'(t) - E(\hat{\Lambda}_n(t))$, theorem 1 implies that $\bar{\xi}(t)$ is not the Hajek projection of $\hat{\Lambda}_n(t) - E(\hat{\Lambda}_n(t))$. However, the following Theorem 2 (proof given in the Appendix) indicates that it is equivalent to the Hajek projection.

Theorem 2. *Let b be any fixed point with $b < b_F$.*

(i) *If $a_G < a_F$, then*

$$\sup_{0 \leqslant t \leqslant b} \mid V'(t) - E(\hat{\Lambda}_n(t)) - \bar{\xi}(t)) \mid \; = O(n(1-\epsilon)^{n-1}) \text{ a.s,}$$

where
$$0 < \epsilon = \alpha^{-1} G(a_F)[1-F(b_G)] < 1.$$

(ii) *If* $a_G = a_F$ *and*

$$\int_0^\infty dF/G < \infty, \qquad \ldots \quad (2.13)$$

then we have

$$\sup_{0 \leqslant t \leqslant b} \mid V'(t) - E(\hat{\Lambda}_n(t)) - \bar{\xi}(t) \mid \; = o_p(n^{-1/2}).$$

3. Hajek projection of $\hat{\Lambda}_n(t)$ under censoring

Assume the censoring model where $(X_i, Y_i), \ldots, (X_n, Y_n)$ are independent copies of (X, Y) and the observations are $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$, where $Z_i = \min(X_i, Y_i) = X_i \wedge Y_i$ and $\delta_i = 1(Z_i = X_i)$. The distribuion function H of Z_i satisfies $1-H(t) = [1-F(t)][1-G(t)]$. Let $H_1(t) = P(Z_i \leqslant t, \delta_i = 1)$ be the sub-distribution function of the uncensored observations. It can be checked easily that

$$\Lambda(t) = \int_0^t \frac{dH_1(x)}{1-H(x-)} dx. \quad \ldots \quad (3.1)$$

Hence Λ can be estimated empirically. Using this fact Nelson (1972) proposed an estimator $\tilde{\Lambda}_n$ of Λ as

$$\tilde{\Lambda}_n(t) = \Sigma 1(Z_i \leqslant t, \delta_i = 1)(n+1-R_i)^{-1}, \quad \ldots \quad (3.2)$$

where R_i is the rank of Z_i.

The corresponding distribution function for $\tilde{\Lambda}_n$ is the well-known product-limit estimator $\tilde{F}_n$ of Kaplan and Meier (1958),

$$\tilde{F}_n(t) = 1 - \prod_{z_i \leqslant t} \left(\frac{n-R_i}{n-R_i+1}\right)^{\delta_i}. \quad \ldots \quad (3.3)$$

In our present setting, U_i in (2.7) is equal to (X_i, Y_i) and $\tilde{\Lambda}_n(t) = \Sigma W_i(t)$ where $W_i(t) = 1(X_i \leqslant t, \delta_i = 1)(n+1-R_i)^{-1}$. For positive z, t and δ taking values 0 or 1, let

$$\eta(z, \delta, t) = 1(z \leqslant t, \delta = 1)[1-H(z)]^{-1} - \int_0^x [1-H(s)]^{-2} dH_1(s).$$

Let T be any point such that $H(T) < 1$, and $\epsilon = 1-H(T) > 0$. Lo and Singh (1986) gave the following i.i.d. representation of $\tilde{\Lambda}_n$:

$$\tilde{\Lambda}_n(t) - \Lambda(t) = n^{-1} \Sigma\, \eta(Z_i, \delta_i, t) + R_n'(t) = \bar{\eta}(t) + R_n'(t), \quad \ldots \quad (3.4)$$

where
$$\sup_{0 \leqslant t \leqslant T} |R_n'(t)| = O(\log n/n) \text{ a.s.} \quad \ldots \quad (3.5)$$

Note that the order of the remainder term R_n' in (3.5) was an improvement over the original order $O((\log n/n)^{3/4})$ and can be derived from Burke *et al.* (1988) and Major and Rejtö (1988).

The Hajek projection W' of $\widetilde{\Lambda}_n$ can be derived as in the previous section and is done in Gaenssler and Stute (1987) under the assumption that F is continuous. We will describe it below and assume that F is continuous for the rest of the section. Note that

$$E(\widetilde{\Lambda}_n(t)) = \Lambda(t) - \int_0^t H^n(y) d\Lambda(y). \quad \ldots \quad (3.6)$$

Theorem 3. *The Hajek projection $W'(t)$ of $\widetilde{\Lambda}_n(t)$ is given as*

$$W'(t) - E(\widetilde{\Lambda}_n(t)) = \bar{\eta}(t) - n \int_0^t H^{n-1}(x)\,[1-H(x)] d\Lambda(x)$$

$$- n^{-1}\Sigma \epsilon_{1i}(t) + n^{-1}\Sigma \epsilon_{2i}(t),$$

where

$$\epsilon_{1i}(t) = H^n(Z_i)\,[1-H(Z_i)]^{-1}\, 1(Z_i \leqslant t, \delta_i = 1),$$

$$\epsilon_{2i}(t) = \int_0^t 1(x \leqslant Z_i)\,\{nH^{n-1}(x) + H^n(x)\,[1-H(x)]^{-1}\} d\Lambda(x).$$

Remark 4. Comparing the Hajek projection in Theorem 2 with that of Theorem 1 of the truncated case, we see that they have the same form except that $1-H(Z_i)$ replaces the role of $C(X_i)$ and the indicator functions involved are slightly different.

Remark 5. It follows from Theorem 3 that $\bar{\eta}(t)$ is not the Hajek projection of $\widetilde{\Lambda}_n(t) - E(\widetilde{\Lambda}_n(t))$. The next theorem gives the equivalence of them.

Theorem 4. *With probability one,*

$$\sup_{0 \leqslant t \leqslant T} |\, W'(t) - E\widetilde{\Lambda}_n(t) - \bar{\eta}(t) \,| = O((1-\epsilon))^n.$$

Proof. Since $H(Z_i) \leqslant 1-\epsilon$ for $Z_i \leqslant t \leqslant T$, we have,

$$\sup_{0 \leqslant t \leqslant T} |\, n^{-1} \Sigma\, \epsilon_{1i}(t) \,| = O((1-\epsilon)^n).$$

Applying (3.1) we obtain

$$\epsilon_{2i}(t) = \int_0^{Z_i \wedge t} \{nH^{n-1}(y)\,[1-H(y)]^{-1} + H^n(y)\,[1-H(y)]^{-2}\}\, dH_1(y)$$

$$\leqslant \int_0^t \{nH^{n-1}(y)\,[1-H(y)]^{-1} + H^n(y)\,[1-H(y)]^{-2}\}\, dH(y)$$

$$= H^n(t)\,[1-H(t)]^{-1}.$$

Therefore,

$$\sup_{0 \leqslant t \leqslant T} |\, n^{-1} \Sigma\, \epsilon_{2i}(y) \,| = O((1-\epsilon)^n).$$

Now consider

$$\int_0^t nH^{n-1}(y)\,[1-H(y)]^{-1}d\Lambda(y)$$
$$= \int_0^t nH^{n-1}(y)\,dH_1(y) \leqslant \int_0^t nH^{n-1}(y)\,dH(y)$$
$$= H^n(t)$$
$$= O((1-\epsilon)^n).$$

The theorem is thus proved.

4. Applications and discussions

In order to utilize the Hajek projection principle to show the local asymptotic normality of $\widetilde{\Lambda}_n(t)$ it remains to check that the standardized versions of $W'(t)$ and $\widetilde{\Lambda}_n(t)$ have the same limiting distribution. From property (2.8) of Hajek projection and standard arguments it suffices to show that :

$$\text{var}\,(W'(t))/\text{var}\,(\widetilde{\Lambda}_n(t)) \to 1. \qquad \ldots \quad (4.1)$$

Instead of evaluating var $(W'(t))$ and var $(\widetilde{\Lambda}_n(t))$ directly as was done in Theorem 3 of Tanner and Wang (1983) and Gaenssler and Stute (1987, p. 66), we shall adopt the following result of Lo, Mack and Wang (1989) :

Let $R_n'(t)$ be the remainder term in (3.4), then for T such that $\epsilon = 1-H(T) > 0$.

$$\sup_{0\leqslant t\leqslant T} E([R_n'(t)]^2) = \sup_{0\leqslant t\leqslant T} E([\widetilde{\Lambda}_n(t)-\Lambda(t)-\overline{\eta}(t)]^2)$$
$$= O((\log n/n)^2). \qquad \ldots \quad (4.2)$$

Letting $S = \overline{\eta}(t)+\Lambda(t)$ in (2.8), we thus obtain from (4.2) that

$$\text{var}\,(W'(t))-\text{var}\,(\widetilde{\Lambda}_n(t)) \leqslant E([R_n'(t)]^2) = O((\log n/n)^2). \qquad \ldots \quad (4.3)$$

This together with the fact that $\text{var}(\widetilde{\Lambda}_n(t)) = \text{constant}\,.n^{-1}$ implies (4.1). We have thus shown :

Corollary 1. *For $0 \leqslant t \leqslant T$ and continuous F, the standardized versions of $W'(t)$ and $\widetilde{\Lambda}_n(t)$ converge weakly to the standard Normal distribution.*

As for weak convergence of the process $n^{1/2}(\widetilde{\Lambda}_n(t)-\Lambda(t))$ for $0 \leqslant t \leqslant T$, we have :

Corollary 2. *For $0 \leqslant t \leqslant T$ and continuous F, the processes $n^{1/2}(\widetilde{\Lambda}_n(t)-\Lambda(t))$, $n^{1/2}(W'(t)-\Lambda(t))$ and $n^{1/2}\overline{\eta}(t)$ all have the same limiting process.*

Proof. First consider the bias of $\tilde{\Lambda}_n(t)$. From (3.6) we have

$$\Lambda(t)-E(\tilde{\Lambda}_n(t)) = \int_0^t H^n(y)d\Lambda(y) \equiv \int_0^t H^n(y)[1-H(y)]^{-1}dH_1(y)$$

$$\leqslant \epsilon^{-1}\int_0^t H^n(y)dH(y) = [(n+1)\,\epsilon]^{-1}H^{n+1}(t) = O((1-\epsilon)^{n+1}). \quad \ldots \quad (4.4)$$

Theorem 4, (3.4) and (4.4) thus imply that all three processes will have the same limiting process provided it exists.

The weak convergence of the process $n^{1/2}\,\bar{\eta}(t)$ can be derived as in Lo and Singh (1986) which implies the weak convergence of $n^{1/2}\,(\tilde{\Lambda}_n(t)-\Lambda(t))$ to a mean zero Gaussian process. Another proof of the weak convergence of $n^{1/2}(\tilde{\Lambda}_n-\Lambda)$ can be found in Gaenssler and Staute (1987, p.69) using Corollary 1, Cramer-Wold device and tightness of the empirical process pertaining to the Z's.

The weak convergence of the process $n^{1/2}(\hat{\Lambda}_n(t)-\Lambda(t))$ can be argued similarly under the assumption that F is continuous and is given below :

(1) If $a_G < a_F$, an immediate consequence of Theorem 2(i) is that $n^{/12}[V'(t)-E(\hat{\Lambda}_n(t))]$ and $n^{1/2}\bar{\xi}(t)$ have the same limiting Gaussian process $Z(t)$ for $0 \leqslant t \leqslant b$. As Chao and Lo (1988) indicated, this limiting Gaussian process $Z(t)$ is the limiting process of $n^{1/2}(\hat{\Lambda}_n(t)-\Lambda(t))$. Since the bias of $\hat{\Lambda}_n(t)$ (cf. (2.10)) is of the order $O((1-\epsilon)^n)$, the limiting process of $n^{1/2}(V'(t)-\Lambda(t))$ is also $Z(t)$.

(2) If $a_G = a_F$, the process $n^{1/2}\,\bar{\xi}(t)$ may diverge, as pointed out by Woodroofe (1985), unless (2.13) holds. Noted here that (2.13) is true if $a_G < a_F$. Theorem 2 then implies :

Corollary 3. *For $0 \leqslant t \leqslant b < b_F$ and continous F, $n^{1/2}[V'(t)-E(\hat{\Lambda}_n(y))]$, $n^{1/2}[V'(t)-\Lambda(t)]$, $n^{1/2}\,\bar{\xi}(t)$, $n^{1/2}[\Lambda_n(t)-\hat{\Lambda}(t)]$ all converge as $n \to \infty$ to the same limiting Gaussian process $Z(t)$ with mean zero and covariance function*

$$cov\,(Z(s),\,Z(t)) = \int_0^{s\wedge t} d\Lambda(x)/C(x),$$

provided that (2.13) *holds.*

Remark 6. Note that the result in Corollary 3 is stated conditionally on the number of observed pairs. However, the statement also holds unconditionally as $N \to \infty$ (cf. Remark 1 in Section 2).

Appendix

A.1 *Proof of Lemma* 1.

$$E(V_j(t)\,|\,X_i, Y_i) = E[E(1(X_j \leqslant t)\,[nC_n(X_j)]^{-1}|X_i, Y_i, X_j, Y_j)\ |X_j, Y_j],$$

$$= E[(1(X_j \leqslant t)\, E(nC_n(X_j)]^{-1}|X_i, Y_i, X_j, Y_j)|X_j, Y_j]. \quad \ldots \quad (A.1)$$

Given X_i, Y_i, X_j, Y_j and n, the conditional distribution of $nC_n(X_j)$ is

$$nC_n(X_j) \begin{cases} 2+\text{Binomial } (n-2, C(X_j)), & \text{if } Y_i \leqslant X_j < X_i \\ 1+\text{Binomial } (n-2, C(X_j)), & \text{otherwise.} \end{cases}$$

Standard calculations then show that, for $Y_i \leqslant X_j < X_i$, writing $p = C(X_j)$,

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

Similarly, for $X_j < Y_i$ or $X_i \leqslant X_j$,

$$E([nC_n(X_j)]^{-1}\ |X_i, Y_i, X_j, Y_j) = [(n-1)p]^{-1}\,(1-(1-p)^{n-1}]. \quad \ldots \quad (A.3)$$

Combining (A.2) and (A.3), we have

$$E[nC_n(X_j)]^{-1}|\,X_i, Y_i, X_j, Y_j) = [(n-1)p]^{-1}\,[1-(1-p)^{n-1}]$$

$$+[n(n-1)p^2]^{-1}[(1-p)^n+np(1-p)^{n-1}-1]1(Y_i \leqslant X_j < X_i) = \mathrm{I}+\mathrm{II} \ \ldots \quad (A.4)$$

Replacing P by $C(X_j)$ in I, we obtain

$$E(1(X_j \leqslant t)\,.\,\mathrm{I}\,|\,X_i, Y_i) = (n-1)^{-1} \int_{a_F}^{t} \{1-[1-C(x)]^{n-1}\}\,[C(x)]^{-1}dF_*(x)$$

$$= (n-1)^{-1} \int_0^t 1-[1-C(x)]^{n-1}d\Lambda(x), \quad \ldots \quad (A.5)$$

where the last step follows from (2.4).

Similarly,

$$E(1(X_j \leqslant t)\,.\,\mathrm{II} \mid X_i, Y_i) = [n(n-1)]^{-1} \int_0^t 1(Y_i \leqslant x < X_i)\{[1-C(x)]^n$$

$$+nC(x)[1-C(x)]^{n-1}-1\}[C(x)]^{-2}dF_*(x)$$

$$= [n(n-1)]^{-1} \int_0^t 1(Y_i \leqslant x < X_i)\{[1-C(x)]^n$$

$$+nC(x)[1-C(x)]^{n-1}-1\}[C(x)]^{-1}d\Lambda(x)$$

$$= (n-1)^{-1}\Big\{ \int_0^t 1(Y_i \leqslant x<X_i)[1-C(x)]^{n-1}d\Lambda(x)$$

$$+ \int_0^t 1(Y_i \leqslant x<X_i)\{[1-C(x)]^n-1\}[nC(x)]^{-1}d\Lambda(x)\Big\}.$$

The lemma now follows from (A.1), (A.4), (A.5), and (A.6).

A.2 *Proof of Theorem* 2. We shall prove (i) first. Note that from (2.3), $C(X_i) \geqslant \epsilon$ a.s. for $X_i \leqslant b$. Hence (2.11) implies that, with probability one

$$\sup_{0 \leqslant t \leqslant b} |n^{-1} \Sigma e_{1i}(t)| = O((1-\epsilon)^n).$$

For $e_{2i}(t)$, (2.12) implies that we only need to consider the integration on those x for which $a_F \leqslant x \leqslant b$, and hence $C(x) \geqslant \epsilon$ and

$$\sup_{t \leqslant t \leqslant b} |n^{-1} \Sigma e_{2i}(t)| = O((1-\epsilon)^{n-1}).$$

Finally, $\sup\limits_{0 \leqslant t \leqslant b} \int_0^t nC(x)[1-C(x)]^{n-1}d\Lambda(x) = O(n(1-\epsilon)^{n-1})$ for the same reason as above. Part (i) is thus completed.

To prove (ii), we assume for convenience that $a_F = a_G = 0$. Notice that

$$V'(t)-E(\hat{\Lambda}_n(t))-\bar{\xi}(t)$$

$$= -n^{-1}\Sigma e_{2i}(t)+n^{-1}\Sigma e_{2i}(t) - n\int_0^t C(x)[1-C(x)]^{n-1}\,d\Lambda(x)$$

$$= -n^{-1}\Sigma[e_{1i}(t)-E(e_{1i}(t))]+n^{-1}\Sigma[e_{2i}(t)-E_n(e_{2i}(t))]$$

$$= -n^{-1}\Sigma X_i(t)+n^{-1}\Sigma Y_i(t), \qquad \dots \quad (A.7)$$

since the expected value of the left hand side is zero.

Consider first,

$$E\,[e_{1i}^2(t)] = \int_0^t [1-C(x)]^{2n}[C(x)]^{-2}dF_*(x)$$

$$= \alpha \int_0^t [1-C(x)]^{2n}[G(x)]^{-1}[1-F(x)]^{-2}dF(x)$$

$$\leqslant \alpha\,.\,[1-F(t)]^{-2}\,.\int_0^t [1-C(x)]^{2n}[G(x)]^{-1}dF(x),$$

where the second equality follows form (2.1) and (2.3).

Lebesgue dominated convergence theorem and (2.13) imply that $\int_0^t [1-C(x)]^{2n}[G(x)]^{-1}dF(x)$ tends to zero as n tends to infinity. We have thus shown that $E\,[e_{1t}^2\,(t)] = o(1)$ for all $0 \leqslant t \leqslant b$, which implies

$$\text{var}\ \ n^{-1}\Sigma X_t\,(t) = n^{-1}\,\text{var}\ [X_t(t)] \leqslant n^{-1}\,E\,[e_{1t}^2]\,(t) = o(n^{-1}).$$

and hence $n^{-1}\,\Sigma X_t\,(t) = o_p(n^{-1/2})$. Since the $o_p(n^{-1/2})$ term above is independent of t for $t \leqslant b$, we therefore conclude that

$$\sup_{0\leqslant t\leqslant b} \mid n^{-1}\Sigma X_t\,(t) \mid\, = o_p(n^{-1/2}).$$

Similarly, one can show that

$$\sup_{0\leqslant t\leqslant b} |n^{-1}\Sigma Y_i(t)| = o_p(n^{-1/2}). \quad \square$$

Part (ii) now follows from (A.7).

References

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Breslow, N. E. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.*, **2**, 437-453.

Bruke, M. D., Csörgo, S. and Horváth, L. (1988). An improved approximation rate for the product-limit process. *Prob. Theory Rel. Fields*, **79**, 51-57.

Chao, M. T. (1987). Influence curves for randomly truncated data. *Biometrika*, **74**, 426-429.

Chao, M. T. and Lo, S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Ann. Statist*, **16**, 661-668.

Gaenssler, P. and Stute, W. (1987). *Seminars on Empirical Processes*, Birkhäuser Verlag, Basel.

Hajek, J. (1968). Asymptotic normality of simple linear rank statistics under alternative. *Ann. Math. Statist*, **39**, 325-346.

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

Lo, S. H., Mack, Y. P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of Kaplan-Meier estimator. *Prob. Theory Rel. Fields*, **80**, 461-473.

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Prob. Theory Rel. Fields*, **71**, 455-465.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astron. Soc.*, **155**, 95-118.

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

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

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

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

Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist*, **14**, 1597-1605.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist*, **13**, 163-177.

Industrial Engineering Department
Bilkent University
06533, Bilkent, Ankara
Turkey.

Division of Statistics
University of California
Davis, CA 95616
U.S.A.