Ann. Inst. Statist. Math.
Vol. 56, No. 4, 683–700 (2004)


# ASYMPTOTICS FOR A WEIGHTED LEAST SQUARES ESTIMATOR OF THE DISEASE ONSET DISTRIBUTION FUNCTION FOR A SURVIVAL-SACRIFICE MODEL

ANTONIO EDUARDO GOMES

*Departamento de Estatística, ICEx/UFMG, 31270-901 - Belo Horizonte-MG, Brazil*

(Received March 11, 2003; revised September 26, 2003)

**Abstract.** In carcinogenicity experiments with animals where the tumor is not palpable it is common to observe only the time of death of the animal, the cause of death (the tumor or another independent cause, as sacrifice) and whether the tumor was present at the time of death. These last two indicator variables are evaluated after an autopsy. A weighted least squares estimator for the distribution function of the disease onset was proposed. Asymptotic properties of that estimator are established here. We demonstrate its strong uniform consistency. A minimax lower bound for the estimation of the disease onset distribution is obtained, as well as the local asymptotic distribution for their estimator.

*Key words and phrases*: Asymptotics, interval censoring, survival-sacrifice, weighted least squares, disease onset estimation.

## 1. Introduction

Suppose that in an experiment for the study of onset and mortality from undetectable moderately lethal incurable diseases (occult tumors, e.g.) we observe the time of death, whether the disease of interest was present at death, and if present, whether the disease was a probable cause of death. Defining the nonnegative variables $T_1$ (time of disease onset), $T_2$ (time of death from the disease) and $C$ (time of death from an unrelated cause), we observe, for the $i$-th individual, $(Y_i, \Delta_{1,i}, \Delta_{2,i})$, where $Y_i = C_i \wedge T_{2,i}$, $\Delta_{1,i} = I(T_{1,i} \leq C_i)$, $\Delta_{2,i} = I(T_{2,i} \leq C_i)$, and $I(\cdot)$ is the indicator function. Variables $T_{1,i}$ and $T_{2,i}$ have an unidentifiable joint distribution function $F$ such that $P(T_{1,i} \leq T_{2,i}) = 1$, $C_i$ has distribution function $G$ and is independent of $(T_{1,i}, T_{2,i})$. Current status data can be seen as a particular case of the survival-sacrifice model above when the disease is nonlethal, i.e., $\Delta_{2,i} = 0$, $i = 1, \dots, n$. In this case, $Y_i = C_i$, and $\hat{F}_2 \equiv 0$ for any estimator $\hat{F}_2$ of $F_2$ (the marginal distribution function of $T_2$). Right-censored data are a special case of the survival-sacrifice model above when a lethal disease is always present at the moment of death, i.e., $\Delta_{1,i} = 1$, $i = 1, \dots, n$. In this case, $\hat{F}_1 \equiv 1$ for any estimator $\hat{F}_1$ of $F_1$ (the marginal distribution function of $T_1$).

An example of a real data set can be found in Turnbull and Mitchell (1984). It contains the ages at death (in days) of 109 female RFM mice. The disease of interest is reticulum cell sarcoma (RCS). These mice formed the control group in a survival experiment to study the effects of prepubertal ovariectomy in mice given 300 R of X-rays (Holland *et al.* (1977)).

The parameter space for the survival-sacrifice model can be taken to be

$$\Theta = \{(F_1, F_2) : F_1 \text{ and } F_2 \text{ are distribution functions with } F_1 <_s F_2\},$$

where $F_1 <_s F_2$ means that $F_1(x) \geq F_2(x)$ for every $x \in \mathbb{R}$ and $F_1(x) > F_2(x)$ for some $x \in \mathbb{R}$. The loglikelihood function for this model is

$$\begin{aligned} \mathcal{L}(F_1, F_2) = \sum_{i=1}^{n} \{ &(1 - \Delta_{1,i})(1 - \Delta_{2,i}) \log(1 - F_1(Y_i)) \\ &+ \Delta_{1,i}(1 - \Delta_{2,i}) \log(F_1(Y_i) - F_2(Y_i)) \\ &+ (\Delta_{1,i}\Delta_{2,i}) \log f_2(Y_i)\} + K(g, G) \end{aligned} \tag{1.1}$$

where $f_2(y) = F_2(y) - F_2(y-)$ and $K(g, G)$ is a term involving only the distribution function $G$ and the probability density function $g$ of $C$. We will assume without loss of generality that $Y_1 \leq Y_2 \leq \cdots \leq Y_n$.

Kodell *et al.* (1982) also studied the nonparametric estimation of $S_1 = 1 - F_1$ and $S_2 = 1 - F_2$, but their work is restricted to the case where $R(t) = S_1(t)/S_2(t)$ is nonincreasing, an assumption that may not be reasonable, for example, for progressive diseases whose incidence is concentrated in the early or middle part of the life span.

Turnbull and Mitchell (1984) proposed an EM algorithm for the joint estimation of $F_1$ and $F_2$ which converges to the nonparametric maximum likelihood estimator of $(F_1, F_2)$ provided the support of the initial estimator contains the support of the maximum likelihood estimator.

Gomes *et al.* (2001) used a faster primal-dual interior point algorithm to calculate the nonparametric maximum likelihood estimator of $(F_1, F_2)$.

Another possible way of estimating $F_1$ is by plugging in the Kaplan-Meier estimator $\hat{F}_{2,n}$ of $F_2$ in (1.1) and calculating the nonparametric maximum pseudolikelihood estimator of $F_1$.

A weighted least squares estimator for $F_1$ making $F_2 = \hat{F}_{2,n}$ was proposed by van der Laan *et al.* (1997). Gomes *et al.* (2001) compared the efficiency of the weighted least squares and nonparametric maximum likelihood estimators of $F_1$. Their simulation studies showed evidence that the local performance of the weighted least squares estimator is superior when $F_1 - F_2$ becomes larger. The authors also pointed out that when $\|F_1 - F_2\|_\infty = 1$ all the estimators of $F_1$ and $F_2$ mentioned above coincide.

The weighted least squares estimator proposed by van der Laan *et al.* (1997) is described in Section 2 and its consistency is established in Section 3. Results about the rate of convergence and the local limit distribution of their estimator are established in Sections 4 and 5, respectively.

The global asymptotic behavior of the weighted least squares estimator is still to be determined as well as that of the joint nonparametric maximum likelihood estimator of $F_1$ and $F_2$. The weighted least squares and nonparametric maximum likelihood estimators of $F_1$ are believed to be asymptotically equivalent, but that still remains to be proved.

## 2. The weighted least squares estimator

A possibility for the estimation of $F_1$ is to calculate a weighted least squares estimator as suggested by van der Laan *et al.* (1997). They showed that estimating $S_1$ can be viewed as a regression of $(1 - \Delta_1)S_2(C)$ on the observed $C_i$'s under the constraint of monotonicity. If we substitute $S_2$ by its Kaplan-Meier estimator $\hat{S}_{2,n}$, we automatically have an estimator for $S_1$ minimizing

$$\frac{1}{n}\sum_{i=1}^{n}[(1-\Delta_{1,i})\hat{S}_{2,n}(Y_i) - S_1(Y_i)]^2(1-\Delta_{2,i})$$

under the constraint that $S_1$ is nonincreasing. This minimization problem can be solved by using results from the theory of isotonic regression (see Barlow *et al.* (1972) or Robertson *et al.* (1988)) with weights inversely proportional to

$$\mathrm{Var}[(1-\Delta_1)S_2(C) \mid C = c, T_2 > C] = S_2^2(c)R(c)[1-R(c)]$$

which depends on $S_1$ and would imply the use of an iterative process. However, if we use weights $w_i = (1-\Delta_{2,i})/\hat{S}_{2,n}^2(Y_i)$ instead, we have an estimator with a closed form, as suggested by van der Laan *et al.* (1997).

Using algorithms for isotonic regression problems, their estimator $\hat{S}_{1,n}$ will be given by the slope of the least concave majorant of the cumulative sum diagram determined by the points $(0,0), (W_1, V_1), \ldots, (W_n, V_n)$, where $W_j = \sum_{i=1}^{j} w_i$ and

$$V_j = \sum_{i=1}^{j} w_i(1-\Delta_{1,i})\hat{S}_{2,n}(Y_i) = \sum_{i=1}^{j}\frac{(1-\Delta_{1,i})(1-\Delta_{2,i})}{\hat{S}_{2,n}(Y_i)} = \sum_{i=1}^{j}\frac{(1-\Delta_{1,i})}{\hat{S}_{2,n}(Y_i)}.$$

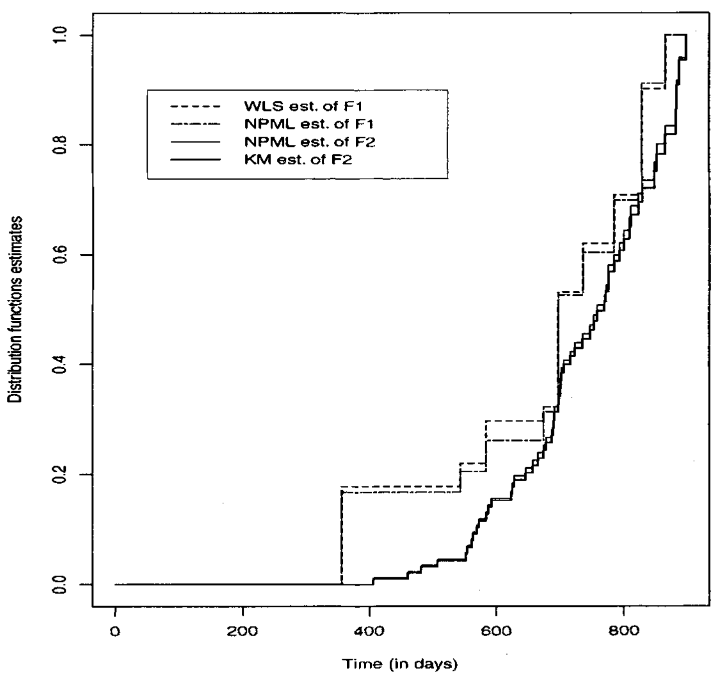


Fig. 1. Weighted least squares estimate of $F_1$, Kaplan-Meier estimate of $F_2$, and nonparametric maximum likelihood estimates of $F_1$ and $F_2$.

So, $\hat{S}_{1,n}(t)$ is the mentioned slope at $W_j$ if $t \in (Y_{j-1}, Y_j]$, and we can write

$$\hat{S}_{1,n}(Y_m) = \min_{l \leq m} \max_{k \geq m} \frac{\sum_{j=l}^{k}(1-\Delta_{1,j})/\hat{S}_{2,n}(Y_j)}{\sum_{j=l}^{k}(1-\Delta_{2,j})/\hat{S}_{2,n}^2(Y_j)}, \tag{2.1}$$

for $m = 1, \ldots, n$. It is easy to see that (2.1) reduces to the expression of the nonparametric maximum likelihood estimator of $F \equiv F_1$ for current status data (Groeneboom and Wellner (1992)) when $\Delta_{2,i} = 0$, $i = 1, \ldots, n$, since in this case $\hat{S}_{2,n} \equiv 1$.

The weighted least squares estimator of $F_1$ and the Kaplan-Meier estimator of $F_2$ do not coincide with the nonparametric maximum likelihood estimators of $F_1$ and $F_2$, respectively. Figure 1 shows those estimators for the real data set in Turnbull and Mitchell (1984). The smoother picture for the estimates of $F_2$ is a consequence of the $n^{1/3}$ rate of convergence for the estimation of $F_1$ (as shown in Section 4) compared to the $\sqrt{n}$-rate for the estimation of $F_2$.

## 3. Consistency

In this section we present a theorem establishing the strong uniform consistency of the weighted least squares estimator proposed by van der Laan *et al.* (1997). In the proof of Theorem 3.1 (in the Appendix) we will use a general method used by Jewell (1982) to prove consistency of the nonparametric maximum likelihood estimator for the mixing distribution in scale mixture of exponential distributions. The same method was used by Groeneboom and Wellner (1992) to prove consistency of the nonparametric maximum likelihood estimator of the disease onset distribution function for interval censoring, cases 1 (current status data) and 2.

THEOREM 3.1. *Suppose $C$, $T_1$ and $T_2$ have continuous distribution functions $G$, $F_1$ and $F_2$, respectively, such that $P_{F_1} \ll P_G$, where $P_{F_1}$ and $P_G$ are the probability measures induced by functions $F_1$ and $G$, respectively. Then*

$$\|\hat{F}_{1,n} - F_1\|_\infty = \sup_{t \in \mathbb{R}} |\hat{F}_{1,n}(t) - F_1(t)| \to 0$$

*almost surely.*

The proof is given in the Appendix.

## 4. Minimax lower bound

We determine here a minimax lower bound for the estimation of $F_1(t_0)$. When $n$ grows, the minimax risk should decrease to zero. The rate $\delta_n$ of this convergence is the best rate of convergence an estimator can have for the estimation problem posed.

Let $T$ be a functional and $q$ a probability density in a class $\mathcal{G}$ with respect to a $\sigma$-finite measure $\mu$ on the measurable space $(\Omega, \mathcal{A})$. Let $Tq$ denote a real-valued parameter and $\{T_n\}$, $n \geq 1$, be a sequence of estimators of $Tq$ based on samples of size $n, X_1, \ldots, X_n$, generated by $q$.

$E_{n,q}[\ell(|T_n - Tq|)]$ is the *risk* of the estimator $T_n$ in estimating $Tq$ when the loss function is $\ell : [0, \infty) \to \mathbb{R}$ ($\ell$ is increasing and convex with $\ell(0) = 0$). $E_{n,q}$ denotes the expectation with respect to the product measure $q^{\otimes n}$ associated with the sample $X_1, \ldots, X_n$. For fixed $n$, the minimax risk

$$\inf_{T_n} \sup_{q \in \mathcal{G}} E_{n,q}[\ell(|T_n - Tq|)]$$

is a way to measure how hard the estimation problem is.

Lemma 4.1 below is quite helpful in deriving asymptotic lower bounds for minimax risks (see Groeneboom (1996), Chapter 4, for the proof). It will be used to prove Theorem 4.1.

LEMMA 4.1. *Let $\mathcal{G}$ be a set of probability densities on a measurable space $(\Omega, \mathcal{A})$ with respect to a $\sigma$-finite measure $\mu$, and let $T$ be a real-valued functional on $\mathcal{G}$. Moreover, let $\ell : [0,\infty) \to \mathbb{R}$ be an increasing convex loss function, with $\ell(0) = 0$. Then, for any $q_1, q_2 \in \mathcal{G}$ such that the Hellinger distance $H(q_1, q_2) < 1$,*

$$\begin{aligned} &\inf_{T_n} \max\{E_{n,q_1}[\ell(|T_n - Tq_1|)], E_{n,q_2}[\ell(|T_n - Tq_2|)]\} \\ &\qquad \geq \ell\left(\tfrac{1}{4}|Tq_1 - Tq_2|[1 - H^2(q_1, q_2)]^{2n}\right). \end{aligned}$$

In our case, let $X_i = (Y_i, \Delta_{1,i}, \Delta_{2,i})$, and

$$\begin{aligned} q_0(y, \Delta_1, \Delta_2) = &\{g(y)[1 - F_1(y)]\}^{(1-\Delta_1)(1-\Delta_2)} \\ &\times \{g(y)[F_1(y) - F_2(y)]\}^{\Delta_1(1-\Delta_2)} \\ &\times \{f_2(y)[1 - G(y)]\}^{\Delta_1\Delta_2}, \end{aligned}$$

$\mu = \lambda \times m$, where $\lambda$ is the Lebesgue measure and $m$ is the counting measure on the set $\{(0,0), (0,1), (1,1)\}$, $Tq_0 = F_1(t_0)$, and $q_n$ is equal to the density corresponding to the perturbation

$$F_{1,n}(x) = \begin{cases} F_1(x) & \text{if } x < t_0 - n^{-1/3}t \\ F_1(t_0 - n^{-1/3}t) & \text{if } x \in [t_0 - n^{-1/3}t, t_0) \\ F_1(t_0 + n^{-1/3}t) & \text{if } x \in [t_0, t_0 + n^{-1/3}t) \\ F_1(x) & \text{if } x \geq t_0 + n^{-1/3}t \end{cases} \tag{4.1}$$

for a suitably chosen $t > 0$.

Using the perturbation (4.1) we show in the proof of Theorem 4.1 that

$$H^2(q_n, q_0) \sim n^{-1} g(t_0) f_1^2(t_0) t^3 S_2(t_0) / \{S_1(t_0)[S_2(t_0) - S_1(t_0)]\}.$$

So, as pointed out by Groeneboom (1996) for current status data, we could say that the Hellinger distance of order $n^{-1/2}$ between $q_n$ and $q_0$ corresponds to a distance of order $n^{-1/3}$ between $Tq_n = F_{1,n}(t_0)$ and $Tq_0 = F_1(t_0)$.

The perturbation (4.1) is the worst possible. When maximizing in $t$, we are taking the worst possible constant.

THEOREM 4.1.

$$\begin{aligned} &n^{1/3} \inf_{\hat{F}_{1,n}} \max\{E_{n,q_0}[|\hat{F}_{1,n}(t_0) - F_1(t_0)|], E_{n,q_n}[|\hat{F}_{1,n}(t_0) - F_{1,n}(t_0)|]\} \\ &\qquad \geq \tfrac{1}{4} n^{1/3} |F_{1,n}(t_0) - F_1(t_0)| [1 - H^2(q_n, q_0)]^{2n} \\ &\qquad \to \tfrac{1}{4} f_1(t_0) t \exp\left\{-\frac{2S_2(t_0)g(t_0)f_1^2(t_0)t^3}{S_1(t_0)[S_2(t_0) - S_1(t_0)]}\right\} \end{aligned}$$

*and the maximum value of this last expression is*

$$k\{f_1(t_0)S_1(t_0)[S_2(t_0) - S_1(t_0)]/[S_2(t_0)g(t_0)]\}^{1/3} \tag{4.2}$$

*where* $k = (1/4)(3e/2)^{-1/3}$ *does not depend on* $f_1$, $F_1$ *or* $g$.

The proof is given in the Appendix.

Groeneboom (1987) applied Lemma 4.1 to obtain a minimax lower bound of the form

$$c\{f(t_0)F(t_0)[1 - F(t_0)]/g(t_0)\}^{1/3} \tag{4.3}$$

for the problem of estimating $F$ with current status data. We can easily see that up to the constants $k$ and $c$, (4.2) reduces to (4.3) if we make $F_1 \equiv F$, $f_1 \equiv f$ and $S_2 \equiv 1$, which are the changes that reduce the survival-sacrifice model studied here to current status data.

## 5. Local limit distribution

Theorem 5.1 below gives the local asymptotic behavior of the weighted least squares estimator of $F_1$ proposed by van der Laan *et al.* (1997).

THEOREM 5.1. *Suppose* $C$, $T_1$ *and* $T_2$ *have continuous distribution functions* $G$, $F_1$ *and* $F_2$, *respectively, such that* $P_{F_1} \ll P_G$. *Additionally, let* $t_0$ *be such that* $0 < F_1(t_0) < 1$, $0 < G(t_0) < 1$, *and let* $F_1$ *and* $G$ *be differentiable at* $t_0$, *with strictly positive derivatives* $f_1(t_0)$ *and* $g(t_0)$, *respectively. Suppose also that* $t_0$ *is such that for some* $\delta > 0$ *and some* $M > 0$, $S_2(t_0 + M) > \delta$. *Then*

$$n^{1/3} \frac{\hat{S}_{1,n}(t_0) - S_1(t_0)}{\left\{\frac{1}{2} f_1(t_0)S_1(t_0)[S_2(t_0) - S_1(t_0)]/[g(t_0)S_2(t_0)]\right\}^{1/3}} \to 2Z \tag{5.1}$$

*in distribution, where* $Z = \arg\max_h\{\mathbb{B}(h) - h^2\}$, *and* $\mathbb{B}$ *is a two-sided standard Brownian motion starting from 0.*

The proof is given in the Appendix.

Groeneboom (1989) studied the distribution of the random variable $Z$, and Groeneboom and Wellner (2001) calculated its quantiles, allowing the construction of confidence intervals for $F_1$.

As noticed in "Introduction", current status data is a particular version of the present problem when we have $\Delta_{2,i} = 0$, $i = 1, \ldots, n$. This is equivalent to have $S_2(t_0) \equiv 1$ in the expression above, which would reduce it to the well known result about the limit distribution of the nonparametric maximum likelihood estimator of $F_1$ when we have current status data (Groeneboom and Wellner (1992)).

Notice that the expression in the denominator of (5.1) is proportional to that in the minimax lower bound in Theorem 4.1.

## Acknowledgements

The author is grateful to Jon A. Wellner for valuable discussions during the development of this work and to two referees whose suggestions led to improvements in the paper.

## Appendix

Proof of Theorem 3.1. Let $(T_{1,1}, T_{2,1}, C_1), \ldots, (T_{1,n}, T_{2,n}, C_n)$ be a sample of random variables in $\mathbb{R}^3_+$, where $C_i$ is independent of $(T_{1,i}, T_{2,i})$ and $C_i, T_{1,i}, T_{2,i}$ have continuous distribution functions $G$, $F_1$ and $F_2$, respectively, satisfying $P_{F_1} \ll P_G$ (the probability measure $P_{F_1}$, induced by $F_1$, is absolutely continuous w.r.t. the probability measure $P_G$, induced by $G$).

The estimator $\hat{F}_{1,n}$ minimizes the function $\psi$ under the constraint of monotonicity, where

$$\begin{aligned}\psi(F_1) &= \int_{\mathbb{R}^3} \frac{\{I(t_1 > c)[1-\hat{F}_{2,n}(c)] - [1-F_1(c)]\}^2 I(t_2 > c)}{[1-\hat{F}_{2,n}(c)]^2} d\mathbb{P}_n(t_1, t_2, c) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\{(1-\Delta_{1,i})[1-\hat{F}_{2,n}(C_i)] - [1-F_1(C_i)]\}^2(1-\Delta_{2,i})}{[1-\hat{F}_{2,n}(C_i)]^2}.\end{aligned}$$

Here $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{(T_{1,i}, T_{2,i}, C_i)}$ is the empirical probability measure.

The fact that $\hat{F}_{1,n}$ minimizes $\psi(F_1)$ implies that for any $0 \le \varepsilon \le 1$,

$$\psi((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1) - \psi(\hat{F}_{1,n}) \ge 0.$$

Dividing by $\varepsilon > 0$ and taking the limit as $\varepsilon \downarrow 0$ this yields

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}[\psi((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1) - \psi(\hat{F}_{1,n})] \ge 0.$$

But

$$\begin{aligned}&\psi((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1) \\ &\quad = \int_{\mathbb{R}^3} \{I(t_1 > c)[1-\hat{F}_{2,n}(c)] - [1-(1-\varepsilon)\hat{F}_{1,n}(c) - \varepsilon F_1(c)]\}^2 \\ &\qquad \times \{I(t_2 > c)/[1-\hat{F}_{2,n}(c)]^2\} d\mathbb{P}_n(t_1, t_2, c).\end{aligned}$$

So,

$$\begin{aligned}\text{(A.1)}\qquad &\frac{d}{d\varepsilon}\psi((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1)\mid_{\varepsilon=0} \\ &\quad = \int_{\mathbb{R}^3} 2\{I(t_1 > c)[1-\hat{F}_{2,n}(c)] - [1-\hat{F}_{1,n}(c)]\} I(t_2 > c) \\ &\qquad \times \{[F_1(c) - \hat{F}_{1,n}(c)]/[1-\hat{F}_{2,n}(c)]^2\} d\mathbb{P}_n(t_1, t_2, c) \ge 0.\end{aligned}$$

Let $\Omega$ be the space of all sequences $\{(T_{1,i}, T_{2,i}, C_i), i = 1, 2, \ldots\}$ endowed with the Borel $\sigma$-algebra generated by the product topology on $\prod_{i=1}^\infty \mathbb{R}^3$. Introducing "$\omega \in \Omega$" in the notation to indicate the dependence on the sequence $\{(T_{1,i}, T_{2,i}, C_i), i = 1, 2, \ldots\}$, $\mathbb{P}_n(\cdot, \cdot, \cdot; \omega)$ converges weakly to $P$, the joint probability distribution of $T_1$, $T_2$ and $C$, by Varadarajan's theorem (Dudley (1989)) for all $\omega$ in a set $B \subset \Omega$ such that $\mathbb{P}(B) = 1$, where $\mathbb{P} = P^\infty$.

There exists $B_2 \subseteq B$ with $\mathbb{P}(B_2) = 1$ such that $\sup_{t \in [0,\tau)} |\hat{F}_{2,n}(t; \omega) - F_2(t)| \to 0$ as $n \to \infty$, for every $\omega \in B_2$, where $\tau = \sup\{t : H(t) < 1\}$ and $H = 1 - (1-F_2)(1-G)$ (Wang (1987)).

For a fixed $\omega \in B_2$, the sequence $\hat{F}_{1,n}(\cdot,\omega)$ has a subsequence $\hat{F}_{1,n_k}(\cdot,\omega)$ converging vaguely to a nondecreasing right continuous function $F_1^*$ taking values in $[0,1]$, by the Helly compactness theorem.

Fix $\varepsilon \in (0,1)$ and choose $b$ such that $F_2(b) = 1-\varepsilon$. Since $\hat{F}_{2,n}(\cdot;\omega) \to F_2(\cdot)$ for $\omega \in B_2$, we have $\hat{F}_{2,n_k}(\cdot;\omega) \to F_2(\cdot)$ for $\omega \in B_2$. Thus, we may assume $1/[1-\hat{F}_{2,n}(t,\omega)]^2$ bounded for $t \in [0,b]$ and $n$ sufficiently large.

By the convergence in distribution of $\hat{F}_{2,n}(\cdot,\omega)$ we may also assume that $1/[1-F_2(t)]^2$ is bounded for $t \in [0,b]$. Hence we assume

$$\frac{1}{[1-F_2(b)]^2} \le M \quad \text{and} \quad \frac{1}{[1-\hat{F}_{2,n}(b,\omega)]^2} \le M \tag{A.2}$$

for a constant $M > 0$ and all $n$ sufficiently large.

We will need the following lemma.

LEMMA A.1. *Let $b$ be chosen such that $F_2(b) = 1-\varepsilon$. Then*

$$\begin{aligned}\lim_{k\to\infty} \int_{\mathbb{R}^2\times[0,b]} & 2\{I(t_1 > c)[1-\hat{F}_{2,n_k}(c)] - [1-\hat{F}_{1,n_k}(c)]\}I(t_2 > c) \\ & \times \{[F_1(c) - \hat{F}_{1,n_k}(c)]/[1-\hat{F}_{2,n_k}(c)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c) \\ = \int_{\mathbb{R}^2\times[0,b]} & 2\{I(t_1 > c)[1-F_2(c)] - [1-F_1^*(c)]\}I(t_2 > c) \\ & \times \{[F_1(c) - F_1^*(c)]/[1-F_2(c)]^2\}dP(t_1,t_2,c).\end{aligned} \tag{A.3}$$

*Moreover,*

$$\begin{aligned}\int_{\mathbb{R}^2\times[0,b]} & 2\{I(t_1 > c)[1-F_2(c)] - [1-F_1^*(c)]\}I(t_2 > c) \\ & \times \{[F_1(c) - F_1^*(c)]/[1-F_2(c)]^2\}dP(t_1,t_2,c) \ge 0.\end{aligned} \tag{A.4}$$

PROOF OF LEMMA A.1. Fix $0 < \delta < 1$ and take a grid of points $0 = u_0 < u_1 < \cdots < u_m = b$ on $[0,b]$ such that $m = 1 + [1/\delta^2]$, where $[\cdot]$ is the integer part of a real number, and $G(u_i) - G(u_{i-1}) = G(b)/m$, $i = 1,\dots,m$. Let $K$ be the set of indices $i$, $i = 1,\dots,m$ such that

$$\frac{1}{[1-F_2(u_i)]^2} - \frac{1}{[1-F_2(u_{i-1})]^2} \ge \delta.$$

The first inequality in (A.2) implies that the number of indices of this type is not bigger than $1 + [M/\delta]$. Let $L$ be the remaining set of indices $i$, $i = 1,\dots,m$.

Denoting the interval $[u_0,u_1]$ by $J_1$ and the intervals $(u_{i-1},u_i]$ by $J_i$, $i = 2,\dots,m$, we have

$$\begin{aligned}\int_{\mathbb{R}^2\times[0,b]} & 2\{I(t_1 > c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2 > c) \\ & \times \{[F_1(c) - \hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\ = \sum_{i=1}^{m} \int_{\mathbb{R}^2\times J_i} & 2\{I(t_1 > c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2 > c) \\ & \times \{[F_1(c) - \hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega).\end{aligned}$$

Since $\hat{F}_{2,n_k}(u_i;\omega)$ converges to $F_2(u_i)$ for each $i$, $0 \le i \le m$, we get, for sufficiently large $k$,

$$\frac{1}{[1-\hat{F}_{2,n_k}(u_i;\omega)]^2} - \frac{1}{[1-\hat{F}_{2,n_k}(u_{i-1};\omega)]^2} < 2\delta, \qquad i \in L.$$

Hence,

$$\begin{aligned}
\text{(A.5)}\quad & \int_{\mathbb{R}^2\times[0,b]} 2\{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-\hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&= \sum_{i\in K}\int_{\mathbb{R}^2\times J_i} 2\{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-\hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&\quad + \sum_{i\in L}\int_{\mathbb{R}^2\times J_i} 2\{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-\hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&= \int_{\mathbb{R}^2\times[0,b]} 2\{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-\hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}dP(t_1,t_2,c) \\
&\quad + r'_k(\omega) + o_p(1),
\end{aligned}$$

where $|r'_k(\omega)| \le c'\delta$, for a constant $c>0$. This can be seen by replacing $\hat{F}_{2,n_k}(t;\omega)$ on each interval $J_i$ by its value $\hat{F}_{2,n_k}(u_i;\omega)$ at the right endpoint of the interval, and by noting that for large $k$

$$\left|\frac{1}{[1-\hat{F}_{2,n_k}(t;\omega)]^2} - \frac{1}{[1-\hat{F}_{2,n_k}(u_i;\omega)]^2}\right| < 2\delta, \qquad i \in L.$$

On the intervals $J_i$ with $i \in K$ we use the second inequality in (A.2).

Letting $P(\mathbb{R}^2\times J_i) = \int I(0\le t_1\le t_2)I(c\in J_i)dP(t_1,t_2,c)$, notice that $\sum_{i\in K}P(\mathbb{R}^2\times J_i) \to 0$, if $\delta \downarrow 0$, since $P(\mathbb{R}^2\times J_i)$ is of order $O(\delta^2)$, while the number of intervals $J_i$ such that $i\in K$ is of order $O(1/\delta)$.

Dominated convergence implies

$$\begin{aligned}
\text{(A.6)}\quad & \lim_{k\to\infty}\int_{\mathbb{R}^2\times[0,b]} \{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-\hat{F}_{1,n_k}(c;\omega)]/[1-\hat{F}_{2,n_k}(c;\omega)]^2\}dP(t_1,t_2,c) \\
&= \int_{\mathbb{R}^2\times[0,b]} \{I(t_1>c)[1-F_2(c)] - [1-F_1^*(c)]\}I(t_2>c) \\
&\qquad \times \{[F_1(c)-F_1^*(c)]/[1-F_2(c)]^2\}dP(t_1,t_2,c).
\end{aligned}$$

Combining (A.5) and (A.6) we obtain

$$\int_{\mathbb{R}^2\times[0,b]} 2\{I(t_1>c)[1-\hat{F}_{2,n_k}(c;\omega)] - [1-\hat{F}_{1,n_k}(c;\omega)]\}I(t_2>c)$$

$$\times \{[F_1(c) - \hat{F}_{1,n_k}(c;\omega)]/[1 - \hat{F}_{2,n_k}(c;\omega)]^2\} d\mathbb{P}_{n_k}(t_1, t_2, c; \omega)$$
$$= \int_{\mathbb{R}^2\times[0,b]} 2\{I(t_1 > c)[1 - F_2(c)] - [1 - F_1^*(c)]\} I(t_2 > c)$$
$$\times \{[F_1(c) - F_1^*(c)]/[1 - F_2(c)]^2\} dP(t_1, t_2, c) + r_k(\omega) + o_p(1),$$

where $|r_k(\omega)| \le c\delta$.

Since $\delta$ can be made arbitrarily small, (A.3) now follows, and relation (A.4) follows from (A.3) and (A.1). □

By monotone convergence and (A.4) we obtain

$$\text{(A.7)} \quad \int_{\mathbb{R}^3} 2\{I(t_1 > c)[1 - F_2(c)] - [1 - F_1^*(c)]\} I(t_2 > c)$$
$$\times \{[F_1(c) - F_1^*(c)]/[1 - F_2(c)]^2\} dP(t_1, t_2, c)$$
$$= \lim_{b\to\infty} \int_{\mathbb{R}^2\times[0,b]} 2\{I(t_1 > c)[1 - F_2(c)] - [1 - F_1^*(c)]\} I(t_2 > c)$$
$$\times \{[F_1(c) - F_1^*(c)]/[1 - F_2(c)]^2\} dP(t_1, t_2, c) \ge 0.$$

This, however, can only happen if $F_1^* = F_1$, since we have

$$\int_{\mathbb{R}^3} 2\{I(t_1 > c)[1 - F_2(c)] - [1 - F_1^*(c)]\} I(t_2 > c)$$
$$\times \{[F_1(c) - F_1^*(c)]/[1 - F_2(c)]^2\} dP(t_1, t_2, c)$$
$$= 2\int_{\mathbb{R}^3} I(t_1 > c)[1 - F_2(c)] \frac{[F_1(c) - F_1^*(c)]}{[1 - F_2(c)]^2} dP(t_1, t_2, c)$$
$$- 2\int_{\mathbb{R}^3} [1 - F_1^*(c)][F_1(c) - F_1^*(c)] \frac{I(t_2 > c)}{[1 - F_2(c)]^2} dP(t_1, t_2, c)$$
$$= 2\left\{\int_{\mathbb{R}} [1 - F_1(c)] \frac{[F_1(c) - F_1^*(c)]}{1 - F_2(c)} dG(c) - \int_{\mathbb{R}} [1 - F_1^*(c)] \frac{[F_1(c) - F_1^*(c)]}{1 - F_2(c)} dG(c)\right\}$$
$$= -2\int_{\mathbb{R}} \frac{[F_1(c) - F_1^*(c)]^2}{1 - F_2(c)} dG(c) \le 0,$$

and the latter expression is strictly negative, unless $F_1^* = F_1$, since by the monotonicity of $F_1^*$, the monotonicity and continuity of $F_1$, and the absolute continuity of $P_{F_1}$ w.r.t. $P_G$, we have $F_1^* \ne F_1 \Rightarrow F_1^*(t) \ne F_1(t)$ on an interval of increase of $G$, which implies

$$-2\int_{\mathbb{R}} \frac{[F_1(c) - F_1^*(c)]^2}{1 - F_2(c)} dG(c) < 0$$

if $F_1^* \ne F_1$, which contradicts (A.7).

Thus we have proved that for each $\omega$ outside a set of probability zero, each subsequence of the sequence $\hat{F}_{1,n}(\cdot;\omega)$ has a vaguely convergent sequence, and that all these convergent subsequences have the same limit $F_1$. This proves that the sequence $\hat{F}_{1,n}$ converges weakly to $F_1$, with probability one. Since $F_1$ is continuous, this is the same as saying that $\hat{F}_{1,n}$ converges with probability one to $F_1$ in the supremum distance on the set of distribution functions, i.e.,

$$\mathbb{P}\left(\lim_{n\to\infty}\sup_{t\in\mathbb{R}}|\hat{F}_{1,n}(t)-F_1(t)|=0\right)=1$$

or

$$\|\hat{F}_{1,n}-F_1\|_\infty\to 0$$

almost surely. □

PROOF OF THEOREM 4.1. Take $\ell(x)=|x|$. Since $q_n$ and $q_0$ coincide when $\Delta_1=\Delta_2=1$, we have

$$\begin{aligned}
H^2(q_n,q_0) &= \int_{t_0-n^{-1/3}}^{t_0} g(c)\{[S_1(t_0-n^{-1/3})]^{1/2}-[S_1(c)]^{1/2}\}^2 dc\\
&\quad+\int_{t_0}^{t_0+n^{-1/3}} g(c)\{[S_1(t_0+n^{-1/3})]^{1/2}-[S_1(c)]^{1/2}\}^2 dc\\
&\quad+\int_{t_0-n^{-1/3}}^{t_0} g(c)\{[S_2(c)-S_1(t_0-n^{-1/3}t)]^{1/2}-[S_2(c)-S_1(c)]^{1/2}\}^2 dc\\
&\quad+\int_{t_0}^{t_0+n^{-1/3}} g(c)\{[S_2(c)-S_1(t_0+n^{-1/3}t)]^{1/2}-[S_2(c)-S_1(c)]^{1/2}\}^2 dc\\
&\cong g(t_0)\{[S_1(t_0-n^{-1/3}t)]^{1/2}-[S_1(t_0)]^{1/2}\}^2 n^{-1/3}t/2\\
&\quad+g(t_0)\{[S_1(t_0+n^{-1/3}t)]^{1/2}-[S_1(t_0)]^{1/2}\}^2 n^{-1/3}t/2\\
&\quad+g(t_0)\{[S_2(t_0)-S_1(t_0-n^{-1/3}t)]^{1/2}-[S_2(t_0)-S_1(t_0)]^{1/2}\}^2 n^{-1/3}t/2\\
&\cong 2g(t_0)\left\{\frac{f_1(t_0)n^{-1/3}t}{2[S_1(t_0)]^{1/2}}\right\}^2\frac{n^{-1/3}t}{2}+2g(t_0)\left\{\frac{f_1(t_0)n^{-1/3}t}{2[S_2(t_0)-S_1(t_0)]^{1/2}}\right\}^2\frac{n^{-1/3}t}{2}\\
&= g(t_0)f_1^2(t_0)n^{-1}t^3S_2(t_0)/\{4S_1(t_0)[S_2(t_0)-S_1(t_0)]\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
&n^{1/3}\inf_{\hat{F}_{1,n}}\max\{E_{n,q_0}[|\hat{F}_{1,n}(t_0)-F_1(t_0)|],E_{n,q_n}[|\hat{F}_{1,n}(t_0)-F_{1,n}(t_0)|]\}\\
&\quad\geq\frac{1}{4}n^{1/3}|F_{1,n}(t_0)-F_1(t_0)|[1-H^2(q_n,q_0)]^{2n}\\
&\quad\cong\frac{1}{4}n^{1/3}|F_{1,n}(t_0)-F_1(t_0)|\left\{1-\frac{S_2(t_0)g(t_0)f_1^2(t_0)t^3n^{-1}}{4S_1(t_0)[S_2(t_0)-S_1(t_0)]}\right\}^{2n}\\
&\quad\cong\frac{1}{4}n^{1/3}f_1(t_0)tn^{-1/3}\left\{1-\frac{S_2(t_0)g(t_0)f_1^2(t_0)t^3n^{-1}}{4S_1(t_0)[S_2(t_0)-S_1(t_0)]}\right\}^{2n}\\
&\quad\to\frac{1}{4}f_1(t_0)t\exp\left\{-\frac{S_2(t_0)g(t_0)f_1^2(t_0)t^3}{2S_1(t_0)[S_2(t_0)-S_1(t_0)]}\right\}\\
&\quad\equiv bt\exp(-at^3).
\end{aligned}$$

The last expression is maximized over $t$ by $t'=(1/3a)^{1/3}$, yielding the minimax lower bound

$$bt'e^{-1/3}=k\{f_1(t_0)S_1(t_0)[S_2(t_0)-S_1(t_0)]/[S_2(t_0)g(t_0)]\}^{1/3}$$

where $k = (1/4)(3e/2)^{-1/3}$ does not depend on $f_1$, $F_1$ or $g$. □

PROOF OF THEOREM 5.1. In the proof of Theorem 5.1 we will use the approach of Groeneboom and Wellner (1992) in establishing the asymptotic distribution of the nonparametric maximum likelihood estimator of a distribution function $F$ with case 1 of interval censoring data. We show also that replacing $F_2$ by its Kaplan-Meier estimator $\hat{F}_{2,n}$ does not affect the asymptotic behavior of $\hat{F}_{1,n}$ since its $n^{1/3}$-rate of convergence is slower than the $\sqrt{n}$-rate of convergence of $\hat{F}_{2,n}$.

In Section 2 we saw that $\hat{S}_{1,n}(t)$ is given by the slope of the least concave majorant at $W_j$ if $t \in (Y_{j-1}, Y_j]$. With $D = \{(t_1,t_2) \in \mathbb{R}^2 : 0 \leq t_1 \leq t_2\}$, $I(D \times (u,t])$ will denote the indicator function of the set

$$\{(t_1,t_2,c) \in \mathbb{R}^3 : 0 \leq t_1 \leq t_2 < \infty, u < c \leq t\}.$$

Let $A = \{(t_1,t_2,c) \in \mathbb{R}^3 : 0 < c < t_1\}$ and $B = \{(t_1,t_2,c) \in \mathbb{R}^3 : 0 < c < t_2\}$, and $Pf = \int f dP$ for any probability measure $P$. Define the processes

$$\begin{aligned} \mathbb{W}_n(t) &= \mathbb{P}_n\left(\frac{I(B)I(D\times(0,t])}{[1-F_2(c)]^2}\right) = \int_0^t \iint_{0<t_1<t_2} \frac{I(t_2>c)}{[1-F_2(c)]^2} d\mathbb{P}_n(t_1,t_2,c) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{I(T_{2,i}>C_i)I(C_i\leq t)}{[1-F_2(C_i)]^2} = W_j \quad \text{for} \quad t\in[Y_j,Y_{j+1}), \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}_n(t) &= \mathbb{P}_n\left(\frac{I(A)I(D\times(0,t])}{1-F_2(c)}\right) = \int_0^t \iint_{0<t_1<t_2} \frac{I(t_1>c)}{1-F_2(c)} d\mathbb{P}_n(t_1,t_2,c) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{I(T_{1,i}>C_i)I(C_i\leq t)}{1-F_2(C_i)} = V_j \quad \text{for} \quad t\in[Y_j,Y_{j+1}). \end{aligned}$$

Then the function $s \mapsto \mathbb{V}_n \circ \mathbb{W}_n^{-1}(s)$ equals the cumulative sum diagram in Section 2. Since $\hat{S}_{1,n}(t)$ is given by the slope of the least concave majorant of the cumulative sum diagram defined by $(\mathbb{W}_n, \mathbb{V}_n)$, we have that if $\hat{S}_{1,n}(t) \leq a$ then a line of slope $a$ moved down vertically from $+\infty$ first hits the cumulative sum diagram to the left of $t$ (see Fig. 2). The point where the line hits the diagram is the point where $\mathbb{V}_n$ is farthest above the line of slope $a$ through the origin. Thus,

$$\hat{S}_{1,n}(t) \leq a \Leftrightarrow \hat{s}_n(a) = \arg\max_s\{\mathbb{V}_n(s) - a\mathbb{W}_n(s)\} \leq t$$

and we can derive the limit distribution of $\hat{S}_{1,n}(t)$ by studying the locations of the maxima of the sequence of processes $s \mapsto \mathbb{V}_n(s) - a\mathbb{W}_n(s)$ since

$$P(n^{1/3}[\hat{S}_{1,n}(t_0) - S_1(t_0)] \leq x) = P(\hat{s}_n(S_1(t_0) + xn^{-1/3}) \leq t_0).$$

Making the change of variables $s \to t_0 + n^{-1/3}t$ we obtain

$$\begin{aligned} &\hat{s}_n(S_1(t_0) + xn^{-1/3}) - t_0 \\ &\quad = n^{-1/3}\arg\max_t\{\mathbb{V}_n(t_0+n^{-1/3}t) - (S_1(t_0)+xn^{-1/3})\mathbb{W}_n(t_0+n^{-1/3}t)\} \\ &\quad = n^{-1/3}\arg\max_t\left\{\int_0^{t_0+n^{-1/3}t}\iint_{0<t_1<t_2}\frac{I(t_1>c)I(t_2>c)}{1-F_2(c)}d\mathbb{P}_n(t_1,t_2,c)\right. \end{aligned}$$

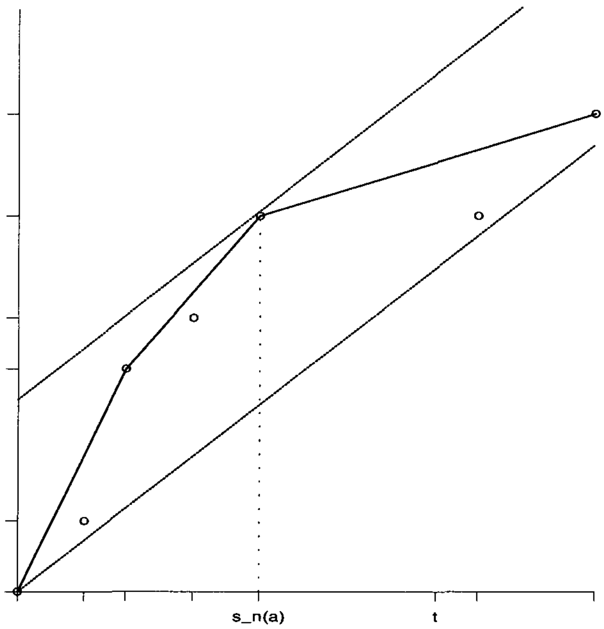


Fig. 2. A cumulative sum diagram and the corresponding least concave majorant.

$$
\begin{aligned}
&\qquad - (S_1(t_0) + xn^{-1/3}) \\
&\qquad \times \int_0^{t_0+n^{-1/3}t} \iint_{0<t_1<t_2} \frac{I(t_2 > c)}{[1 - F_2(c)]^2} d\mathbb{P}_n(t_1, t_2, c)\Bigg\} \\
&= n^{-1/3} \arg\max_t \{\mathbb{P}_n(I(A)I(D \times (0, t_0 + n^{-1/3}t])/[1 - F_2(c)]) \\
&\qquad - S_1(t_0)\mathbb{P}_n(I(B)I(D \times (0, t_0 + n^{-1/3}t])/[1 - F_2(c)]^2) \\
&\qquad - xn^{-1/3}\mathbb{P}_n(I(B)I(D \times (0, t_0 + n^{-1/3}t])/[1 - F_2(c)]^2)\}.
\end{aligned}
$$

The location of the maximum of a function does not change when the function is multiplied by a positive constant or shifted vertically. Thus, the arg max above is also a point of maximum of the process

$$
\begin{aligned}
\text{(A.8)} \qquad & n^{2/3}\Bigg\{(\mathbb{P}_n - P)\left(\left[\frac{I(A)\,S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&+ P\left(\left[\frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&- xn^{-1/3}\mathbb{P}_n\left(\frac{I(B)I(D \times (t_0, t_0 + n^{-1/3}t])}{S_2^2(c)}\right)\Bigg\} \\
&= n^{2/3}(M_1 + M_2 + M_3).
\end{aligned}
$$

So the probability of interest is

$$
P(\hat{s}_n(S_1(t_0) + xn^{-1/3}) \le t_0) = P\left(\arg\max_t\{n^{2/3}(M_1 + M_2 + M_3)\} \le 0\right).
$$

Notice that $F_2$ is unknown and should be substituted by $\hat{F}_{2,n}$. We will rewrite the arg max of (A.8) as

$$\begin{aligned}
\arg\max_t n^{2/3}\Bigg\{ &(\mathbb{P}_n - P)\Bigg(\Bigg[\frac{I(A)\hat{S}_{2,n}(c) - I(B)S_1(t_0)}{\hat{S}^2_{2,n}(c)} \\
&\qquad - \frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\Bigg] I(D \times (t_0, t_0 + n^{-1/3}t])\Bigg) \\
&+ (\mathbb{P}_n - P)\left(\left[\frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&+ P\Bigg(\Bigg[\frac{I(A)\hat{S}_{2,n}(c) - I(B)S_1(t_0)}{\hat{S}^2_{2,n}(c)} \\
&\qquad - \frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\Bigg] I(D \times (t_0, t_0 + n^{-1/3}t])\Bigg) \\
&+ P\left(\left[\frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&- xn^{-1/3}(\mathbb{P}_n - P)\left(I(B)\left[\frac{1}{\hat{S}^2_{2,n}(c)} - \frac{1}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&- xn^{-1/3}\mathbb{P}_n\left(\frac{I(B)I(D \times (t_0, t_0 + n^{-1/3}t])}{S_2^2(c)}\right)\Bigg\} \\
&- xn^{-1/3}P\left(I(B)\left[\frac{1}{\hat{S}^2_{2,n}(c)} - \frac{1}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right)
\end{aligned}$$

$$= \arg\max_t n^{2/3}\{I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7\}. \tag{A.9}$$

We will now analyze each of the terms in the arg max expression separately. For term $I_1$ we have

$$\begin{aligned}
n^{2/3}&(\mathbb{P}_n - P)\left(I(t_1 > c)\left[\frac{1}{\hat{S}_{2,n}(c)} - \frac{1}{S_2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&+ n^{2/3}(\mathbb{P}_n - P)\left(I(t_2 > c)S_1(t_0)\left[\frac{1}{\hat{S}^2_{2,n}(c)} - \frac{1}{S_2^2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right)
\end{aligned}$$

and each term converges uniformly in probability to 0. In fact,

$$\begin{aligned}
&\sup_{0 \le t \le t_0 + M}\left|n^{2/3}(\mathbb{P}_n - P)\left(I(A)\left[\frac{1}{\hat{S}_{2,n}(c)} - \frac{1}{S_2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right)\right| \\
&\quad + \sup_{0 \le t \le t_0 + M}\Bigg|n^{2/3}(\mathbb{P}_n - P)\Bigg(I(B)S_1(t_0)\left[\frac{1}{\hat{S}^2_{2,n}(c)} - \frac{1}{S_2^2(c)}\right] \\
&\qquad \times I(D \times (t_0, t_0 + n^{-1/3}t])\Bigg)\Bigg| \\
&= \sup_{0 \le t \le t_0 + M}\Bigg|n^{1/2}(\mathbb{P}_n - P)\Bigg(n^{1/6}I(A)\left[\frac{S_2(c) - \hat{S}_{2,n}(c)}{S_2(c)\hat{S}_{2,n}(c)}\right]
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad \times I(c \in (t_0, t_0 + n^{-1/3}t])\Big)\Big| \\
&+ \sup_{0 \le t \le t_0 + M} \Big| n^{1/2}(\mathbb{P}_n - P)\Big( n^{1/6} I(B) S_1(t_0) \left[ \frac{S_2^2(c) - \hat{S}_{2,n}^2(c)}{S_2^2(c)\hat{S}_{2,n}^2(c)} \right] \\
&\qquad\qquad \times I(c \in (t_0, t_0 + n^{-1/3}t])\Big)\Big| \\
&\le \frac{n^{1/2}\|\hat{S}_{2,n}(c) - S_2(c)\|_0^{t_0+M}}{\hat{S}_{2,n}(t_0+M)S_2(t_0+M)} n^{1/6} S_1(t_0)(\mathbb{P}_n - P) I(c \in (t_0, t_0 + n^{-1/3}t]) \\
&\quad + \frac{n^{1/2}\|\hat{S}_{2,n}^2(c) - S_2^2(c)\|_0^{t_0+M}}{\hat{S}_{2,n}^2(t_0+M)S_2^2(t_0+M)} n^{1/6}(\mathbb{P}_n - P) I(c \in (t_0, t_0 + n^{-1/3}t]) \\
&= O_p(1)o_p(1) + O_p(1)o_p(1)
\end{aligned}
$$

since the rate of convergence of $\hat{S}_{2,n}$ is $\sqrt{n}$.

Consider $I_2$ now. Taking

$$f_{n,t}(t_1, t_2, c) = n^{1/6}\{[I(A)S_2(c) - I(B)S_1(t_0)]/S_2^2(c)\} I(D \times (t_0, t_0 + n^{-1/3}t])$$

and $F_n(t_1, t_2, c) = (n^{1/6}/\delta^2) I(D \times (t_0, t_0 + n^{-1/3}t])$ we have, by Theorems 2.11.23 and 2.7.11 in van der Vaart and Wellner (1996),

$$n^{2/3}(\mathbb{P}_n - P)(\{[I(A)S_2(c) - I(B)S_1(t_0)]/S_2^2(c)\} I(D \times (t_0, t_0 + n^{-1/3}t]))$$

converging to a mean zero Gaussian process with covariance function (for $0 < s < t$) given by

$$
\begin{aligned}
&(n^{2/3}/n^{1/2})^2 \boldsymbol{E}[\boldsymbol{E}(\{[I(A)S_2(C) - I(B)S_1(t_0)]/S_2^2(C)\}^2 \\
&\qquad\qquad \times I(D \times (t_0 + n^{-1/3}s, t_0 + n^{-1/3}t]) \mid C)] \\
&= n^{1/3} \boldsymbol{E}[(\{[S_2(C) - S_1(t_0)]/S_2^2(C)\}^2 S_1(C) \\
&\qquad + [S_1(t_0)/S_2^2(C)]^2 P(T_1 < C < T_2)) I(C \in (t_0 + n^{-1/3}s, t_0 + n^{-1/3}t])] \\
&= n^{1/3} \int_{t_0 + n^{-1/3}s}^{t_0 + n^{-1/3}t} (\{[S_2(u) - S_1(t_0)]/S_2^2(u)\}^2 S_1(u) \\
&\qquad\qquad + [S_1(t_0)/S_2^2(u)]^2 [S_2(u) - S_1(u)]) g(u) du \\
&\cong n^{1/3}(\{[S_2(t_0) - S_1(t_0)]/S_2^2(t_0)\}^2 S_1(t_0) \\
&\qquad + [S_1(t_0)/S_2^2(t_0)]^2 [S_2(t_0) - S_1(t_0)]) g(t_0) n^{-1/3}|t - s| \\
&= \{[S_2(t_0) - S_1(t_0)] S_1(t_0) [S_2(t_0) - S_1(t_0) + S_1(t_0)] g(t_0) n^{-1/3}|t - s|\}/S_2^4(t_0) \\
&= [S_2(t_0) - S_1(t_0)] S_1(t_0) g(t_0) |t - s| / S_2^3(t_0).
\end{aligned}
$$

For term $I_3$ we have

$$
\begin{aligned}
n^{2/3} P\Big(\Big[ &\frac{I(A)\hat{S}_{2,n}(c) - I(B)S_1(t_0)}{\hat{S}_{2,n}^2(c)} \\
&- \frac{I(A)S_2(c) - I(B)S_1(t_0)}{S_2^2(c)} \Big] I(D \times (t_0, t_0 + n^{-1/3}t])\Big)
\end{aligned}
$$

$$
\begin{aligned}
&= n^{2/3} P\left(\left[\frac{I(A)}{\hat{S}_{2,n}(c)} - \frac{I(A)}{S_2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&\quad - n^{2/3} P\left(\left[\frac{I(B)S_1(t_0)}{\hat{S}^2_{2,n}(c)} - \frac{I(B)S_1(t_0)}{S^2_2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right).
\end{aligned}
$$

For the first term in the sum above, assuming $S_1$ and $g$ continuous, we have

$$
\begin{aligned}
&n^{2/3} P\left(\left[\frac{I(A)}{\hat{S}_{2,n}(c)} - \frac{I(A)}{S_2(c)}\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&\le n^{2/3} \int_{t_0}^{t_0+n^{-1/3}t} \iint_{c<t_1<t_2} \frac{|S_2(c) - \hat{S}_{2,n}(c)|}{\hat{S}_{2,n}(t_0 + n^{-1/3}t) S_2(t_0 + n^{-1/3}t)} dF(t_1, t_2) dG(c) \\
&= n^{2/3} \int_{t_0}^{t_0+n^{-1/3}t} \frac{S_1(c)|S_2(c) - \hat{S}_{2,n}(c)|}{\hat{S}_{2,n}(t_0 + n^{-1/3}t) S_2(t_0 + n^{-1/3}t)} dG(c) \\
&\le n^{2/3} \frac{\|S_1(t)g(t)\|_{t_0}^{t_0+n^{-1/3}t} n^{-1/3}t \|S_2(t) - \hat{S}_{2,n}(t)\|_{t_0}^{t_0+M}}{\hat{S}_{2,n}(t_0 + n^{-1/3}t) S_2(t_0 + n^{-1/3}t)} \\
&= \frac{n^{2/3} O(n^{-1/3}) O_p(n^{-1/2})}{\hat{S}_{2,n}(t_0 + n^{-1/3}t) S_2(t_0 + n^{-1/3}t)} = O_p(n^{-1/6}) = o_p(1).
\end{aligned}
$$

Similarly, for the second term, assuming $S_2$ continuous,

$$
\begin{aligned}
&n^{2/3} P\left(\left[I(B)S_1(t_0)\left(\frac{1}{\hat{S}^2_{2,n}(c)} - \frac{1}{S^2_2(c)}\right)\right] I(D \times (t_0, t_0 + n^{-1/3}t])\right) \\
&\le n^{2/3} \int_{t_0}^{t_0+n^{-1/3}t} \iint_{c<t_2} \frac{|S^2_2(c) - \hat{S}^2_{2,n}(c)|}{\hat{S}^2_{2,n}(t_0 + n^{-1/3}t) S^2_2(t_0 + n^{-1/3}t)} dF(t_1, t_2) dG(c) \\
&= n^{2/3} \int_{t_0}^{t_0+n^{-1/3}t} \frac{S_2(c)|S^2_2(c) - \hat{S}^2_{2,n}(c)|}{\hat{S}^2_{2,n}(t_0 + n^{-1/3}t) S^2_2(t_0 + n^{-1/3}t)} dG(c) \\
&\le n^{2/3} \frac{\|S_2(t)g(t)\|_{t_0}^{t_0+n^{-1/3}t} n^{-1/3}t \|S^2_2(t) - \hat{S}^2_{2,n}(t)\|_{t_0}^{t_0+M}}{\hat{S}^2_{2,n}(t_0 + n^{-1/3}t) S^2_2(t_0 + n^{-1/3}t)} \\
&= \frac{n^{2/3} O(n^{-1/3}) O_p(n^{-1/2})}{\hat{S}_{2,n}(t_0 + n^{-1/3}t) S_2(t_0 + n^{-1/3}t)} = O_p(n^{-1/6}) = o_p(1).
\end{aligned}
$$

The limit of term $I_4$ can be easily calculated as

$$
\begin{aligned}
&n^{2/3} \boldsymbol{E}[\boldsymbol{E}(\{[I(A)S_2(C) - I(B)S_1(t_0)]/S^2_2(C)\} I(D \times (t_0, t_0 + n^{-1/3}t]) \mid C)] \\
&\quad = n^{2/3} \boldsymbol{E}\{[S_1(C)S_2(C) - S_2(C)S_1(t_0)] I(C \in (t_0, t_0 + n^{-1/3}t])/S^2_2(C)\} \\
&\quad = n^{2/3} \int_{t_0}^{t_0+n^{-1/3}t} \{[F_1(t_0) - F_1(u)] g(u)/S_2(u)\} du \\
&\quad \cong n^{2/3} \frac{1}{2} \left(\frac{F_1(t_0) - F_1(t_0 + n^{-1/3}t)}{S_2(t_0 + n^{-1/3}t)}\right) g(t_0 + n^{-1/3}t) n^{-1/3}t \\
&\quad \cong -n^{2/3} f_1(t_0) n^{-1/3} t g(t_0) n^{-1/3} t/[2S_2(t_0)] = -f_1(t_0) g(t_0) t^2/[2S_2(t_0)].
\end{aligned}
$$

The uniform convergence in probability to zero of terms $I_5$ and $I_7$ in (A.9) have been established (up to the constant $S_1(t_0)$) in the evaluation of the limit of term $I_1$ and in the evaluation of the limit of the second part of term $I_3$, respectively.

And finally the limit behavior of term $I_6$ is calculated below.

$$\begin{aligned}
&-n^{2/3}xn^{-1/3}\mathbb{P}_n(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c))\\
&\quad=-n^{2/3}xn^{-1/3}(\mathbb{P}_n-P)(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c))\\
&\qquad-n^{2/3}xn^{-1/3}P(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c))\\
&\quad=-n^{1/2}\frac{n^{1/3}}{n^{1/2}}x(\mathbb{P}_n-P)(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c))\\
&\qquad-n^{1/3}xP(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c)).
\end{aligned}$$

Taking $F_n(t_1,t_2,c)=xI(D\times(t_0,t_0+n^{-1/3}t])/(n\delta^2)$, Theorems 2.11.23 and 2.7.11 in van der Vaart and Wellner (1996) imply that the first part of the sum above converges to a mean zero Gaussian process with covariance function given by

$$\begin{aligned}
&x^2(n^{2/3}/n)\boldsymbol{E}(\boldsymbol{E}\{[I(B)/S_2^2(C)]^2I(D\times(t_0+n^{-1/3}s,t_0+n^{-1/3}t])\mid C\})\\
&\quad=(x^2/n^{1/3})\boldsymbol{E}[S_2(C)I(C\in(t_0+n^{-1/3}s,t_0+n^{-1/3}t])/S_2^4(C)]\\
&\quad=(x^2/n^{1/3})\int_{t_0+n^{-1/3}s}^{t_0+n^{-1/3}t}[g(u)/S_2^3(u)]du\cong x^2n^{-2/3}\frac{g(t_0)}{S_2^3(t_0)}|t-s|
\end{aligned}$$

which converges to 0 as $n\to\infty$. The second part gives

$$\begin{aligned}
&-xn^{1/3}P(I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(c))\\
&\quad=-xn^{1/3}\boldsymbol{E}\{\boldsymbol{E}[I(B)I(D\times(t_0,t_0+n^{-1/3}t])/S_2^2(C)\mid C]\}\\
&\quad=-xn^{1/3}\int_{t_0}^{t_0+n^{-1/3}t}[S_2(u)g(u)/S_2^2(u)]du\\
&\quad\cong-xn^{1/3}g(t_0)tn^{-1/3}/S_2(t_0)=-xg(t_0)t/S_2(t_0).
\end{aligned}$$

Exercise 3.2.5, p. 308, in van der Vaart and Wellner (1996) states that the random variables $\arg\max_t\{a\mathbb{B}(t)-bt^2-ct\}$ and $(a/b)^{2/3}\arg\max_t\{\mathbb{B}(t)-t^2\}-c/(2b)$ are equal in distribution, where $\{\mathbb{B}(t):t\in\mathbb{R}\}$ is a standard two-sided Brownian motion with $\mathbb{B}(0)=0$, and $a$, $b$ and $c$ are positive constants. Thus, making $a=\{S_1(t_0)g(t_0)[S_2(t_0)-S_1(t_0)]/S_2^3(t_0)\}^{1/2}$, $b=f_1(t_0)g(t_0)/[2S_2(t_0)]$ and $c=g(t_0)x/S_2(t_0)$ we have

$$\begin{aligned}
P(n^{1/3}[\hat{S}_{1,n}(t_0)-S_1(t_0)]\le x)&=P(\hat{s}_n(S_1(t_0)+xn^{-1/3})\le t_0)\\
&=P\left(\arg\max_t\{n^{2/3}(M_1+M_2+M_3)\}\le 0\right)\\
&\to P\left(\arg\max_t\{a\mathbb{B}(t)-bt^2-ct\}\le 0\right)\\
&=P\left((a/b)^{2/3}\arg\max_t\{\mathbb{B}(t)-t^2\}-c/(2b)\le 0\right)\\
&=P\left(2Z\le\left\{\frac{2S_2(t_0)g(t_0)}{f_1(t_0)S_1(t_0)[S_2(t_0)-S_1(t_0)]}\right\}^{1/3}x\right)
\end{aligned}$$

which implies that

$$P\left(n^{1/3}\frac{\hat{S}_{1,n}(t_0)-S_1(t_0)}{\{f_1(t_0)S_1(t_0)[S_2(t_0)-S_1(t_0)]/[2S_2(t_0)g(t_0)]\}^{1/3}}\le z\right)\to P(2Z\le z).\qquad\square$$

## References


Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*, Wiley, New York.

Dudley, R. M. (1989). *Real Analysis and Probability*, Wadsworth and Brooks/Cole, Pacific Grove, California.

Gomes, A. E., Groeneboom, P. and Wellner, J. A. (2001). Nonparametric estimation of the lifetime and disease onset distributions for a survival-sacrifice model, (pre-print), http://ssor.twi.tudelft.nl/~pietg/surv.pdf

Groeneboom, P. (1987). *Asymptotics for the Interval Censored Observations*, Tech. Report, No. 87-18, Department of Mathematics, University of Amsterdam.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions, *Probability Theory and Related Fields*, **81**, 79–109.

Groeneboom, P. (1996). *Lectures on Inverse Problems*, Lecture Notes in Mathematics, **1648**, 67–164, Springer, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser, Boston.

Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution, *Journal of Computational and Graphical Statistics*, **10**, 388–400.

Holland, J. M., Mitchell, T. J. and Walburg, H. E. (1977). Effects of prepubertal ovariectomy on survival and specific diseases in female RFM mice given 300 R of X rays, *Radiation Research*, **69**, 317–327.

Jewell, N. P. (1982). Mixtures of exponential distribution, *The Annals of Statistics*, **10**, 479–484.

Kodell, R., Shaw, G. and Johnson, A. (1982). Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments, *Biometrics*, **38**, 43–58.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

Turnbull, B. W. and Mitchell, T. J. (1984). Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments, *Biometrics*, **40**, 41–50.

van der Laan, M. J., Jewell, N. P. and Peterson, D. (1997). Efficient estimation of the lifetime and disease onset distribution, *Biometrika*, **84**, 539–554.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.

Wang, J.-G. (1987). A note on the uniform consistency of the Kaplan-Meier estimator, *The Annals of Statistics*, **15**, 1313–1316.