Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Consistency of a Weighted Least Squares Estimator of the Disease Onset Distribution Function for a Survival-Sacrifice Model

Antonio Eduardo Gomes

Relatório Técnico
RTP-11/2001

Relatório Técnico
Série Pesquisa

Consistency of a Weighted Least Squares Estimator of the Disease Onset Distribution Function for a Survival-Sacrifice Model

Antonio Eduardo Gomes

Departamento de Estatística - Universidade Federal de Minas Gerais

Abstract

One of the possible data structures that may occur in carcinogenicity experiments with animals where the tumor is not palpable consists of 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. Defining the non-negative variables T_1 (time of tumor onset), T_2 (time of death from the tumor) and C (time of death from an unrelated cause), we observe (Y, Δ_1, Δ_2), where $Y = \min\{T_2, C\}$, $\Delta_1 = 1_{\{T_1 \leq C\}}$, and $\Delta_2 = 1_{\{T_2 \leq C\}}$, T_1 and T_2 have a joint distribution function F such that $P(T_1 \leq T_2) = 1$, and are independent of C. This structure is a "survival-sacrifice model". van der Laan, Jewell, and Peterson (1997) proposed a Weighted Least Squares estimator for F_1 (the marginal distribution of T_1) using the Kaplan-Meier estimator of F_2 (the marginal distribution of T_2). Strong uniform consistency of their estimator is established.

1 Introduction

In experiments for the study of onset and mortality from undetectable irreversible diseases (occult tumors, e.g.) a possible data structure consists of 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. This data structure is related to moderately lethal incurable diseases when the cause of death is known. Defining the non-negative 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 ith individual, $(Y_i, \Delta_{1,i}, \Delta_{2,i})$, where $Y_i = C_i \wedge T_{2,i} = \min\{C_i, T_{2,i}\}$, $\Delta_{1,i} = 1_{\{T_{1,i} \leq C_i\}}$, $\Delta_{2,i} = 1_{\{T_{2,i} \leq C_i\}}$, $T_{1,i}$ and $T_{2,i}$ have a 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 $(T_{1,i}, T_{2,i})$. Some

authors call this model a "survival-sacrifice model" (see GROENEBOOM (1998)). Current status data can be seen as a particular case of this data structure when the disease is non-lethal, i.e., $\Delta_{2,i} = 0, i = 1, \ldots, n$ (in this case, $Y_i = C_i$, and $\hat{F}_2 \equiv 0$ for any estimator of F_2). The right censoring problem can also be considered as a special case of data with the structure above when a lethal disease is always present at the moment of death, i.e., $\Delta_{1,i} = 1, i = 1, \ldots n$ (in this case, $\hat{F}_1 \equiv 1$ for any estimator of F_1).
An example of data with the structure considered here is given in HOLLAND, MITCHELL, AND WALBURG (1977) and is shown in Talbe 1. These data were studied by DINSE AND LAGAKOS (1982) and TURNBULL AND MITCHELL (1984) and represent 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.

Table 1: *Ages at death (in days) in unexposed female RFM mice.*

$\Delta_1 = 1, \Delta_2 = 1$	406,461,482,508,553,555,562,564,570,574,585,588,593, 624,626, 629,647,658,666,675,679,688,690,691,692,698,699,701,702,703, 707,717,724,736,748,754,759,770,772,776,776,785,793,800,809, 811,823,829,849,853,866,883,884,888,889
$\Delta_1 = 1, \Delta_2 = 0$	356,381,545,615,708,750,789,838,841,875
$\Delta_1 = 0, \Delta_2 = 0$	192,234,243,300,303,330,339,345,351,361,368,419,430,430,464, 488,494,496,517,552,554,555,563,583,629,638,642,656,668,669, 671,694,714,730,731,732,756,756,782,793,805,821,828,853

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

$$\Theta = \{(F_1, F_2) : F_1 \text{ and } F_2 \text{ are d.f.'s with } F_1 <_s F_2\} \ ,$$

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 log-likelihood function for this data structure is

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

where $K(g, G)$ is a term involving only the distribution G of C. KODELL, SHAW, AND JOHNSON (1982) also studied 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 non-increasing, 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 very slowly to the NPMLE of (F_1, F_2) (provided the support of the initial estimator contains the support of the NPMLE). It should be noticed that the two-dimensional nature of their method enables us to avoid the use of Lagrange multipliers.
Another possible way of estimating F_1 is by plugging in the Kaplan-Meier estimator of F_2 and calculating the Nonparametric Maximum Pseudo Likelihood Estimator (NPMPLE) of F_1. The part of the log-likelihood involving F_1 is

$$\sum_{i=1}^{n} \left(1-\Delta_{2,(i)}\right)\left[\Delta_{1,(i)}\log(x_i-\hat{F}_{2,KM}(Y_{(i)}))+\left(1-\Delta_{1,(i)}\right)\log\left(1-x_i\right)\right] \tag{1.1}$$

where $x_i = F_1(Y_{(i)})$, $Y_{(i)}$ is the ith order statistic of $(Y_1, \ldots, Y_n)$, $\Delta_{1,(i)}$, and $\Delta_{2,(i)}$ are the values of $\Delta_{1,i}$ and $\Delta_{2,i}$ observed at $Y_{(i)}$ respectively. Since (1.1) can be written as

$$\sum_{i=1}^{n}\left\{\Phi(f(Y_{(i)}))+\left[g(Y_{(i)})-f(Y_{(i)})\right]\phi(f(Y_{(i)}))\right\}w(Y_{(i)})$$

with $f = F_1$, $\phi = d\Phi/df$, $g = 1-(1-\hat{F}_{2,KM})(1-\Delta_1)$, $w = (1-\Delta_2)/(1-\hat{F}_{2,KM})$ and $\Phi(y) = (y-F_2)\log(y-F_2)+(1-y)\log(1-y)$, DINSE AND LAGAKOS (1982) concluded that the values of $F_1(Y_{(i)}), i = 1, \ldots, n$, maximizing the log-likelihood (1.1) could be obtained applying Theorem 1.10 in BARLOW, BARTHOLOMEW, BREMNER, AND BRUNK (1972), i.e., the NPMPLE of F_1 would be given by the isotonic regression g^* of $g(Y_{(i)})$ with weights $w(Y_{(i)}), i = 1, \ldots, n$. However, Theorem 1.10 in BARLOW, BARTHOLOMEW, BREMNER, AND BRUNK (1972) is applicable to a real convex function Φ defined on $\mathbb{R}$ while in the application above the function Φ is in fact defined on $\mathbb{R}^2$ since the value of F_2 is not supposed to be constant. It should be mentioned here that, although the Kaplan-Meier estimator $\hat{F}_2$ is uniquely defined, except possibly at times exceeding the largest observation, the pseudo NPMLE $\hat{F}_1$ is uniquely defined only over certain data-determined intervals. Specifically, $\hat{F}_1$ is always uniquely defined at the observed C_i's, i.e., the observations for which $\Delta_{2,i} = 0$.
VAN DER LAAN, JEWELL, AND PETERSON (1997) proposed a weighted least squares estimator for F_1 using $F_2 = \hat{F}_{2,KM}$. Simulations studies performed by GOMES (1999) showed evidence that their estimator tend to be more efficient than the NPMLE of F_1 when the distribution functions F_1 and F_2 are far apart. On the other hand, the opposite seems to be true when F_1 and F_2 are close. In their paper, VAN DER LAAN, JEWELL, AND PETERSON (1997) did not establish consistency or the asymptotic distribution of their estimator. This study is a first effort to obtain such properties. Their estimator is described in Section 2 and its consistency is proved in Section 3.

2 The Weighted Least Squares Estimator of F_1

Another possibility for estimation of F_1 is to calculate a weighted least squares estimator as suggested by VAN DER LAAN, JEWELL, AND PETERSON (1997). Making $S_1 = 1 - F_1$ and

$S_2 = 1 - F_2$, in terms of populations, $R(c) = S_1(c)/S_2(c)$ is the proportion of subjects alive at time c who are disease free (i.e., $1 - R(c)$ is the prevalence function at time c). It can be written as

$$\begin{aligned} R(c) &= \frac{S_1(c)}{S_2(c)} = \frac{1 - F_1(c)}{1 - F_2(c)} = \frac{P(T_1 > c)}{P(T_2 > c)} \\ &= \frac{P(T_1 > c, T_2 > c)}{P(T_2 > c)} = P(T_1 > C \mid C = c, T_2 > C) \\ &= \mathbb{E}\left[1_{\{T_1 > C\}} \mid C = c, T_2 > C\right] = \mathbb{E}\left[1 - \Delta_1 \mid C = c, T_2 > C\right]. \end{aligned}$$

So, it is possible to rewrite

$$\begin{aligned} S_1(c) &= R(c)S_2(c) = S_2(c)\mathbb{E}\left[1 - \Delta_1 \mid C = c, T_2 > C\right] \\ &= \mathbb{E}\left[S_2(C)\left(1 - \Delta_1\right) \mid C = c, T_2 > C\right]. \end{aligned}$$

The estimation of S_1 can be viewed, then, as a regression of $S_2(C)\left(1 - \Delta_1\right)$ on the observed C_i's under the constraint of monotonicity. If we substitute S_2 by its Kaplan-Meier estimator $\hat{S}_{2,n} = \hat{S}_{2,KM}$ we automatically have an estimator for S_1 minimizing

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

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, Bartholomew, Bremner, and Brunk (1972)) and its solution is given by

$$\hat{S}_1(Y_{(m)}) = \min_{l \le m}\max_{k \ge m}\frac{\sum_{j=l}^{k}\hat{S}_{2,KM}(Y_{(j)})\left(1 - \Delta_{1(j)}\right)\left(1 - \Delta_{2(j)}\right)}{\sum_{j=l}^{k}\left(1 - \Delta_{2(j)}\right)},$$

$m = 1, \ldots, n$.
However, $\mathrm{Var}\left[S_2(C)\left(1 - \Delta_1\right) \mid C = c, T_2 > C\right]$ is not constant. In fact,

$$\begin{aligned} & \mathrm{Var}\left[S_2(C)\left(1 - \Delta_1\right) \mid C = c, T_2 > C\right] \\ = \; & S_2^2(c)\mathrm{Var}\left[\left(1 - \Delta_1\right) \mid C = c, T_2 > C\right] \\ = \; & S_2^2(c)P(T_1 > C \mid C = c, T_2 > C)\left\{1 - P(T_1 > C \mid C = c, T_2 > C)\right\} \\ = \; & S_2^2(c)\mathbb{E}\left[1 - \Delta_1 \mid C = c, T_2 > C\right]\left\{1 - \mathbb{E}\left[1 - \Delta_1 \mid C = c, T_2 > C\right]\right\} \\ = \; & S_2^2(c)R(c)\left(1 - R(c)\right). \end{aligned}$$

We may, then, use a weighted least squares estimator with weights $w_i, i = 1, \ldots, n$, inversely proportional to the variance $S_2^2(c)R(c)\left[1 - R(c)\right]$. This expression for the variance involves the unknown value $S_1(C_i)$ that we want to estimate, suggesting the use of an iterative procedure. In each step, the estimate would be given by

$$\hat{S}_1(Y_{(m)}) = \min_{l \leq m} \max_{k \geq m} \frac{\sum_{j=l}^{k} \hat{S}_{2,KM}(Y_{(j)}) \left[1 - \Delta_{1(j)}\right] \left(\frac{1-\Delta_{2(j)}}{\hat{S}^2_{2,KM}(Y_{(j)}) R(Y_{(j)}) \left[1 - R(Y_{(j)})\right]} \right)}{\sum_{j=l}^{k} \left(\frac{1-\Delta_{2(j)}}{\hat{S}^2_{2,KM}(Y_{(j)}) R(Y_{(j)}) \left[1 - R(Y_{(j)})\right]} \right)}, \tag{2.2}$$

for $m = 1, \ldots, n$. If we use $w_j = (1 - \Delta_{2(j)})/\hat{S}^2_{2,KM}(Y_{(j)})$ instead, we have an estimator with a closed form, as suggested by van der Laan, Jewell, and Peterson (1997).

The estimators expressed as the solution for an isotonic regression problem have a nice geometric interpretation. Consider the least concave majorant 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

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

$\hat{S}_1(t)$ is the slope of the Least Concave Majorant at W_j if $t \in \left(Y_{(j-1)}, Y_{(j)}\right]$. See Barlow, Bartholomew, Bremner, and Brunk (1972) or Robertson, Wright, and Dykstra (1988) for a more detailed description of these equivalent representations.

So, we can write

$$\hat{S}_1(Y_{(m)}) = \min_{l \leq m} \max_{k \geq m} \frac{\sum_{j=l}^{k} (1 - \Delta_{1(j)})(1 - \Delta_{2(j)})/\hat{S}_{2,KM}(Y_{(j)})}{\sum_{j=l}^{k} (1 - \Delta_{2(j)})/\hat{S}^2_{2,KM}(Y_{(j)})}, \tag{2.3}$$

for $m = 1, \ldots, n$. Figure 1 shows $\hat{F}_1$ and $\hat{F}_{2,KM}$ for the data in Table 1, and Figure 2 shows $\hat{F}_1$ and the real distribution function F_1 for increasing sample sizes of simulated data generated with $T_1 \sim \exp(0.5)$, $T_2 - T_1 \sim \exp(1)$, and $C \sim \exp(0.4)$.

It is easy to see that (2.3) reduces to the expression of the NPMLE of $F \equiv F_1$ for current status data (see Groeneboom and Wellner (1992)) when $\Delta_{2,i} = 0, i = 1, \ldots, n$ (in this case, $\hat{S}_{2,KM} \equiv 1$).

3 Consistency

Theorem 3.1 *Suppose C, T_1 and T_2 have continuous distribution functions G, F_1 and F_2, respectively, satisfying $P_{F_1} \ll P_G$. Then*

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

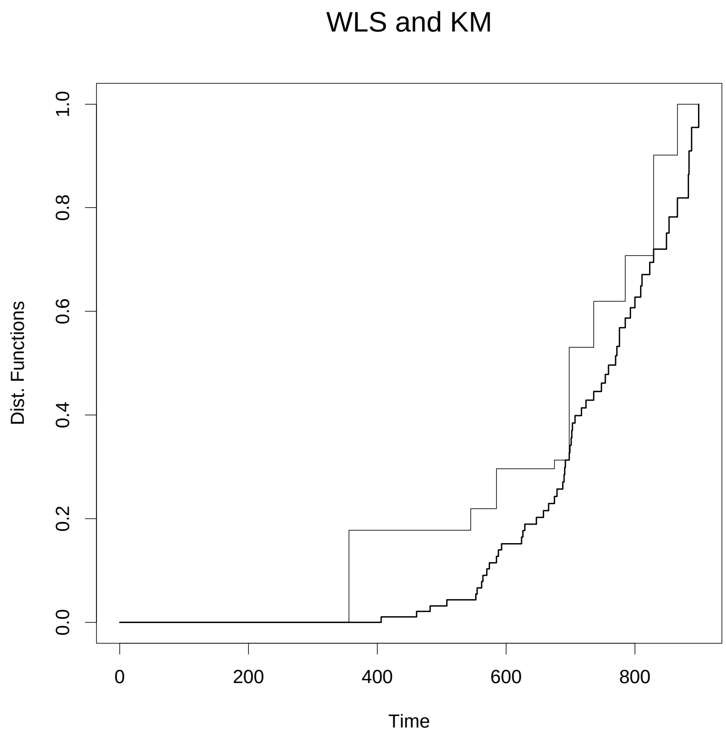

Figure 1: **Weighted Least Square estimate of F_1 and Kaplan-Meier estimate of F_2.**

Proof: 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 ψ under the constraint of monotonicity, where

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

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 $\psi\left(F_1\right)$ implies that for any $0 \le \varepsilon \le 1$,

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

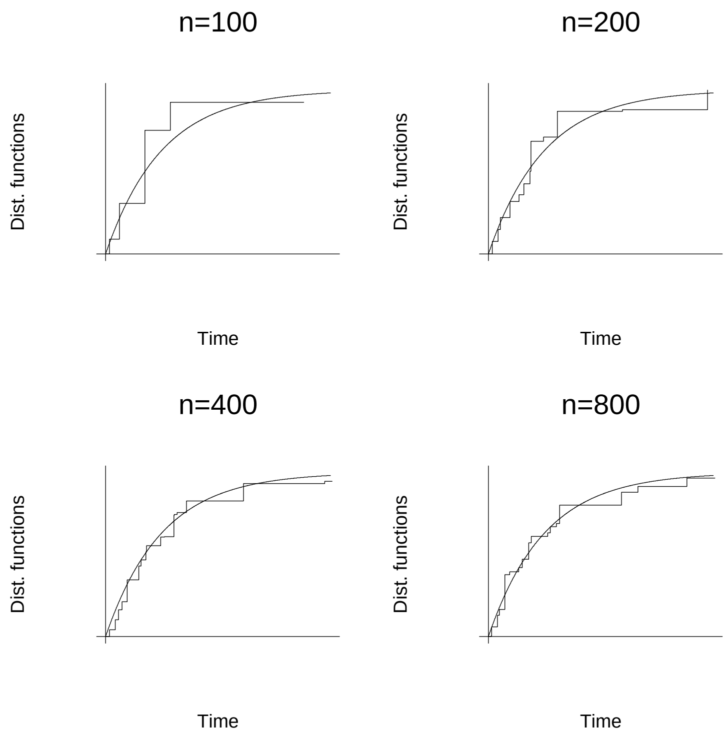

Figure 2: **Weighted Least Squares estimate of F_1 and the real d.f. F_1.**

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

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left\{ \psi\left((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1 \right) - \psi\left(\hat{F}_{1,n} \right) \right\} \geq 0 \quad .$$

But

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

So,

$$\begin{aligned} \frac{d}{d\varepsilon} \psi\left((1-\varepsilon)\hat{F}_{1,n} + \varepsilon F_1 \right) \Big|_{\varepsilon=0} &= \int_{\mathbb{R}^3} 2 \left[1_{\{t_1 > c\}} \left(1 - \hat{F}_{2,n}(c) \right) - \left(1 - \hat{F}_{1,n}(c) \right) \right] 1_{\{t_2 > c\}} \\ &\quad \times \frac{\left[F_1(c) - \hat{F}_{1,n}(c) \right]}{\left[1 - \hat{F}_{2,n}(c) \right]^2} d\mathbb{P}_n\left(t_1, t_2, c \right) \;\geq\; 0 \,. \end{aligned} \tag{3.4}$$

Let Ω be the space of all sequences $\{(T_{1,i}, T_{2,i}, C_i), i = 1, 2, \ldots\}$ endowed with the Borel σ-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 (see Dudley (1989)) for all ω 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 $\sup_{t \in [0,\tau)} \mid \hat{F}_{2,n}(t; \omega) - F_2(t) \mid \longrightarrow 0$ as $n \longrightarrow \infty$, for every $\omega \in B_2$, where $\tau = \sup\{t : H(t) < 1\}$ and $H = 1 - (1 - F_2)(1 - G)$ (see 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 $\hat{F}_{2,n}(\cdot; \omega) \longrightarrow F_2(\cdot)$ for $\omega \in B_2$, we have

$$\hat{F}_{2,n_k}(\cdot; \omega) \longrightarrow F_2(\cdot) \quad \text{for} \quad \omega \in B_2 \quad .$$

Thus, we may assume $1/\left[1 - \hat{F}_{2,n}(t, \omega)\right]^2$ 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 $1/\left[1 - F_2(t)\right]^2$ is bounded for $t \in [0, b]$. Hence we assume

$$\frac{1}{[1 - F_2(b)]^2} \leq M \qquad \text{and} \qquad \frac{1}{\left[1 - \hat{F}_{2,n}(b, \omega)\right]^2} \leq M \tag{3.5}$$

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

By monotone convergence and (3.8) from lemma 3.1 (stated and proved in the appendix) we obtain:

$$\begin{aligned} &\int_{\mathbb{R}^3} 2\left[1_{\{t_1 > c\}}(1 - F_2(c)) - (1 - F_1^*(c))\right] \\ &\qquad \times [F_1(c) - F_1^*(c)] \frac{1_{\{t_2 > c\}}}{[1 - F_2(c)]^2} dP(t_1, t_2, c) \\ &= \lim_{b \to \infty} \int_{\mathbb{R}^2 \times [0,b]} 2\left[1_{\{t_1 > c\}}(1 - F_2(c)) - (1 - F_1^*(c))\right] \\ &\qquad \times [F_1(c) - F_1^*(c)] \frac{1_{\{t_2 > c\}}}{[1 - F_2(c)]^2} dP(t_1, t_2, c) \quad \geq \quad 0 \, . \end{aligned} \tag{3.6}$$

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

$$\int_{\mathbb{R}^3} 2\left[1_{\{t_1 > c\}}(1 - F_2(c)) - (1 - F_1^*(c))\right]$$

$$
\begin{aligned}
& \times \left[F_1(c) - F_1^*(c)\right] \frac{1_{\{t_2>c\}}}{\left[1 - F_2(c)\right]^2} dP\left(t_1, t_2, c\right) \\
= \quad & 2 \int_{\mathbb{R}^3} 1_{\{t_1>c\}} \left(1 - F_2(c)\right) \frac{\left[F_1(c) - F_1^*(c)\right]}{\left[1 - F_2(c)\right]^2} dP\left(t_1, t_2, c\right) \\
& - 2 \int_{\mathbb{R}^3} \left(1 - F_1^*(c)\right) \left[F_1(c) - F_1^*(c)\right] \frac{1_{\{t_2>c\}}}{\left[1 - F_2(c)\right]^2} dP\left(t_1, t_2, c\right) \\
= \quad & 2 \left(\int_{\mathbb{R}} \left[1 - F_1(c)\right] \frac{\left[F_1(c) - F_1^*(c)\right]}{1 - F_2(c)} dG(c) \right. \\
& \left. - \int_{\mathbb{R}} \left[1 - F_1^*(c)\right] \frac{\left[F_1(c) - F_1^*(c)\right]}{\left[1 - F_2(c)\right]} dG(c) \right) \\
= \quad & -2 \int_{\mathbb{R}} \frac{\left[F_1(c) - F_1^*(c)\right]^2}{1 - F_2(c)} dG(c) \ \le \ 0 \ ,
\end{aligned}
$$

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 $F_1^* \neq F_1 \Rightarrow F_1^*(t) \neq F_1(t)$ on an interval of increase of G, which implies

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

if $F_1^* \neq F_1$, which contradicts (3.6).

Thus we have proved that for each ω 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.,

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

or

$$
\| \hat{F}_{1,n} - F_1 \|_\infty \longrightarrow_{a.s.} 0.
$$

References

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

Dinse, G.E. and Lagakos, S.W. (1982). Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. *Biometrics* **38**, 921-932.

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

GOMES, A.E. (1999). Lifetime and Disease Onset Distributions from Incomplete Observations, Ph.D. Thesis, Department of Statistics, University of Washington, Seattle, USA.

GROENEBOOM, P. AND WELLNER, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser Verlag.

GROENEBOOM, P. (1998). *Special Topics Course 593C: Nonparametric Estimation for Inverse Problems; algorithms and asymptotics*, Technical Report no. 344, Department of Statistics, University of Washington, Seattle, USA.

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.

KAPLAN, E.L. AND MEIER, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**, 457-481.

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.* John Wiley and Sons, 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.

Appendix

Lemma 3.1 *Let b be chosen such that $F_2(b) = 1 - \varepsilon$. Then*

$$\lim_{k\to\infty} \int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c)\right) - \left(1-\hat{F}_{1,n_k}(c)\right)\right]$$

$$\times \left[F_1(c) - \hat{F}_{1,n_k}(c)\right] \frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c)\right]^2} d\mathbb{P}_{n_k}(t_1,t_2,c)$$

$$= \int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}(1-F_2(c)) - (1-F_1^*(c))\right] \times \left[F_1(c) - F_1^*(c)\right] \frac{1_{\{t_2>c\}}}{[1-F_2(c)]^2} dP(t_1,t_2,c) \tag{3.7}$$

Moreover,

$$\int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}(1-F_2(c)) - (1-F_1^*(c))\right] \times \left[F_1(c) - F_1^*(c)\right] \frac{1_{\{t_2>c\}}}{[1-F_2(c)]^2} dP(t_1,t_2,c) \geq 0. \tag{3.8}$$

Proof: Fix $0 < \delta < 1$ and take a grid of points $0 = u_0 < u_1 < \ldots < u_m = b$ on $[0,b]$ such that $m = 1 + [1/\delta^2]$ and

$$G(u_i) - G(u_{i-1}) = \frac{G(b)}{m}, \quad i = 1,\ldots,m.$$

Let K be the set of indices $i, i = 1,\ldots,m$ such that

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

The first inequality in (3.5) implies that the number of indices of this type is not bigger than $1 + [M/\delta]$. Let L be the remaining set of indices $i, i = 1,\ldots,m$.

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

$$\int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right) - \left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \times \left[F_1(c) - \hat{F}_{1,n_k}(c;\omega)\right] \frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2} d\mathbb{P}_{n_k}(t_1,t_2,c;\omega)$$

$$= \sum_{i=1}^m \int_{\mathbb{R}^2\times J_i} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right) - \left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \times \left[F_1(c) - \hat{F}_{1,n_k}(c;\omega)\right] \frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2} d\mathbb{P}_{n_k}(t_1,t_2,c;\omega).$$

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,

$$\frac{1}{\left[1-\hat{F}_{2,n_k}(u_i;\omega)\right]^2} - \frac{1}{\left[1-\hat{F}_{2,n_k}(u_{i-1};\omega)\right]^2} < 2\delta,\ i \in L. \tag{3.9}$$

Hence,

$$\begin{aligned}
&\int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \\
&\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&= \sum_{i\in K}\int_{\mathbb{R}^2\times J_i} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \\
&\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&\quad+\sum_{i\in L}\int_{\mathbb{R}^2\times J_i} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \\
&\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}d\mathbb{P}_{n_k}(t_1,t_2,c;\omega) \\
&= \int_{\mathbb{R}^2\times[0,b]} 2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right] \qquad (3.10)\\
&\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}dP(t_1,t_2,c)\ +\ r_k'(\omega)\ +\ o_p(1),
\end{aligned}$$

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

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

On the intervals J_i with $i \in K$ we use the second inequality in (3.5).

Note that $\sum_{i\in K} P(\mathbb{R}^2 \times J_i) \longrightarrow 0$, 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

$$
\begin{aligned}
&\lim_{k\to\infty}\int_{\mathbb{R}^2\times[0,b]}\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right]\\
&\qquad\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}dP\left(t_1,t_2,c\right)\\
=&\int_{\mathbb{R}^2\times[0,b]}\left[1_{\{t_1>c\}}\left(1-F_2(c)\right)-\left(1-F_1^*(c)\right)\right] \qquad (3.11)\\
&\qquad\qquad\times\left[F_1(c)-F_1^*(c)\right]\frac{1_{\{t_2>c\}}}{\left[1-F_2(c)\right]^2}dP\left(t_1,t_2,c\right)\quad .
\end{aligned}
$$

Combining (3.10) and (3.11) we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^2\times[0,b]}2\left[1_{\{t_1>c\}}\left(1-\hat{F}_{2,n_k}(c;\omega)\right)-\left(1-\hat{F}_{1,n_k}(c;\omega)\right)\right]\\
&\qquad\qquad\times\left[F_1(c)-\hat{F}_{1,n_k}(c;\omega)\right]\frac{1_{\{t_2>c\}}}{\left[1-\hat{F}_{2,n_k}(c;\omega)\right]^2}d\mathbb{P}_{n_k}\left(t_1,t_2,c;\omega\right)\\
=&\int_{\mathbb{R}^2\times[0,b]}2\left[1_{\{t_1>c\}}\left(1-F_2(c)\right)-\left(1-F_1^*(c)\right)\right] \qquad (3.12)\\
&\qquad\qquad\times\left[F_1(c)-F_1^*(c)\right]\frac{1_{\{t_2>c\}}}{\left[1-F_2(c)\right]^2}dP\left(t_1,t_2,c\right)\ +\ r'_k(\omega)\ +\ o_p(1),
\end{aligned}
$$

where $|r'_k(\omega)|\le c'\delta$.

Since δ can be made arbitrarily small, (3.7) now follows, and relation (3.8) follows from (3.7) and (3.4). □