Nonparametric Estimation of the Lifetime and Disease Onset Distributions for a Survival - Sacrifice Model

Antonio Eduardo Gomes, Piet Groeneboom, and Jon A. Wellner

Universidade Federal de Minas Gerais, Delft University of Technology, and University of Washington

August 27, 2001

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. 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 $Y = \min\{T_2, C\}$, $\Delta_1 = 1_{\{T_1 \le C\}}$, and $\Delta_2 = 1_{\{T_2 \le C\}}$. The random variables T_1 and T_2 are independent of C and have a joint distribution such that $P(T_1 \le T_2) = 1$. Some authors call this model a "survival-sacrifice model". van der Laan, Jewell, and Peterson (1997) (to be denoted by LJP (1997)) proposed a Weighted Least Squares estimator for F_1 (the marginal distribution function of T_1), using the Kaplan-Meier estimator of F_2 (the marginal distribution function of T_2). The authors claimed that their estimator is more efficient than the MLE (maximum likelihood estimator) of F_1 and that the Kaplan-Meier estimator is more efficient than the MLE of F_2. However, we show that they did not compute the MLE of F_1 correctly, and that their (claimed) MLE estimate of F_1 is even undefined in the case of active constraints. In our simulation study we used a primal-dual interior point algorithm to obtain the true MLE of F_1. The results showed a better performance of the MLE of F_1 over the weighted least squares estimator in LJP (1997) for points where F_1 is close to F_2. Moreover, application to the model, used in the simulation study of LJP (1997), showed smaller variances of the MLE estimators of the first and second moments for both F_1 and F_2, and sample sizes from 100 up to 5000, in comparison to the estimates, based on the weighted least squares estimator for F_1, proposed in LJP (1997), and the Kaplan-Meier estimator for F_2.

Outline

1 Introduction

Suppose that (T_1, T_2) is a pair of nonnegative random variables with joint distribution F concentrated on $\{(t_1, t_2) : 0 \le t_1 \le t_2 < \infty\}$. Here we think of T_1 as the "time of disease onset", and T_2 as the "time of death from the disease", and let F_1 and F_2 denote their respective (marginal) distribution functions. Suppose that C is a nonnegative random variable with distribution function G which is independent of (T_1, T_2). We think of C as the "time of death from an unrelated cause". Furthermore, we can only observe the triple

$$X \equiv (C \wedge T_2, 1_{[T_1 \le C]}, 1_{[T_2 \le C]}) \equiv (Y, \Delta_1, \Delta_2)\,.$$

If G has density g with respect to Lebesgue measure, and the marginal distribution function F_2 of T_2 has density f_2 with respect to Lebesgue measure, then it is easily seen that the joint density of X with respect to the product of Lebesgue measure on $\mathbb{R}^+$ and counting measure on $D \equiv \{(0,0), (1,0), (1,1)\}$ is given by

$$p(y, \delta_1, \delta_2) = \begin{cases} (1 - F_1(y))g(y), & \text{if } (\delta_1, \delta_2) = (0,0)\,, \\ (F_1(y) - F_2(y))g(y), & \text{if } (\delta_1, \delta_2) = (1,0)\,, \\ (1 - G(y))f_2(y), & \text{if } (\delta_1, \delta_2) = (1,1)\,. \end{cases} \tag{1.1}$$

Let $P = P_{F,G}$ denote the corresponding probability measure on $\mathbb{R}^+ \times D$. Note that the marginal density of (Y, Δ_2) is given by

$$p_2(y, \delta_2) = \begin{cases} (1 - F_2(y))g(y), & \text{if } \delta_2 = 0\,, \\ (1 - G(y))f_2(y), & \text{if } \delta_2 = 1\,, \end{cases}$$

which is exactly that of random right censoring of $T_2 \sim F_2$ by $C \sim G$. On the other hand, the marginal density of (Y, Δ_1) is

$$p_1(y, \delta_1) = \begin{cases} (1 - F_1(y))g(y)\,, & \text{if } \delta_1 = 0\,, \\ (F_1(y) - F_2(y))g(y) + (1 - G(y))f_2(y)\,, & \text{if } \delta_1 = 1\,, \end{cases}$$

which is *not* the same as "current status data" for $T_1 \sim F_1$ with observation time $C \sim G$ since the $\delta_1 = 1$ component of this density only reduces to $F_1(y)g(y)$ if F_2 puts all its mass at $+\infty$ (corresponding to a non-lethal disease). While the resulting "survival-sacrifice model" is very much related to right-censored data via its marginal distribution P_2, and to current status data via its marginal distribution P_1, the model as a whole is more complicated than either of these simpler models, especially so because of the restriction $F_1 \le_s F_2$ which results from $T_1 \le T_2$ a.s. F.

Our goal is to construct nonparametric estimators of F_1 and F_2 based on observation of $X_1, \ldots, X_n$ i.i.d. as $X \equiv (C \wedge T_2, \Delta_1, \Delta_2)$.

This model has been proposed for experiments involving the study of onset and mortality from undetectable irreversible diseases (e.g. occult tumors). The model is reasonable when the disease is moderately lethal but incurable and when the cause of death is known. It has a long history in the biometrics literature: see e.g. Dinse and Lagakos (1982); Kodell, Shaw, and Johnson (1982); Turnbull and Mitchell (1984); and, more recently, van der Laan, Jewell, and Peterson (1997).

The parameter space 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 Maximum Likehood Estimation method for this problem is based on maximization of the log-likelihood function

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

where $f_2(x) \equiv F_2(x) - F_2(x-)$ and $K(g, G)$ is a term involving only the distribution G of C.

Kodell, Shaw, and Johnson (1982) 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 allows them to avoid the use of Lagrange multipliers.

Another possible way of estimating F_1 is plugging in the Kaplan-Meier estimator of F_2 and calculating the pseudo NPMLE of F_1. The part of the log-likelihood involving F_1 is

$$\sum_{i=1}^{n} \left(1-\Delta_{2,(i)}\right) \left[\Delta_{1,(i)} \log(x_i - \hat{F}_{2,KM}(Y_{(i)})) + \left(1-\Delta_{1,(i)}\right) \log(1-x_i)\right] \tag{1.2}$$

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 (1.2) 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 (1.2) could be obtained applying theorem 1.10 in Barlow, Bartholomew, Bremner, and Brunk (1972), i.e., the pseudo NPMLE 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 $\hat{F}_{2,KM}$ is uniquely defined, except possibly at times exceeding the largest observation, the pseudo NPMLE $\hat{F}_{1,n}$ is uniquely defined only over certain data-determined intervals. Specifically, $\hat{F}_{1,n}$ is always uniquely defined at the observed C_i's, i.e., the observations for which $\Delta_{2,i} = 0$.

2 A Weighted Least Squares Estimator

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 in the population alive at time c who are disease free (i.e., $1 - R(c)$ is the prevalence function at time c), and it can be written as

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

So, it is possible to rewrite

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

Estimating 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 is given by

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

However,

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

is not constant. 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

$$\hat{S}_1(Y_{(m)}) = \min_{l\le m}\max_{k\ge m} \frac{\sum_{j=l}^{k} \hat{S}_{2,KM}(Y_{(j)})\left[1-\Delta_{1(j)}\right]\left(\frac{1-\Delta_{2(j)}}{\hat{S}^2_{2,KM}(Y_{(j)})R(Y_{(j)})\left[1-R(Y_{(j)})\right]}\right)}{\sum_{j=l}^{k}\left(\frac{1-\Delta_{2(j)}}{\hat{S}^2_{2,KM}(Y_{(j)})R(Y_{(j)})\left[1-R(Y_{(j)})\right]}\right)} \tag{2.3}$$

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 that can be calculated as the left derivative of the least concave majorant of the cumulative sum diagram $(0,0),(W_1,G_1),\ldots,(W_n,G_n)$, where $W_i = \sum_{j=1}^{i} w_j$ and

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

$\hat{S}_1(t)$ is the slope of the Least Concave Majorant at W_i for $t \in \left(Y_{(i-1)}, Y_{(i)}\right]$. van der Laan, Jewell, and Peterson (1997) claim that the estimator above is more efficient than the NPMLE of F_1 which would be calculated, according to them, as follows. Let S^0_{1n} and S^0_{2n} be initial estimators of S_1 and S_2. Let $k = 0$. For a given estimator S^k_{1n}, use the EM-algorithm to compute the estimator S^{k+1}_{2n} solving the equation

$$F_{2n}(t) = \frac{\sum_{i=1}^{n} \frac{\int_{C_i}^{t} dF_{2n}}{(S_{2n}-S_{1n})(C_i)} 1_{\{\delta_i=(1,0)\}} + 1_{T_{2,i}\le t}1_{\{\delta_i=(1,1)\}}}{\sum_{i=1}^{n} 1_{\{\delta_i=(1,1)\}} + \frac{S_{2n}(C_i)}{(S_{2n}-S_{1n})(C_i)} 1_{\{\delta_i=(1,0)\}}}, \tag{2.4}$$

adjusting the estimate so that $S^{k+1}_{2n} \ge S^k_{1n}$ in case this restriction is violated. For a given S^{k+1}_{2n}, use the weighted least squares estimator proposed by van der Laan, Jewell, and Peterson (1997) to obtain a new estimator S^{k+1}_{1n}, which is adjusted so that $S^{k+1}_{1n} \le S^{k+1}_{2n}$ if this restriction is violated. By repeating this joint algorithm, the authors claim that the actual NPMLE of S_1 and S_2 is obtained. One would indeed expect this to be true in situations like this, but the weighted least squares estimator that they introduced and claim to represent the NPMLE is in fact *undefined* at crucial points. The trouble with their approach is that they completely neglect the Lagrangian terms in deriving their "score equations" on page 544 of their paper.

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{1n}(C_i)}-\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)-S_{1n}(C_i)}\right\}=0, \tag{2.5}$$

where $(t_j,t_{j+1}]$ is an interval on which F_{1n} is constant and where t_j and t_{j+1} are points of jump of F_{1n}. From this they derive the following formula for the survival function $S_{1n}=1-F_{1n}$:

$$S_{1n}(t)=\frac{1}{n}\sum_{i=1}^{n}\frac{\{\Delta_{1,i}=0,\ \Delta_{2,i}=0,\ C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)R_n(C_i)(1-R_n(C_i))}\Bigg/\sum_{i=1}^{n}\frac{\{\Delta_{2,i}=0,\ C_i\in(t_j,t_{j+1}]\}}{S_{2n}^2(C_i)R_n(C_i)(1-R_n(C_i))}\,, \tag{2.6}$$

for $t\in(t_j,t_{j+1}]$ (in fact they say that $S_{1n}(t_j)=S_{1n}(C_i)$, for $C_i\in(t_j,t_{j+1}]$, but the point t_j does not belong to the interval $(t_j,t_{j+1}]$, so there is also some definition problem here). However, at points t where the constraint $F_{1n}(t)\ge F_{2n}(t)$ is active (these are precisely the points where we will get a Lagrange multiplier $\lambda_i>0$ in the primal-dual interior point algorithm, described in section 3), the expression above is not defined, since we get zeros in the denominators!

As an example, in one of our simulations studies we found the following values for the NPMLE of the pair (F_1,F_2) between points $t_j=1.141807$ and $t_{j+1}=1.567906$:

$$F_{1n}(t)=0.632059,\ t\in(t_j,t_{j+1}]\,,$$

and for part F_{2n} of the NPMLE we found the successive values

$$0.526933,\ 0.561975,\ 0.597017,\ \text{and } 0.632059$$

on the same interval. So F_{2n} becomes equal to F_{1n} at the upper part of this interval. Exactly at the point where this happens, the constraint $F_{1n}\ge F_{2n}$ becomes active, and this means that the equation (2.5) is *not* satisfied. In fact, the sum on the left-hand side of (2.5) was precisely the value of the Lagrange multiplier divided by n, missing in (2.5)! And, as noted above, the expression on the right-hand side of (2.6) is undefined in this situation, because of the fact that the denominators become zero if the constraint is active.

The conclusion is that the NPMLE is *not* the reweighted least squares estimator as introduced by van der Laan, Jewell, and Peterson (1997), because that estimator is not well-defined, their (not iteratively defined) weighted least squares estimator, claimed to be superior to the NPMLE, actually often *coincides* with the NPMLE. One might perhaps hope that maximizing the likelihood, ignoring the constraint $F_1\ge F_2$ (if it is not explicitly forced by the likelihood), and "resetting" values $F_1(t)$ and $F_2(t)$ if a violation $F_1(t)<F_2(t)$ is encountered, will produce the maximum likelihood estimator in the end. But it is well-known that this method will in fact not work and lead to a procedure that may "hang" somewhere far from the maximizing value, since an algorithm of this type tries to move the estimators to values they cannot move to and resets the estimators to the same (non-maximizing) values

each time the active constraints would be violated. In fact, it is quite easy to give numerical examples of this behavior for such a method in the present model, where we get a stationary point that will not correspond to the NPMLE, even if we start the procedure with positive masses at each point, as recommended in van der Laan, Jewell, and Peterson (1997). Use of Lagrange multipliers or a similar technique is the only way out here.

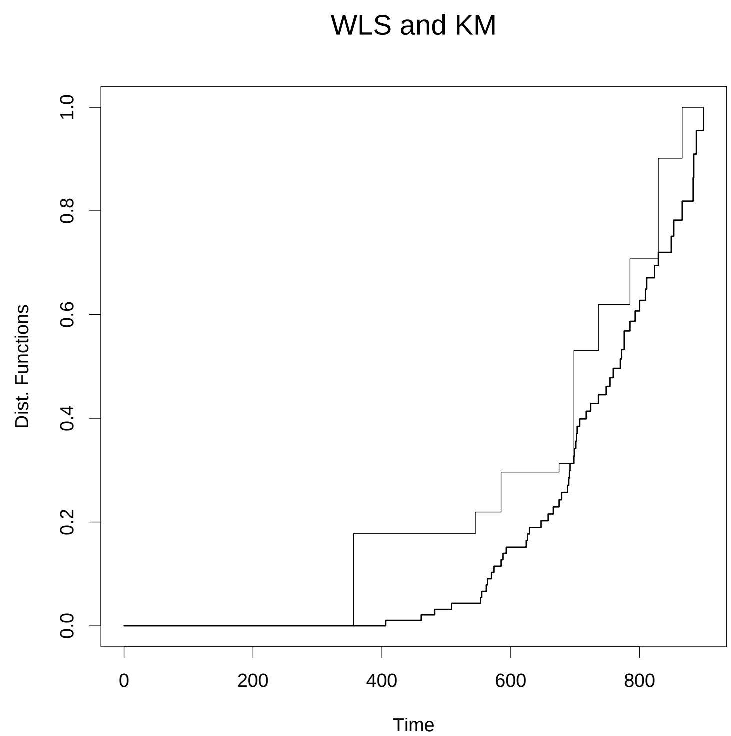

Figure 1: **Weighted Least Square estimate of F_1 and Kaplan-Meier estimate of F_2.**

We close this section by showing the Kaplan-Meier estimator and the (non-iterative) weighted least squares estimator introduced in van der Laan, Jewell, and Peterson (1997) for a real data set studied by Dinse and Lagakos (1982) and Turnbull and Mitchell (1984) representing the ages at death (in days) of 109 female RFM mice (table 1). 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. The smoother picture for the estimate of F_2 in figure 1 is a consequence of the fact that the estimators of F_1 have a $n^{-1/3}$ rate of convergence.
The value of the log-likelihood for these data set at the estimates obtained through the algorithm proposed by van der Laan, Jewell, and Peterson (1997) above is smaller than that obtained at the true NPMLE of F_1 and F_2 obtained through the Primal-Dual Interior Point algorithm (-262.7964 and -262.5468, respectively).

3 Primal-Dual Interior Point Algorithm

Our problem here is to maximize the log likelihood function

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{y}) &= \sum_{i=1}^{n} \{(1-\delta_{1,(i)})(1-\delta_{2,(i)})\log(1-x_i) \\ &\quad + \delta_{1,(i)}(1-\delta_{2,(i)})\log(x_i-y_i) \\ &\quad + \delta_{1,(i)}\delta_{2,(i)}\log(y_i-y_{i-1})\}\,, \end{aligned} \tag{3.7}$$

where $\delta_{1,(i)}$ and $\delta_{2,(i)}$ correspond to the ith (realized) order statistic $u_{(i)}$, $u_i = c_i \wedge t_{2,i}$, and where the vectors $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\mathbf{y} = (y_1, \ldots, y_n)'$ have to satisfy the restrictions $0 \le x_1 \le \ldots \le x_n \le 1$, $0 \le y_1 \le \ldots \le y_n \le 1$, and $x_i \ge y_i$, for $i = 1, \ldots, n$, i.e., vector $\mathbf{x}$ contains the values of $F_1(u_{(i)})$ and vector $\mathbf{y}$ has the values of $F_2(u_{(i)})$ as their components. This corresponds to the "full" maximum likelihood estimators of the lifetime and disease onset distributions for the model, considered in Dinse and Lagakos (1982) and LJP (1997).

Let $\mathbf{z} = (x_1, y_1, \ldots, x_n, y_n)'$ and let $\phi(\mathbf{z})$ be defined by

$$\begin{aligned}\phi(\mathbf{z}) &= -\sum_{i=1}^{n} \{\left(1-\delta_{1,(i)}\right)\left(1-\delta_{2,(i)}\right)\log\left(1-x_i\right) \\ &\quad + \delta_{1,(i)}\left(1-\delta_{2,(i)}\right)\log\left(x_i-y_i\right) \\ &\quad + \delta_{1,(i)}\delta_{2,(i)}\log\left(y_i-y_{i-1}\right)\}\;. \end{aligned}$$

Furthermore, we define the vector $g(\mathbf{z}) = (g_1(\mathbf{z}), \ldots, g_{3n}(\mathbf{z}))'$ by

$$\begin{aligned} g_1(\mathbf{z}) &= -z_1, \\ g_i(\mathbf{z}) &= z_{2i-3} - z_{2i-1},\ i = 2, \ldots, n, \\ g_{n+i}(\mathbf{z}) &= z_{2i} - z_{2i+2},\ i = 1, \ldots, n-1 \\ g_{2n}(\mathbf{z}) &= z_{2n} - 1, \\ g_{2n+i}(\mathbf{z}) &= z_{2i-1} - z_{2i},\ i = 1, \ldots, n\,, \end{aligned}$$

and the matrix G by

$$G = \left(\frac{\partial g_i(\mathbf{z})}{\partial z_j}\right)_{i=1,\ldots,3n;\, j=1,\ldots,2n}.$$

Note that the matrix G does not depend on $\mathbf{z}$. The original maximization problem now becomes a problem of minimizing $\phi(\mathbf{z})$ over $\mathbb{R}^{2n}$ (adopting the convention of making $\phi(\mathbf{z}) = \infty$ if we encounter the logarithm of an argument less or equal to 0), under the restriction that all components of the vector $g(\mathbf{z})$ are less or equal to 0. The latter restriction will be denoted by

$$g(\mathbf{z}) \;\le\; 0,\ \mathbf{z} \in \mathbb{R}^{2n}\,. \tag{3.8}$$

Theorem 1: Let $\hat{\mathbf{z}} = (\hat{x}_1, \hat{y}_1, \ldots, \hat{x}_n, \hat{y}_n)'$ be a vector in $\mathbb{R}^{2n}$ such that $\phi(\hat{\mathbf{z}}) < \infty$. Then $\hat{\mathbf{z}}$ minimizes $\phi(\hat{\mathbf{z}})$ over the set of vectors $\mathbf{z}$, satisfying (3.8), if and only if the following conditions are satisfied:

$$\nabla\phi(\hat{\mathbf{z}}) + G'\lambda = 0 \tag{3.9}$$

$$g(\hat{\mathbf{z}}) + \mathbf{w} = 0 \tag{3.10}$$

$$\langle \lambda, \mathbf{w} \rangle = 0, \tag{3.11}$$

for vectors λ and $\mathbf{w}$ in $\mathbb{R}^{3n}_+$.

Remark: The vector λ is the vector of Lagrange multipliers and $\mathbf{w}$ is called a vector of "slack variables" for the constraints; see, e.g., Wright (1997), page 164. Defining the function ϕ_λ by

$$\phi_\lambda(\mathbf{z}) \;=\; \phi(\mathbf{z}) + \langle \lambda, g(\mathbf{z}) \rangle,$$

we can write condition (3.9) in the form

$$\nabla\phi_\lambda(\hat{\mathbf{z}}) = 0.$$

Note that $g(\hat{\mathbf{z}}) + \mathbf{w} = 0$, for $\mathbf{w} \in \mathbb{R}^{3n}_+$, implies $g(\hat{\mathbf{z}}) \le 0$, and that $\lambda_i > 0$ implies $w_i = 0$, by (3.11), and hence $g_i(\hat{\mathbf{z}}) = 0$, by (3.10). Thus:

$$\langle \lambda, g(\hat{\mathbf{z}}) \rangle = 0.$$

The proof of Theorem 1 can be found in Luenberger (1969), pp. 249-250.

The primal-dual interior point method for finding the solution to this minimization problem is now formulated (the method is called "primal-dual" because we solve the primal problem for the vector $\mathbf{z}$ and simultaneously the dual problem for the vectors λ and $\mathbf{w}$).

A peculiar difficulty is that not all variables appear in the object function that we want to maximize (the log likelihood). For example, if $\delta_{1,(i)} = \delta_{2,(i)} = 0$, then only x_i figures in the (3.7) and not y_i or y_{i-1} (y_{i-1} could appear in a preceding term, though). For this reason the log likelihood will never have $2n$ arguments, unless only terms $\log(x_i - y_i)$ occur. Nevertheless, it is advantageous to work with the "overparametrized" set of $2n$ variables, since (after the inclusion of the constraints) this produces a Hessian which is a "band matrix" that can easily be inverted (see below). This structure is lost if we first perform a preliminary reduction to the variables that really appear in the likelihood.

Note that the constraints are not linear but affine; we can write them in the form

$$g(\mathbf{z}) \le 0,$$

or

$$G\mathbf{z} \le \mathbf{a},$$

where $\mathbf{a}$ is a vector having zero components, except for the $2n$-th component, which is 1.

We now start the computation of the NPMLE with a vector $\mathbf{z}_0 \in \mathbb{R}^{2n}$, strictly satisfying all constraints, i.e., $g(\mathbf{z}_0) < 0$. An easy choice is:

$$z_{2i} = i/(n+1),\ z_{2i-1} = 0.9i/(n+1),\ i = 1, \ldots, n.$$

For λ and $\mathbf{w}$ we take as starting values $\lambda_0 = \mathbf{w}_0 = 0.5e$, where e denotes the vector in $\mathbb{R}^m, m = 3n$, with all components equal to 1. For a vector $\mathbf{a}$ we denote the diagonal matrix with component a_i as its ith diagonal element by A; for example, Λ is the diagonal matrix with element λ_i as its ith diagonal element. The first (Newton) iteration step now solves the system of equations:

$$\begin{pmatrix} \nabla_{\mathbf{z}_0\mathbf{z}_0}\phi(\mathbf{z}_0) & G' & 0 \\ G & 0 & I \\ 0 & W_0 & \Lambda_0 \end{pmatrix} \begin{pmatrix} \mathbf{z} - \mathbf{z}_0 \\ \lambda - \lambda_0 \\ \mathbf{w} - \mathbf{w}_0 \end{pmatrix} = - \begin{pmatrix} \nabla\phi_{\lambda_0}(\mathbf{z}_0) \\ g(\mathbf{z}_0) + \mathbf{w}_0 \\ (\Lambda_0 W_0 - \sigma\mu_0)e \end{pmatrix} \tag{3.12}$$

in $(\mathbf{z}, \mathbf{w}, \lambda)$, where μ_0 and $\sigma \in (0,1)$ are tuning parameters, which we take $\mu_0 = \sigma = 0.5$. The notation $\nabla_{\mathbf{z}\mathbf{z}}\phi(\mathbf{z})$ will be used to denote the matrix of second derivatives of $\phi_\lambda(\mathbf{z})$ w.r.t. $\mathbf{z}$, the so-called *Hessian* of the function ϕ_λ with respect to $\mathbf{z}$. We now define, for fixed $\beta > 0$, the set

$$\begin{aligned} \mathcal{N}(\mu) \;\; = \;\; & \{(\mathbf{z}, \lambda, \mathbf{w}) : \| \nabla_{\mathbf{z}}\phi(\mathbf{z}) \| \le \beta\mu, \| g(\mathbf{z}) + \mathbf{w} \| \le \beta\mu, \\ & \lambda \ge 0, \mathbf{w} \ge 0, \lambda_i w_i \ge \mu, 1 \le i \le m\} \end{aligned}$$

where $\| \cdot \|$ denotes the Euclidean norm, and we require the first iterate to be in this set. The parameter μ is called the *duality measure*, and defined by

$$\mu = \frac{1}{m}\langle \lambda, \mathbf{w} \rangle.$$

By taking $\lambda = \mathbf{w} = 0.5e$, we have made this parameter equal to 0.5 at the start of the iterations.

We now take a final parameter $\gamma \in (0,1)$, and take α_1 as the first number in the sequence

$$1, \gamma, \gamma^2, \gamma^3, \ldots,$$

such that

$$(\mathbf{z}(\alpha), \lambda(\alpha), \mathbf{w}(\alpha)) \stackrel{\text{def}}{=} (\mathbf{z}_0, \lambda_0, \mathbf{w}_0) + \alpha(\mathbf{z} - \mathbf{z}_0, \lambda - \lambda_0, \mathbf{w} - \mathbf{w}_0) \in \mathcal{N}(\mu_0),$$

where $(\mathbf{z}, \lambda, \mathbf{w})$ solves the system of equations above, and such that

$$\mu(\alpha) \stackrel{\text{def}}{=} \frac{1}{m}\langle \lambda(\alpha), \mathbf{w}(\alpha) \rangle \le (1 - 0.01\alpha)\mu_0.$$

We then take $(\mathbf{z}_1, \lambda_1, \mathbf{w}_1) = (\mathbf{z}(\alpha), \lambda(\alpha), \mathbf{w}(\alpha))$ and $\mu_1 = \mu(\alpha)$, and repeat the procedure for the new values of the parameters, i.e., we solve the system

$$\begin{pmatrix} \nabla_{\mathbf{z}_1\mathbf{z}_1}\phi(\mathbf{z}_1) & G' & 0 \\ G & 0 & I \\ 0 & W_1 & \Lambda_1 \end{pmatrix} \begin{pmatrix} \mathbf{z} - \mathbf{z}_1 \\ \lambda - \lambda_1 \\ \mathbf{w} - \mathbf{w}_1 \end{pmatrix} = - \begin{pmatrix} \nabla\phi_{\lambda_1}(\mathbf{z}_1) \\ g(\mathbf{z}_1) + \mathbf{w}_1 \\ (\Lambda_1 W_1 - \sigma\mu_1)e \end{pmatrix} \tag{3.13}$$

and find the new $(\mathbf{z}(\alpha), \lambda(\alpha), \mathbf{w}(\alpha))$, required to lie in $\mathcal{N}(\mu_1)$, for this system, denoted by $(\mathbf{z}_2, \lambda_2, \mathbf{w}_2)$, and the new μ_2. This is repeated until the duality measure μ_k is below a certain criterion, say 10^{-10} or 10^{-15}.

Generally we start the kth iteration step with the system

$$\begin{pmatrix} \nabla_{\mathbf{zz}}\phi(\mathbf{z}) & G' & 0 \\ G & 0 & I \\ 0 & W & \Lambda \end{pmatrix} \begin{pmatrix} \Delta\mathbf{z} \\ \Delta\lambda \\ \Delta\mathbf{w} \end{pmatrix} = - \begin{pmatrix} \nabla\phi_{\lambda}(\mathbf{z}) \\ g(\mathbf{z}) + \mathbf{w} \\ (\Lambda W - \sigma\mu)e \end{pmatrix} \tag{3.14}$$

where the vector $(\Delta\mathbf{z}, \Delta\lambda, \Delta\mathbf{w})$, denotes the vector

$$(\mathbf{z} - \mathbf{z}_k, \lambda - \lambda_k, \mathbf{w} - \mathbf{w}_k),$$

if $(\mathbf{z}_k, \lambda_k, \mathbf{w}_k)$ is the value at the start of the kth iteration, and we solve for $\Delta\mathbf{z}$, $\Delta\lambda$ and $\Delta\mathbf{w}$. We now transform this system into a system that is better suited for numerical computation. We first solve for $\Delta\mathbf{w}$. This yields:

$$\Delta\mathbf{w} = -\Lambda^{-1}(\Lambda W e - \sigma\mu e + W\Delta\lambda).$$

Let D denote the diagonal matrix $W^{-1}\Lambda$. Then we can write the remaining part of the system in the form

$$\begin{pmatrix} \nabla_{\mathbf{zz}}\phi(\mathbf{z}) & G' \\ G & -D^{-1} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{z} \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} \nabla\phi_{\lambda}(\mathbf{z}) \\ g(\mathbf{z}) + \sigma\mu\Lambda^{-1}e \end{pmatrix}. \tag{3.15}$$

We then solve the system above for $\Delta\lambda$. This gives:

$$\Delta\lambda = W^{-1}\left\{\Lambda G\Delta\mathbf{z} + \Lambda g(\mathbf{z}) + \sigma\mu e\right\}.$$

Using this result to solve for $\Delta\mathbf{z}$, we obtain the system

$$\begin{aligned} \left(\nabla_{\mathbf{zz}}\phi(\mathbf{z}) + G'DG\right)\Delta\mathbf{z} &= -\nabla\phi_{\lambda}(\mathbf{z}) - G'Dg(\mathbf{z}) - \sigma\mu G'W^{-1}e \\ \Delta\lambda &= W^{-1}\left\{\Lambda G\Delta\mathbf{z} + \Lambda g(\mathbf{z}) + \sigma\mu e\right\}, \\ \Delta\mathbf{w} &= -\Lambda^{-1}\left(\Lambda W e - \sigma\mu e + W\Delta\lambda\right) \end{aligned}$$

which we first solve for $\Delta\mathbf{z}$, next for $\Delta\lambda$, and finally for $\Delta\mathbf{w}$. The only matrix for which inversion is not trivial is the matrix

$$\nabla_{\mathbf{zz}}\phi(\mathbf{z}) + G'DG \ , \tag{3.16}$$

and this matrix is a symmetric positive definite matrix at each step. The matrices W and Λ are diagonal matrices, so inversion of these is trivial. In our case, the matrix (3.16) is a "sparse" band matrix (by the particular parametrization we chose!). This fact can be used for fast and efficient inversion methods, where we only have to reserve computer memory for the elements that can be non-zero. Figure 2 illustrates the application of the primal-dual

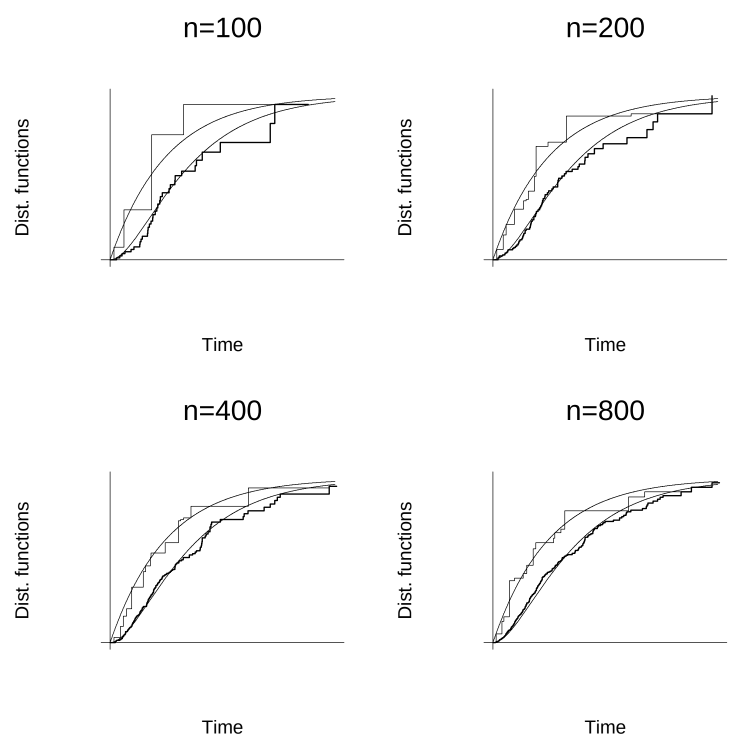

Figure 2: **Joint NPML estimates of F_1 and F_2 (simulated data for example 1).**

interior point algorithm for the estimation of the NPMLE of F_1 and F_2 for simulated data following the distribution functions of Example 1 (in Section 4.2), with several sample sizes. The Primal-Dual Interior Point algorithm for the estimation of the NPMLE of F_1 and F_2 was also applied to the data set presented in Table 1. Figure 3 shows the estimates, and we can notice the larger number of jumps that the estimate of F_2 has compared to the estimate of F_1, a fact related to the difference in the rate of convergence of the estimators of each distribution function.

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

4 Moment Functionals

4.1 Functionals of F_1 or F_2

If we want to estimate a functional of $F = (F_1, F_2)$, like the first moment of F_1, the score operator is given by (where we temporarily abuse notation by writing (X, Y) for (T_1, T_2), and T for Y),

$$\begin{aligned}
[Aa](t, \delta_1, \delta_2) &= E[a(X, Y) \,|\, Y \wedge C = t, \Delta_1 = \delta_1, \Delta_2 = \delta_2] \\
&= \frac{(1-\delta_1)(1-\delta_2) \iint_{t<x\le y} a(x,y)\, dF(x,y)}{1 - F_1(t)} \\
&\quad + \frac{\delta_1(1-\delta_2) \iint_{x\le t\le y} a(x,y)\, dF(x,y)}{F_1(t) - F_2(t)} \\
&\quad + \frac{\delta_1\delta_2 \int_{x\le t} a(x,t)\, h(x,t)\, dx}{f_2(t)} \quad \text{a.e. } [P_{F,G}]. \qquad (4.1)
\end{aligned}$$

where G is the distribution of the censoring time C. This operator may be defined on $L_2(F)$, with range in $L_2(P_{F,G})$. Since it relates scores, our main interest lies in the domain $L_2^0(F)$. Then its range is contained in $L_2^0(P_{F,G})$. The adjoint of A on $L_2^0(P_{F,G})$ can be written as $[A^*b](x, y) = E[b(T, \Delta_1, \Delta_2) \mid (X, Y) = (x, y)]$ and we get

$$\begin{aligned}
[A^*b](x, y) &= E\left[b(T, \Delta_1, \Delta_2) \mid (X, Y) = (x, y)\right] \\
&= \int_0^x b(u, 0, 0)\, g(u)\, du + \int_x^y b(u, 1, 0)\, g(u)\, du \\
&\qquad + b(y, 1, 1)\, \{1 - G(y)\} \quad \text{a.e. } [F]. \qquad (4.2)
\end{aligned}$$

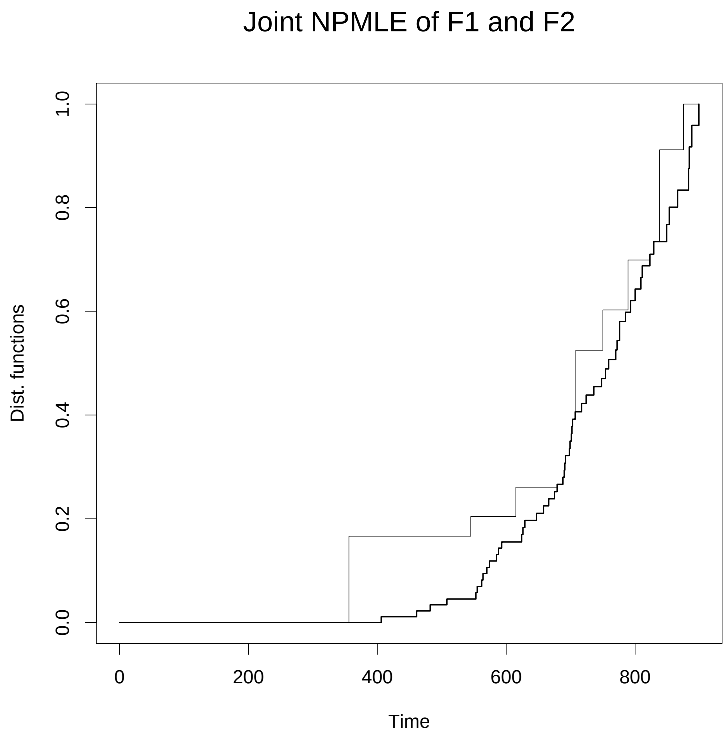

Figure 3: **Joint NPML estimates of F_1 and F_2.**

We have pathwise differentiability of a functional κ of F with canonical gradient $\tilde{\kappa} = \tilde{\kappa}_F$ if and only if

$$\tilde{\kappa} \in \mathcal{R}(A^*)$$

and if this holds, then the canonical gradient in the observation space is the unique element $\tilde{l}_\kappa$ in $\overline{\mathcal{R}(A)} \subset L_2^0(P_{F,G})$ satisfying

$$A^* \tilde{l}_\kappa = \tilde{\kappa}_F \,; \tag{4.3}$$

see Van der Vaart (1991), and Bickel, Klaassen, Ritov, and Wellner (1993), Theorem 5.4.1, page 202.

We now investigate under what conditions the expectation of a function $\Psi(T_1)$ of the time of onset $T_1 \sim F_1$ is a smooth functional. We will assume smoothness of the distribution functions and the function Ψ, allowing us to differentiate in the relations we get below. We have:

$$\tilde{\kappa}_{F_1}(x) = \Psi(x) - E_{F_1}\Psi(X),$$

and we have to solve the equation

$$\begin{aligned} \Psi(x) - E_{F_1}\Psi(X) &= \int_0^x b(u,0,0)\,g(u)\,du + \int_x^y b(u,1,0)\,g(u)\,du \\ &\quad + b(y,1,1)\,\{1 - G(y)\}\,. \end{aligned} \tag{4.4}$$

Differentiating with respect to x we obtain

$$b(x,0,0)g(x) - b(x,1,0)g(x) = \psi(x).$$

where $\psi = \Psi'$. So, defining

$$b_2(x) = b(x,1,0),\ x \ge 0, \tag{4.5}$$

we obtain

$$b(x,0,0) = \frac{\psi(x) + b_2(x)\,g(x)}{g(x)}\,,\ x \ge 0. \tag{4.6}$$

By letting $x \downarrow 0$ in (4.4), we obtain

$$b(y,1,1) = -\frac{\int_0^y b_2(u)\,dG(u) + E_{F_1}\Psi(X)}{1-G(y)}\,,\ y \ge 0. \tag{4.7}$$

We note here, that, if

$$\lim_{t\to\infty} b(t,1,1)\{1-G(t)\} = 0, \tag{4.8}$$

we get

$$\int_0^\infty b_2(u)\,dG(u) = -E_{F_1}\Psi(X)\,. \tag{4.9}$$

But we will see an example below of a smooth functional for which (4.8) is not satisfied and therefore also (4.9) does not hold.

We now determine the null space of A^*. Suppose $A^*\phi = 0$. Then we get:

$$\begin{aligned}[A^*\phi](x,y) &= E\left[\phi(T,\Delta_1,\Delta_2) \mid (X,Y)=(x,y)\right]\\ &= \int_0^x \phi(u,0,0)\,g(u)\,du + \int_x^y \phi(u,1,0)\,g(u)\,du\\ &\qquad + \phi(y,1,1)\,\{1-G(y)\} = 0,\ \text{ a.e. } [F].\end{aligned}$$

By taking the derivative with respect to x we obtain:

$$\phi(x,0,0) = \phi(x,1,0),\ \text{ a.e. } [F_1]. \tag{4.10}$$

Letting $x \downarrow 0$ yields:

$$\phi(y,1,1) = \frac{-\int_0^y \phi(u,1,0)\,dG(u)}{1-G(y)}\,,\ \text{ a.e. } [F_2]. \tag{4.11}$$

Note that if ϕ has compact support, we have:

$$\lim_{y\to\infty} \phi(y,1,1)\{1-G(y)\} = 0, \tag{4.12}$$

and hence, in that case,

$$\phi(y,1,1) = \frac{-\int_0^y \phi(u,1,0)\,dG(u)}{1-G(y)} = \frac{\int_y^\infty \phi(u,1,0)\,dG(u)}{1-G(y)}\,,\ \text{ a.e. } [F_2]. \tag{4.13}$$

Claim: $\mathcal{N}(A^*) = \{\phi \in L_2(P_{F,G}) : (4.10),\ \text{and } (4.11)\ \text{ hold}\}$.

Proof of claim: It is clear that $\phi \in \mathcal{N}(A^*)$ implies (4.10) and (4.11). Conversely, if (4.10) and (4.11) hold, then

$$[A^*\phi]\,(x,y) = \int_0^y \phi(x,1,0)dG(u) + (1-G(y))\phi(y,1,1) = 0,\ \text{a.e. } [F],$$

and hence $\phi \in \mathcal{N}(A^*)$. □

Remark. If $\phi \in L_2(P_{F,G})$ satisfies (4.10), and (4.11), then also $\phi \in L_2^0(P_{F,G})$, since

$$\begin{aligned}
\int \phi\, dQ_{H,G} &= \int \phi(t,0,0)\{1-F_1(t)\}\,dG(t) + \int \phi(t,1,0)\{F_1(t)-F_2(t)\}\,dG(t) \\
&\qquad\qquad + \int \phi(t,1,1)\{1-G(t)\}\,dF_2(t) \\
&= \int \phi(t,1,0)\{1-F_2(t)\}\,dG(t) - \int\int_0^t \phi(u,1,0)\,dG(u)\,dF_2(t) \\
&= \int \phi(t,1,0)\{1-F_2(t)\}\,dG(t) - \int \phi(t,1,0)\{1-F_2(t)\}\,dG(t) = 0,
\end{aligned}$$

where we use Fubini's theorem on the last line of the preceding relations.

We now have:

$$\begin{aligned}
E\tilde{l}_\kappa(T,\Delta_1,\Delta_2)\phi(T,\Delta_1,\Delta_2) &= Eb(C,0,0)\phi(C,0,0)1_{\{C<X\}} + Eb(C,1,0)\phi(C,1,0)1_{\{X\le C\le Y\}} \\
&\qquad + Eb(Y,1,1)\phi(Y,1,0)1_{\{Y<C\}} \\
&= \int_0^\infty b(t,0,0)\phi(t,0,0)\,g(t)\{1-F_1(t)\}\,dt \\
&\qquad + \int_0^\infty b(t,1,0)\phi(t,1,0)\,g(t)\{F_1(t)-F_2(t)\}\,dt \\
&\qquad + \int_0^\infty b(t,1,1)\phi(t,1,1)\,f_2(t)\{1-G(t)\}\,dt.
\end{aligned}$$

By (4.10), this can be written

$$\begin{aligned}
E\tilde{l}_\kappa(T,\Delta_1,\Delta_2)\phi(T,\Delta_1,\Delta_2) &= \int_0^\infty \{b(t,0,0)-b(t,1,0)\}\phi(t,0,0)\,g(t)\{1-F_1(t)\}\,dt \\
&\qquad + \int_0^\infty b(t,1,0)\phi(t,0,0)\,g(t)\{1-F_2(t)\}\,dt \\
&\qquad + \int_0^\infty b(t,1,1)\phi(t,1,1)\,f_2(t)\{1-G(t)\}\,dt.
\end{aligned}$$

Inserting the values of the function b yields:

$$
\begin{aligned}
E\tilde{l}_\kappa(T,\Delta_1,\Delta_2)\phi(T,\Delta_1,\Delta_2) &= \int_0^\infty \psi(t)\phi(t,0,0)\{1-F_1(t)\}\,dt \\
&\quad + \int_0^\infty b_2(t)\,\phi(t,0,0)\,g(t)\{1-F_2(t)\}\,dt \\
&\quad - \int_0^\infty \left\{E_{F_1}\Psi(X) + \int_0^t b_2 dG\right\}\phi(t,1,1)\,f_2(t)\,dt \\
&= 0.
\end{aligned}
$$

Define the R-operator, corresponding to the distribution G, by

$$
Rh(t) = h(t) - \frac{\int_t^\infty h(u)\,dG(u)}{1-G(t)}\,.
$$

for $h \in L_2(G)$. Moreover, define

$$
\alpha(t) = \phi(t,0,0) = \phi(t,1,0), \qquad t \geq 0. \tag{4.14}
$$

Then, if ϕ has compact support, we can write the preceding relation in the following form

$$
\begin{aligned}
&\int_0^\infty \alpha(t)\psi(t)\{1-F_1(t)\}\,dt + \int_0^\infty b_2(t)\,\alpha(t)\,g(t)\{1-F_2(t)\}\,dt \\
&\qquad + \int_0^\infty \frac{E_{F_1}\Psi(X) + \int_0^t b_2 dG}{1-G(t)} \int_0^t \alpha(u)\,dG(u)\,f_2(t)\,dt \\
&= \int_0^\infty \alpha(t)\psi(t)\{1-F_1(t)\}\,dt \\
&\qquad + \int_0^\infty \left\{b_2(t) + \frac{E_{F_1}\Psi(X) + \int_0^t b_2(x)\,dG(x)}{1-G(t)}\right\}[R\alpha](t)\{1-F_2(t)\}\,g(t)\,dt \\
&= 0,
\end{aligned}
$$

where we use (4.13). So we obtain the relation

$$
\begin{aligned}
&\int_0^\infty \alpha(t)\left\{\frac{\psi(t)\{1-F_1(t)\}}{g(t)}\right. \\
&\qquad \left. + R^*\left[\{1-F_2(t)\}\left\{b_2(t) + \frac{E_{F_1}\Psi(X) + \int_0^t b_2(x)\,dG(x)}{1-G(t)}\right\}\right](t)\right\} dG(t) \\
&= 0,
\end{aligned}
$$

for all $\alpha = \phi(\cdot,0,0)$, where $\phi \in \mathcal{N}(A^*)$ and has compact support, and R^* is the adjoint of R in $L_2(G)$. Since the functions ϕ with compact support are dense in $L_2(G)$, we get the equation

$$
\{1-F_2(t)\}\left\{b_2(t) + \frac{E_{F_1}\Psi(X) + \int_0^t b_2(x)\,dG(x)}{1-G(t)}\right\} = -\left[R\left(\frac{\psi\{1-F_1\}}{g}\right)\right](t)\,, \tag{4.15}
$$

since $R \circ R^* = I$; see, e.g., Proposition 8, Part B, p. 421, in Bickel, Klaassen, Ritov, and Wellner (1993).

So b_2 can be determined from (4.15). In fact, dividing by $1 - F_2(t)$ yields:

$$b_2(t) + \frac{E_{F_1}\Psi(X) + \int_0^t b_2(x)\, dG(x)}{1 - G(t)} = -\{1 - F_2(t)\}^{-1} \left[R\left(\frac{\psi\{1 - F_1\}}{g} \right) \right](t),$$

and next, by multiplying by $1 - G(t)$, we get the following integral equation:

$$\{1 - G(t)\}b_2(t) + \int_0^t b_2(x)\, dG(t) + E_{F_1}\Psi(X) = -\frac{1 - G(t)}{1 - F_2(t)} \left[R\left(\frac{\psi\{1 - F_1\}}{g} \right) \right](t), \quad (4.16)$$

which becomes, if one can differentiate g, F_1 and F_2,

$$b_2'(t) = -\{1 - G(t)\}^{-1} \frac{d}{dt} \left\{ \frac{1 - G(t)}{1 - F_2(t)} \left[R\left(\frac{\psi\{1 - F_1\}}{g} \right) \right](t) \right\}, \quad (4.17)$$

and

$$b_2(0) = -\frac{\psi(0)}{g(0)} - E_{F_1}\Psi(X) + \int_0^\infty \psi(t)\{1 - F_1(t)\}\, dt\, . \quad (4.18)$$

Note that when $\Psi(0) = 0$, (4.18) reduces to

$$b_2(0) = -\frac{\psi(0)}{g(0)}\, . \quad (4.19)$$

The complete solution b can now be found by first getting $b_2 = b(\cdot, 1, 0)$ from (4.16) (or (4.17) and (4.18)), and next getting $b(\cdot, 0, 0)$ and $b(\cdot, 1, 1)$ from relations (4.6) and (4.7).

A summary of the above calculations is as follows: the efficient influence for estimation of $E_{F_1}\Psi(X)$ is $\tilde{l}_\kappa(t, \Delta_1, \Delta_2)$ given by

$$\tilde{l}_\kappa(t, \Delta_1, \Delta_2) = \begin{cases} (\psi(t) + b_2(t)g(t))/g(t)\, , & \text{if } (\Delta_1, \Delta_2) = (0, 0), \\ b_2(t)\, , & \text{if } (\Delta_1, \Delta_2) = (1, 0), \\ -\dfrac{E_{F_1}\Psi(X) + \int_0^t b_2(x)\, dG(x)}{1 - G(t)}\, , & \text{if } (\Delta_1, \Delta_2) = (1, 1), \end{cases} \quad (4.20)$$

where b_2 is determined by (4.16) (or (4.17) and (4.18)). The information bound is just

$$\begin{aligned} I_\kappa^{-1} &= E(\tilde{l}_\kappa(Y, \Delta_1, \Delta_2)^2) \\ &= \int_0^\infty \left(\frac{\psi(y) + b_2(y)g(y)}{g(y)} \right)^2 g(y)(1 - F_1(y))dy \\ &\quad + \int_0^\infty b_2^2(y)g(y)(F_1(y) - F_2(y))dy \\ &\quad + \int_0^\infty \frac{\left\{E_{F_1}\Psi(X) + \int_0^y b_2 dG\right\}^2}{(1 - G(y))^2}(1 - G(y))dF_2(y). \end{aligned} \quad (4.21)$$

Remark: Note that when the d.f. F_2 is concentrated at $+\infty$ and $\Psi(x) = x$, then (4.16) is solved by $b_2(t) = -(1 - F_1(t))/g(t)$, so that (with probability one)

$$\tilde{l}_\kappa(y, \Delta_1, \Delta_2) = -\,\frac{\Delta_1 - F_1(y)}{g(y)}\,,$$

which agrees with the influence function of the mean for interval censored case 1 (current status) data; see e.g. BKRW (1993), page 209; Groeneboom and Wellner (1992), page 115, and Huang and Wellner (1995), page 157.

Now consider the estimation of the expectation $E_{F_2}\Psi(Y)$ for a function $\Psi(Y)$ of the time of death $Y \sim F_2$, satisfying $\Psi(0) = 0$. Note that the moments $E_{F_2}(Y^k)$ of the time of death distribution are of this type. We have:

$$\tilde{\kappa}_F(y) = \Psi(y) - E_{F_2}(\Psi(Y)),$$

and to determine whether $\kappa(F) = E_{F_2}\Psi(Y)$ is a differentiable functional, we have to solve the equation

$$\begin{aligned} \Psi(y) - E_{F_2}\Psi(Y) &= \int_0^x b(u,0,0)\,g(u)\,du + \int_x^y b(u,1,0)\,g(u)\,du \\ &\qquad + b(y,1,1)\,\{1 - G(y)\}\,. \end{aligned} \tag{4.22}$$

By differentiating with respect to x we get:

$$b(x,0,0) = b(x,1,0), \ \text{ a.e. } [G]\,. \tag{4.23}$$

But this implies that the calculation collapses to the same calculation as for random right censoring problem (of $T_2 \sim F_2$ by $C \sim G$) since the marginal distribution $P_2 \equiv P_{2,F,G}$ of (Y, Δ_2) is exactly that of random right-censored data. For these calculations, see e.g. Bickel, Klaassen, Ritov, and Wellner (1993), pages 272 - 280, and especially page 276. Hence the efficient influence function $\tilde{l}_\kappa$ for functionals of this type is given by

$$\begin{aligned} \tilde{l}_\kappa(y, \Delta_1, \Delta_2) &= \begin{cases} 0 - \displaystyle\int_0^y \frac{R_{F_2}(\Psi)(s)}{1 - G(s)} \frac{dF_2(s)}{1 - F_2(s)}\,, & \text{if } (\Delta_1, \Delta_2) = (0,0) \ \text{ or } \ (\Delta_1, \Delta_2) = (1,0) \\[2ex] \displaystyle\frac{R_{F_2}(\Psi)(y)}{1 - G(y)}\,, & \text{if } (\Delta_1, \Delta_2) = (1,1)\,. \end{cases} \\ &= \int_0^\infty \frac{R_{F_2}(\Psi)}{1 - G} d\mathbb{M}_{uc} \end{aligned}$$

where

$$\mathbb{M}_{uc}(t) \equiv 1_{[T \le t, \Delta_2 = 1]} - \int_0^t 1_{[T \ge s]} d\Lambda_{F_2}(s)\,.$$

Then the information bound for estimation of $\kappa(F_2) = E_{F_2}\Psi(Y)$ is

$$I_\kappa^{-1} = \int_0^\infty \frac{[R_{F_2}(\Psi)]^2}{\overline{G}^2} \overline{F_2}\,\overline{G} d\Lambda_F = \int_0^\infty \frac{[R_{F_2}(\Psi)]^2}{\overline{G}}\, dF_2\,. \tag{4.24}$$

This is in agreement with the results of Gill (1983), Stute (1995), and Akritas (2000).

4.2 Examples

Example 1. In the particular case studied by van der Laan, Jewell, and Peterson (1997), we have

$$F_1(x) = 1 - e^{-x/2}, \quad F_2(x) = 1 + e^{-x} - 2e^{-x/2}, \quad g(x) = \tfrac{2}{5}e^{-2x/5}.$$

Note that F_2 is the distribution of the sum of a standard exponential random variable U and an exponential random variable V with scale parameter 2, where U and V are independent. In this case we would get, if $\Psi(x) = x$ (the functional, corresponding to the first moment of F_1),

$$\left[R\left(\frac{1-F_1}{g}\right)\right](t) = \tfrac{1}{2}e^{-t/10}\,,$$

$$\frac{1-G(t)}{1-F_2(t)}\left[R\left(\frac{1-F_1}{g}\right)\right](t) = -\frac{1}{4-2e^{-t/2}}\,,$$

and

$$\frac{d}{dt}\left\{\frac{1-G(t)}{1-F_2(t)}\left[R\left(\frac{1-F_1}{g}\right)\right](t)\right\} = -\frac{e^{-t/2}}{4\left(2-e^{-t/2}\right)^2}\,.$$

Hence

$$b_2'(t) = \frac{e^{-t/10}}{4\left(2-e^{-t/2}\right)^2}\,,$$

and b_2 is given by:

$$b_2(t) = -\frac{5}{2} + \int_0^t \frac{e^{-x/10}}{4\left(2-e^{-x/2}\right)^2}\,dx, \qquad t \ge 0.$$

Note that $b_2 \in L_2(G)$.

Furthermore, $b(t,1,1)$ satisfies the relation

$$b(t,1,1) = -\,\frac{2+\int_0^t b_2(u)\,dG(u)}{1-G(t)}\,, \qquad t \ge 0\,;$$

(see (4.7)). We have:

$$\begin{aligned}\int_0^t b_2(u)\,dG(u) &= b_2(0) - \{1-G(t)\}b_2(t) + \int_0^t \{1-G(u)\}b_2'(u)\,du \\ &= b_2(0) - \{1-G(t)\}b_2(t) + \int_0^t \frac{e^{-u/2}}{4\left(2-e^{-u/2}\right)^2}\,du \\ &= -2 - \{1-G(t)\}b_2(t) - \frac{1}{2\left(2-e^{-t/2}\right)}\,,\end{aligned}$$

and hence

$$\begin{aligned} b(t,1,1) &= b_2(t) + \frac{e^{2t/5}}{2\,(2-e^{-t/2})} = -\frac{5}{2} + \frac{e^{2t/5}}{2\{2-e^{-t/2}\}} + \int_0^t \frac{e^{-u/10}}{4\{2-e^{-u/2}\}^2}\,du \\ &= -2 + \frac{1}{5}\int_0^t \frac{e^{2u/5}}{2-e^{-u/2}}\,du. \end{aligned}$$

Note that $b(\cdot,1,1) \notin L_2(G)$, but that

$$\int b(t,1,1)^2\{1-G(t)\}\,dF_2(t) < \infty,$$

and hence $b \in L_2(P_{F,G})$.

The efficient asymptotic variance is given by:

$$\begin{aligned} \|b\|^2_{P_{F,G}} &= \int_0^\infty \frac{1-F_1(t)}{g(t)}\,dt + 2\int_0^\infty \{1-F_1(t)\}b_2(t)\,dt \\ &\quad + \int b_2(t)^2\{1-F_2(t)\}\,dG(t) + \int b(t,1,1)^2\{1-G(t)\}\,dF_2(t), \end{aligned}$$

and numerical evaluation of this expression yields in the present case:

$$I_\kappa^{-1} = \|b\|^2_{P_{F,G}} \approx 27.19....$$

If $\Psi(x) = x^2$ (corresponding to the second moment of F_1), we have to solve the equation

$$b_2'(t) = -\{1-G(t)\}^{-1}\frac{d}{dt}\left\{\frac{1-G(t)}{1-F_2(t)}\left[R\left(\frac{\psi\{1-F_1\}}{g}\right)\right](t)\right\},$$

where $\psi(t) = 2t$, under the side condition $b_2(0) = 0$, see (4.17) and (4.18). We get:

$$\left[R\left(\psi\frac{1-F_1}{g}\right)\right](t) = e^{-t/10}\{t-8\},$$

and hence

$$b_2'(t) = -\frac{4e^{2t/5} + (6-t)e^{-t/10}}{2\{2-e^{-t/2}\}^2},$$

implying

$$\begin{aligned} b_2(t) &= -\int_0^t \frac{4e^{2u/5} + (6-u)e^{-u/10}}{2\{2-e^{-u/2}\}^2}\,du \\ &= -\int_0^t \frac{2e^{2u/5}\{2-e^{-u/2}\} + (8-u)e^{-u/10}}{2\{2-e^{-u/2}\}^2}\,du \\ &= -\int_0^t \frac{e^{2u/5}}{2-e^{-u/2}}\,du - \frac{2}{5}\int_0^t \frac{(8-u)e^{2u/5}}{2-e^{-u/2}}\,du \\ &\quad + \int_0^t \frac{e^{2u/5}}{2-e^{-u/2}}\,du + \frac{(8-t)e^{2t/5}}{2-e^{-t/2}} - 8 \\ &= -\frac{2}{5}\int_0^t \frac{(8-u)e^{2u/5}}{2-e^{-u/2}}\,du + \frac{(8-t)e^{2t/5}}{2-e^{-t/2}} - 8. \end{aligned} \tag{4.25}$$

Moreover,

$$b(t,1,1) = -\frac{2}{5}\int_0^t \frac{(8-u)e^{2u/5}}{2-e^{-u/2}}\,du - 8. \tag{4.26}$$

This can perhaps be most easily seen from the fact that $b(\cdot,1,1)$ has to satisfy the differential equation

$$\tfrac{5}{2}y'(t) - y(t) = -b_2(t),$$

under the side condition $b(0,1,1) = -8$, which follows from (4.7), and by using (4.25). Using these expressions for b_2 and $b(\cdot,1,1)$, we obtain as estimate of the efficient asymptotic variance for the estimation of the second moment of F_1:

$$I_\kappa^{-1} = \|b\|^2_{P_{F,G}} \approx 19705.$$

The largest contribution is coming from the first term in the information lower bound, that is, the integral

$$\int b(\cdot,0,0)^2\{1-F_1\}dG = \int \left\{\frac{\psi^2}{g^2} + b_2^2\right\}\{1-F_1\}dG.$$

Note that

$$\int \frac{\psi^2}{g^2}\{1-F_1\}dG = 10\int_0^\infty x^2 e^{-x/10}\,dx = 2\cdot 10^4,$$

showing that the largest contribution is in fact coming from this part of the first term. Also note that, for example $10\int_0^{10} x^2e^{-x/10}\,dx \approx 1606.03$. This indicates that the finite sample variances differ considerably from the values predicted by asymptotic theory, and this is confirmed by our experimental results in Section 5.2; see Table 5 and compare with the truncated integrals shown in Table 2 in the case of $E_{F_2}\Psi(Y)$.

Example 2. If we take

$$F_1(x) = 1-e^{-x},\quad F_2(x) = 1+e^{-x}-2e^{-x/2},\quad g(x) = \tfrac{2}{5}e^{-2x/5},$$

so only changing the distribution of X to a standard exponential and leaving the other distributions the same, we get:

$$b_2(t) = -\frac{5}{2} + \int_0^t \frac{6e^{9x/10}}{\{4e^{x/2}-2\}^2}\,dx,$$

and

$$\int b_2\,dG = -1 = -E_{F_1}(X),$$

since now $\lim_{t\to\infty} b(t,1,1)\{1-G(t)\} = 0$ and $b(\cdot,1,1)\in L_2(G)$, in contrast with the preceding example.

In this case we calculate

$$\int_0^\infty \frac{1-F_1(y))}{g(y)}dy = 25/6\,,$$

$$\int_0^\infty b_2(y)(1-F_1(y))dy = -1.63852...\,,$$

$$\int_0^\infty b_2^2(y)g(y)(1-F_2(y))dy = 1.55032...\,,$$

$$\int_0^\infty \left\{\int_y^\infty b_2 dG\right\}^2 \frac{1}{1-G(y)} dF_2(y) = 0.0904487...\,,$$

Thus $I_\kappa^{-1} = 2.5304....$

For $\kappa(F_2) = E_{F_2}(T_2)$ we find from (4.24) that $I_\kappa^{-1} = 35.38...$. On the other hand, it is interesting to note that if we let $\tau_n \equiv H_2^{-1}(\frac{n-1}{n})$ where $H_2(x) = 1-(1-F_2(x))(1-G(x))$ is the distribution function corresponding to the minimum of $T_2 \sim F_2$ and $C \sim G$, then the contribution to the integral in (4.24) from the interval $[0, \tau_n]$ can be considerably smaller than I_κ^{-1}. The following table gives the values of

$$I_{\kappa,n}^{-1} \equiv \int_0^{\tau_n} \frac{[R_{F_2}(\Psi)]^2}{\overline{G}^2} \overline{F_2}\,\overline{G} d\Lambda_F = \int_0^{\tau_n} \frac{[R_{F_2}(\Psi)]^2}{\overline{G}} dF_2\,. \tag{4.27}$$

Table 2: *Truncated Information Bound Integrals*

n	100	200	300	400	500	
τ_n	5.86	6.64	7.09	7.41	7.66	
$I_{\kappa_1,n}^{-1}$	13.22	14.85	15.74	16.36	16.83	
$I_{\kappa_2,n}^{-1}$	1543.9	1989.6	2270.2	2482.48	2651.3	
n	1000	2000	3000	4000	5000	∞
τ_n	8.44	9.21	9.66	9.98	10.23	∞
$I_{\kappa_1,n}^{-1}$	18.20	19.27	20.17	20.65	21.01	35.38 ...
$I_{\kappa_2,n}^{-1}$	3205.8	3801.1	4167.7	4435.5	4645.7	38819.6

These estimates seem to nicely fit the simulation results of the MLE estimates for corresponding sample sizes.

5 Simulation Studies

5.1 Estimation of F_1 or F_2

In order to have a proper view of the behavior of the NPMLE of F_1 and the Weighted Least Squares estimator proposed by van der Laan, Jewell, and Peterson (1997) to solve the problem of estimating F_1, two examples of distributions for T_1, T_2 and C were considered. Example 1 was exactly as described in section 4.2: $T_1 \sim \exp(0.5)$, $\quad T_2 - T_1 \sim \exp(1)$, $\quad C \sim \exp(0.4)$, i.e. $F_1(t) = 1 - e^{-t/2}, \quad t \geq 0$, $F_2(t) = 1 - 2e^{-t/2} + e^{-t}, \quad t \geq 0$, e $G(t) = 1 - e^{-2t/5}, \quad t \geq 0$. This is the primary example considered by van der Laan, Jewell, and Peterson (1997). It will make a comparison of our results with theirs possible. Here is our further example:

Example 3. In this example we have

$$(T_1, T_2) \sim U(A_1) \quad \text{with probability } 1/3, \text{ where} \quad A_1 = \{(t_1, t_2) : 0 \leq t_1 \leq t_2 \leq 1\},$$

$$(T_1, T_2) \sim U(A_2) \quad \text{with probability } 1/3, \text{ where} \quad A_2 = \{(t_1, t_2) : 1 \leq t_1 = t_2 \leq 2\},$$

$$(T_1, T_2) \sim U(A_3) \quad \text{with probability } 1/3, \text{ where} \quad A_3 = \{(t_1, t_2) : 2 \leq t_1 \leq t_2 \leq 3\}.$$

$C \sim U(0, 3)$.

In this case, the (marginal) distribution functions are

$$F_1(t) = \begin{cases} (1 - (1-t)^2)/3 & \text{if } 0 \leq t \leq 1 \\ t/3 & \text{if } 1 < t \leq 2 \\ 1 - (3-t)^2/3 & \text{if } 2 < t \leq 3 \end{cases},$$

$$F_2(t) = \begin{cases} t^2/3 & \text{if } 0 \leq t \leq 1 \\ t/3 & \text{if } 1 < t \leq 2 \\ (2 + (t-2)^2)/3 & \text{if } 2 < t \leq 3 \end{cases}.$$

In each case studied, we generated $r = 625$ samples of sizes 100 and 400. Those are the number of samples and sample sizes used by van der Laan, Jewell, and Peterson (1997) in their paper.

The summary statistics include the Mean Squared Error at 9 quantiles ($F_1^{-1}(j/10)$ and $F_2^{-1}(j/10)$ for $j = 1, \ldots, 9$) estimated by

$$\widehat{MSE}(t_j) = \frac{1}{r} \sum_{i=1}^{r} \left(\hat{F}_{1,i}(t_j) - \left(\frac{j}{10} \right) \right)^2$$

$$= \frac{1}{r} \left(\sum_{i=1}^{r} \left[\hat{F}_{1,i}(t_j) \right]^2 - \frac{j}{5} \sum_{i=1}^{r} \hat{F}_{1,i}(t_j) \right) + \left(\frac{j}{10} \right)^2$$

and similarly for $\hat{F}_2$.

In tables 7, 8 and 9, the bias and variance of the estimators at each quantile are presented as well as the standard error of the Mean Squared Error, estimated by the square root of

$$\begin{aligned}
\widehat{Var}(\widehat{MSE}(t_j)) &= \frac{1}{r}\left\{\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^4 - \frac{1}{r}\left(\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^2\right)^2\right. \\
&\quad + \left(\frac{j}{5}\right)^2\left[\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^2 - \left(\frac{1}{r}\sum_{i=1}^{r}\hat{F}_{1,i}(t_j)\right)^2\right] \\
&\quad \left. -\frac{2j}{5}\left[\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^3 - \left(\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^2\right)\left(\frac{1}{r}\sum_{i=1}^{r}\hat{F}_{1,i}(t_j)\right)\right]\right\}.
\end{aligned}$$

In order to assess the variability of the ratio of the averages of the MSE of the Weighted Least Squares and NPML estimators of F_1 (and the Kaplan-Meier and NPML estimators of F_2), confidence intervals ($\pm 2s.e.$) for that ratio were calculated for each quantile. The delta method gives us an expression for the asymptotic variance of the ratio of two averages:

$$\sqrt{r}\left(\frac{\overline{X}_r}{\overline{Y}_r} - \frac{m_X}{m_Y}\right) = \sqrt{r}\left(\frac{\overline{X}_r - m_X}{\overline{Y}_r}\right) + \sqrt{r}m_X\left(\frac{1}{\overline{Y}_r} - \frac{1}{m_Y}\right)$$

$$\sqrt{r}\left(\frac{\overline{X}_r - m_X}{\overline{Y}_r}\right) \xrightarrow{\mathcal{D}} \frac{\sigma_X Z_1}{m_Y}, \text{ where } Z_1 \sim N(0,1)$$

$$\sqrt{r}m_X\left(\frac{1}{\overline{Y}_r} - \frac{1}{m_Y}\right) \xrightarrow{\mathcal{D}} -m_X\frac{1}{m_Y^2}\sigma_Y Z_2, \text{where } Z_2 \sim N(0,1)$$

$$\text{Var}\left(\frac{\sigma_X Z_1}{m_Y} - \frac{m_X\sigma_Y}{m_Y^2}Z_2\right) = \frac{\sigma_X^2}{m_X^2} + \frac{m_X^2\sigma_Y^2}{m_Y^4} - \frac{2\sigma_X m_X\sigma_Y}{m_Y^3}\rho$$

Consequently we can estimate the variance of $\overline{X}_r/\overline{Y}_r$ by

$$\frac{\hat{\sigma}_X^2/r}{\overline{X}_r^2} + \frac{\overline{Y}_r^2\hat{\sigma}_Y^2/r}{\overline{X}_r^4} - 2\frac{\hat{\rho}(\hat{\sigma}_X/\sqrt{r})\overline{X}_r(\hat{\sigma}_Y/\sqrt{r})}{\overline{Y}_r^3}$$

where ρ is the correlation between Z_1 and Z_2, (i.e., the correlation between estimators being compared).

Figures 4 to 5 show the relative efficiency (ratio between the average MSE's) with confidence bands between the Weighted Least Squares and NPML estimators of F_1 and between the Kaplan-Meier and the NPML estimators of F_2 for each quantile (x-axis).

Figure 4 shows that our results for example 1, obtained using the Primal-Dual Interior Point algorithm, do not support the conclusions in van der Laan *et al.* (1997) since the

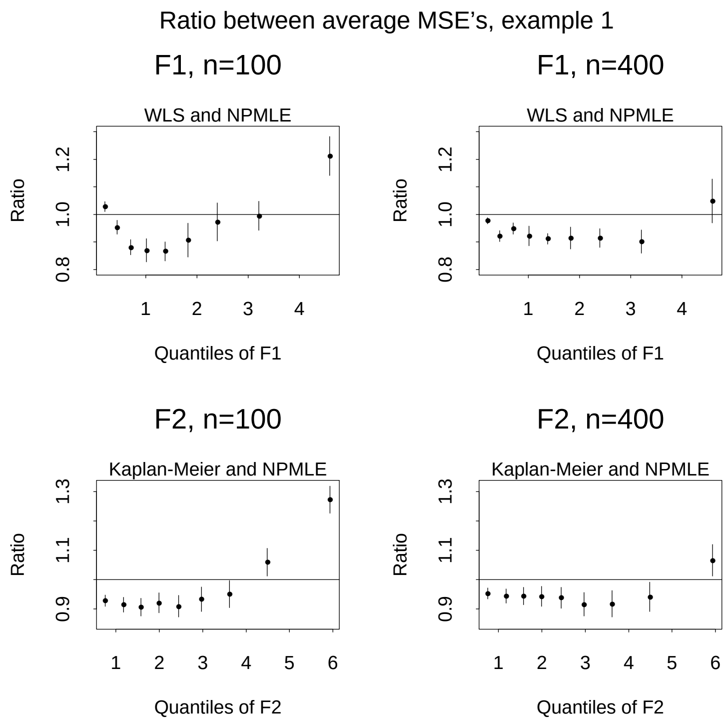

Figure 4: **Ratio between average MSE's of WLSE and NPMLE of F_1 and between the Kaplan-Meier and NPMLE of F_2 (example 1).**

NPMLE seems to estimate F_1 better in the right-hand tail for both sample sizes (efficiency > 1), while in their results they claim to have observed a relative efficiency of 0.92 for both sample sizes in the last quantile considered. The difference in the results is caused by the way they calculated the NPMLE of F_1, which may not yield the correct estimate. Looking at columns (2), (3) and (4) for both estimators in table 7 we see that the lower variance of the NPMLE of F_1 at the last quantile is the reason for its superior performance there, and the variance seems to explain also the better performance of the WLS estimator at the other quantiles since the variance is the biggest component of the MSE for both estimators.

For Example 3 (figure 5), the WLS estimator is beaten by the NPMLE in the central quantiles for all sample sizes. These results indicate that the good performance of their estimator claimed by van der Laan, Jewell, and Peterson (1997) tends to happen when F_1 and F_2 are far apart since it was generally beaten by the NPMLE in the tails for Example 1 or in the central quantiles in Example 3, where $F_1 = F_2$.

Figure 6 shows the plots of the logarithm of the variance of each estimator against the logarithm of the sample size. The slopes of the curves give us information about the rate of convergence of the estimators since

$$\text{Var}(\hat{\theta}_n) \doteq cn^{-r} \quad \text{implies} \quad \log(\text{Var}(\hat{\theta}_n)) \doteq \log c - r \log n.$$

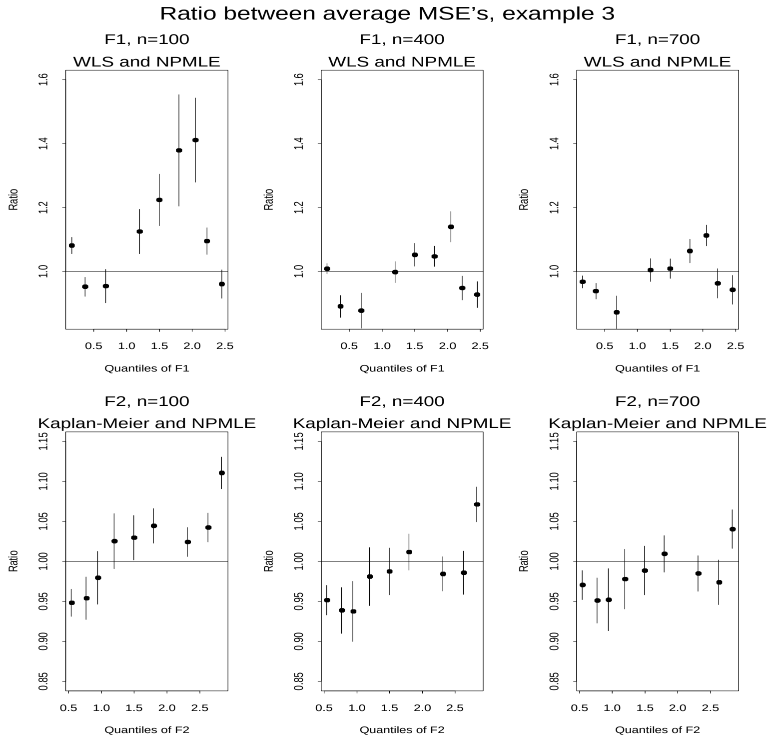

Figure 5: **Ratio between average MSE's of WLSE and NPMLE of F_1 and between Kaplan-Meier and NPMLE of F_2 (example 3).**

In each plot, the solid line refers to the second quantile, the dotted line refers to the fifth quantile, and the dashed line refers to the eighth quantile. In the second and eighth quantiles, F_1 and F_2 are far apart, while in the fifth one (in the central part of the range of T_1 and T_2) we have $F_1 = F_2$. We can see in the plots and in table 3 that the slopes of the curves for the estimators of F_1 are around $-2/3$ in the second and eighth quantiles (where $F_1 > F_2$) and around -1 in the fifth one (where $F_1 = F_2$), suggesting that when $F_1 = F_2$ we probably have $n^{-1/2}$ as the rate of convergence of the estimators of F_1, a property that the estimators of F_2 have for the whole real line, as we can see in the plots for the Kaplan-Meier and NPML estimators of F_2 in figure 6 and table 3, where all the lines have slopes close to -1. This

Table 3: *Slopes of the log-log plots at each quantile.*

quantile	WLS	NPMLE of F1	Kaplan-Meier	NPMLE of F2
2nd	-0.673	-0.597	-0.984	-0.987
5th	-1.048	-0.970	-0.980	-0.970
8th	-0.786	-0.671	-1.062	-1.037

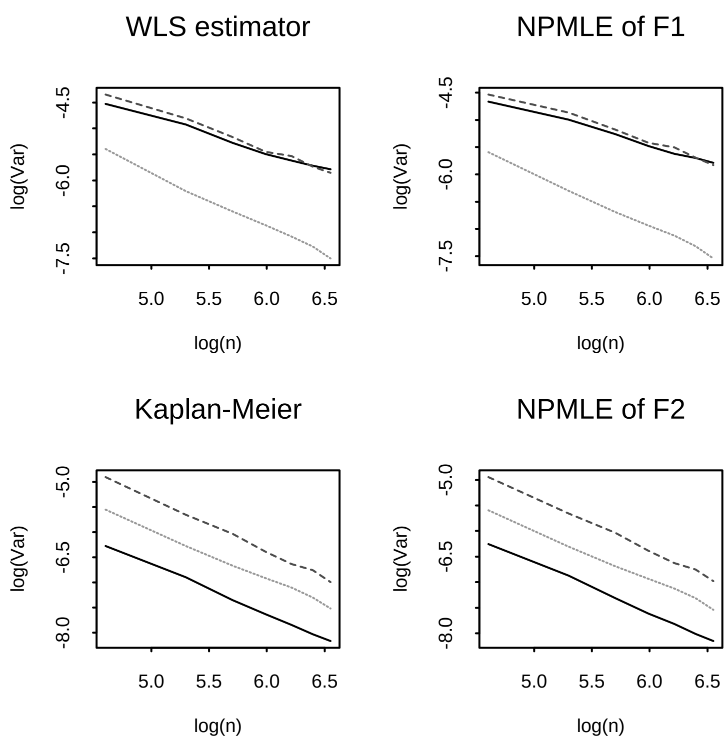

Figure 6: **log-log plot of variance of the estimators versus sample size.**

happens because $F_1(t) = F_2(t)$ for $t \in (a, b)$ means that the disease kills instantly in the interval (a, b) and thus the time of death is also the time of disease onset, which is then actually observed in that situation.

It should be noticed that when $\| F_1 - F_2 \|_\infty = 1$ the NPMLE, the pseudo NPMLE and the Weighted Least Squares estimator of F_1 proposed by van der Laan, Jewell, and Peterson (1997) coincide. The same is true for the NPMLE, the pseudo NPMLE and the Kaplan-Meier estimator for F_2. That happens because $\| F_1 - F_2 \|_\infty = \sup_{t\in\mathbb{R}} | F_1 - F_2 | = 1$ implies that the ordered observed death times $U_i = C_i \wedge T_{2,i}$ are such that there are two blocks of observations, the first one with $\Delta_{2,i} = 0$, and the second one with $\Delta_{2,i} = 1$. That makes $\hat{F}_{2,KM}(U_{(i)}) = \hat{F}_{2,n}(U_{(i)}) = 0$ where $\hat{F}_{2,KM}$ is the Kaplan-Meier estimator and $\hat{F}_{2,n}$ is the NPML estimator of F_2 for the observations in the first block. Also, in that block the weights w_i become equal to 1 and the cumulative sum diagram that will be used to calculate the WLS estimator is similar to the one used to calculate the NPMLE of $F \equiv F_1$ in the Interval Censoring, case 1 (see Groeneboom and Wellner (1992)), implying that the NPMLE of F_1, the pseudo NPMLE of F_1 and the weighted least squares estimators will coincide. On the other hand, in the second block of observations we have $\Delta_{1,(i)} = 1$. A look at the expression of the log-likelihood shows that it will be maximized making $\hat{F}_{1,n} = 1$ for all the observations in the second block, and hence the log-likelihood coincides with the log-likelihood for the right censoring problem, which is maximized in F_2 by the Kaplan-

Meier estimator. Also, the cumulative sum diagram for the calculation of the weighted least squares estimator has $G_i = 0$ for those observations, making $\hat{F}_{1,WLS}(U_{(i)}) = 1$, where $\hat{F}_{1,n}$ is the NPMLE of F_1 and $\hat{F}_{1,WLS}$ is the estimator proposed by van der Laan *et al.* (1997). Putting all these facts together we see that in the case where $\| F_1 - F_2 \|_\infty = 1$, all the estimators of F_1 and F_2 considered here coincide.

5.2 Estimation of moment functionals

In the computations for the following tables, the estimators of the first and second moments were of the form

$$\int_0^{Y_{(n)}} (1 - \hat{F}_n(x))\, dx \qquad \text{or} \qquad \int_0^{Y_{(n)}} 2x(1 - \hat{F}_n(x))\, dx \tag{5.28}$$

respectively where $\hat{F}_n$ represents any one of the possible estimators of F_1 or F_2. This amounts to using the estimators

$$\int_0^{Y_{(n)}} x\, d\hat{F}_n(x) \qquad \text{or} \qquad \int_0^{Y_{(n)}} x^2\, d\hat{F}_n(x), \tag{5.29}$$

but putting $\hat{F}_n$ equal to 1 at $Y_{(n)}$. It is also possible to allow defective distribution functions $\hat{F}_n$ in (5.29), but then these estimators will have a large downward bias and a much bigger variance and for this reason we prefer to use the estimators (5.28).

Table 4: *Monte-Carlo Results for estimation of moments, Example 1, Means*

Meth, d.f.	MLE, 1	JLP, 1	MLE, 2	KM, 2	MLE, 1	JLP, 1	MLE, 2	KM, 2
Moment	1	1	1	1	2	2	2	2
n								
100	1.9301	1.8948	2.7907	2.8969	6.4141	6.3781	11.3654	12.2310
200	1.9368	1.9153	2.8374	2.9306	6.6686	6.6963	11.8521	12.6867
300	1.9443	1.9303	2.8455	2.9278	6.8093	6.8653	11.9804	12.7501
400	1.9647	1.9539	2.8814	2.9578	7.0611	7.1224	12.2997	13.0534
500	1.9592	1.9513	2.8856	2.9540	7.0773	7.1386	12.3897	13.0735
1000	1.9734	1.9715	2.9178	2.9707	7.3012	7.3727	12.7669	13.3431
2000	1.9790	1.9794	2.9403	2.9803	7.4848	7.5575	13.0730	13.5484
3000	1.9825	1.9834	2.9487	2.9820	7.5519	7.6156	13.1678	13.5800
4000	1.9840	1.9863	2.9563	2.9860	7.5939	7.6594	13.2748	13.6555
5000	1.9874	1.9886	2.9591	2.9861	7.6418	7.7001	13.3079	13.6598
Theory	2	2	3	3	8	8	14	14

Table 5: *Monte-Carlo Results for estimation of moments, Example 1, Variances*

Meth, d.f.	MLE, 1	JLP, 1	MLE, 2	KM, 2	MLE, 1	JLP, 1	MLE, 2	KM, 2
Moment	1	1	1	1	2	2	2	2
n								
100	8.57	9.53	12.07	14.29	649.2	749.7	1337.3	1703.8
200	8.69	9.57	13.45	15.87	708.4	825.8	1720.4	2215.4
300	9.54	10.42	13.11	15.37	866.2	1036.2	1645.3	2141.1
400	9.56	10.27	13.57	16.05	915.4	1064.4	1743.9	2313.6
500	10.34	11.04	14.23	16.48	1049.9	1224.0	2059.2	2602.4
1000	10.94	15.70	11.62	17.86	1304.2	1499.2	2555.7	3191.5
2000	11.91	13.01	17.12	19.29	1599.8	1878.1	3017.3	3775.4
3000	10.80	11.74	17.16	19.25	1583.8	1831.3	3259.8	4032.7
4000	11.62	12.46	17.51	19.77	1658.8	1887.0	3410.2	4252.3
5000	12.70	13.43	19.26	21.30	1819.3	2003.9	3659.4	4428.0
Theory	27.19		35.38	35.38	19705		38819.6	38819.6

Table 6: *Monte-Carlo Results for estimation of moments, Example 1, Ratios of Variances*

Variance ratio	$\frac{Var(MLE1)}{Var(JLP)}$	$\frac{Var(MLE2)}{Var(KM)}$	$\frac{Var(MLE1)}{Var(JLP)}$	$\frac{Var(MLE2)}{Var(KM)}$
Moment	1	1	2	2
n				
100	0.90	0.84	0.87	0.78
200	0.91	0.85	0.86	0.78
300	0.92	0.85	0.84	0.77
400	0.93	0.85	0.86	0.75
500	0.86	0.86	0.86	0.80
1000	0.94	0.88	0.87	0.80
2000	0.92	0.89	0.85	0.80
3000	0.92	0.89	0.86	0.81
4000	0.93	0.89	0.88	0.81
5000	0.95	0.90	0.91	0.83

6 Conclusions

The Weighted Least Squares of F_1 proposed by van der Laan *et al.* (1997) tends to be more efficient than the NPMLE of F_1 for the survival - sacrifice model when F_1 and F_2 are far apart. When F_1 and F_2 are close, however, the opposite seems to happen. This was observed after the calculation of the joint NPMLE of F_1 and F_2 using the Primal-Dual Interior Point algorithm since the algorithm LJP (1997) applied for that purpose does not yield the true NPML estimator of F_1 and F_2, making their results about the relative efficiency of their estimator and the NPMLE of F_1 unreliable. The magnitude of the variance of each estimator at the quantiles seems to explain the differences between the NPMLE and the Weighted Least Squares estimator of F_1, since the bias of both estimators are similar.

When $F_1 = F_2$ we have $n^{-1/2}$ as the rate of convergence for the estimator of F_1 instead of $n^{-1/3}$ which is the rate when $F_1 > F_2$. That happens because when $F_1 = F_2$ in an interval, the disease kills instantly in that interval, making T_1 observable then.

When $\| F_1 - F_2 \|_\infty = 1$, the NPMLE, the pseudo NPMLE, and the Weighted Least Squares estimator of F_1 proposed by LJP (1997) coincide. The same is true for the NPMLE, the pseudo NPMLE and the Kaplan-Meier estimator of F_2.

Finally, we computed information bounds for the estimation of smooth functionals of F_1 and F_2. Application to the model, used in the simulation study of LJP (1997), showed smaller variances of the MLE estimators of the first and second moments for both F_1 and F_2, and sample sizes from 100 up to 5000, in comparison to the estimates, based on the weighted least squares estimator for F_1, proposed in LJP (1997), and the Kaplan-Meier estimator for F_2 (see table 6). We obtained similar results for another model (Example 2 above), but the results of that simulation study is not given here for reasons of space.

The information bounds are given by slowly converging integrals on $(0, \infty)$, in particular for higher moments. Table 2 shows the values of these integrals on the intervals $(0, \tau_n)$, instead of $(0, \infty)$, where $\tau_n \equiv H_2^{-1}(\frac{n-1}{n})$, where H_2 is the distribution function of the minimum of $Y \sim F_2$ and $C \sim G$. These values are in agreeement with the simulation results of the MLE estimates for corresponding sample sizes.

The better performance of the MLE estimators in the estimation of the first and second moments is probably due to the fact that the MLE's do a better job in estimating the tails of the distributions than the weighted least squares estimator proposed in van der Laan, Jewell, and Peterson (1997), combined with the Kaplan-Meier estimator. This, in turn, is probably due to the better performance of the MLE's in regions where the constraints become active.

Table 7: *Summary results of the simulations for the WLS and NPML estimators of F_1 (Example 1, n=100, 400).*

(1) MSE of each estimator at the quantile (2) Bias of each estimator at the quantile (3) Variance of each estimator at the quantile (4) Standard error of the MSE of each estimator at the quantile								
n	WLS				NPMLE			
100	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.0075	-0.0468	0.0053	0.000236	0.0073	-0.0526	0.0045	0.000212
2	0.0110	-0.0357	0.0097	0.000454	0.0115	-0.0507	0.0090	0.000445
3	0.0125	-0.0207	0.0120	0.000626	0.0142	-0.0388	0.0127	0.000669
4	0.0118	-0.0058	0.0118	0.000656	0.0136	-0.0234	0.0130	0.000702
5	0.0112	-0.0063	0.0116	0.000604	0.0135	-0.0218	0.0130	0.000671
6	0.0107	0.0005	0.0107	0.000621	0.0118	-0.0155	0.0116	0.000629
7	0.0106	0.0026	0.0106	0.000559	0.0109	-0.0116	0.0108	0.000554
8	0.0091	0.0121	0.0090	0.000476	0.0092	0.0007	0.0092	0.000447
9	0.0066	0.0247	0.0060	0.000261	0.0055	0.0081	0.0054	0.000251
400	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.0035	-0.0208	0.0031	0.000148	0.0036	-0.0248	0.0030	0.000145
2	0.0043	-0.0114	0.0041	0.000237	0.0046	-0.0179	0.0043	0.000251
3	0.0044	-0.0027	0.0044	0.000238	0.0047	-0.0085	0.0046	0.000249
4	0.0042	-0.0022	0.0042	0.000224	0.0045	-0.0081	0.0045	0.000249
5	0.0041	-0.0014	0.0041	0.000227	0.0046	-0.0057	0.0045	0.000250
6	0.0040	-0.0039	0.0039	0.000214	0.0043	-0.0082	0.0043	0.000239
7	0.0033	-0.0031	0.0033	0.000184	0.0037	-0.0057	0.0036	0.000201
8	0.0030	-0.0005	0.0030	0.000159	0.0033	-0.0041	0.0033	0.000176
9	0.0022	0.0037	0.0021	0.000121	0.0021	-0.0016	0.0020	0.000117

Table 8: *Summary results of the simulations for the WLS and NPML estimators of* F_1 *(Example 3, n=100, 400).*

(1) MSE of each estimator at the quantile (2) Bias of each estimator at the quantile (3) Variance of each estimator at the quantile (4) Standard error of the MSE of each estimator at the quantile								
n	WLS				NPMLE			
100	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.0086	-0.0458	0.0065	0.000289	0.0079	-0.0535	0.0051	0.000237
2	0.0118	-0.0313	0.0108	0.000478	0.0124	-0.0547	0.0094	0.000456
3	0.0074	-0.0016	0.0074	0.000370	0.0077	-0.0318	0.0067	0.000358
4	0.0038	0.0183	0.0035	0.000233	0.0034	0.0122	0.0032	0.000192
5	0.0046	0.0085	0.0045	0.000262	0.0037	0.0067	0.0037	0.000204
6	0.0060	0.0070	0.0059	0.000462	0.0043	0.0025	0.0043	0.000235
7	0.0080	-0.0188	0.0076	0.000505	0.0056	-0.0260	0.0050	0.000339
8	0.0143	-0.0370	0.0129	0.000662	0.0131	-0.0485	0.0107	0.000619
9	0.0097	-0.0011	0.0097	0.000479	0.0101	-0.0254	0.0094	0.000496
400	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.0041	-0.0238	0.0035	0.000158	0.0040	-0.0287	0.0032	0.000150
2	0.0043	-0.0158	0.0041	0.000221	0.0049	-0.0271	0.0041	0.000238
3	0.0026	0.0003	0.0026	0.000130	0.0030	-0.0185	0.0026	0.000147
4	0.0008	0.0019	0.0008	0.000046	0.0008	0.0058	0.0008	0.000047
5	0.0010	0.0023	0.0010	0.000057	0.0010	0.0056	0.0010	0.000053
6	0.0012	0.0016	0.0012	0.000067	0.0011	0.0033	0.0011	0.000064
7	0.0023	-0.0237	0.0018	0.000123	0.0021	-0.0216	0.0016	0.000109
8	0.0044	-0.0136	0.0043	0.000250	0.0047	-0.0174	0.0044	0.000251
9	0.0028	0.0010	0.0028	0.000153	0.0031	-0.0025	0.0031	0.000161

Table 9: *Summary results of the simulations for the WLS and NPML estimators of F_1 (Example 3, n=700).*

(1) MSE of each estimator at the quantile (2) Bias of each estimator at the quantile (3) Variance of each estimator at the quantile (4) Standard error of the MSE of each estimator at the quantile								
n	WLS				NPMLE			
700	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.0028	-0.0194	0.0024	0.000132	0.0030	-0.0237	0.0024	0.000130
2	0.0031	-0.0091	0.0030	0.000161	0.0034	-0.0170	0.0030	0.000171
3	0.0017	-0.0005	0.0017	0.000092	0.0020	-0.0151	0.0017	0.000116
4	0.0005	0.0009	0.0005	0.000026	0.0005	0.0047	0.0005	0.000025
5	0.0005	0.0011	0.0005	0.000029	0.0005	0.0042	0.0005	0.000028
6	0.0007	0.0015	0.0007	0.000037	0.0006	0.0027	0.0006	0.000034
7	0.0018	-0.0214	0.0014	0.000090	0.0016	-0.0195	0.0012	0.000084
8	0.0029	-0.0069	0.0028	0.000162	0.0030	-0.0091	0.0029	0.000162
9	0.0017	0.0027	0.0017	0.000086	0.0018	0.0019	0.0018	0.000091

References

Akritas, M. G. (2000). The central limit theorem under censoring. *Bernoulli* **6**, 1109–1120.

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.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

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

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

Groeneboom, P. (1998). *Special topics course 593C: Nonparametric Estimation for Inverse Problems: Algorithms and Asymptotics*. Technical Report 344, Department of Statistics, University of Washington, Seattle, USA.

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

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.

Luenberger, D.G. (1969). *Optimization by Vector Space Methods*. John Wiley and Sons, New York.

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

Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422 - 439.

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. (1991). On differentiable functionals. *Ann. Statist.* **19**, 178 - 204.

Wright, S.G. (1997). *Primal-Dual Interior Point Methods*. SIAM, Philadelphia.

Departamento de Estatística
Universidade Federal de Minas Gerais
Caixa Postal 702 - CEP 30123-970
Belo Horizonte-MG, Brazil
e-mail: aegomes@est.ufmg.br

Delft University of Technology
Department of Mathematics
Faculty of Information Technology and Systems
Mekelweg 4, 2628 CD Delft
The Netherlands
e-mail: p.groeneboom@twi.tudelft.nl

Department of Statistics
University of Washington
P.O. Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu