Ann. Inst. Statist. Math.
Vol. 44, No. 2, 313–333 (1992)

A RANK ESTIMATOR IN THE TWO-SAMPLE TRANSFORMATION MODEL WITH RANDOMLY CENSORED DATA

HIDEATSU TSUKAHARA*

Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

(Received November 1, 1990; revised April 15, 1991)

Abstract. We consider the transformation model which is a generalization of Lehmann alternatives model. This model contains a parameter θ and a nonparametric part F_1 which is a distribution function. We propose a kind of M-estimator of θ based on ranks in the presence of random censoring. It is nonparametric in the sense that we do not have to know F_1. Moreover, it is simple and asymptotically normal. For the proportional hazards model with special censoring, we obtain the asymptotic relative efficiency of our estimator with respect to the best nonparametric estimator for this model. It is quite efficient for special values of θ. We also make a comparison between our estimator and other proposed estimators with real data.

Key words and phrases: Transformation models, M-estimator based on ranks, proportional hazards model, censored data, product-limit estimator, empirical processes.

1. Introduction

We consider the transformation model which is a generalization of Lehmann alternatives model (Lehmann (1953)). In this model, we have samples from distributions with distribution functions (df's) F_1 and $F_2 = D(F_1; \theta)$ where $D(\cdot\,; \theta)$ is a parametrized transformation. We then make inference about the parameter θ without knowing F_1. In this sense, transformation model is a semiparametric model (Wellner (1986)).

For the first sample, let $X_1^\circ, X_2^\circ, \dots, X_m^\circ$ be independently identically distributed (iid) positive random variables (rv's) with df F_1. For each X_i°, there corresponds a positive rv C_{1i} which is independent of X_i°'s and iid with df G_1. X_i°'s are survival times and C_{1i}'s are censoring times. We can only observe $(X_1, \delta_1), (X_2, \delta_2), \dots, (X_m, \delta_m)$, where

$$X_i \triangleq X_i^\circ \wedge C_{1i}, \quad \delta_i \triangleq I_{[X_i = X_i^\circ]}, \qquad i = 1, 2, \dots, m,$$

* Now at Department of Statistics, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, U.S.A.

$x \wedge y$ denotes $\min(x, y)$ and I_A is the indicator function of a set A.

Similarly for the second sample, let $Y_1^\circ, Y_2^\circ, \ldots, Y_n^\circ$ be iid positive rv's denoting survival times with df F_2. Independent of the Y_j's, let C_{2j}, $j = 1, 2, \ldots, n$ be also iid positive rv's denoting censoring times with df G_2. Then $(Y_1, \epsilon_1), (Y_2, \epsilon_2), \ldots, (Y_n, \epsilon_n)$ are observed where

$$Y_j \triangleq Y_j^\circ \wedge C_{2j}, \quad \epsilon_j \triangleq I_{[Y_j = Y_j^\circ]}, \qquad j = 1, 2, \ldots, n.$$

Further we assume all df's F_1, F_2, G_1, G_2 are continuous and have pdf's f_1, f_2, g_1, g_2 respectively.

Under the above assumptions X_i's and Y_j's are iid with df's H_1 and H_2 defined by

$$(1.1) \qquad 1 - H_1(x) = P\{X_i > x\} = (1 - F_1(x))(1 - G_1(x)),$$

$$(1.2) \qquad 1 - H_2(y) = P\{Y_j > y\} = (1 - F_2(y))(1 - G_2(y)).$$

We define the following sub-df's for later use.

$$(1.3) \qquad H_1^u(t) \triangleq P\{X_i \le t, \delta_i = 1\} = \int_0^t (1 - G_1) dF_1,$$

$$(1.4) \qquad H_1^c(t) \triangleq P\{X_i \le t, \delta_i = 0\} = \int_0^t (1 - F_1) dG_1,$$

$$(1.5) \qquad H_2^u(t) \triangleq P\{Y_j \le t, \epsilon_j = 1\} = \int_0^t (1 - G_2) dF_2,$$

$$(1.6) \qquad H_2^c(t) \triangleq P\{Y_j \le t, \epsilon_j = 0\} = \int_0^t (1 - F_1) dG_2.$$

Then clearly we have $H_1 = H_1^u + H_1^c$ and $H_2 = H_2^u + H_2^c$.

Now we shall define the two-sample transformation model precisely. Following Miura (1985), the model is expressed as

$$(1.7) \qquad F_2(t) = D(F_1(t); \theta), \qquad \theta \in \Theta \subset \mathrm{R}^1,$$

where Θ is a parameter space and for each θ $D(u; \theta)$ is a continuous df on $(0, 1)$ whose functional form is known and has pdf $d(u; \theta)$. Furthermore, we assume that $D(u; \theta)$ is monotonically increasing in θ and continuously differentiable in both u and θ.

In this model we consider inference for the parameter θ. In the next section, for the case when F_1 is unknown, we suggest a nonparametric estimator of θ based on ranks.

A particular example of our framework is the proportional hazards model.

Example 1. (proportional hazards model) Let $D(u; \theta) = 1 - (1 - u)^\theta$, $0 < \theta < \infty$. Then

$$(1.8) \qquad \lambda_1(t) = \theta \lambda_2(t),$$

where λ_1 and λ_2 denote hazard functions corresponding to F_1 and F_2 respectively (Cox (1972)).

For other examples, see Dabrowska *et al.* (1989) (hereafter DDM). In the sequel we mainly use the proportional hazards model for illustration of our procedure.

For the above model, without censoring DDM suggest two estimators based on ranks and prove their asymptotic normality. In the present paper, we consider an extension of one of them to the model with randomly censored data. In Section 2, our estimator is defined in the presence of censoring. The estimator is a kind of M-estimator and obtained by solving an estimating equation. It is rather simple, especially for the proportional hazards model. Moreover, in Section 3, our estimator turns out to be asymptotically normal under certain mild regularity conditions. For the proportional hazards model with special censoring, we obtain the asymptotic relative efficiency of our estimator with respect to the maximum partial likelihood estimator, which is known to be the best nonparametric estimator for this model. We see that our estimator is quite efficient near $\theta = 1$.

2. RAM estimator

First we shall consider the case when F_1 is known, and introduce the M-estimator which depends on F_1. For this purpose let us calculate the likelihood of $(Y_1, \epsilon_1), (Y_2, \epsilon_2), \ldots, (Y_n, \epsilon_n)$. If F_1 is known we only have to consider the second sample, because by transforming

$$U_i \triangleq F_1(X_i), \qquad W_j \triangleq F_1(Y_j),$$

the joint distribution of $U_1, \ldots, U_m, W_1, \ldots, W_n$ has the following survival function

$$\prod_{i=1}^{m}(1-u_i)[1-G_1(F_1^{-1}(u_i))] \cdot \prod_{j=1}^{n}(1-D(w_j;\theta))[1-G_2(F_1^{-1}(w_j))],$$

and therefore $W_1, \ldots, W_n$ are sufficient for θ.

Denoting the likelihood of a single observation (Y_j, ϵ_j) by $L(y_j, \epsilon_j)$, we have

$$\begin{aligned} L(y_j, \epsilon_j) &= \begin{cases} f_2(y_j)[1-G_2(y_j)] & \text{if } \epsilon_j = 1 \\ g_2(y_j)[1-F_2(y_j)] & \text{if } \epsilon_j = 0 \end{cases} \\ &= \{f_2(y_j)[1-G_2(y_j)]\}^{\epsilon_j} \cdot \{g_2(y_j)[1-F_2(y_j)]\}^{1-\epsilon_j}. \end{aligned}$$

Hence the total likelihood L of $(Y_1, \epsilon_1), (Y_2, \epsilon_2), \ldots, (Y_n, \epsilon_n)$ becomes

$$\begin{aligned} L &= \prod_{j=1}^{n} L(y_j, \epsilon_j) \\ &= \prod_{u} f_2(y_j)[1-G_2(y_j)] \cdot \prod_{c} g_2(y_j)[1-F_2(y_j)], \qquad j = 1, \ldots, n, \end{aligned}$$

where $\prod_u$ and $\prod_c$ denote products over uncensored observations and over censored ones respectively. Now the terms concerning g_2 and G_2 may be considered as constants for maximum likelihood estimation since they do not depend on θ, and hence we may regard L as

$$L = \prod_u f_2(y_j) \cdot \prod_c [1 - F_2(y_j)].$$

Taking logarithms gives

$$\log L = \sum_u \log f_2(y_j) + \sum_c \log[1 - F_2(y_j)],$$

where, similarly as above, $\sum_u$ and $\sum_c$ denote the sums over uncensored and censored samples, respectively. Further,

$$\frac{\partial \log L}{\partial \theta} = \sum_u \frac{\dot{f}_2(y_j)}{f_2(y_j)} + \sum_c \left[-\frac{\dot{F}_2(y_j)}{1 - F_2(y_j)} \right],$$

where

$$\dot{f}_2(y_j) \triangleq \frac{\partial}{\partial \theta} f_2(y_j), \qquad \dot{F}_2(y_j) \triangleq \frac{\partial}{\partial \theta} F_2(y_j).$$

By (1.7)

$$f_2 = f_1 \cdot d(F_1; \theta), \qquad \dot{F}_2 = \dot{D}(F_1; \theta),$$
$$d(u; \theta) = \frac{\partial}{\partial u} D(u; \theta), \qquad \dot{D}(u; \theta) \triangleq \frac{\partial}{\partial \theta} D(u; \theta),$$

hence we have

$$\frac{\partial \log L}{\partial \theta} = \sum_{j=1}^{n} \left[\epsilon_j \frac{\dot{d}(w_j; \theta)}{d(w_j; \theta)} - (1 - \epsilon_j) \frac{\dot{D}(w_j; \theta)}{1 - D(w_j; \theta)} \right], \tag{2.1}$$

where $w_j \triangleq F_1(y_j)$. To obtain further expression, we state the following lemma (c.f. James (1986)).

LEMMA 2.1. *Assume that for $d(w; \theta)$ the order of differentiation in θ and integration in w are interchangeable. If we set*

$$\Delta(w; \theta) \triangleq \frac{\dot{d}(w; \theta)}{d(w; \theta)} = \frac{\partial}{\partial \theta} \log d(w; \theta), \tag{2.2}$$

then

$$-\frac{\dot{D}(w; \theta)}{1 - D(w; \theta)} = E[\Delta(W^\circ; \theta) \mid W^\circ > w], \tag{2.3}$$

where Y° is a rv with df F_2 and $W^\circ \triangleq F_1(Y^\circ)$.

PROOF.

$$\begin{aligned}
E[\Delta(W^\circ;\theta) \mid W^\circ > w] &= \frac{1}{P\{W^\circ > w\}} \int_w^\infty \Delta(w^\circ;\theta) d(w^\circ;\theta) dw^\circ \\
&= \frac{1}{P\{W^\circ > w\}} \int_w^\infty \dot{d}(w^\circ;\theta) dw^\circ \\
&= \frac{1}{1 - D(w;\theta)} \cdot \frac{\partial}{\partial\theta} \int_w^\infty d(w^\circ;\theta) dw^\circ \\
&= \frac{1}{1 - D(w;\theta)} \cdot \frac{\partial}{\partial\theta} [1 - D(w;\theta)] \\
&= -\frac{\dot{D}(w;\theta)}{1 - D(w;\theta)}. \qquad \square
\end{aligned}$$

Using (2.2) and (2.3), (2.1) can be written as

$$\frac{\partial \log L}{\partial\theta} = \sum_{j=1}^n [\epsilon_j \Delta(w_j;\theta) + (1-\epsilon_j) E\{\Delta(W^\circ;\theta) \mid W^\circ > w_j\}].$$

Note that $E[\partial \log L/\partial\theta] = 0$ and MLE is defined by $\partial \log L/\partial\theta = 0$. Now let η be any function satisfying $E_\theta[\eta(W^\circ;\theta)] = 0$. Then as proved in Lemma 3.1 below, we can show

$$E[\epsilon_j \eta(W_j;\theta) + (1-\epsilon_j) E\{\eta(W^\circ;\theta) \mid W^\circ > W_j\}] = E[\eta(W^\circ;\theta)] = 0. \tag{2.4}$$

Therefore a natural score function ψ for M-estimation may be defined by

$$\psi(w_j, \epsilon_j;\theta) \triangleq \epsilon_j \psi_0(w_j;\theta) + (1-\epsilon_j) E\{\psi_0(W^\circ;\theta) \mid W^\circ > w_j\} \tag{2.5}$$

where ψ_0 is any function satisfying $E_\theta[\psi_0(W^\circ;\theta)] = 0$. For brevity, let us write

$$\psi_1(w_j;\theta) \triangleq E\{\psi_0(W^\circ;\theta) \mid W^\circ > w_j\}, \tag{2.6}$$

so that

$$\psi(w_j, \epsilon_j;\theta) = \epsilon_j \psi_0(w_j;\theta) + (1-\epsilon_j) \psi_1(w_j;\theta).$$

Here we further assume that ψ_k, $k = 0, 1$ are monotonically increasing in θ. We then define an M-estimator (Huber (1981)) of θ as a solution of the equation

$$\sum_{j=1}^n \psi(W_j, \epsilon_j;\theta) = 0$$

Now we turn to the main case where F_1 is unknown. Then we cannot transform Y_j by F_1. However we can replace F_1 by some estimator of F_1. For the censored case we shall take Kaplan-Meier PL(product limit)-estimator $\boldsymbol{F}_{1m}$ as an estimator of F_1 (Kaplan and Meier (1958)). It is defined as follows: let $X_{(1)} < X_{(2)} < \cdots <$

$X_{(m)}$ be order statistics in the first sample, and $\delta_{(i)}$ be the δ_i corresponding to $X_{(i)}$. Then $\boldsymbol{F}_{1m}$ is defined by

$$1 - \boldsymbol{F}_{1m}(t) \triangleq \begin{cases} \prod_{X_{(i)} \leq t} \left(1 - \frac{1}{m-i+1}\right)^{\delta_{(i)}} & \text{if } t < \tau_m, \\ 0 & \text{if } t \geq \tau_m, \end{cases}$$

where $\tau_m \triangleq X_{(m)}$. In order that the proposed estimator is applicable for small samples, we define

$$\hat{\boldsymbol{F}}_{1m}(t) \triangleq \begin{cases} \frac{1}{N+1} & \text{if } 0 < t < X^u_{(1)}, \\ \boldsymbol{F}_{1m}(t) & \text{if } X^u_{(1)} \leq t < \tau_m, \\ \frac{N}{N+1} & \text{if } \tau_m \leq t, \end{cases}$$

where $N \triangleq m + n$ and

$$X^u_{(1)} \triangleq \min_{\substack{1 \leq i \leq m \\ \delta_i = 1}} X_i.$$

Further we set $\hat{W}_j \triangleq \hat{\boldsymbol{F}}_{1m}(Y_j)$ and then define our RAM (Rank Approximate M) estimator $\hat{\theta}_N$ of θ by a solution of the equation

$$\sum_{j=1}^{n} \psi(\hat{W}_j, \epsilon_j; \theta) = 0.$$

More precisely, if we set

$$\theta^*_N \triangleq \sup\left\{\theta : \sum_{j=1}^{n} \psi(\hat{W}_j, \epsilon_j; \theta) > 0\right\} \quad \text{and}$$

$$\theta^{**}_N \triangleq \inf\left\{\theta : \sum_{j=1}^{n} \psi(\hat{W}_j, \epsilon_j; \theta) < 0\right\},$$

then the RAM estimator is defined by

$$\hat{\theta}_N \triangleq \frac{\theta^*_N + \theta^{**}_N}{2}.$$

If $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ are order statistics of $Y_1, Y_2, \ldots, Y_n$, then $\hat{\boldsymbol{F}}_{1m}(Y_{(j)})$ has the same information about θ as rank of $Y_{(j)}$. Therefore we can interpret $\hat{\theta}_N$ as an approximate M-estimator based on ranks. That is why we call it RAM estimator as in DDM.

In the absence of censoring, Cuzick (1988) developed a similar estimation procedure in the linear transformation model which is related to, but essentially different from our model.

Example 1. (continued) Under the proportional hazards model, we have

$$d(u;\theta) = \theta(1-u)^{\theta-1}, \qquad \dot{d}(u;\theta) = [1+\theta\log(1-u)](1-u)^{\theta-1}.$$

Hence we compute

$$\psi_0(u;\theta) \triangleq \frac{\dot{d}(u;\theta)}{d(u;\theta)} = \frac{1}{\theta} + \log(1-u). \tag{2.7}$$

Similarly the equation $\dot{D}(u;\theta) = -\log(1-u)\cdot(1-u)^{\theta}$ leads to

$$\psi_1(u;\theta) \triangleq -\frac{\dot{D}(u;\theta)}{1-D(u;\theta)} = \log(1-u). \tag{2.8}$$

Then both $\psi_0(u;\theta)$ and $\psi_1(u;\theta)$ are nonincreasing in θ, and hence $\hat{\theta}_N$ becomes a solution of the equation

$$\sum_{j=1}^{n}[\epsilon_j(\theta^{-1} + \log(1-\hat{\boldsymbol{F}}_{1m}(Y_j))) + (1-\epsilon_j)\log(1-\hat{\boldsymbol{F}}_{1m}(Y_j))] = 0.$$

Solving this we obtain

$$\hat{\theta}_N = \frac{\sum_{j=1}^{n}\epsilon_j}{\sum_{j=1}^{n}[-\log(1-\hat{\boldsymbol{F}}_{1m}(Y_j))]}. \tag{2.9}$$

Note that the numerator represents the number of uncensored observations in the second sample, and that $-\log\hat{\boldsymbol{F}}_{1m}$ in the denominator can be viewed as an estimator of the cumulative hazard function corresponding to F_1.

3. Asymptotic theory

3.1 *Notations and assumptions*

Let us denote the empirical (sub-)df's of H_2, H_2^u, H_2^c by

$$\boldsymbol{H}_{2n}(t) \triangleq \frac{1}{n}\sum_{j=1}^{n} I_{[Y_j\le t]}, \qquad \boldsymbol{H}_{2n}^u(t) \triangleq \frac{1}{n}\sum_{j=1}^{n} I_{[Y_j\le t]}\epsilon_j,$$
$$\boldsymbol{H}_{2n}^c(t) \triangleq \frac{1}{n}\sum_{j=1}^{n} I_{[Y_j\le t]}(1-\epsilon_j),$$

respectively. Next, we shall define the (sub-)df and empirical one of $W_j \triangleq F(Y_j)$. Hereafter we denote the composition of two functions f and g by fg. With this rule we have

$$W_j^\circ \triangleq F_1(Y_j^\circ) \overset{d}{\sim} D(\,\cdot\,;\theta), \qquad F_1(C_{2j}) \overset{d}{\sim} G_2F_1^{-1},$$

and so we define the (sub-)df of W_j by

$$\begin{aligned}1-E_2 &\triangleq (1-D)(1-G_2F_1^{-1}) = H_2F_1^{-1},\\ E_2^u(t) &\triangleq \int_0^t (1-G_2F_1^{-1})dD = H_2^uF_1^{-1}(t),\\ E_2^c(t) &\triangleq \int_0^t (1-D)dG_2F_1^{-1} = H_2^cF_1^{-1}(t).\end{aligned}$$

Empirical ones are also defined by

$$\begin{aligned}\boldsymbol{E}_{2n}(t) &\triangleq \frac{1}{n}\sum_{j=1}^n I_{[W_j\le t]} = \boldsymbol{H}_{2n}F_1^{-1}(t),\\ \boldsymbol{E}_{2n}^u(t) &\triangleq \frac{1}{n}\sum_{j=1}^n I_{[W_j\le t]}\epsilon_j = \boldsymbol{H}_{2n}^uF_1^{-1}(t),\\ \boldsymbol{E}_{2n}^c(t) &\triangleq \frac{1}{n}\sum_{j=1}^n I_{[W_j\le t]}(1-\epsilon_j) = \boldsymbol{H}_{2n}^cF_1^{-1}(t).\end{aligned}$$

Next let us define empirical processes which we need for the proof of the asymptotic normality of $\hat{\theta}_N$. For the second sample, the processes

$$\boldsymbol{V}_n^u(H_2^u) \triangleq \sqrt{n}(\boldsymbol{H}_{2n}^u - H_2^u), \qquad \boldsymbol{V}_n^c(H_2^c) \triangleq \sqrt{n}(\boldsymbol{H}_{2n}^c - H_2^c) \tag{3.1}$$

appear and for the first sample we need the Kaplan-Meier process:

$$\boldsymbol{X}_m \triangleq \sqrt{m}(\boldsymbol{F}_{1m} - F_1). \tag{3.2}$$

For the asymptotic argument of this process we define the following functions;

$$C(t) \triangleq \int_0^t \frac{1}{(1-H_1)^2}dH_1^u, \tag{3.3}$$

$$K(t) \triangleq \frac{C(t)}{1+C(t)}. \tag{3.4}$$

Furthermore let $\boldsymbol{V}_0$ and $\boldsymbol{V}_1$ be independent Brownian bridges and $\boldsymbol{X} \triangleq (1-F)\boldsymbol{B}$, where $\boldsymbol{B} \cong \boldsymbol{S}(C)$ and $\boldsymbol{S}$ is the standard Brownian motion on $[0,\infty)$. Here $X \cong Y$ means that X and Y have the same distribution. We denote the sup-norm over an interval $[a,b]$ by $\|\cdot\|_a^b$ and when a and b are omitted it represents the sup-norm over $[0,\infty)$.

We now state our assumptions.

Assumption 1. (1) If we set

$$\begin{aligned}\tau_{F_1} &\triangleq \inf\{t : F_1(t)=1\}, \qquad \tau_{G_1} \triangleq \inf\{t : G_1(t)=1\},\\ \tau_{H_1} &\triangleq \inf\{t : H_1(t)=1\} = \tau_{F_1}\wedge\tau_{G_1},\end{aligned}$$

we assume that $\tau_{F_1} = \tau_{G_1} = \tau_{H_1} = \infty$.

(2) Let $\lambda_N \triangleq m/N$. Then, for some $\lambda_0 \in (0, 1/2)$,

$$\lambda_0 \leq \lambda_N \leq 1 - \lambda_0, \qquad \text{for all} \quad N \geq 1,$$

and there exists $\lambda > 0$ satisfying

$$\lambda_N \to \lambda, \qquad \text{as} \quad N \to \infty.$$

On the score functions ψ_0 and ψ_1 and the (sub-)df's E_2, E_2^u and E_2^c, we put the following assumption.

ASSUMPTION 2. Define

$$\psi_k'(t;\theta) \triangleq \frac{\partial}{\partial t}\psi_k(t;\theta), \quad \dot{\psi}_k(t;\theta) \triangleq \frac{\partial}{\partial \theta}\psi_k(t;\theta), \qquad k = 0, 1.$$

Assume that for θ in a neighborhood of the true parameter value θ_0, the following conditions (1)–(4) hold.

(1) For some δ satisfying $0 < \delta < 1/2$,

$$\text{a)}\ |\psi_k(t;\theta)| \leq M[t(1-t)]^{-1/2+\delta}, \qquad \text{b)}\ |\psi_k'(t;\theta)| \leq M[t(1-t)]^{-3/2+\delta},$$

where M is a universal constant.

(2) For the same δ in (1),

$$\int_0^1 t^{-2+\delta}[CF_1^{-1}(t)]^{1-\delta/2} dE_2(t) < \infty,$$

uniformly in θ.

(3) For the same δ,

$$\int_{F_1(\tau_m)}^1 (1-t)^{-1/2+\delta} dE_2(t) = o_p(m^{-1/2}).$$

(4)

$$\text{a)} \int_0^1 \dot{\psi}_0(t;\theta) dE_2^u(t) < \infty, \qquad \text{b)} \int_0^1 \dot{\psi}_1(t;\theta) dE_2^c(t) < \infty,$$

uniformly in θ.

Let us interpret our assumptions. Assumption 2(1) is an ordinary assumption on the smoothness of the score functions which has been very often used in non-parametrics, and Assumption 2(4) has the same meaning on the smoothness in the argument θ. Assumption 2(2) is a technical and direct assumption ensuring the convergence of the process, and so seems with Assumption 2(3). But, as easily verified, it is concerned with the censoring distribution G_1 and a measure which reflects "the lightness of censoring"; it can be satisfied if the convergence of $F_1(\tau_m)$ to 1 is appropriately rapid. Noting that $F_1(\tau_m)$ is the largest order statistic from the distribution with df $E_1 = H_1 F_1^{-1}$ and (1.1), E_1 should have a density whose derivative at 1 does not exist, or at least sufficiently large. Therefore G_1 puts mass on the part of the positive real line bounded away from the origin. A simple sufficient condition for this is $1 - F_1(\tau_m) = o_p(m^{-1/2})$.

Remark 3.1. If $G_1 \equiv 0$, that is, if there is no censoring, easy calculation shows that $C = F_1/(1-F_1)$, so that Assumption 2(2) reduces to Assumption 2(2) in DDM. Also, Assumption 2(3) is satisfied since $F_1(\tau_m) \overset{d}{\sim} U(0,1)$.

3.2 *Asymptotic normality*

We begin with the integral representation of the statistic $\sum_{j=1}^n \psi(\hat{W}_j, \epsilon_j; \theta)$ which is used to define our estimator:

$$S_N(\theta) \triangleq \frac{1}{n}\sum_{j=1}^{n}[\epsilon_j\psi_0(\hat{W}_j;\theta) + (1-\epsilon_j)\psi_1(\hat{W}_j;\theta)] \tag{3.5}$$
$$= \int_0^\infty \psi_0(\hat{\boldsymbol{F}}_{1m};\theta)d\boldsymbol{H}_{2n}^u + \int_0^\infty \psi_1(\hat{\boldsymbol{F}}_{1m};\theta)d\boldsymbol{H}_{2n}^c.$$

The following lemma shows that the expectation of $S_N(\theta_0)$ is 0.

LEMMA 3.1. *If the true value of θ is θ_0, then*

$$\mu \triangleq \int_0^\infty \psi_0(F_1;\theta_0)dH_2^u + \int_0^\infty \psi_1(F_1;\theta_0)dH_2^c = 0.$$

PROOF. By the change of variable,

$$\begin{aligned}
\mu &= \int_0^1 \psi_0(t;\theta_0)dE_2^u(t) + \int_0^1 \psi_1(t;\theta_0)dE_2^c(t) \\
&= \int_0^1 \psi_0(t;\theta_0)[1 - G_2F_1^{-1}(t)]dD(t;\theta) \\
&\quad + \int_0^1 \psi_1(t;\theta_0)[1 - D(t;\theta_0)]dG_2F_1^{-1}(t).
\end{aligned}$$

Using (2.6) and the identity $1 - G_2F_1^{-1}(t) = \int_t^1 dG_2F_1^{-1}(s)$, we obtain

$$\begin{aligned}
\mu &= \int_0^1 \psi_0(t;\theta_0)\left[\int_t^1 dG_2F_1^{-1}(s)\right]dD(t;\theta_0) \\
&\quad + \int_0^1\int_t^1 \psi_0(s;\theta_0)dD(s;\theta_0)dG_2F_1^{-1}(t) \\
&= \int_0^1 \psi_0(t;\theta_0)\left[\int_t^1 dG_2F_1^{-1}(s)\right]dD(t;\theta_0)
\end{aligned}$$

$$+\int_0^1 \psi_0(s;\theta_0)\left[\int_0^s dG_2F_1^{-1}(t)\right]dD(s;\theta_0)$$

$$=\int_0^1 \psi_0(t;\theta_0)\left[\int_0^1 dG_2F_1^{-1}(s)\right]dD(t;\theta_0)$$

$$=\int_0^1 \psi_0(t;\theta_0)dD(t;\theta_0),$$

which is 0 because of the condition on $\psi_0 : E_\theta[\psi_0(W^\circ;\theta)] = 0,\ W^\circ \sim D(\cdot\,;\theta)$. □

We are now ready to prove the asymptotic normality of RAM estimator.

THEOREM 3.1. *Under Assumptions* 1 *and* 2, *as* $N \to \infty$,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, \sigma^2(\theta_0)),$$

where

$$\sigma^2(\theta) \triangleq \frac{\frac{1}{1-\lambda}\tau_1(\theta) + \frac{1}{\lambda}\tau_2(\theta)}{\left[\int_0^\infty \dot{\psi}_0(F_1;\theta)dH_2^u + \int_0^\infty \dot{\psi}_1(F_1;\theta)dH_2^c\right]^2},$$

and

$$\begin{aligned}\tau_1(\theta) &\triangleq \int_0^\infty [\psi_0(F_1;\theta)]^2 dH_2^u + \int_0^\infty [\psi_1(F_1;\theta)]^2 dH_2^c \\ &\quad - \left\{\int_0^\infty \psi_0(F_1;\theta)dH_2^u + \int_0^\infty \psi_1(F_1;\theta)dH_2^c\right\}^2, \\ \tau_2(\theta) &\triangleq 2\biggl[\iint_{s<t} (1-F_1(s))(1-F_1(t))C(s) \\ &\qquad\qquad \cdot \psi_0'(F_1(s);\theta)\psi_0'(F_1(t);\theta)dH_2^u(s)dH_2^u(t) \\ &\quad + \iint (1-F_1(s))(1-F_1(t))C(s\wedge t) \\ &\qquad\qquad \cdot \psi_0'(F_1(s);\theta)\psi_1'(F_1(t);\theta)dH_2^u(s)dH_2^c(t) \\ &\quad + \iint_{s<t} (1-F_1(s))(1-F_1(t))C(s)\psi_1'(F_1(s);\theta) \\ &\qquad\qquad \cdot \psi_1'(F_1(t);\theta)dH_2^c(s)dH_2^c(t)\biggr],\end{aligned}$$

provided $\sigma^2(\theta) > 0$.

PROOF. Let $\theta \triangleq \theta_0 + N^{-1/2}b$, and decompose $S_N(\theta)$ as follows:

$$S_N(\theta) = S_N^u(\theta) + S_N^c(\theta),$$

where

$$S_N^u(\theta) \triangleq \int_0^\infty \psi_0(\hat{\boldsymbol{F}}_{1m};\theta)d\boldsymbol{H}_{2n}^u, \qquad S_N^c(\theta) \triangleq \int_0^\infty \psi_1(\hat{\boldsymbol{F}}_{1m};\theta)d\boldsymbol{H}_{2n}^c.$$

In view of the symmetry of the assumptions on ψ_0 and ψ_1, we have to show only the convergence of $S_N^u(\theta)$, for the convergence of $S_N^c(\theta)$ can then be proved quite similarly.

As a method of proof, we shall adopt the so-called Pyke-Shorack approach (Pyke and Shorack (1968)), by which we obtain almost sure convergences of the processes on the specially constructed probability space. However, it should be remarked that only convergences in distribution is true on the original space.

We note that the convergences in the sequel are all true uniformly in b satisfying $|b| \le B$ for any fixed (sufficiently large) $B < \infty$.

Now let us show the convergence of $S_N^u(\theta)$. A simple algebra leads to the expression:

$$\sqrt{N}\, S_N^u(\theta) = A_{1N} + A_{2N} + A_{3N} + r_N,$$

where

$$\begin{aligned}
A_{1N} &\triangleq \sqrt{N} \int_0^\infty \psi_0(F_1;\theta) d\{\boldsymbol{H}_{2n}^u - H_2^u\}, \\
A_{2N} &\triangleq \sqrt{N} \int_0^\infty \psi_0'(F_1;\theta)(\boldsymbol{F}_{1m} - F_1) dH_2^u, \\
A_{3N} &\triangleq \sqrt{N} \int_0^\infty [\psi_0(F_1;\theta) - \psi_0(F_1;\theta_0)] dH_2^u, \\
r_N &\triangleq \sqrt{N} \int_0^\infty [\psi_0(\hat{\boldsymbol{F}}_{1m};\theta) - \psi_0(F_1;\theta) - \psi_0'(F_1;\theta)(\boldsymbol{F}_{1m} - F_1)] d\boldsymbol{H}_{2n}^u \\
&\quad + \sqrt{N} \int_0^\infty \psi_0'(F_1;\theta)(\boldsymbol{F}_{1m} - F_1) d\{\boldsymbol{H}_{2n}^u - H_2^u\}.
\end{aligned}$$

Here we used Lemma 3.1 implicitly. Using the results of Shorack and Wellner (1986) and Gill (1983) (see the Appendix), we can show the weak convergences of A_{1N}, A_{2N} and A_{3N} and asymptotic negligibility of r_N.

(i) A_{1N}.

$$\begin{aligned}
A_{1N} &= \sqrt{\frac{N}{n}} \int_0^\infty \psi_0(F_1;\theta) d\{\sqrt{n}(\boldsymbol{H}_{2n}^u - H_2^u)\} \\
&= \frac{1}{\sqrt{1-\lambda_N}} \int_0^1 \psi_0(t;\theta) d\{\sqrt{n}(\boldsymbol{E}_{2n}^u(t) - E_2^u(t))\} \\
&= \frac{1}{\sqrt{1-\lambda_N}} \int_0^1 \psi_0(t;\theta) d\{\boldsymbol{V}_n^u(E_2^u(t)) - \boldsymbol{V}_0(E_2^u(t))\} \\
&\quad + \frac{1}{\sqrt{1-\lambda_N}} \int_0^1 \psi_0(t;\theta) d\boldsymbol{V}_0(E_2^u(t)).
\end{aligned}$$

By Proposition A.2 in the Appendix, the first term converges to 0 almost surely as $N \to \infty$. Thus we get

$$A_{1N} \xrightarrow{a.s.} \frac{1}{\sqrt{1-\lambda_N}} \int_0^1 \psi_0(t;\theta_0) d\boldsymbol{V}_0(E_2^u(t)). \tag{3.6}$$

(ii) A_{2N}.

$$\begin{aligned}
A_{2N} &= \sqrt{\frac{1}{\lambda_N}} \int_0^\infty \psi_0'(F_1;\theta)\sqrt{m}(\boldsymbol{F}_{1m} - F_1)dH_2^u \\
&= \sqrt{\frac{1}{\lambda_N}} \int_0^{\tau_m} \psi_0'(F_1;\theta)\boldsymbol{X} dH_2^u + \sqrt{\frac{1}{\lambda_N}} \int_0^{\tau_m} \psi_0'(F_1;\theta)(\boldsymbol{X}_m - \boldsymbol{X})dH_2^u \\
&\quad + \sqrt{\frac{1}{\lambda_N}} \int_{\tau_m}^\infty \psi_0'(F_1;\theta)\sqrt{m}(1 - F_1)dH_2^u \\
&= A_{21N} + A_{22N} + A_{23N}, \qquad \text{say.}
\end{aligned}$$

Then

$$\begin{aligned}
A_{21N} &\xrightarrow{a.s.} \frac{1}{\sqrt{\lambda}} \int_0^\infty \psi_0'(F_1;\theta_0)\boldsymbol{X} dH_2^u \qquad \text{and} \\
|A_{22N}| &\le M \int_0^{\tau_m} [F_1(1 - F_1)]^{-3/2+\delta} \frac{1-K}{1-F_1}(\boldsymbol{X}_m - \boldsymbol{X})\frac{1-F_1}{1-K} dH_2^u \\
&\le \left\| \frac{1-K}{1-F_1} \frac{\boldsymbol{X}_m - \boldsymbol{X}}{[K(1-K)]^{(1-\delta)/2}} \right\|_0^{\tau_m} \\
&\quad \cdot \int_0^{\tau_m} K^{(1-\delta)/2} F_1^{-3/2+\delta}(1-K)^{-(1+\delta)/2}(1-F_1)^{-1/2+\delta} dH_2^u.
\end{aligned}$$

If we see the integral on the right is finite, then $A_{22N} \xrightarrow{P} 0$ since the factor before the integral is $o_p(1)$ by Proposition A.4. In fact, using Proposition A.5,

$$\begin{aligned}
&\int_0^{\tau_m} K^{(1-\delta)/2} F_1^{-3/2+\delta}(1-K)^{-(1+\delta)/2}(1-F_1)^{-1/2+\delta} dH_2^u \\
&\quad \le \int_0^{\tau_m} K^{(1-\delta)/2} F_1^{-3/2+\delta}(1-K)^{-1+\delta/2} dH_2^u \\
&\quad \le \int_0^{\tau_m} F_1^{-2+\delta} C^{1-\delta/2} dH_2^u \\
&\quad \le \int_0^1 t^{-2+\delta}(CF_1^{-1}(t))^{1-\delta/2} dE_2^u(t),
\end{aligned}$$

which is finite due to Assumption 2(2). Here the first inequality follows from the fact $1 - K \le 1 - F_1$ and the second one from $F_1 \le K$ and $K = C/(1+C)$. Therefore, we obtain $A_{22N} \xrightarrow{P} 0$. Finally,

$$\begin{aligned}
|A_{23N}| &\le M \int_{\tau_m}^\infty [F_1(1-F_1)]^{-3/2+\delta}\sqrt{m}(1-F_1)dH_2^u \\
&\le M \int_{\tau_m}^\infty \sqrt{m}(1-F_1)^{-1/2+\delta} dH_2^u \\
&= M\sqrt{m} \int_{F_1(\tau_m)}^\infty (1-t)^{-1/2+\delta} dE_2^u(t).
\end{aligned}$$

This converges to 0 in probability by virtue of Assumption 2(3), and therefore we conclude that

$$A_{2N} \xrightarrow{P} \frac{1}{\sqrt{\lambda}} \int_0^\infty \psi_0'(F_1;\theta_0)\boldsymbol{X} dH_2^u. \tag{3.7}$$

(iii) A_{3N}. By the mean value theorem, we can write

$$A_{3N} = \sqrt{N} \int_0^\infty \dot{\psi}_0(F_1;\theta^*)(\theta-\theta_0) dH_2^u,$$

where θ^* assumes a value between θ and θ_0. Noting that $\theta = \theta_0 + N^{-1/2}b$, it follows from Assumption 2(4) that

$$\begin{aligned} A_{3N} &= b\int_0^\infty \dot{\psi}_0(F_1;\theta^*) dH_2^u = b\int_0^\infty \dot{\psi}_0(t;\theta^*) dE_2^u(t) \\ &\to b\int_0^1 \dot{\psi}_0(t;\theta_0) dE_2^u(t). \end{aligned} \tag{3.8}$$

(iv) r_N. We shall decompose r_N as follows: $r_N = r_{1N} + r_{2N}$, where

$$\begin{aligned} r_{1N} &\triangleq \sqrt{N} \int_0^\infty [\psi_0(\hat{\boldsymbol{F}}_{1m};\theta) - \psi_0(F_1;\theta) - \psi_0'(F_1;\theta)(\boldsymbol{F}_{1m} - F_1)] d\boldsymbol{H}_{2n}^u, \\ r_{2N} &\triangleq \sqrt{N} \int_0^\infty \psi_0'(F_1;\theta)(\boldsymbol{F}_{1m} - F_1) d\{\boldsymbol{H}_{2n}^u - H_2^u\}. \end{aligned}$$

(iv-a) r_{1N}. Let $\gamma > 0$ be sufficiently small, and $r_{1N} = r_{11N} + r_{12N} + r_{13N}$, where the ranges of integration in r_{11N}, r_{12N} and r_{13N} are $(0, F_1^{-1}(\gamma))$, $[F_1^{-1}(\gamma), F_1^{-1}(1-\gamma)\wedge \tau_m]$ and $(F_1^{-1}(1-\gamma)\wedge\tau_m, \infty)$ respectively, while their integrands are the same. And F_1^* denotes a random function assuming the value between $\boldsymbol{F}_{1m}$ and F_1. For r_{11N}, let $\epsilon > 0$ be given. Then the fact that there exists some N_0 such that for all $N \geq N_0$, with probability greater than $1-\epsilon$,

$$\begin{aligned} |\psi_0'(F_1^*;\theta) - \psi_0'(F_1;\theta)| &\leq 2[(F_1^* \wedge F_1)(1 - F_1^* \wedge F_1)]^{-3/2+\delta} \\ &\leq 2M[F_1(1-F_1)]^{-3/2+\delta}, \end{aligned} \tag{3.9}$$

and the obvious relation $|\hat{\boldsymbol{F}}_{1m} - \boldsymbol{F}_{1m}| \leq 1/(m+1)$ together show that

$$|r_{11N}| \leq M \int_0^{F_1^{-1}(\gamma)} [F_1(1-F_1)]^{-3/2+\delta} |\boldsymbol{X}_m| d\boldsymbol{H}_{2n}^u.$$

It follows from an application of Proposition A.4 that

$$\left\| \frac{1-K}{1-F_1} \frac{\boldsymbol{X}_m}{[K(1-K)]^{(1-\delta)/2}} \right\| = O_p(1). \tag{3.10}$$

And then, as in the argument of A_{22N}, we obtain

$$|r_{11N}| \le O_p(1) \int_0^\gamma t^{-2+\delta} C^{1-\delta/2} dE_2^u,$$

which converges to 0 in probability as $\gamma \downarrow 0$ and $N \to \infty$.

Next we work on r_{12N}. Since ψ_0' is uniformly continuous on $[F_1^{-1}(\gamma), F_1^{-1}(1-\gamma)]$, $\|F_1^* - F_1\| \le \|\boldsymbol{F}_{1m} - F_1\| \xrightarrow{a.s.} 0$, and $\|\boldsymbol{X}_m\| = O_p(1)$, we have, for any fixed γ,

$$\begin{aligned} |r_{12N}| &= \left| \int_{F_1^{-1}(\gamma)}^{F_1^{-1}(\gamma)\wedge\tau_m} \sqrt{N}[\psi_0'(F_1^*;\theta) - \psi_0'(F_1;\theta)](\boldsymbol{F}_{1m} - F_1) d\boldsymbol{H}_{2n}^u \right| \\ &\le M O_p(1) \int_{F_1^{-1}(\gamma)}^{F_1^{-1}(\gamma)\wedge\tau_m} |\psi_0'(F_1^*;\theta) - \psi_0'(F_1;\theta)| d\boldsymbol{H}_{2n}^u. \end{aligned}$$

Hence, as $N \to \infty$, $r_{12N} \xrightarrow{P} 0$.

Concerning r_{13N}, it is obvious that it should be $o_p(1)$ on $(F_1^{-1}(1-\gamma), \tau_m]$ as before, so that it suffices to show that it converges to 0 in probability as $N \to \infty$ on $[\tau_m, \infty)$. But this follows immediately from (3.9), Proposition A.1 and Assumption 2(3). Therefore, we conclude that $r_{1N} \xrightarrow{P} 0$ as $\gamma \downarrow 0$ and $N \to \infty$.

(iv-b) r_{2N}. Let r_{21N} and r_{22N} equal r_{2N} with change of the range of integration into $[0, \tau_m)$ and $[\tau_m, \infty)$ respectively. Then, using (3.10), the same argument as for A_{22N} leads to

$$\begin{aligned} |r_{21N}| &\le \frac{1}{\sqrt{\lambda_N}} \int_0^{\tau_m} [F_1(1-F_1)]^{-3/2+\delta} |\boldsymbol{X}_m| d|\boldsymbol{H}_{2n}^u - H_2^u| \\ &\le O_p(1) \int_0^1 t^{-2+\delta} (C F_1^{-1}(t))^{1-\delta/2} d|\boldsymbol{E}_{2n}^u(t) - E_2^u(t)|, \end{aligned}$$

which converges to 0 in probability as $N \to \infty$ by Proposition A.1. As for A_{23N}, it is easy to show that r_{22N} also converges to 0 in probability.

Therefore, we finally conclude that, uniformly in b ($|b| \le B$),

$$\begin{aligned} \sqrt{N} S_N^u(\theta) \xrightarrow{P} &\frac{1}{\sqrt{1-\lambda}} \int_0^1 \psi_0(t;\theta_0) d\boldsymbol{V}_0(E_2^u(t)) \\ &+ \frac{1}{\sqrt{\lambda}} \int_0^1 \psi_0'(t;\theta_0) \boldsymbol{X}(F_1^{-1}(t)) dE_2^u(t) \\ &+ b \int_0^1 \dot{\psi}_0(t;\theta_0) dE_2^u(t). \end{aligned}$$

As mentioned at the beginning of the proof, convergence of $S_N^c(\theta)$ can be shown quite similarly; that is,

$$\begin{aligned} \sqrt{N} S_N^c(\theta) \xrightarrow{P} &\frac{1}{\sqrt{1-\lambda}} \int_0^1 \psi_1(t;\theta_0) d\boldsymbol{V}_1(E_2^c(t)) \\ &+ \frac{1}{\sqrt{\lambda}} \int_0^1 \psi_1'(t;\theta_0) \boldsymbol{X}(F_1^{-1}(t)) dE_2^c(t) \\ &+ b \int_0^1 \dot{\psi}_1(t;\theta_0) dE_2^c(t) \end{aligned}$$

uniformly in b satisfying $|b| \leq B$. Uniformity in convergence may be easily verified by the monotonicity of $S_N(\theta)$ in b and the compactness of the interval $[-B, B]$. Now set

$$T_1 \triangleq \frac{1}{\sqrt{1-\lambda}}\left[\int_0^1 \psi_0(t;\theta_0)d\boldsymbol{V}_0(E_2^u(t)) + \int_0^1 \psi_1(t;\theta_0)d\boldsymbol{V}_1(E_2^c(t))\right],$$
$$T_2 \triangleq \frac{1}{\sqrt{\lambda}}\left[\int_0^1 \psi_0'(t;\theta_0)\boldsymbol{X}(F_1^{-1}(t))dE_2^u(t) + \int_0^1 \psi_1'(t;\theta_0)\boldsymbol{X}(F_1^{-1}(t))dE_2^c(t)\right],$$
$$T_3 \triangleq \int_0^1 \dot{\psi}_0(t;\theta_0)dE_2^u(t) + \int_0^1 \dot{\psi}_1(t;\theta_0)dE_2^c(t).$$

Then the result above is rewritten in a form of well-known asymptotic linearity: uniformly in b ($|b| \leq B$),

$$\sqrt{N}\, S_N(\theta) \xrightarrow{P} T_1 + T_2 + bT_3. \tag{3.11}$$

To show the convergence of $\sqrt{N}(\hat{\theta}_N - \theta_0)$, according to the spirit of Shorack (1970), we have to check that $|\sqrt{N}S_N(\theta)|$ is bounded away from 0 for b outside $[-B, B]$ with probability sufficiently close to 1. For if this is proved, then, for sufficiently large N, a unique minimizer of $|\sqrt{N}S_N(\theta)|$ is $\sqrt{N}(\hat{\theta}_N - \theta_0)$, and this converges to $-(T_1+T_2)/T_3$ which minimizes $|T_1+T_2+bT_3|$. Now set $\mu_{b,N} \triangleq A_{3N}$. Then, by triangle inequality,

$$|\sqrt{N}S_N(\theta)| \geq \left|\sqrt{N}|S_N(\theta) - \mu_{b,N}| - \sqrt{N}|\mu_{b,N}|\right|. \tag{3.12}$$

We have already seen that $\sqrt{N}(S_N(\theta) - \mu_{b,N}) \xrightarrow{P} T_1 + T_2$ uniformly in b, and hence it is $O_p(1)$. On the other hand, it was shown that $\sqrt{N}\mu_{b,N} \xrightarrow{P} bT_3$ also uniformly in b. Thus, for an arbitrary sufficiently large $M > 0$, we can take $B_M > 0$ so that $B_M|T_3| > M + 1$, and then for this B_M we can choose n_M such that $|\sqrt{N}\mu_{b,N} - bT_3| \leq 1$ for all $|b| \leq B_M$ and all $N > n_M$. It then follows from the monotonicity of $\mu_{b,N}$ in b that $\sqrt{N}|\mu_{b,N}| > M$ for all $|b| > B_M$ and all $N > n_M$. This, together with (3.12), implies that for a given $\epsilon > 0$, there exist a $B_\epsilon > 0$ and an integer n_ϵ such that for all $N > n_\epsilon$

$$P\left\{|\sqrt{N}S_N(\theta)| > \frac{1}{\epsilon} \text{ for all } |b| > B_\epsilon\right\} > 1 - \epsilon,$$

which shows the assertion stated above.

Consequently we arrive at the following:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{P} -\frac{T_1+T_2}{T_3}. \tag{3.13}$$

Clearly the mean of the rv on the right hand side is zero. For variance, note that

$$\mathrm{Cov}[\boldsymbol{X}(s), \boldsymbol{X}(t)] = (1 - F_1(s))(1 - F_1(t))C(s \wedge t) \quad \text{and}$$
$$\mathrm{Cov}[\boldsymbol{V}_0(E_2^u(s)), \boldsymbol{V}_1(E_2^c(s))] = -E_2^u(s)E_2^c(t).$$

These follow from Proposition A.2 and A.3. Since $\boldsymbol{X}$ and $\boldsymbol{V}_k$ $(k = 0, 1)$ are mutually independent, direct calculation shows that the variance of the rv on the right in (3.13) is given by $\sigma^2(\theta_0)$. □

Remark 3.2. By the above proof, it is easy to see that the effect of estimating F_1 by $\hat{\boldsymbol{F}}_{1m}$ appears in the asymptotic variance in the form of $\tau_2(\theta)$.

Remark 3.3. If $G_1 \equiv 0$, that is, there is no censoring, asymptotic variance $\sigma^2(\theta_0)$ of $\hat{\theta}_N$ coincides with the one obtained in DDM. This is easily verified by noticing that the PL-estimator agrees with the empirical df in this case.

3.3 *Estimation of variance*

Apparently it seems possible to estimate asymptotic variance by replacing F_1, H_2, H_2^u, H_2^c and C in the expression of $\sigma^2(\theta)$ by their empiricals. In the case of continuous F_1 a consistent estimator of C is given in Shorack and Wellner (1986) by

$$\boldsymbol{C}_{1m}(t) \triangleq \frac{1}{m}\sum_{i=1}^{m}\left(1-\frac{i}{m}\right)^{-1}\left(1-\frac{i-1}{m}\right)^{-1} I_{[X_{(i)}\leq t]}\delta_{(i)}.$$

However, here one is faced with difficulty: the value of $\boldsymbol{C}_{1m}(t)$ for $t \geq \tau_m$ is always infinite, that is, ∞. This fact makes it impossible to estimate C by $\boldsymbol{C}_{1m}$ if the largest order statistic $Y_{(n)}$ is greater than τ_m (see the expression of $\tau_2(\theta)$). In this case, we shall modify $\boldsymbol{C}_{1m}$ slightly by

$$\hat{\boldsymbol{C}}_{1m}(t) \triangleq \frac{1}{m}\sum_{i=1}^{m}\left(1-\frac{i-1}{m}\right)^{-2} I_{[X_{(i)}\leq t]}\delta_{(i)}.$$

This may underestimate C in comparison with $\boldsymbol{C}_{1m}$ because of replacing $(1 - i/m)^{-1}$ by $(1-(i-1)/m)^{-1}$, but considering the definition of C in the case that F_1 is not necessarily continuous (see Shorack and Wellner (1986)), our definition of $\hat{\boldsymbol{C}}_{1m}$ seems natural.

Example 1. (continued) In the proportional hazards model (1.8), the asymptotic variance $\sigma^2(\theta)$ of $\hat{\theta}_N$ of (2.9) is given by

$$\begin{aligned}
\tau_1(\theta) &= \frac{1}{\theta^2}[H_2^u(\infty) - H_2^u(\infty)^2] \\
&\quad + \frac{2}{\theta}\left[\int_0^\infty \log(1-F_1)dH_2^u - H_2^u(\infty)\int_0^\infty \log(1-F_1)dH_2\right] \\
&\quad + \int_0^\infty [\log(1-F_1)]^2 dH_2 - \left[\int_0^\infty \log(1-F_1)dH_2\right]^2, \\
\tau_2(\theta) &= 2\int_0^\infty C(s)[1-H_2(s)]dH_2(s),
\end{aligned}$$

and

$$\left[\int_0^\infty \dot{\psi}_0(F_1;\theta)dH_2^u + \int_0^\infty \dot{\psi}_1(F_1;\theta)dH_2^c\right]^2 = \frac{1}{\theta^4}H_2^u(\infty)^2.$$

For this model, it is known that the maximum partial likelihood estimator (MPLE) (Cox (1975)) is the best nonparametric estimator of θ (for precise statement, see Begun and Wellner (1983)), so that one would like to compare the RAM estimator with the MPLE. Set

$$\frac{1}{\sigma_*^2} \triangleq \int_0^\infty \frac{1-H_1}{(1-\lambda)^{-1}(1-H_1)+\theta\lambda^{-1}(1-H_2)} dH_2^u,$$

then the asymptotic variance of the MPLE is given by $\theta^2\sigma_*^2$. In order to obtain tractable asymptotic relative efficiency (ARE), let us suppose that the censoring mechanism is given by $1-G_k = (1-F_k)^{\gamma_k}$, $(k=1,2)$ as in Kalbfleisch and Prentice (1981). Then we find that the asymptotic variances of the RAM estimator and MPLE are given by

$$\theta^2 \frac{\gamma_2+1}{1-\lambda} + \frac{\theta^4}{\lambda} \frac{(\gamma_2+1)^2}{2\theta+2\theta\gamma_2-\gamma_1-1}$$

and

$$\theta^2 \left[\int_0^1 \{((1-\lambda)\theta)^{-1} t^{1-\theta-\theta\gamma_2} + \lambda^{-1} t^{-\gamma_1}\}^{-1} dt \right]^{-1}$$

respectively for $\theta \geq 1$. For $\lambda = 1/2$ the ARE of the RAM estimator with respect to the MPLE is tabulated in Table 1 for various values of θ, γ_1, γ_2. We see from Table 1 that the RAM estimator is quite efficient for $\theta \in [1,2]$.

Table 1. Asymptotic relative efficiency of RAM-estimator with respect to MPLE for special censoring.

	$\theta=1$			$\theta=2$			$\theta=4$		
$\gamma_1\backslash\gamma_2$	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
0.0	1.000	0.986	0.978	0.951	0.934	0.927	0.864	0.850	0.844
0.5	0.863	1.000	0.991	0.981	0.951	0.938	0.886	0.864	0.854
1.0	$***$	0.951	1.000	1.000	0.971	0.951	0.913	0.878	0.864

	$\theta=8$			$\theta=16$		
$\gamma_1\backslash\gamma_2$	0.0	0.5	1.0	0.0	0.5	1.0
0.0	0.756	0.746	0.741	0.647	0.640	0.637
0.5	0.772	0.756	0.749	0.657	0.647	0.642
1.0	0.789	0.766	0.756	0.668	0.654	0.647

Table 2, from Pike (1966), gives the times from insult with the carcinogen DMBA to mortality from vaginal cancer in rats for two pretreatment regimens. Table 3 gives the results of estimation of θ for the data of Table 2 using four different estimates. Besides the RAM and MPLE estimates we consider the two-step estimate of Begun and Reid (1983) and the average hazard ratio estimate of Kalbfleisch and Prentice (1981).

Table 2. Days to cancer mortality in rats.

Group 1	143, 164, 188, 190, 192, 206, 209, 213, 216, 216+, 220, 227, 230, 234, 244+, 246, 265, 304
Group 2	142, 156, 163, 198, 204+, 205, 232, 232, 233, 233, 233, 233, 239, 240, 261, 280, 280, 296, 296, 323, 344+

+ indicates censored.

Table 3. Estimation for the data of Table 1.

	Estimate of θ	Estimate of $\beta = \log\theta$	Estimated standard error of β
Two-step estimate	.553	−.593	1.165
MPLE	.551	−.596	1.086
Average hazard ratio estimate	.547	−.603	1.045
RAM estimate	.615	−.486	1.153

Begun and Reid make use of the relation

$$(1 - H_1)dH_2^u = \theta(1 - H_2)dH_1^u$$

and proposes an estimator, using some score function J,

$$\hat{\theta}(J) \triangleq \frac{\int_0^\infty J(1 - \boldsymbol{H}_{1m}, 1 - \boldsymbol{H}_{2n})(1 - \boldsymbol{H}_{1m})d(1 - \boldsymbol{H}_{2n}^u)}{\int_0^\infty J(1 - \boldsymbol{H}_{1m}, 1 - \boldsymbol{H}_{2n})(1 - \boldsymbol{H}_{2n})d(1 - \boldsymbol{H}_{1m}^u)}.$$

Two-step estimate is constructed by taking $\hat{J}_*(s,t) = [\lambda s + \hat{\theta}^\circ(1-\lambda)t]^{-1}$, where $\hat{\theta}^\circ$ is a preliminary estimate obtained with $J \equiv 1$ in the first step. It is fully efficient with respect to the MPLE.

Kalbfleisch and Prentice define the average hazard ratio by

$$\nu = \int_0^\infty \lambda_{F_1}(t)/[\lambda_{F_1}(t) + \lambda_{F_2}(t)]d(1 - F_1F_2)^\delta.$$

The parameter $\delta \geq 0$ is a weight to be chosen. Under the proportional hazards model, we have $\theta = \nu/(1 - \nu)$. The average hazard ratio estimate is obtained by replacing F_1 and F_2 with their PL-estimators.

The estimation principle of RAM estimator is intelligible and compared with these estimators, RAM estimator is easy to compute and so it is of practical use.

Acknowledgements

This paper is based on the author's M.A. thesis written under the direction of Ryozo Miura, whom the author would like to thank for his patience and for his suggestions on this problem. He is also grateful to Kei Takeuchi and Akimichi Takemura for their helpful comments.

Appendix

In this appendix, we summarize the results on PL-estimator in Shorack and Wellner (1986) which we used for the proof of Theorem 2.1. Proofs of these propositions are not given, so the reader should refer to the above book or Gill (1983). Notations are the same as in Subsection 3.1.

PROPOSITION A.1. (Glivenko-Cantelli) *As $n \to \infty$, we have*

$$\|\boldsymbol{H}_{2n}^u - H_2^u\| \xrightarrow{a.s.} 0, \qquad \|\boldsymbol{H}_{2n}^c - H_2^c\| \xrightarrow{a.s.} 0, \qquad \|\boldsymbol{H}_{2n} - H_2\| \xrightarrow{a.s.} 0.$$

PROPOSITION A.2. *On the specially constructed probability space we have, as $n \to \infty$,*

$$\|\boldsymbol{V}_n^u(H_2^u) - \boldsymbol{V}_0(H_2^u)\| \xrightarrow{a.s.} 0, \qquad \|\boldsymbol{V}_n^c(H_2^c) - \boldsymbol{V}_1(H_2^c)\| \xrightarrow{a.s.} 0,$$

where $\boldsymbol{V}_0$ and $\boldsymbol{V}_1$ are Brownian bridges whose covariance is given by

$$\mathrm{Cov}[\boldsymbol{V}_0(H_2^u(s)), \boldsymbol{V}_1(H_2^c(t))] = -H_2^u(s)H_2^c(t).$$

PROPOSITION A.3. *On the specially constructed probability space, for any fixed $T < H_1(1)$, it follows that*

$$\|\boldsymbol{X}_m - \boldsymbol{X}\|_0^T \xrightarrow{a.s.} 0, \qquad m \to \infty.$$

Covariance function of $\boldsymbol{X}$ is given by

$$\mathrm{Cov}[\boldsymbol{X}(s), \boldsymbol{X}(t)] = (1 - F_1(s))(1 - F_1(t))C(s \wedge t).$$

PROPOSITION A.4. *On the specially constructed probability space we have*

$$\left\| \frac{1-K}{1-F_1} \cdot \frac{\boldsymbol{X}_m - \boldsymbol{X}}{q(K)} \right\|_0^{\tau_m} \xrightarrow{P} 0, \qquad m \to \infty,$$

provided the function $q(t)$ on $[0,1]$ satisfies the following conditions:

$$q(t) \nearrow \text{ on } [0,1/2], \text{ and symmetric about } t = 1/2;$$
$$q(t)/\sqrt{t} \searrow \text{ on } [0,1/2]; \qquad \int_0^1 [q(t)]^{-2} dt < \infty.$$

PROPOSITION A.5. *We have $F_1 \leq K$. Moreover if both F_1 and G_1 are continuous, then*

$$\frac{1-F_1}{1-K} = 1 + \int_0^{\cdot} \tilde{C} dF_1,$$

where

$$\tilde{C}(t) \triangleq \int_0^t \frac{dG_1}{(1-F_1)(1-G_1)^2} = \int_0^t \frac{dH_1^c}{(1-H_1)^2},$$

so that $(1-F_1)/(1-K)$ is monotonically increasing.

REFERENCES

Begun, J. M. and Reid, N. (1983). Estimating the relative risk with censored data, *J. Amer. Statist. Assoc.*, **78**, 337–341.

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

Cox, D. R. (1972). Regression models and life table, *J. Roy. Statist. Soc. Ser. B*, **34**, 187–220.

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

Cuzick, J. (1988). Rank regression, *Ann. Statist.*, **16**, 1369–1389.

Dabrowska, D. M., Doksum, K. A. and Miura, R. (1989). Rank estimates in a class of semiparametric two-sample models, *Ann. Inst. Statist. Math.*, **41**, 63–79.

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

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

James, I. R. (1986). On estimating equations with censored data, *Biometrika*, **73**, 35–42.

Kalbfleisch, J. D. and Prentice, R. L. (1981). Estimation of the average hazard ratio, *Biometrika*, **68**, 105–112.

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

Lehmann, E. L. (1953). The power of rank tests, *Ann. Math. Statist.*, **24**, 23–43.

Miura, R. (1985). Hodges-Lehmann type estimates and Lehmann's alternative models: a special lecture presented at the annual meeting of Japanese Mathematical Society, the Division of Statistical Mathematics, April, 1985 (in Japanese).

Pike, M. C. (1966). A method of analysis of certain class of experiments in carcinogenesis, *Biometrics*, **22**, 142–161.

Pyke, R. and Shorack, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems, *Ann. Math. Statist.*, **39**, 755–771.

Shorack, G. R. (1970). The one-sample symmetry problem, Tech. Report, No. 21, Department of Mathematics, University of Washington.

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

Wellner, J. A. (1986). Semiparametric models: progress and problems, *Proc. ISI.*, Amsterdam.