Ann. Inst. Statist. Math.
Vol. 40, No. 3, 541-553 (1988)

ASYMPTOTIC THEOREMS FOR ESTIMATING THE DISTRIBUTION FUNCTION UNDER RANDOM TRUNCATION

WEI-MIN HUANG[1] AND WEI-YANN TSAI[2]

[1]*Department of Mathematics, Lehigh University, Bethlehem, PA 18015, U.S.A.*
[2]*Applied Mathematics Department, Brookhaven National Lab., Upton, NY 11973, U.S.A. and Division of Biostatistics, School of Public Health, Columbia Univ.*

(Received January 16, 1987; revised May 25, 1987)

Abstract. Representation theorem and local asymptotic minimax theorem are derived for nonparametric estimators of the distribution function on the basis of randomly truncated data. The convolution-type representation theorem asserts that the limiting process of any regular estimator of the distribution function is at least as dispersed as the limiting process of the product-limit estimator. The theorems are similar to those results for the complete data case due to Beran (1977, *Ann. Statist.*, **5**, 400–404) and for the censored data case due to Wellner (1982, *Ann. Statist.*, **10**, 595–602). Both likelihood and functional approaches are considered and the proofs rely on the method of Begun *et al.* (1983, *Ann. Statist.*, **11**, 432–452) with slight modifications.

Key words and phrases: Nonparametric (conditional and unconditional) maximum likelihood estimator, product limit estimator, truncated data, representation theorem, asymptotic minimax.

1. Introduction and motivation

Let (X^0, T^0) be a bivariate (X^0 and T^0 are not necessarily independent) random variable and let $G(x, t)$ be the joint cumulative distribution of X^0 and T^0. Let $G_1(x) = G(x, \infty)$ and $G_2(t) = G(\infty, t)$ denote the marginal distribution function of X^0 and T^0, respectively. Let $G^*(x, t)$ be the truncated (on $X^0 \leq T^0$) joint cumulative distribution of X^0 and T^0, i.e.,

$$G^*(x, t) = P(X^0 \leq x, T^0 \leq t \mid X^0 \leq T^0) . \tag{1.1}$$

Also let $G_1^*(x) = G^*(x, \infty)$ and $G_2^*(t) = G^*(\infty, t)$. Then it is obvious that $G_1^*(x) = P(X^0 \leq x \mid X^0 \leq T^0)$ and $G_2^*(t) = P(T^0 \leq t \mid X^0 \leq T^0)$.

In randomly truncated data problems, we assume that the bivariate random variable (X^0, T^0) is not observable if $X^0 > T^0$ and the whole vector

(X^0, T^0) is observed if $X^0 \le T^0$. Let $(X_1, T_1), (X_2, T_2), \ldots, (X_n, T_n)$ be independent observations with each (X_i, T_i) having distribution function $G^*(x, t)$. A basic problem with truncated data is how to estimate the distribution function $G_1(x)$ of X^0 based on the truncated observations (X_i, T_i), $i = 1, 2, \ldots, n$.

The joint distribution function $G^*(x, t)$ can be written as

$$G^*(x, t) = \alpha^{-1} \int_{-\infty}^{t} \int_{-\infty}^{x} I(u \le v) dG(u, v) , \tag{1.2}$$

where $\alpha = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(u \le v) dG(u, v)$ and $I(u \le v) = 1$ if $u \le v$ and 0 otherwise.

Set $U_i = \inf\{x: G_i(x) = 1\}$, and define $U_i = \infty$ if the set is empty, for $i = 1, 2$. Note that we cannot hope to estimate G_1 if $U_1 > U_2$, when we observe only the (X, T) pairs, since the probability $X_i \le T_i \le U_2$, for all i, is one. In this case we will be able to estimate only the conditional distribution $G_1^c(\cdot) = G_1(\cdot)/G_1(T^*)$ with $T^* < U_2$. In order to simplify our notations and proofs, we will consider the estimation of G_1 under the assumption $U_1 < U_2 \le \infty$.

The product-limit estimator $\hat{G}_1$ of G_1 is defined by

$$\hat{G}_1(x) = \prod_{i: X_i \ge x} \left\{ 1 - \left[\sum_{j=1}^{n} I(X_j = X_i) \Big/ \sum_{j=1}^{n} I(X_j \le X_i \le T_j) \right] \right\} . \tag{1.3}$$

Under the assumption that X^0 and T^0 are independent, the product-limit estimator was originally derived by Lynden-Bell (1971). Wang and Jewell (1985) also derived $\hat{G}_1$ as the "nonparametric conditional maximum likelihood estimator" (NCMLE) of G_1. Recently Woodroofe (1985) also mentioned an alternative derivation of $\hat{G}_1$ by first deriving the nonparametric maximum likelihood estimators (NMLE) of G_1^* and G_2^* and then utilizing the following equations:

$$G_1^*(x) = \alpha^{-1} \int_{-\infty}^{x} [1 - G_2(u^-)] dG_1(u), \qquad -\infty < x < \infty , \tag{1.4a}$$

$$G_2^*(t) = \alpha^{-1} \int_{-\infty}^{t} G_1(u^-) dG_2(u), \qquad -\infty < t < \infty , \tag{1.4b}$$

to solve for G_1 and G_2, so that G_1 and G_2 become functionals of G_1^* and G_2^* and the associated empirical functionals are natural estimators of G_1^*. Woodroofe (1985) argued that the resulting estimators are NMLE by the invariance properties of MLES. The estimator of G_1 derived this way coincides with the product-limit estimator $\hat{G}_1$ in (1.3). Since $\hat{G}_1$ is the NMLE of G_1 we may expect that $\hat{G}_1$ possesses some optimal properties; but a general theory in nonparametric maximum likelihood is not available. The main objective of the current paper is to show that $\hat{G}_1$ does possess some optimalities. Since the current truncated data model is a variation of a semiparametric model (as explained later), the projection technique in Begun *et al.* (1983) with slight modifications applies.

The truncated model has also been studied by Segal (1975), Nicoll and Segal (1980), Bhattacharya (1983), Bhattacharya *et al.* (1983) and Wang *et al.* (1986).

The following investigations provided some of the motivations that led us to the current study (for expositional ease, we assume in this section that the joint probability density function $g(x,t)$ of $G(x,t)$ exists).

1.1 *A useful class of functionals and nonunique representation of* G_1

For every fixed x_0, define a class of functionals $\{T_c(G;x_0): -\infty < c < \infty\}$ where

$$(1.5)\qquad T_c(G;x_0) = \exp\left[-\int_{x_0}^{\infty} \frac{\int_c^{\infty} g(x,v)dv}{\int_{-\infty}^{x}\int_c^{\infty} g(u,v)dvdu}\, dx \right].$$

Let $h_c(G;t) = d[\log T_c(G;t)]/dt$, i.e., $h_c(G;t) = P(X^0 = t\,,\, T^0 \geq c \mid T^0 \geq X^0)$. The following results can easily be derived from (1.1), (1.2) and (1.5).

(i) If $c \geq x_0$, then $h_c(G;x_0) = h_c(G^*;x_0)$. This means that the functional $h_c(G;x_0)$ is invariant under the truncation (on $X^0 \leq T^0$).

(ii) If X^0 and T^0 are independent, then $T_c(G;X_0) = G_1(x_0)$ for all c. It says that all the functionals $T_c(G;x_0)$, $-\infty < c < \infty$ are equal to the distribution function G_1 at x_0 that we are trying to estimate, hence the functional expression of G_1 in terms of G is not unique. This is a special structure under the assumption of independence between X^0 and T^0, but it is not surprising since G_1 is a marginal distribution of G.

(iii) If X^0 and T^0 are independent, then for all $c \in (-\infty, \infty)$

$$\tilde{T}_c(G;x_0) = \tilde{T}_c(G^*;x_0) = G_1(x_0)\,,$$

where $\tilde{T}_c(G;X_0) \equiv \exp\left[-\int_{x_0}^{\infty} h_{\max(c,x_0)}(G;x)dx \right]$. This shows that the functional expression of G_1 in terms of the truncated distribution G^* is also not unique, and it explains that the truncated model is a variation of the semiparametric model (a variation since it has no finite dimensional parameters and has only two infinite dimensional parameters G_1 and G_2). See Stein (1956), Bickel (1982) and Begun *et al.* (1983) for a definition of the semiparametric model and the discussion on p. 434 of the third paper.

Properties (i), (ii) and (iii) indicate that there are infinitely many ways of estimating G_1 (or G_2) and the NMLE is one of them. In fact, it can be shown that if X^0 and T^0 are independent, then the $G_1(x)$ functional derived from (1.4a) and (1.4b) is equal to $T_c(G^*;x)$ with the choice $c = x$. Note that if $G_i^*(x)$ are not continuous, then the resulting functional $G_1(x)$ from (1.4a) and (1.4b) is slightly different from $T_x(G; x)$, in this case

$$G_1(x) = T_x(G_s; x) \cdot \left[\prod_{u:u>x} \left(1 - \frac{G_1^*(u) - G_1^*(u^-)}{G_1^*(u) - G_2^*(u^-)} \right) \right], \tag{1.6}$$

where G_s is the continuous part of G. The second term on the right hand side is obviously one if $G_i^*(x)$ are continuous.

1.2 *Optimal estimate of $T_c(G; x_0)$ without independent assumption*

The empirical functional $T_c(\hat{G}^*; x_0)$ (with $\hat{G}^*$ an estimate of G^* based on truncated data) can still be used to estimate the functional $T_c(G^*; x_0)$ even if X_i^0 and T_i^0 are dependent but then $T_c(G^*; x_0)$ may not be equal to $G_1(t)$. Standard theory (see for example, Beran (1977), Bickel and Wellner (1983), among others) implies that under mild regularity conditions any regular estimator $\hat{T}_c(x_0)$ of $T_c(G; x_0)$ (note that x_0 is fixed) satisfies.

$$n^{1/2}[\hat{T}_c(x_0) - T_c(G; x_0) \xrightarrow{d} Z_c(x_0) + W(x_0) , \tag{1.7}$$

where $Z_c(x_0)$ is $N(0, \sigma_c^2(x_0))$ and $Z_c(x_0)$ and $W(x_0)$ are independent. The random variable $W(x_0)$ depends on the estimator $\hat{T}_c(x_0)$. The $\sigma_c^2(x_0)$ can be determined by

$$\sigma_c^2(x_0) = E\{[IC(X_i, T_i; G, x_0, c)]^2\} , \tag{1.8}$$

where $IC(s, w; G, x_0, c)$ is the influence (curve) function of the functional $T_c(G; x_0)$ at (s, w). By some careful and long calculations, it can be shown that

$$IC(s, w; G, x_0, c) = \begin{cases} T_c(G; x_0) \left\{ \int_{x_0}^{c} \frac{f_c(x)}{M_c^x} dx + \int_c^w \frac{f_x(x)}{(M_x^x)^2} dx \right\} \\ \qquad \text{if} \quad s < w, \quad w > c \geq x_0 \quad \text{and} \quad s < x_0 , \\ -T_c(G; x_0) \left\{ \frac{1}{M_s^s} - \int_s^w \frac{f_x(x)}{(M_x^x)^2} dx \right\} \\ \qquad \text{if} \quad s < w, \quad x > c \geq x_0 , \end{cases} \tag{1.9}$$

where $M_c^x \equiv \int_{-\infty}^{x} \int_c^{\infty} g(u, v) dv du$ and $f_c(x) = \int_c^{\infty} g(x, y) dy$. Note that it suffices to consider $c \geq x_0$ since $\tilde{T}_c(G; x_0) = \tilde{T}_t(G; x_0)$ for all $c < x_0$. Thus

$$\sigma_c^2(x_0) = \iint T_c^2(G; x_0) \left[\int_{x_0}^{c} \frac{f_c(x)}{(M_c^x)^2} dx + \int_c^w \frac{f_x(x)}{(M_x^x)^2} dx \right]^2 g(s, w) ds dw \quad (1) \tag{1.10}$$

$$+\iint T_c^2(G;x_0)\left\{\frac{1}{M_s^s}-\int_s^w \frac{f_x(x)}{(M_x^x)^2}\,dx\right\}^2 g(s,w)dsdw\,, \tag{2}$$

where the integrals in (1) and (2) are over the two regions in (1.9). It can be shown that the estimator $\hat{T}_c(x_0)=T_c(\hat{G}_n^*;x_0)$ with $\hat{G}_n^*$ being the joint empirical distribution based on truncated data (X_i, T_i), $i=1,2,\ldots,n$, has asymptotic variance $\sigma_c^2(x_0)$. It means asymptotically $T_c(\hat{G}_n^*, x_0)$ is optimal in the sense that any other estimator is at least as dispersed as $T_c(\hat{G}_n^*, x_0)$. This is obvious from (1.7). We may want to use a smooth version of $\hat{G}_n^*$ to estimate G^*.

1.3 *The best possible functional under independence assumption*

Under the assumption that X_i^0 and T_i^0 are independent we have $T_c(G;x_0)=G_1(x_0)$ for all c and $T_c(G;x_0)=\tilde{T}_c(G;x_0)=G_1(x_0)$ for all $c\geq x_0$. Consider the estimation of $\tilde{T}_c(G;x_0)$ based on the truncated data: it is natural, according to (iii) in Subsection 1.1 and the estimation of $T_c(G;x_0)$ discussed above, to estimate $G_1(x_0)=\tilde{T}_c(G;x_0)$ by $\hat{\tilde{T}}_c(x_0)=\tilde{T}_c(\hat{G}_n^*;x_0)$ with c being any constant not less than x_0. It can be shown that $T_c(G;x_0)$ is constant for all $c\geq x_0$ and $k'(c)>0$ for $c>x_0$, where

$$k(c)=\int_{x_0}^{c}\frac{f_c(x)}{(M_c^x)^2}\,dx+\int_c^w \frac{f_c(x)}{(M_x^x)^2}\,dx\,. \tag{1.11}$$

These with (1.10) imply by some careful calculations that

$$\sigma_{x_0}^2(x_0)=\min \sigma_c^2(x_0) \qquad \text{over} \qquad c\geq x_0\,, \tag{1.12}$$

and

$$\begin{aligned}\sigma_{x_0}^2(x_0) &= G_1^2(x_0)\cdot\int_{x_0}^{\infty}\frac{f_x(x)}{(M_x^x)^2}\,dx \\ &= G_1^2(x_0)\cdot\int_{x_0}^{\infty}\frac{\int_x^\infty g(x,y)dy}{\int_{-\infty}^{x}\int_x^\infty g(u,v)dvdu}\,dx\,.\end{aligned} \tag{1.13}$$

We call $T_{x_0}(G;x_0)$ the "best functional representation" of $G_1(x_0)$ since (a) $\sigma_{x_0}^2(x_0)$ minimizes the value of $\sigma_c^2(x_0)$ over $c\geq x_0$ and (b) the best possible asymptotic variance $\sigma_{x_0}^2(x_0)$ in (1.13) coincides with the covariance structure of PLE (or NCMLE) given in Wang *et al.* (1986), since if $c=x_0$ and X_i^0 and T_i^0 are independent, then $T_{x_0}(G^*;x_0)$ coincides with the functional $G_1(x_0)$ derived from (1.4a) and (1.4b) and this functional produces the PLE in (1.3), but the starting viewpoint of our derivation is different from theirs.

So far we have shown that in the class of estimators which are induced by $\tilde{T}_c(G;x)$, an invariant functional under truncation (on $X^0 \le T^0$), $\tilde{T}_t(\hat{G}_n^*,x)$ is asymptotically optimal as an estimator of $G_1(x)$ for every given x. Two questions remain to be answered: (1) is $\tilde{T}_x(\hat{G}_n^*;x)$ still an optimal estimator among all possible estimators of $G_1(x)$ and (2) what are the properties of the entire process $\{\tilde{T}_t(\hat{G}_n^*;t): t > -\infty\}$, for instance, what is the covariance structure of this process? The main objective is to develop lower bounds for limiting processes of estimators of G_1 and to show that the product-limit estimator is asymptotically optimal in some senses. Some properties of the process (or its modification, the PLE) have been studied by Woodroofe (1985), Wang *et al.* (1986) and Chou and Lo (1988). Our previous discussions are based on the "functional approach". In the following section we derive some optimalities of the PLE (or NCMLE) based on the "likelihood approach".

The current result can also be considered as a generalization, specially in estimating distribution function under random truncation, of convolution type results for limiting distributions of regular estimators developed (among others) by LeCam (1969), Inagaki (1970), Hajek (1970) and Beran (1977).

2. Main results

Let ν be a measure on $R = (-\infty,\infty)$ with respect to which G_1 and G_2 have densities g_1 and g_2, respectively (the measure ν induced by $G_1 + G_2$ always works). Then it is easy to see that the observed pairs (X_i, T_i) have a common density f with respect to $\mu = \nu \times \nu$ on $S = \{(x,t) | x \le t\}$ given by

$$f(x,t;g_1,g_2) = g_1(x)g_2(t)/\alpha_0 \qquad (x,t) \in S\,, \tag{2.1}$$

where $\alpha_0 = \alpha_0(g_1,g_2) = P(X \le T) = \iint_S g_1(x)g_2(t)d\mu(x,t)$. When it causes no confusion we will usually write $f(x,t)$ and α_0 for $f(x,t;g_1,g_2)$ and $\alpha_0(g_1,g_2)$, respectively. In this section all the results are based on the assumption that X and T are independent.

Let $L^2(\mu) = L^2(S,\mu)$ and $L^2(\nu) = L^2(R,\nu)$ denote the usual L^2-space of square-integrable functions and let $\langle\cdot,\cdot\rangle_\mu$ ($\|\cdot\|_\mu$) and $\langle\cdot,\cdot\rangle_\nu$ ($\|\cdot\|_\nu$) denote the usual inner products (and norms) in $L^2(\mu)$ and $L^2(\nu)$, respectively. Thus $f^{1/2} \in L^2(\mu)$, $g_i^{1/2} \in L^2(\nu)$, $i = 1,2$ and $\|f^{1/2}\|_\mu = \|g_i^{1/2}\| = 1$, $i = 1,2$. We will usually write $\langle\cdot,\cdot\rangle$ for $\langle\cdot,\cdot\rangle_\mu$ and $\langle\cdot,\cdot\rangle_\nu$ (and similarly for norms) when it causes no confusion.

Let $\mathscr{F}(\mu)$ ($\mathscr{F}(\nu)$) denote the set of all densities with respect to $\mu(\nu)$ on $S(R)$. Let $\mathscr{C}(f,\alpha)$ denote the set of all sequences $\{f_n\}_{n\ge1}$ with $f_n \in \mathscr{F}(\mu)$ such that

$$\|n^{1/2}(f_n^{1/2} - f^{1/2}) - \alpha\| \to 0 \quad \text{as} \quad n \to \infty,$$

for some $\alpha \in L^2(\mu)$. Similarly let $\mathscr{C}_1(g_1, g_2; \beta_1, \beta_2)$ denote the set of sequences $\{(g_{1n}, g_{2n})\}_{n\geq 1}$ with each $g_{in} \in \mathscr{F}(\nu)$ and

$$\|n^{1/2}(g_{in}^{1/2} - g_i^{1/2}) - \beta_i\| \to 0 \quad \text{as} \quad n \to \infty,$$

for some $\beta_i \in L^2(\nu)$. It is obvious that $\langle \alpha, f^{1/2}\rangle = 0$ and $\langle \beta_i, g_i^{1/2}\rangle = 0$ for $i = 1, 2$. Also let $\mathscr{B} = \{(\beta_1, \beta_2): \beta_i \in L^2(\nu), \langle \beta_i, g_i^{1/2}\rangle = 0, i = 1, 2\}$, $\mathscr{H} = \{\alpha: \alpha \in L^2(\mu), \langle f^{1/2}, \alpha\rangle = 0\}$, $\mathscr{C}_1(g_1, g_2) = U\{\mathscr{C}_1(g_1, g_2; \beta_1, \beta_2): (\beta_1, \beta_2) \in \mathscr{B}\}$, and $\mathscr{C}(f) = U\{\mathscr{C}(f, \alpha): \alpha \in \mathscr{H}\}$.

We say an estimator $\tilde{G}_{1n}$ of G_1 is regular at f if, for every sequence $\{f_n\} = \{f(\cdot, \cdot; g_{1n}, g_{2n})\}$ with $(g_{1n}, g_{2n}) \in \mathscr{C}_1(g_1, g_2)$, the process $n^{1/2}(\tilde{G}_{1n} - G_{1n})$ converges weakly on $C(-\infty, U_1]$ to the same limit process $\tilde{Z}$: $n^{1/2}(\tilde{G}_{1n} - G_{1n}) \Rightarrow \tilde{Z}$ on $C(-\infty, U_1]$ (under f_n) where the law of $\tilde{Z}$ on $C(-\infty, U_1]$ depends only on f for all sequences $\{f_n\}$. The sequence of distribution functions G_{1n} corresponds to the sequence of densities g_{1n}.

Let $Z = \{Z(x): -\infty < x \leq U_1\}$ be a mean zero Gaussian process on $(-\infty, U_1]$ with covariance function

$$\text{Cov}\,(Z(s), Z(t)) = G_1(s)G_1(t)\int_{s\vee t}^{\infty} \frac{dG_1^*(x)}{[G_1^*(x) - G_2^*(x^-)]^2}. \tag{2.2}$$

The following theorem extends the result of Beran (1977) and Wellner (1982) to the case of randomly truncated data.

THEOREM 2.1. *Suppose that $\tilde{G}_{1n}$ is a regular estimator of G_1 in the random truncation model with limit process $\tilde{Z}$, then*

$$\tilde{Z} = Z + W, \tag{2.3}$$

in distribution where Z is a mean zero Gaussian process with covariance function given by (2.2) *and the process W is independent of Z.*

To see that the product-limit estimator $\hat{G}_{1n}$ defined in (1.3) is asymptotically optimal, recall that, by Theorem 6 of Wang *et al.* (1986), the process $\hat{Z}_n = n^{1/2}(\hat{G}_{1n} - G_1)$ converges weakly to the process Z. Hence the product-limit estimator is optimal in the sense of the convolution-type representation theorem above. We omit the discussion of "regularity" of $\hat{G}_{1n}$.

To state a local asymptotic minimax bound, we let l: $C(-\infty, U_1] \to [0, \infty)$ be a subconvex loss function such as $l(x) = \|x\|_\infty = \sup_t \|x(t)\|$ or $l(x) = \int x^2(t)dt$.

THEOREM 2.2. *Let l be subconvex and $B_n(c) = \{f_n \in \mathscr{C}(f)$: $n^{1/2} \cdot \| f_n^{1/2} - f^{1/2} \| \le c\}$, then*

$$\lim_{c\to\infty} \lim_{n\to\infty} \inf_{\tilde{G}_n} \sup_{f_n \in B_n(c)} E_{f_n} l[n^{1/2}(\tilde{G}_{1n} - G_1)] \ge El(Z)\,, \tag{2.4}$$

where Z is the mean zero Gaussian process on $(-\infty, U_1]$ with covariance function given by (2.2). *The infimum in* (2.4) *is taken over the class of "generalized procedures", the closure of the class of randomized procedures as, in Millar* ((1979), *p.* 235).

3. Proofs of Theorems 2.1 and 2.2

We begin with two lemmas. The density g_2 in $f(\cdot,\cdot; g_1, g_2)$ is treated as a "nuisance nonparametric component" of the model and the function G_1 (or equivalently g_1) is the parameter of interest. The first lemma provides the "derivative" of f at (g_1, g_2) and the second lemma gives an orthogonal decomposition of the "partial derivative" of f with respect to g_1.

LEMMA 3.1. *Let bounded operators A_i: $L^2(\nu) \to L^2(\mu)$, $i = 1, 2$, be defined, respectively, by*

$$A_1\beta_1(x,t) = f^{1/2}(x,t)\{g_1^{-1/2}(x)\beta_1(x) - \alpha_0^{-1} D_1(\beta_1; g_1, g_2)\} \tag{3.1}$$

and

$$A_2\beta_2(x,t) = f^{1/2}(x,t)\{g_2^{-1/2}(t)\beta_2(t) - \alpha_0^{-1} D_2(\beta_2; g_1, g_2)\}\,, \tag{3.2}$$

where $D_1(\beta; g_1, g_2) \equiv \iint_{u\le v} g_1^{1/2}(u)\beta(u) g_2(v)\, d\mu(u,v)$, $D_2(\beta; g_1, g_2) \equiv \iint_{u \le v} g_1(u) \cdot g_2^{1/2}(v)\beta(v)\, d\mu(u,v)$, and β_1 and β_2 are in $L^2(\nu)$. Then with $f_n = f(\cdot,\cdot; g_{1n}, g_{2n})$,

$$\frac{\| f_n^{1/2} - f^{1/2} - A_1(g_{1n}^{1/2} - g_1^{1/2}) - A_2(g_{2n}^{1/2} - g_2^{1/2})\|_\mu}{\|g_{1n}^{1/2} - g_1^{1/2}\|_\nu + \|g_{2n}^{1/2} - g_2^{1/2}\|_\nu} \to 0 \quad as \quad n \to \infty\,, \tag{3.3}$$

for all sequences $\|g_{in}^{1/2} - g_i^{1/2}\| \to 0$, $i = 1, 2$, as $n \to \infty$, in $L^2(\nu)$.

PROOF. First note that

$$\begin{aligned} &\| f^{1/2}(\cdot,\cdot; g_{1n}, g_2) - f^{1/2}(\cdot,\cdot; g_1, g_2) - A_1(g_{1n}^{1/2} - g_1^{1/2})\| \\ &\quad = o(\|g_{1n}^{1/2} - g_1^{1/2}\|) \end{aligned} \tag{3.4}$$

and

$$\|f^{1/2}(\cdot,\cdot;g_1,g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_2)-A_2(g_{2n}^{1/2}-g_2^{1/2})\|$$
$$=o(\|g_{2n}^{1/2}-g_2^{1/2}\|)\,,$$

for every fixed g_1 and fixed g_2, respectively. The verification of (3.4) is similar to Lemma 1 of Begun and Wellner (1983), so we omit the details. Let $(A_{1n}\beta_1)(\cdot,\cdot)$ be the sequence of $(A_1\beta_1)(\cdot,\cdot)$ operators with g_2 replaced by g_{2n} in $A_1\beta_1(\cdot,\cdot)$ and let $G_{2n}(\cdot)=\int_{-\infty}^{\cdot}g_{2n}dv$. Then

$$(3.5)\quad \|A_{1n}\beta-A_1\beta\|$$
$$\leq\|[f^{1/2}(\cdot,\cdot;g_1,g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_2)]\cdot[g_1^{-1/2}\beta-\alpha_0^{-1}D_1(\beta;g_1,g_2)]\|$$
$$+\|f^{1/2}(\cdot,\cdot;g_1,g_{2n})[D_1(\beta;g_1,g_{2n})-D_1(\beta;g_1,g_2)]/\alpha_0(g_1,g_{2n})\|$$
$$+\|f^{1/2}(\cdot,\cdot;g_1,g_{2n})D_1(\beta;g_1,g_{2n})\cdot[\alpha_0^{-1}(g_1,g_2)-\alpha_0^{-1}(g_1,g_{2n})]\|\,.$$

It can be shown that the first term goes to zero as $\|g_{2n}^{1/2}-g_2^{1/2}\|\to 0$ and the sum of the last two terms is less than or equal to

$$\sup_t|G_{2n}(t)-G_2(t)|\cdot\|\beta\|\alpha_0^{-1}(g_1,g_{2n})$$
$$+\sup_t|G_{2n}(t)-G_2(t)|\cdot\|\beta\|\cdot\alpha_0^{-1}(g_1,g_2)\cdot\alpha_0^{-1}(g_1,g_{2n})\to 0\,,$$

for any $\beta\in L^2(\nu)$, as $\|g_{2n}^{1/2}-g_2^{1/2}\|\to 0$ by the facts that $\alpha_0(g_1,g_{2n})\to\alpha_0(g_1,g_2)=\alpha_0>0$ and $\sup_t|G_{2n}(t)-G_2(t)|\to 0$, as $\|g_{2n}^{1/2}-g_2^{1/2}\|\to 0$. Therefore, for any given $\varepsilon>0$, there exists δ_1 such that $\|A_{1n}-A_1\|<\varepsilon$ for every g_{2n} with $\|g_{2n}^{1/2}-g_2^{1/2}\|<\delta_1$. Write $f_n^{1/2}-f^{1/2}=f^{1/2}(\cdot,\cdot;g_{1n},g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_{2n})+f^{1/2}(\cdot,\cdot;g_1,g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_2)$. For the given $\varepsilon>0$, (3.4) implies that there exists $\delta_2>0$ such that for $\|g_{2n}^{1/2}-g_2^{1/2}\|<\delta_2$,

$$\|f^{1/2}(\cdot,\cdot;g_1,g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_2)-A_2(g_{2n}^{1/2}-g_2^{1/2})\|<\varepsilon\|g_{2n}^{1/2}-g_2^{1/2}\|\,.$$

On the other hand, for each g_{2n} with $\|g_{2n}^{1/2}-g_n^{1/2}\|<\delta_3$ we can apply (3.5) and the mean value theorem (see (8.5.2) and (8.6.2) of Dieudonn'e (1969), pp. 160–162) to the mapping $\beta\to f(\cdot,\cdot;\beta^2,g_{2n})-A_{1n}\beta$. This gives $\|f^{1/2}(\cdot,\cdot;g_{1n},g_{2n})-f^{1/2}(\cdot,\cdot;g_1,g_{2n})-A_{1n}(g_{1n}^{1/2}-g_1^{1/2})\|<\varepsilon\|g_{1n}^{1/2}-g_1^{1/2}\|$ for all g_{2n} with $\|g_{2n}^{1/2}-g_2^{1/2}\|<\delta_3$. Also from (4.5), we have

$$\|A_{1n}(g_{1n}^{1/2}-g_1^{1/2})-A_1(g_{1n}^{1/2}-g_1^{1/2})\|<\varepsilon\|g_{1n}^{1/2}-g_1^{1/2}\|\,,$$

for all g_{2n} with $\|g_{2n}^{1/2}-g_2^{1/2}\|<\delta_4$.

Hence, for all g_{2n} with $\|g_{2n}^{1/2}-g_2^{1/2}\|<\min(\delta_1,\delta_2,\delta_3,\delta_4)$, we have

$$
\begin{aligned}
&\| f_n^{1/2} - f^{1/2} - A_1(g_{1n}^{1/2} - g_1^{1/2}) - A_2(g_{2n}^{1/2} - g_2^{1/2}) \| \\
&\quad \le \| f^{1/2}(\cdot,\cdot; g_{1n}, g_{2n}) - f^{1/2}(\cdot,\cdot; g_1, g_{2n}) - A_{1n}(g_{1n}^{1/2} - g_1^{1/2}) \| \\
&\qquad + \| A_{1n}(g_{1n}^{1/2} - g_1^{1/2}) - A_1(g_{1n}^{1/2} - g_1^{1/2}) \| \\
&\qquad + \| f^{1/2}(\cdot,\cdot; g_1, g_{2n}) - f^{1/2}(\cdot,\cdot; g_1, g_2) - A_2(g_{2n}^{1/2} - g_2^{1/2}) \| \\
&\quad < 2\varepsilon \cdot (\| g_{1n}^{1/2} - g_1^{1/2} \| + \| g_{2n}^{1/2} - g_2^{1/2} \|) .
\end{aligned}
$$

This completes the proof of Lemma 3.1.

Remark 1. In order to apply the results of Begun *et al.* (1983) to our model, we make the following slight modifications of the statements in Section 4 of Begun *et al.* (1983): the "score" A_1 for g_1, the "nonparametric component" of interest, must first be projected onto the "score" A_2 for the "nuisance nonparametric component" g_2. Define $A_{1\cdot 2} \equiv (A_1 - A_2(A_2^* A_2)^{-1} \cdot A_2^* A_1)$, then $A_{1\cdot 2}$ and A_2 are orthogonal and $A_{1\cdot 2}$ is the "effective score" for g_1 in the presence of the nuisance nonparametric component g_2, where A_2^*: $L^2(\mu) \to L^2(\nu)$ denote the adjoint of the linear operator A_2. It can also be argued in terms of the notion of tangent space (in the sense of Pfanzagl (1982) and the forthcoming monograph by Bickel *et al.* (1987)).

LEMMA 3.2. *Let B be a linear operator from $L^2(\nu)$ to $L^2(\mu)$ which is defined by*

$$
(3.6) \qquad (B\beta)(x,t) = f^{1/2}(x,t; g_1, g_2) \left\{ g_1^{-1/2}(x)\beta(x) - \int_{-\infty}^{t} g_1^{1/2}\beta d\nu / G_1(t) \right\} .
$$

Then $A_{1\cdot 2} = B$.

PROOF. For any $\beta_1, \beta_2 \in L^2(\nu)$,

$$
\begin{aligned}
(3.7) \qquad \langle B\beta_1, A_2\beta_2 \rangle &= \iint_{x \le t} f(x,t) \left\{ g_1^{-1/2}(x)\beta_1(x) - \int_{-\infty}^{t} g_1^{1/2}\beta_1 d\nu / G_1(t) \right\} \\
&\qquad \cdot \{ g_2^{-1/2}(t)\beta_2(t) - \alpha_0^{-1} D_1(\beta_2, g_1, g_2) \} d\mu(x,t) \\
&= \alpha_0^{-1} \int_{-\infty}^{\infty} g_2(t) [g_2^{-1/2}(t)\beta_2(t) - \alpha_0^{-1} D(\beta_2, g_1, g_2)] \\
&\qquad \cdot \int_{-\infty}^{t} g_1(x) f^{1/2}(x,t)(B\beta_1)(x,t) d\nu(x) d\nu(t) \\
&= 0 ,
\end{aligned}
$$

since $\int_{-\infty}^{t} \left[g_1^{1/2}\beta_1 - g_1 \int_{-\infty}^{t} g_1^{1/2}\beta_1 d\nu / G_1(t) \right] d\nu(x) = 0$. This implies B and A_2 are orthogonal.

For every $\beta_1 \in L^2(v)$, let $\beta_1^0(t) \equiv g_2^{1/2}(t) \int_{-\infty}^t g_1^{1/2}\beta_1 dv / G_1(t)$. By direct calculation we have

$$
\begin{aligned}
A_2\beta_1^0(x,t) &= f(x,t)\{g_2^{-1/2}(t)\beta_1^0(t) - D_2(\beta_1^0; g_1, g_2)/\alpha_0\} \\
&= f(x,t)\left\{\int_{-\infty}^t g_1^{1/2}\beta_1 dv - D_1(\beta_1; g_1, g_2)/\alpha_0\right\} \\
&= (A_1 - B)\beta_1(x,t)\,. \qquad (3.8)
\end{aligned}
$$

Now (3.7) and (3.8) imply that $\langle B\beta_1, A_2\beta_1\rangle = 0$ and $\langle B\beta_1, (A_1 - B)\beta_1\rangle = 0$ for all $\beta_1 \in L^2(v)$. Hence the lemma follows from the uniqueness assertion of the projection theorem.

PROOFS OF THEOREMS 2.1 AND 2.2. It is easy to see that Assumption S of Begun *et al.* (1983) holds and our Lemma 3.1 implies Proposition 2.1 of Begun *et al.* (1983). From Remark 1 after Lemma 3.1 and the effective score given in Lemma 3.2, the main work to complete the proof is the computation of $(B^*B)^{-1}$ and $K(s,t) \equiv \langle G_{1s}, (B^*B)^{-1}G_{1t}\rangle$, where $G_{1u}(\cdot) = [I(-\infty < \cdot \le u) - G_1(u)]g_1^{1/2}(\cdot)$.

Direct calculations (as in Luenberger (1969), pp. 150–153), yield

$$
(B^*B)\beta(t) = \left\{ R\beta(t)\frac{M(t)}{G_1(t)} - \int_t^\infty R\beta \frac{M}{G_1^2}\, dG_1 \right\} g_1^{1/2}(t)\,,
$$

where $R\beta(t) \equiv \beta(t)g_1^{-1/2}(t) - \int_{-\infty}^t \beta g_1^{1/2} dv / G_1(t)$ and $M(t) \equiv G_1^*(t) - G_2^*(t^-)$. It is straightforward to verify (as in equation (6.4) of Begun *et al.* (1983)) that

$$
(B^*B)^{-1}\beta(t) = \left\{ R\beta(t)\frac{G_1(t)}{M(t)} - \int_t^\infty R\beta \frac{G_1}{M}\frac{dG_1}{G_1} \right\} g_1^{1/2}(t)\,.
$$

By some careful calculations we find that

$$
\begin{aligned}
&(B^*B)^{-1}G_{1t}(\cdot) \\
&\quad = -G_1(t)\left\{ \frac{I(t < \cdot \le U_1)}{M(\cdot)} - \int_{\cdot}^{U_1} \frac{I(t < x \le U_1)}{M(x)G_1(x)}\, dG_1(x) \right\} g_1^{1/2}(\cdot)\,,
\end{aligned}
$$

and hence

$$
\begin{aligned}
&K(s,t) \\
&\quad = G_1(t)\int_{-\infty}^{U_1} [1 - I(s \le y \le U_1) - G_1(s)]
\end{aligned}
$$

$$\cdot \left[\int_y^{U_1} \frac{-I(t < x \leq U_1)}{M(x)G_1(x)} \, dG_1(x) \right] g_1(y) dv(y)$$

$$= G_1(t) \left\{ \int_{s \vee t}^{U_1} \frac{dG_1}{M} - \int_{s \vee t}^{U_1} [G_1 - G_1(s)] \frac{1}{M} \frac{dG_1}{G_1} \right\}$$

$$= G_1(s) G_1(t) \int_{s \vee t}^{U_1} \frac{dG_1}{MG_1}$$

$$= G_1(s) G_1(t) \int_{s \vee t}^{\infty} \frac{dG_1^*}{M^2}, \tag{3.9}$$

by the facts that $dG_1^*/M = dG_1/G_1$ and $\inf\{x\colon G_1^*(x) = 1\} = U_1$. The function $K(s,t)$ equals the covariance function in (2.2). Theorems 2.1 and 2.2 follow by similar arguments used in Begun *et al.* (1983).

Remark 2. Note that $\sigma_s^2(s)$ in (1.13) is equal to $K(s,s)$ in (3.9). That is, both the "functional approach" and the "likelihood approach" provide the same covariance structure of the Gaussian process $\{Z(x)\colon -\infty < x < U_1\}$ in the representation theorem. Also $K(s,t)$ coincides with the covariance function of the asymptotic distribution of process $n^{1/2}(\hat{G}_{1n} - G_1)$ where $\hat{G}_{1n}$ is the product-limit estimator of G_1 defined in (1.3).

Remark 3. Under independence assumption, the effective score in (3.6) can be used to set up the nonparametric effective (or efficient) score equation. The equation can then provide functional representation (1.5) (in Section 1) and its estimate, and representations (7) and (8) of Wang *et al.* ((1986), p. 1601), and representations (4) and (7) of Woodroofe (1985). See Huang (1986) for a discussion of this relationship.

REFERENCES

Begun, J. M. and Wellner, J. A. (1983). Asymptotic efficiency of relative risk estimates, *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, (ed. P. K. Sen), 47-62, North-Holland Publishing Company.

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

Beran, R. (1977). Estimating a distribution function, *Ann. Statist.*, **5**, 400-404.

Bhattacharya, P. K. (1983). Justification for a K-S type test for the slope of a truncated regression, *Ann. Statist.*, **11**, 697-701.

Bhattacharya, P. K., Chernoff, H. and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression, *Ann. Statist.*, **11**, 505-514.

Bickel, P. J. (1982). On adaptive estimation, *Ann. Statist.*, **10**, 647-671.

Bickel, P. J. and Wellner, J. A. (1983). Lecture Notes of the Mathematical Sciences Lectures at Johns Hopkins University in June, 1983.

Bickel, P. J., Klaassen, C. A. J., Rifov, Y. and Wellner, J. A. (1987). *Efficient and Adaptive Inference in Semiparametric Models*, Forthcoming monograph, Johns Hopkins University Press, Baltimore.

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

Dieudonn'e, J. (1969). *Foundation of Modern Analysis* (enlarged and corrected printing), Academic Press, New York.

Hajek, J. (1970). A characterization of limiting distributions of regular estimates, *Z. Wahrsch. Verw. Gebiete.*, **14**, 323–330.

Huang, Wei-Min (1986). Estimating a distribution via effective score (complete and incomplete data), *1986 International Statistical Symp. Volumes*, **I**, 391–406.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniform property, *Ann. Inst. Statist. Math.*, **22**, 1–13.

LeCam, L. (1969). *Theorie Asymptotique de la Decision Statistique*, Les Press de l'Universite de Montreal.

Luenberger, D. G. (1969). *Optimization by Vector Space Method*, Wiley, New York.

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

Millar, P. W. (1979). Asymptotic minimax theorems for the sample distribution function, *Z. Wahrsch. Verw. Gebiete.*, **48**, 233–252.

Nicoll, J. F. and Segal, I. E. (1980). Nonparametric elimination of the observational cut-off bias, *Astronom. and Astrophys.*, **82**, L3–L6.

Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory*, Lecture Notes in Statistics, **13**, Springer-Verlag.

Segal, I. E. (1975). Observational validation of the chronometric cosmology: I. Preliminaries and the red shift-magnitude relation, *Proc. Nat. Acad. Sci. U.S.A.*, **72**, 2437–2477.

Stein, C. (1956). Efficient nonparametric testing and estimation, *Proc. Third Berkeley Symp. Math. Statist. Prob.*, Vol. 1, 187–195, University of California Press, Berkeley.

Wang, M.-C. and Jewell, N. P. (1985). The product limit estimate of a distribution function under random truncation, Technical Report, Program in Statistics, University of California, Berkeley.

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.

Wellner, J. A. (1982). Asymptotic optimality of the product-limit estimator, *Ann. Statist.*, **10**, 595–602.

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