ISSN: 1935-7524

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

Antonio Eduardo Gomes

Universidade de Brasília, Departamento de Estatística, 70910-900, Brasília-DF, Brazil
e-mail: aegomes@unb.br

Piet Groeneboom

Delft University of Technology, Building 28, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands.
e-mail: P.Groeneboom@tudelft.nl

Jon A. Wellner

University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195-4322, USA
e-mail: jaw@stat.washington.edu

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".

[20] (generally 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 the MLE of F_1 was not computed correctly, and that the (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. R scripts are provided for computing the estimates either with the primal-dual interior point method or by the EM algorithm.

In spite of the long history of the model in the biometrics literature (since about 1982), basic properties of the real maximum likelihood estimator (MLE) were still unknown. We give necessary and sufficient conditions for the MLE (Theorem 3.1), as an element of a cone, where the number of generators of the cone increases quadratically with sample size. From this and a self -consistency equation, turned into a Volterra integral equation, we derive the consistency of the MLE (Theorem 4.1). We conjecture that (under some natural conditions) one can extend the methods, used to prove consistency, to proving that the MLE is $\sqrt{n}$ consistent for F_2 and cube root n convergent for F_1, but this has presently not yet been proved.

AMS 2000 subject classifications: Primary 62G09, 62N01.
Keywords and phrases: MLE, survival sacrifice model, self-consistency equation, Volterra integral equation, primal-dual interior point algorithm, EM algorithm, smooth functionals.

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

$$X \equiv (C \wedge T_2, 1_{[T_1 \le C]}, 1_{[T_2 \le C]}) \equiv (Y, \Delta_1, \Delta_2)\,. \tag{1.1}$$

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

$$p(y,\delta_1,\delta_2) = \begin{cases} (1-F_1(y))g(y), & \text{if } (\delta_1,\delta_2)=(0,0)\,, \\ (F_1(y)-F_2(y))g(y), & \text{if } (\delta_1,\delta_2)=(1,0)\,, \\ (1-G(y))f_2(y), & \text{if } (\delta_1,\delta_2)=(1,1)\,. \end{cases} \tag{1.2}$$

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. [4],[14],[19] and, more recently, [20].

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 $F_1(x) \ge F_2(x)$ 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.

[14] 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.

[19] proposed an EM algorithm for the joint estimation of F_1 and F_2 which converges very slowly to the MLE of (F_1, F_2) (provided the support of the initial estimator contains the support of the MLE). We discuss an implementation of this method in Section 5.2, and give an `R` script for this implementation in [10].

A more efficient way of computing the full MLEs is given by the primal-dual interior point algorithm, as, for example, also discussed in [15]. The latter authors first transform their model before applying the algorithm, but such a transformation is not needed in our present approach for the model considered here. Our implementation is discussed in Section 5.1 and an R script for the implementation is also given in [10].

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

$$\sum_{i=1}^{n} \left(1-\Delta_{2,(i)}\right) \left[\Delta_{1,(i)} \log(x_i - \hat{F}_{2,KM}(Y_{(i)})) + \left(1-\Delta_{1,(i)}\right) \log\left(1-x_i\right)\right] \tag{1.3}$$

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 (1.3) 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 $\Phi(y) = (y-F_2)\log(y-F_2) + (1-y)\log(1-y)$, [4] concluded that the values of $F_1(Y_{(i)}), i = 1,\ldots,n$, maximizing the log-likelihood (1.3) could be obtained applying theorem 1.10 in [2], i.e., the pseudo MLE 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 [2] 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 MLE $\hat{F}_{1,n}$ 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 $\Delta_{2,i} = 0$.

[6] established uniform consistency of the pseudo - likelihood estimator $\hat{F}_{1,n}^{pseudo}$ obtained by maximizing (1.3) over the polyhedron $\{\mathbf{x}:\ 0 \le x_1 \le \cdots \le x_n \le 1\}$ under the assumption that F_1 and F_2 are continuous distribution functions and that $P_{F_1} \prec P_G$. Recently [17] established a global rate of convergence result for the pseudo-likelihood estimator of F_1 for a slightly different model under reasonable assumptions concerning a preliminary estimator $\widetilde{F}_{2,n}$ of F_2.

For the Weighted Least Squares estimator proposed by [20], [7] established uniform consistency under the same assumptions used to study the pseudo-likelihood estimator in [6]. [7] also established an asymptotic minimax lower bound for estimation of $F_1(t_0)$ under the following assumptions:
(i) $0 < F_2(t_0) < F_1(t_0) < 1$.
(ii) F_1 and G are continuously differentiable at t_0 with derivatives $f_1(t_0) > 0$ and $g(t_0) > 0$.
Specifically, [7] used Lemma 4.1 of Groeneboom (1996) (see also Theorem 6.1 of [11]) to show that

$$\begin{aligned} & n^{1/3} \inf_{\hat{F}_{1,n}} \max\left\{E_{n,p_0}|\hat{F}_{1,n}(t_0) - F_1(t_0)|, E_{n,p_n}|\hat{F}_{1,n}(t_0) - F_{1,n}(t_0)|\right\} \\ \ge \quad & \frac{1}{4} n^{1/3} |F_{1,n}(t_0) - F_1(t_0)| \{1 - H^2(p_n,p_0)\}^{2n} \\ \to \quad & \frac{1}{4} f_1(t_0) t \exp\left(-\frac{2S_2(t_0)g(t_0)f_1^2(t_0)t^3}{S_1(t_0)[S_2(t_0)-S_1(t_0)]}\right) \end{aligned} \tag{1.4}$$

and the maximum value of this last expression (with respect to $t > 0$) is

$$k\left\{\frac{f_1(t_0)S_1(t_0)[S_2(t_0)-S_1(t_0)]}{S_2(t_0)g(t_0)}\right\}^{1/3} \equiv kC(S_1,S_2,f_1,g,t_0) \tag{1.5}$$

where $k = (1/4)(3e/2)^{-1/3}$ does not depend on F_1, F_2, f_1, or g. Here $p_0(y,\delta_1,\delta_2; F_1,F_2,G)$ is given by (1.2) and p_n is given by $p_n(y,\delta_1,\delta_2; F_{1,n},F_2,G)$ where $F_{1,n}(x) = F_1(x)$ if $x \in [t_0 - n^{-1/3}t, t_0 + n^{-1/3}t)^c$ and

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

Note that the dependence on n in p_n is just through $F_1 = F_{1,n}$ and not via F_2 or G.

[7] also showed that the WLS estimator achieves this bound in the sense that (under slightly stronger conditions than those imposed in the minimax lower bound),

$$n^{1/3}(\widehat{S}_{1,n}^{WLS}(t_0) - S_1(t_0)) \to_d 2^{-1/3} C(S_1, S_2, f_1, g) 2Z$$

where $Z = \text{argmax}\{B(t) - t^2\}$ where B is two-sided Brownian motion starting at 0.

Thus we see that virtually all of the progress to date concerning estimation of F_1 and F_2 in the model given by (1.2) concerns either the pseudo MLE of F_1 or the Weighted Least Squares estimator $\hat{F}_{1,n}$.

The current challenge is to understand the behavior of the (real, joint) MLE. Here we establish consistency of the sequence of MLEs $\{(\widehat{F}_{n,1}, \widehat{F}_{n,2}):\ n \geq 1\}$. We also give some new lower bounds for smooth functionals of the distribution functions, which partially explain why we expect that the full MLE performs better than the pseudo MLE or the WLS estimators for the distribution function F_1 and also better than the Kaplan-Meier estimator of the distribution function F_2, based on the marginal likelihood, with the advantage of the real MLE appearing mostly when the distribution functions F_1 and F_2 are nearly equal.

The plan of the paper is as follows. In Section 2 we explain the difficulties in the approach of [20] (LJP 1997). In Section 3 we give the characterization of the (real) MLE as the solution of a maximization on a cone and derive from this a self-consistency equation in Corollary 3.3 which is used to prove consistency in Section 4. In Section 5 we describe the primal-dual interior point method and the EM algorithm for computing the MLE. We also give in subsection 5.3 a comparison of the behavior of primal-dual interior point method and the EM algorithm. These results can be reproduced by running the `R` scripts for the algorithms in [10]. Results and examples for the estimation of smooth functionals are given in Section 6. Quantile estimation is considered in Section 7 and we end by some concluding remarks in Section 8.

2. A Weighted Least Squares Estimator

The main point of the present section is that the MLE is *not* the reweighted least squares estimator as introduced by [20] (LJP 1997), because that estimator is not well-defined. Their (not iteratively defined) weighted least squares estimator, claimed to be superior to the MLE, can actually *coincide* with the MLE. One might perhaps hope that maximizing the likelihood, ignoring the constraint $F_1 \geq F_2$ (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 MLE, even if we start the procedure with positive masses at each point, as recommended in [20]. Use of Lagrange multipliers or methods using Section 3 below is the only way out here.

Indeed, a possibility for estimation of F_1 is to calculate a weighted least squares estimator as suggested by [20]. 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

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

So, it is possible to rewrite

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

Estimating S_1 can be viewed, then, as a regression of $S_2(C)\,(1-\Delta_1)$ 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

$$\frac{1}{n}\sum_{i=1}^{n}\Big[\big(1-\Delta_{1(i)}\big)\,\hat S_{2,KM}(Y_{(i)}) - S_1(Y_{(i)})\Big]^2\big(1-\Delta_{2(i)}\big)$$

under the constraint that S_1 is nonincreasing. This minimization problem can be solved by using results from the theory of isotonic regression (see [2]) and is given by

$$\hat S_1(Y_{(m)}) = \min_{l\le m}\max_{k\ge m}\frac{\sum_{j=l}^{k}\hat S_{2,KM}(Y_{(j)})\,\big(1-\Delta_{1(j)}\big)}{\sum_{j=l}^{k}\big(1-\Delta_{2(j)}\big)},\quad m=1,\ldots,n\,.$$

However,

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

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 $S_2^2(c)R(c)\,(1-R(c))$. 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

$$\hat S_1(Y_{(m)}) = \min_{l\le m}\max_{k\ge m}\frac{\sum_{j=l}^{k}\hat S_{2,KM}(Y_{(j)})\,\big[1-\Delta_{1(j)}\big]\left(\frac{1-\Delta_{2(j)}}{\hat S^2_{2,KM}(Y_{(j)})R_n(Y_{(j)})\big[1-R_n(Y_{(j)})\big]}\right)}{\sum_{j=l}^{k}\left(\frac{1-\Delta_{2(j)}}{\hat S^2_{2,KM}(Y_{(j)})R_n(Y_{(j)})\big[1-R_n(Y_{(j)})\big]}\right)} \tag{2.1}$$

for $m=1,\ldots,n$, where $R_n(Y_{(j)}) = S_{1n}(Y_{(j)})/S_{2,KM}(Y_{(j)})$ is also iteratively updated.

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 $t\in\big(Y_{(i-1)},Y_{(i)}\big]$. [20] claim that the estimator above is more efficient than the MLE 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

$$F_{2n}(t) = \frac{\sum_{i=1}^{n}\frac{\int_{C_i}^{t} dF_{2n}}{(S_{2n}-S_{1n})(C_i)}1_{\{\delta_i=(1,0)\}} + 1_{T_{2,i}\le t}1_{\{\delta_i=(1,1)\}}}{\sum_{i=1}^{n}1_{\{\delta_i=(1,1)\}} + \frac{S_{2n}(C_i)}{(S_{2n}-S_{1n})(C_i)}1_{\{\delta_i=(1,0)\}}}, \tag{2.2}$$

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 [20] 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 MLE 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 MLE is in fact *undefined* at crucial points.

The trouble with this approach is that the Lagrangian terms in deriving the "score equations" on page

544 of their paper is completely neglected:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{1n}(C_i)}\right.\\
\left.-\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)-S_{1n}(C_i)}\right\}=0, \tag{2.3}$$

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 $S_{1n}=1-F_{1n}$:

$$S_{1n}(t)\\
=\frac{1}{n}\sum_{i=1}^{n}\frac{\{\Delta_{1,i}=0,\,\Delta_{2,i}=0,\,C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)R_n(C_i)(1-R_n(C_i))}\Bigg/\sum_{i=1}^{n}\frac{\{\Delta_{2,i}=0,\,C_i\in(t_j,t_{j+1}]\}}{S_{2n}^2(C_i)R_n(C_i)(1-R_n(C_i))}\,, \tag{2.4}$$

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 5.1), 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 MLE 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 MLE 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 (2.3) is *not* satisfied. In fact, the sum on the left-hand side of (2.3) was precisely the value of the Lagrange multiplier divided by n, missing in (2.3)! And, as noted above, the expression on the right-hand side of (2.4) is undefined in this situation, because of the fact that the denominators become zero if the constraint is active.

We close this section by showing the Kaplan-Meier estimator and the (non-iterative) weighted least squares estimator introduced in [20] for a real data set studied by [4] and [19] 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 [20] above is smaller than that obtained at the true MLE of F_1 and F_2 obtained through the Primal-Dual Interior Point algorithm (-262.7964 and -262.5468, respectively).

3. Characterization of the MLE

In this section we derive the so-called Fenchel duality conditions for the (real) MLE. Our problem is to maximize the function

$$\mathcal{L}(\mathbf{x},\mathbf{y})=\sum_{i=1}^{n}\Big\{(1-\delta_{1,(i)})(1-\delta_{2,(i)})\log(1-x_i)+\delta_{1,(i)}(1-\delta_{2,(i)})\log(x_i-y_i)\\
+\delta_{1,(i)}\delta_{2,(i)}\log(y_i-y_{i-1})\Big\}, \tag{3.1}$$

where $\delta_{1,(i)}$ and $\delta_{2,(i)}$ correspond to the ith order statistics of the observations, and where the vectors $\mathbf{x}=(x_1,\dots,x_n)$ and $\mathbf{y}=(y_1,\dots,y_n)$ have to satisfy $0\le x_1\le\dots\le x_n\le 1$, $0\le y_1\le\dots\le y_n\le 1$, and

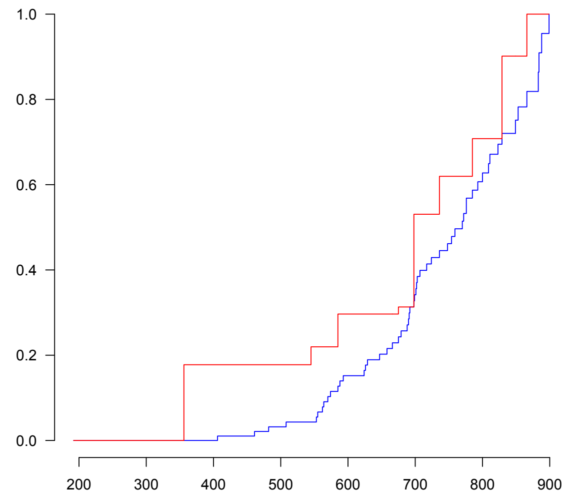


Fig 1: Weighted Least Square estimate of F_1 and Kaplan-Meier estimate of F_2.

$x_i \geq y_i$, for $i = 1, \ldots, n$, i.e., vector $\mathbf{x}$ contains the values of F_1 and vector $\mathbf{y}$ contains the values of F_2 as their components. This corresponds to the "full" maximum likelihood estimators of the lifetime and disease onset distributions for the model, considered in [4] and LJP (1997).

To allow for ties in the observations, we set $\mathbf{z} = (x_1, \ldots, x_m, y_1, \ldots, y_m)$, where m is the number of distinct points, and we define $\phi(\mathbf{z})$ by:

$$\phi(\mathbf{z}) = -\sum_{i=1}^{m} \left\{ f_{i1} \log(1 - x_i) + f_{i2} \log(x_i - y_i) + f_{i3} \log(y_i - y_{i-1}) \right\}, \tag{3.2}$$

where $\sum_{i=1}^{m} \sum_{j=1}^{3} f_{ij} = n$, and where f_{i1} is the frequency of the observations where both delta's are zero, f_{i2} the frequency of observations where the first delta equals 1 and the second zero, and f_{i3} the frequency of observations where both delta's are equal to 1. Note that (3.2) is the same as (3.1), but with a changed sign, so we have minimization problem instead of a maximization problem.

We are going to reduce the minimization problem to the problem of minimizing on a cone $\mathcal{C}$.

Definition 3.1. The cone $\mathcal{C}$ is the set of vectors $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$ such that $0 \leq x_1 \leq \cdots \leq x_m$, $0 \leq y_1 \leq \cdots \leq y_m$ and $x_i \geq y_i$, $i = 1, \ldots, m$. This is a polyhedral convex cone, with generators $(\sum_{i=1}^{m} \boldsymbol{e}_i, \mathbf{0}), (\sum_{i=2}^{m} \boldsymbol{e}_i, \mathbf{0}), \ldots, (\boldsymbol{e}_m, \mathbf{0})$ and $(\sum_{j=i}^{m} \boldsymbol{e}_j, \sum_{j=k}^{m} \boldsymbol{e}_j)$ for $1 \leq i \leq k \leq m$, where the $\boldsymbol{e}_i$ and $\mathbf{0}$ are the unit vectors and the vector of zeros in $\mathbb{R}^m$, respectively. Note that the size of the set of generators grows quadratically with m.

Before being able to reduce the problem to a problem of minimizing on the cone $\mathcal{C}$, we have to remove the restriction $x_m \leq 1$. We do this by introducing Lagrange terms to the criterion.

Definition 3.2. The modified criterion for the minimization problem is:

$$\phi_{\boldsymbol{\lambda}}(\boldsymbol{z}) = -\sum_{i=1}^{m} \left\{ f_{i1} \log(1 - x_i) + f_{i2} \log(x_i - y_i) + f_{i3} \log(y_i - y_{i-1}) \right\} + \lambda_1(x_m - 1) + \lambda_2(y_m - 1), \tag{3.3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, and where the Lagrange multipliers λ_1 and λ_2 are nonnegative. If the constraints at the right end are not active, the Lagrange multipliers are equal to zero. We have at the solution:

$$\lambda_1 = -\sum_{i:x_i=1} \frac{\partial}{\partial x_i}\phi(\boldsymbol{z}), \qquad \lambda_2 = -\sum_{i:y_i=1} \frac{\partial}{\partial y_i}\phi(\boldsymbol{z}),$$

where ϕ is defined by (3.2). Note that the Lagrange multipliers, as just defined, are zero if the constraints are not active. Furthermore,

(i) If $f_{i1} > 0$, $f_{i2} = f_{i3} = 0$, for $i = 1, \ldots, k-1$, while $f_{k2} > 0$ or $f_{k3} > 0$, we set $x_i = y_i = 0$, $i < k$, and define

$$\frac{\partial \phi_{\boldsymbol{\lambda}}(\boldsymbol{z})}{\partial x_i} = \frac{\partial \phi_{\boldsymbol{\lambda}}(\boldsymbol{z})}{\partial y_i} = 0, \qquad i < k. \tag{3.4}$$

(ii) If (i) occurs and $f_{k2} > 0$, we set $y_k = 0$.

We disregard the observations with index i for $i < k$ in condition (i) from the minimization problem. Note that (i) and (ii) make the criterion as small as possible without influencing the remaining minimization problem.

We define $x_0 = y_0 = 0$, and, apart from the definitions in Definition 3.2, we define

$$\frac{\partial \phi_{\boldsymbol{\lambda}}(\boldsymbol{z})}{\partial x_i} = \begin{cases} \dfrac{f_{i1}}{1-x_i} - \dfrac{f_{i2}}{x_i - y_i} + \lambda_1\,, & i = m \\[2ex] \dfrac{f_{i1}}{1-x_i} - \dfrac{f_{i2}}{x_i - y_i}\,, & i < m, \end{cases}$$

and

$$\frac{\partial \phi_{\boldsymbol{\lambda}}(\boldsymbol{z})}{\partial y_i} = \begin{cases} \dfrac{f_{i2}}{x_i - y_i} - \dfrac{f_{i3}}{y_i - y_{i-1}} + \lambda_2\,, & i = m \\[2ex] \dfrac{f_{i2}}{x_i - y_i} - \dfrac{f_{i3}}{y_i - y_{i-1}} + \dfrac{f_{i+1,3}}{y_{i+1} - y_i}\,, & i < m, \end{cases}$$

where the nonnegative Lagrange multipliers λ_1 and λ_2 are defined in Definition 3.2. The following lemma from [8] will be used.

Lemma 3.1 (Fenchel duality Lemma 2.1, section 2.3.1 of [8])**.** *Let $\phi : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be a continuous concave function. Let $\mathcal{C} \subset \mathbb{R}^m$ be a convex cone, and let $\mathcal{C}_0 = \mathcal{C} \cap \phi^{-1}(\mathbb{R})$. Suppose that $\mathcal{C}_0$ is nonempty and that ϕ is differentiable on $\mathcal{C}_0$. Then $\hat{\boldsymbol{z}} \in \mathcal{C}_0$ satisfies*

$$\phi(\hat{\boldsymbol{z}}) = \min_{z \in \mathcal{C}} \phi(\boldsymbol{z}),$$

if and only if

$$\langle \boldsymbol{z}, \nabla\phi(\hat{\boldsymbol{z}})\rangle \ge 0, \qquad \text{for all } \boldsymbol{z} \in \mathcal{C}, \tag{3.5}$$

and

$$\langle \hat{\boldsymbol{z}}, \nabla\phi(\hat{\boldsymbol{z}})\rangle = 0. \tag{3.6}$$

A solution of the minimization problem for ϕ is now characterized in the following theorem.

Theorem 3.1. *Let the function ϕ be defined by (3.2) and the function $\phi_{\boldsymbol{\lambda}}$ by (3.3). Then the vector $\hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ minimizes $\phi(\boldsymbol{z})$ over the set of vectors $(x_1, \ldots, x_m, y_1, \ldots, y_m) : 0 \le y_i \le x_i \le 1\}$ iff the following conditions are satisfied:*

(i)

$$\sum_{j=i}^{m} \frac{\partial \phi(\hat{\boldsymbol{z}})}{\partial \hat{x}_j} + \hat{\lambda}_1 \ge 0, \qquad i = 1, \ldots, m, \tag{3.7}$$

and

$$\sum_{j=i}^{m} \frac{\partial \phi_{\hat{\boldsymbol{\lambda}}}(\hat{\boldsymbol{z}})}{\partial \hat{x}_j} + \sum_{j=k}^{m} \frac{\partial \phi_{\hat{\boldsymbol{\lambda}}}(\hat{\boldsymbol{z}})}{\partial \hat{y}_j} \ge 0, \qquad \text{for } i = 1, \ldots, m,\ k = i, \ldots, m. \tag{3.8}$$

Equality holds in (3.8) if $\hat{x}_i$*, and* $\hat{y}_k$ *are points of increase of* $\hat{x}$ *and* $\hat{y}$*, respectively, and if* $\hat{x}_i > \hat{y}_k$*.*

(ii)

$$\sum_{i=1}^{m} \left\{ \hat{x}_i \frac{\partial \phi_{\hat{\boldsymbol{\lambda}}}(\hat{\boldsymbol{z}})}{\partial \hat{x}_i} + \hat{y}_i \frac{\partial \phi_{\hat{\boldsymbol{\lambda}}}(\hat{\boldsymbol{z}})}{\partial \hat{y}_i} \right\} = 0,$$

where $\hat{\boldsymbol{\lambda}}$ *is defined as in Definition 3.2 for the solution* $\hat{\boldsymbol{z}}$*.*

Proof. By Definition 3.2 we reduced the minimization problem to the minimization problem on the cone $\mathcal{C}$ of Definition 3.1. We now apply Lemma 3.1. Since the inequalities (3.5) only have to be checked on the generators of the cone $\mathcal{C}$, we can use the generators of Definition 3.1. This gives the inequality conditions of part (i). The equality condition (ii) is (3.6) in another notation. □

Corollary 3.1. *Let conditions (i) and (ii) of Definition 3.2 be satisfied and let the points where* $x_i = y_i = 0$ *be excluded from the minimization problem. Let the function* ϕ *be defined by (3.3). Furthermore, let the nonnegative Lagrange multipliers* $\hat{\lambda}_1$ *and* $\hat{\lambda}_2$ *be defined as in Definition 3.2 for the solution* $\hat{\boldsymbol{z}}$*. Then the vector* $\hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ *minimizes* $\phi(\boldsymbol{z})$ *iff the following conditions are satisfied:*

(a)

$$\sum_{j=i}^{m} \left\{ \frac{f_{j1}}{1 - \hat{x}_j} - \frac{f_{j2}}{\hat{x}_j - \hat{y}_j} \right\} + \hat{\lambda}_1 \ge 0, \qquad i = 1, \ldots, m, \tag{3.9}$$

and, for $1 \le i \le k \le m$*,*

$$\sum_{j=i}^{m} \left\{ \frac{f_{j1}}{1 - \hat{x}_j} - \frac{f_{j2}}{\hat{x}_j - \hat{y}_j} \right\} + \sum_{j=k}^{m} \left\{ \frac{f_{j2}}{\hat{x}_j - \hat{y}_j} - \frac{f_{i3}}{\hat{y}_j - \hat{y}_{j-1}} + 1_{\{j<m\}} \frac{f_{j+1,3}}{\hat{y}_{j+1} - \hat{y}_j} \right\} + \hat{\lambda}_1 + \hat{\lambda}_2 \ge 0. \tag{3.10}$$

Moreover, equality holds in (3.10) if (i, k) *corresponds to an* $\hat{x}_i$ *and an* $\hat{y}_k$ *such that* $\hat{x}_i$ *and* $\hat{y}_k$ *are points of increase of* $\hat{\boldsymbol{x}}$ *and* $\hat{\boldsymbol{y}}$*, respectively, and if* $\hat{x}_i > \hat{y}_k$*.*

(b)

$$\sum_{i=1}^{m} \hat{x}_i \left\{ \frac{f_{i1}}{1 - \hat{x}_i} - \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} \right\} + \sum_{i=1}^{m} \hat{y}_i \left\{ \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} - \frac{f_{i3}}{\hat{y}_i - \hat{y}_{i-1}} + 1_{\{i<m\}} \frac{f_{i+1,3}}{\hat{y}_{i+1} - \hat{y}_i} \right\} + \hat{\lambda}_1 + \hat{\lambda}_2 = 0. \tag{3.11}$$

(c)

$$\sum_{i=1}^{m} \left\{ \frac{f_{i1}}{1 - \hat{x}_i} - \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} \right\} + \sum_{i=1}^{m} \left\{ \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} - \frac{f_{i3}}{\hat{y}_i - \hat{y}_{i-1}} + 1_{\{i<m\}} \frac{f_{i+1,3}}{\hat{y}_{i+1} - \hat{y}_i} \right\} + \hat{\lambda}_1 + \hat{\lambda}_2 = n. \tag{3.12}$$

Proof. Parts (a) and (b) are straightforward consequences of Theorem 3.1. As to part (c), the expression on the left of (3.12) is equal to:

$$\sum_{i=1}^{m} \frac{f_{i1}}{1-\hat{x}_i} + \hat{\lambda}_1 + \hat{\lambda}_2.$$

We can write:

$$\sum_{i=1}^{m} \frac{f_{i1}}{1-\hat{x}_i} = \sum_{i=1}^{m} \frac{f_{i1}\{1-\hat{x}_i\}}{1-\hat{x}_i} + \sum_{i=1}^{m} \frac{\hat{x}_i f_{i1}}{1-\hat{x}_i} = \sum_{i=1}^{m} f_{i1} + \sum_{i=1}^{m} \frac{\hat{x}_i f_{i1}}{1-\hat{x}_i}\,.$$

So we get:

$$\begin{aligned}
\sum_{i=1}^{m} \frac{f_{i1}}{1-\hat{x}_i} &= \sum_{i=1}^{m} f_{i1} + \sum_{i=1}^{m} \frac{\hat{x}_i f_{i1}}{1-\hat{x}_i} \\
&= \sum_{i=1}^{m} f_{i1} + \sum_{i=1}^{m} f_{i2} + \sum_{i=1}^{m} \hat{x}_i \left\{ \frac{f_{i1}}{1-\hat{x}_i} - \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} \right\} + \sum_{i=1}^{m} \frac{\hat{y}_i f_{i2}}{\hat{x}_i - \hat{y}_i} \\
&= \sum_{i=1}^{m} f_{i1} + \sum_{i=1}^{m} f_{i2} + \sum_{i=1}^{m} f_{i3} + \sum_{i=1}^{m} \hat{x}_i \left\{ \frac{f_{i1}}{1-\hat{x}_i} - \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} \right\} \\
&\qquad + \sum_{i=1}^{m} \hat{y}_i \left\{ \frac{f_{i2}}{\hat{x}_i - \hat{y}_i} - \frac{f_{i3}}{\hat{y}_i - \hat{y}_{i-1}} + 1_{\{i<m\}} \frac{f_{i+1,3}}{\hat{y}_{i+1} - \hat{y}_i} \right\},
\end{aligned}$$

using

$$\sum_{i=1}^{m} \hat{y}_i \left\{ \frac{f_{i3}}{\hat{y}_i - \hat{y}_{i-1}} - 1_{\{i<m\}} \frac{f_{i+1,3}}{\hat{y}_{i+1} - \hat{y}_i} \right\} = \sum_{i=1}^{m} f_{i3}.$$

Since

$$\sum_{i=1}^{m} f_{i1} + \sum_{i=1}^{m} f_{i2} + \sum_{i=1}^{m} f_{i3} = n,$$

the result now follows from part (b). □

The convergence criterion of Corollary 3.1 was used to check the convergence of the EM algorithm. Since the EM algorithm has such slow convergence, differences between successive iterations are not the right criterion for checking convergence. The primal-dual interior point algorithm uses another criterion, directly based on the gradient with Lagrange multipliers (see Section 5.1), but it was independently checked that it indeed satisfies the conditions of Corollary 3.1 with high accuracy (10^{-10}).

We also have:

Corollary 3.2. *Let conditions of Corollary 3.1 be satisfied and let the nonnegative Lagrange multipliers* $\hat{\lambda}_1$ *and* $\hat{\lambda}_2$ *also be defined as in Corollary 3.1. Let* $\hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ *minimize* $\phi(\boldsymbol{z})$*. Then:*

(a)

$$\sum_{1\le j<i} \hat{x}_j \left\{ \frac{f_{j1}}{1-\hat{x}_j} - \frac{f_{j2}}{\hat{x}_j - \hat{y}_j} \right\} + \sum_{1\le j<k} \frac{f_{j2}\,\hat{y}_j}{\hat{x}_j - \hat{y}_j} - \sum_{1\le j<k} f_{j3} + \frac{f_{k3}\,\hat{y}_{k-1}}{\hat{y}_k - \hat{y}_{k-1}} = 0, \tag{3.13}$$

if (i,k)*,* $i \le k$*, corresponds to an* $\hat{x}_i$ *and an* $\hat{y}_k$ *such that* $\hat{x}_i$ *and* $\hat{y}_k$ *are points of increase of* $\hat{\boldsymbol{x}}$ *and* $\hat{\boldsymbol{y}}$*, respectively, and if* $\hat{x}_i > \hat{y}_k$*, that is: we are in the interior of the cone.*

(b)

$$\sum_{1\le j<i}\frac{f_{j1}}{1-\hat x_j}+\sum_{i\le j<k}\frac{f_{j2}}{\hat x_j-\hat y_j}+\frac{f_{k,3}}{\hat y_k-\hat y_{k-1}}=n, \tag{3.14}$$

for indices (i,k) as in part (a).

Proof. The equality relations in (3.10), together with part (b) of Corollary 3.1 imply:

$$\sum_{1\le j<i}\hat x_j\left\{\frac{f_{j1}}{1-\hat x_j}-\frac{f_{j2}}{\hat x_j-\hat y_j}\right\}+\sum_{1\le j<k}\hat y_j\left\{\frac{f_{j2}}{\hat x_j-\hat y_j}-\frac{f_{i3}}{\hat y_j-\hat y_{j-1}}+\frac{f_{j+1,3}}{\hat y_{j+1}-\hat y_j}\right\}=0,$$

if (i,k) corresponds to an $\hat x_i$ and an $\hat y_k$ such that $\hat x_i$ and $\hat y_k$ are points of increase of $\hat{\boldsymbol x}$ and $\hat{\boldsymbol y}$, respectively, and if $\hat x_i>\hat y_k$. The second sum can be written:

$$\begin{aligned}&\sum_{1\le j<k}\hat y_j\left\{\frac{f_{j2}}{\hat x_j-\hat y_j}-\frac{f_{j3}}{\hat y_j-\hat y_{j-1}}+\frac{f_{j+1,3}}{\hat y_{j+1}-\hat y_j}\right\}\\&=\sum_{1\le j<k}\frac{f_{j2}\,\hat y_j}{\hat x_j-\hat y_j}-\sum_{1\le j<k}\{\hat y_j-\hat y_{j-1}\}\sum_{j\le l<k}\left\{\frac{f_{l3}}{\hat y_l-\hat y_{l-1}}-\frac{f_{l+1,3}}{\hat y_{l+1}-\hat y_l}\right\}\\&=\sum_{1\le j<k}\frac{f_{j2}\,\hat y_j}{\hat x_j-\hat y_j}-\sum_{1\le j<k}f_{j3}+\frac{f_{k3}\,\hat y_{k-1}}{\hat y_k-\hat y_{k-1}}\,.\end{aligned}$$

The equality relations in (3.10), together with part (c) of Corollary 3.1 imply:

$$\sum_{1\le j<i}\left\{\frac{f_{j1}}{1-\hat x_j}-\frac{f_{j2}}{\hat x_j-\hat y_j}\right\}+\sum_{1\le j<k}\left\{\frac{f_{j2}}{\hat x_j-\hat y_j}-\frac{f_{i3}}{\hat y_j-\hat y_{j-1}}+\frac{f_{j+1,3}}{\hat y_{j+1}-\hat y_j}\right\}=n,$$

if (i,k) corresponds to an $\hat x_i$ and an $\hat y_k$ such that $\hat x_i$ and $\hat y_k$ are points of increase of $\hat{\boldsymbol x}$ and $\hat{\boldsymbol y}$, respectively, and if $\hat x_i>\hat y_k$, which in turn implies, by telescoping sums:

$$\sum_{1\le j<i}\frac{f_{j1}}{1-\hat x_j}+\sum_{i\le j<k}\frac{f_{j2}}{\hat x_j-\hat y_j}+\frac{f_{k,3}}{\hat y_k-\hat y_{k-1}}=n,$$

for such points. □

Example 3.1. Consider the following very simple example of a sample, generated by the model used in the simulations in Section 5.3:

$$(0.76,1,1),(0.86,0,0),(1.34,1,0),(1.67,1,1),(2.32,1,0).$$

Here $n=m$ (no ties) and the MLE is given by

$$\hat F_{n1}(x)=\begin{cases}1/5, & x<1.34\\ 1, & \text{otherwise,}\end{cases}$$

and

$$\hat F_{n2}(x)=\begin{cases}1/5, & x<1.67\\ 3/5, & \text{otherwise,}\end{cases}$$

The pair $(t,u)=(1.34,1.67)$ gives a pair of points, as described in Corollary 3.2, and it can be verified that $\hat F_{n1}(t)=1>\hat F_{n2}(u)=3/5$, and that (3.14) is satisfied, since for $(i,k)=(3,4)$:

$$\sum_{1\le j<i}\frac{f_{j1}}{1-\hat x_j}+\sum_{i\le j<k}\frac{f_{j2}}{\hat x_j-\hat y_j}+\frac{f_{k,3}}{\hat y_k-\hat y_{k-1}}=\frac54+\frac54+\frac52=5=n.$$

In the present case:

$$\hat{\lambda}_1 = \frac{15}{4}, \qquad \hat{\lambda}_2 = 0.$$

To verify part (c) of Corollary 3.1 we compute:

$$\sum_{i=1}^{n} \frac{f_{i1}}{1-\hat{x}_i} + \hat{\lambda}_1 + \hat{\lambda}_2 = \frac{5}{4} + \frac{15}{4} + 0 = 5.$$

In empirical process notation, (3.13) means:

$$\begin{aligned}
&\int_{v<t} \frac{(1-\delta_1)(1-\delta_2)}{1-\hat{F}_{n1}(v)}\,\hat{F}_{n1}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2) - \int_{v<t} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,\hat{F}_{n1}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2)\\
&\quad + \int_{v<u} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,\hat{F}_{n2}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2) - \int_{v<u} \delta_1\delta_2\,d\mathbb{P}_n(v,\delta_1,d_2)\\
&\qquad + \frac{f_{u3}\,\hat{F}_{n2}(u-)}{n\{\hat{F}_{n2}(u)-\hat{F}_{n2}(u-)\}} = 0,
\end{aligned}$$

where $t \le u$ and f_{u3} is the number of observations such that $\delta_1 = \delta_2 = 1$ at u, and where t and u are points of increase of $\hat{F}_{n1}$ and $\hat{F}_{n2}$, respectively, satisfying $\hat{F}_{n1}(t) > \hat{F}_{n2}(u)$. Relation (3.14) says:

$$\begin{aligned}
&\frac{f_{u3}}{n\{\hat{F}_{n2}(u)-\hat{F}_{n2}(u-)\}}\\
&= 1 - \int_{v<t} \frac{(1-\delta_1)(1-\delta_2)}{1-\hat{F}_{n1}(v)}\,d\mathbb{P}_n(v,\delta_1,\delta_2) - \int_{t\le v<u} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,d\mathbb{P}_n(v,\delta_1,\delta_2).
\end{aligned} \tag{3.15}$$

Using this, we get the equation

$$\begin{aligned}
&\left\{1 - \int_{v<t} \frac{(1-\delta_1)(1-\delta_2)}{1-\hat{F}_{n1}(v)}\,d\mathbb{P}_n(v,\delta_1,\delta_2)\right.\\
&\qquad\qquad \left. - \int_{t\le v<u} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,d\mathbb{P}_n(v,\delta_1,\delta_2)\right\}\hat{F}_{n2}(u-)\\
&= -\int_{v<t} \frac{(1-\delta_1)(1-\delta_2)}{1-\hat{F}_{n1}(v)}\,\hat{F}_{n1}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2)\\
&\qquad\qquad + \int_{v<t} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,\hat{F}_{n1}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2)\\
&\quad - \int_{v<u} \frac{\delta_1(1-\delta_2)}{\hat{F}_{n1}(v)-\hat{F}_{n2}(v)}\,\hat{F}_{n2}(v)\,d\mathbb{P}_n(v,\delta_1,\delta_2) + \int_{v<u} \delta_1\delta_2\,d\mathbb{P}_n(v,\delta_1,\delta_2),
\end{aligned} \tag{3.16}$$

where t and u are as described above.

We define the following random measures:

$$dP_n^{(1)} = (1-\delta_1)(1-\delta_2)\,d\mathbb{P}_n, \qquad dP_n^{(2)} = \delta_1(1-\delta_2)\,d\mathbb{P}_n, \qquad dP_n^{(3)} = \delta_1\delta_2\,d\mathbb{P}_n.$$

and similarly the non-random measures

$$\begin{aligned}
&dP^{(1)} = \{1-F_{01}(v)\}\,dG(v), \qquad dP^{(2)} = \{F_{01}(v)-F_{02}(v)\}\,dG(v),\\
&dP^{(3)} = \{1-G(v)\}\,dF_{02}(v).
\end{aligned}$$

We can now write (3.16) in the form

$$\left\{1-\int_{v<t}\frac{1}{1-\hat F_{n1}(v)}\,dP_n^{(1)}-\int_{t\le v<u}\frac{1}{\hat F_{n1}(v)-\hat F_{n2}(v)}\,dP_n^{(2)}\right\}\hat F_{n2}(u-)$$
$$=-\int_{v<t}\frac{\hat F_{n1}(v)}{1-\hat F_{n1}(v)}\,dP_n^{(1)}+\int_{v<t}\frac{\hat F_{n1}(v)}{\hat F_{n1}(v)-\hat F_{n2}(v)}\,dP_n^{(2)}$$
$$-\int_{v<u}\frac{\hat F_{n2}(v)}{\hat F_{n1}(v)-\hat F_{n2}(v)}\,dP_n^{(2)}+\int_{v<u}dP_n^{(3)}$$
$$=-\int_{v<t}\frac{1}{1-\hat F_{n1}(v)}\,dP_n^{(1)}+\int_{v<t}dP_n^{(1)}+\int_{v<t}dP_n^{(2)}$$
$$-\int_{t\le v<u}\frac{\hat F_{n2}(v)}{\hat F_{n1}(v)-\hat F_{n2}(v)}\,dP_n^{(2)}+\int_{v<u}dP_n^{(3)}\,. \tag{3.17}$$

So we get the following self-consistency property.

Corollary 3.3. *Let t and u be points of increase of $\hat F_{n1}$ and $\hat F_{n2}$, respectively, so that $t\le u$ and $\hat F_{n1}(t)>\hat F_{n2}(u)$. Then:*

$$\hat F_{n2}(u-)$$
$$=-\int_{v<t}\frac{1-\hat F_{n2}(u-)}{1-\hat F_{n1}(v)}\,dP_n^{(1)}+\int_{v<t}d\big(P_n^{(1)}+P_n^{(2)}\big)$$
$$+\int_{t\le v<u}\frac{\hat F_{n2}(u-)-\hat F_{n2}(v)}{\hat F_{n1}(v)-\hat F_{n2}(v)}\,dP_n^{(2)}+\int_{v<u}dP_n^{(3)}. \tag{3.18}$$

4. Consistency of the MLE

In this section we derive consistency of the (real) MLE from the Fenchel duality conditions in Section 3. For clarity, we use the notation F_{01} and F_{02} instead of F_1 and F_2, respectively, for the underlying distribution functions and use F_1 and F_2 for limits of sequences of $\hat F_{n1}$ and $\hat F_{n2}$.

Assuming that that $\hat F_{n1}$ and $\hat F_{n2}$ tend along subsequences to limits F_1 and F_2, we get from (3.16) in the limit:

$$\left\{1-\int_{v<t}\frac{1-F_{01}(v)}{1-F_1(v)}\,dG(v)-\int_{t\le v<u}\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,dG(v)\right\}F_2(u)$$
$$=-\int_{v<t}\left\{\frac{1-F_{01}(v)}{1-F_1(v)}\,F_1(v)-\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,F_1(v)\right\}dG(v)$$
$$-\int_{v<u}F_2(v)\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,dG(v)+\int_{v<u}\{1-G(v)\}\,dF_{02}(v),$$

where $t\le u$ and $F_1(t)\ge F_2(u)$.

For proving consistency, we now will use the following lemma.

Lemma 4.1 (Uniqueness lemma)**.** *Let F_{01}, F_{02} and G be distribution functions on $[0,\infty)$, with continuous derivatives f_{01}, f_{02} and g, respectively, and let F_{01} and F_{02} have the same support, contained in the support of g. Moreover, let the set S_0 be defined by:*

$$S_0=\{(t,u):t\le u,\,F_{01}(t)\ge F_{02}(u)\}. \tag{4.1}$$

Assume that for all t in the interior of the support of F_{01} there is a nondegenerate interval of points u such that $(t,u)\in S_0$ and $F_{01}(t)>F_{02}(u)$. Moreover, assume:

$$\mathbb{P}\{\Delta_1=1|T_1=t\}>0\qquad\text{and}\qquad\mathbb{P}\{(\Delta_1,\Delta_2)=(1,1)|T_2=u\}>0,$$

for all (t,u) in the interior of the set S_0. Finally, let the following relation for the nondecreasing right-continuous functions F_1 and F_2 be satisfied for (t,u) in the interior of the set S_0:

$$\begin{aligned}&\left\{1-\int_{v<t}\frac{1-F_{01}(v)}{1-F_1(v)}\,dG(v)-\int_{t\le v<u}\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,dG(v)\right\}F_2(u)\\&=-\int_{v<t}\left\{\frac{1-F_{01}(v)}{1-F_1(v)}\,F_1(v)-\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,F_1(v)\right\}dG(v)\\&\qquad-\int_{v<u}F_2(v)\frac{F_{01}(v)-F_{02}(v)}{F_1(v)-F_2(v)}\,dG(v)+\int_{v<u}\{1-G(v)\}\,dF_{02}(v),\end{aligned}\tag{4.2}$$

where we define $0/0=0$. Then we must have: $F_1\equiv F_{01}$ and $F_2\equiv F_{02}$ on the support of g.

Remark 4.1. Note that our conditions are somewhat similar to those of [12] for the model of doubly censored data. In fact, in their approach also a self-consistency equation is used to derive the results.

Proof of lemma 4.1. Replacing F_{0i} and F_i, $i=1,2$ in (4.2) we get an identity. So we can subtract this identity from (4.2), and defining $H_1=F_{01}-F_1$ and $H_2=F_{02}-F_2$ we then get, for (t,u) in the interior of S_0:

$$\begin{aligned}&\left\{\int_{v<t}\frac{H_1(v)}{1-F_1(v)}\,dG(v)-\int_{t\le v<u}\frac{H_1(v)-H_2(v)}{F_1(v)-F_2(v)}\,dG(v)\right\}F_2(u)\\&=\int_{v<t}\left\{\frac{H_1(v)}{1-F_1(v)}\,F_1(v)+\frac{H_1(v)-H_2(v)}{F_1(v)-F_2(v)}\,F_1(v)\right\}dG(v)\\&\qquad-\int_{v<u}F_2(v)\frac{H_1(v)-H_2(v)}{F_1(v)-F_2(v)}\,dG(v)+\int_{v<u}\{1-G(v)\}\,dH_2(v).\end{aligned}$$

Differentiating w.r.t. t we get:

$$\left\{\frac{H_1(t)}{1-F_1(t)}+\frac{H_1(t)-H_2(t)}{F_1(t)-F_2(t)}\right\}F_2(u)\,g(t)=\left\{\frac{H_1(t)}{1-F_1(t)}+\frac{H_1(t)-H_2(t)}{F_1(t)-F_2(t)}\right\}F_1(t)\,g(t).$$

If $F_1(t)>F_2(u)$ and $g(t)>0$, this can only be true if

$$\frac{H_1(t)}{1-F_1(t)}+\frac{H_1(t)-H_2(t)}{F_1(t)-F_2(t)}=0,$$

implying

$$\frac{H_1(t)-H_2(t)}{F_1(t)-F_2(t)}=-\frac{H_1(t)}{1-F_1(t)}=-\frac{H_2(t)}{1-F_2(t)}.\tag{4.3}$$

Since we can differentiate (4.2) on the right-hand side, and since also the factor of $F_2(u)$ can be differentiated w.r.t. u on the left-hand side, we can differentiate $F_2(u)$ and define its derivative by $f_2(u)$. We get:

$$\begin{aligned}&\left\{\int_{v<t}\frac{H_1(v)}{1-F_1(v)}\,dG(v)-\int_{t\le v<u}\frac{H_1(v)-H_2(v)}{F_1(v)-F_2(v)}\,dG(v)\right\}f_2(u)\\&\qquad\qquad-\frac{H_1(u)-H_2(u)}{F_1(u)-F_2(u)}\,g(u)\,F_2(u)\\&=-F_2(u)\frac{H_1(u)-H_2(u)}{F_1(u)-F_2(u)}\,g(u)+\{1-G(u)\}h_2(u).\end{aligned}$$

Using (4.3), with t replaced by u, yields:

$$\begin{aligned}&\left\{\int_{v<t}\frac{H_2(v)}{1-F_2(v)}\,dG(v)+\int_{t\le v<u}\frac{H_2(v)}{1-F_2(v)}\,dG(v)\right\}f_2(u)+\frac{H_2(u)}{1-F_2(u)}\,g(u)\,F_2(u)\\&=F_2(u)\frac{H_2(u)}{1-F_2(u)}\,g(u)+\{1-G(u)h_2(u).\end{aligned}$$

So we end up with the equation

$$h_2(u) = \{1 - G(u)\}^{-1} \left\{ \int_{v<u} \frac{H_2(v)}{1 - F_2(v)}\, dG(v) \right\} f_2(u).$$

Integrating on both sides yields:

$$H_2(u) = \int_{v<u} \left\{ \int_{v<w<u} \frac{1}{1 - G(w)}\, dF_2(w) \right\} \frac{H_2(v)}{1 - F_2(v)}\, dG(v).$$

This is a homogeneous Volterra equation which can only have the solution $H_2 \equiv 0$. Hence also $H_1 \equiv 0$ (see (4.3)).

□

The uniqueness lemma gives the consistency of the MLE.

Theorem 4.1. *Let the conditions of Lemma 4.1 be satisfied. Then the MLE* $(\hat F_{n1}, \hat F_{n2})$ *of* (F_{01}, F_{02}) *satisfies:*

$$\sup_{t\in[0,\infty)} \left\{ \left|\hat F_{n1}(t) - F_{01}(t)\right| + \left|\hat F_{n2}(t) - F_{02}(t)\right| \right\} \xrightarrow{a.s.} 0, \qquad n \to \infty. \tag{4.4}$$

Proof. By the self-consistency equation (3.16), all limit points of subsequences of $(\hat F_{n1}, \hat F_{n2})$ must satisfy (4.2) for (t, u) in the interior of the set S_0, defined in Lemma 4.1. The result now follows from the uniqueness of the solution of this equation. □

5. The primal-dual interior point algorithm and the EM algorithm for the real MLE

In this section we explain the primal-dual interior point method and the EM algorithm, as for example used in [19], for computing the real maximum likelihood estimate of the pair (F_1, F_2) and compare their behavior. The EM algorithm is vastly inferior to the primal-dual interior point method and leads to prohibitive computing times for bigger samples. One of the reasons for its inferiority is that it needs many iteration steps, but the other more fundamental reason is that it needs a number of parameters that increases quadratically with sample size, whereas the primal-dual interior point method only needs a number of parameters which increases linearly with sample size.

5.1. Primal-Dual Interior Point Algorithm

Instead of using the notation $\boldsymbol{z} = (x_1, \dots, x_m, y_1, \dots, y_m)$ as in (3.2), it is more convenient in this section to define the vector $\boldsymbol{z}$ by $\boldsymbol{z} = (y_1, x_1, \dots, y_m, x_m)$; the relations for the x_i and y_i remain the same. We define the vector $\boldsymbol{g}(\mathbf{z}) = (g_1(\mathbf{z}), \dots, g_{3m}(\mathbf{z}))'$ by

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

and the matrix $\boldsymbol{G}$ by

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

Note that the matrix $\boldsymbol{G}$ does not depend on $\mathbf{z}$. The original maximization problem now becomes a problem of minimizing $\phi(\mathbf{z})$ over $\mathbb{R}^{2m}$ (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

$$\boldsymbol{g}(\mathbf{z}) \quad \leq \quad \mathbf{0}, \ \mathbf{z} \in \mathbb{R}^{2m} \ . \tag{5.1}$$

Theorem 1: Let $\hat{\mathbf{z}} = (\hat{x}_1, \hat{y}_1, \dots, \hat{x}_m, \hat{y}_m)'$ be a vector in $\mathbb{R}^{2m}$ such that $\phi(\hat{\mathbf{z}}) < \infty$. Then $\hat{\mathbf{z}}$ minimizes $\phi(\hat{\mathbf{z}})$ over the set of vectors $\mathbf{z}$, satisfying (5.1), if and only if the following conditions are satisfied:

$$\nabla\phi(\hat{\mathbf{z}}) + \boldsymbol{G}'\boldsymbol{\lambda} \quad = \quad 0 \tag{5.2}$$

$$\boldsymbol{g}(\hat{\mathbf{z}}) + \mathbf{w} \quad = \quad 0 \tag{5.3}$$

$$\langle \boldsymbol{\lambda}, \mathbf{w} \rangle \quad = \quad 0, \tag{5.4}$$

for vectors $\boldsymbol{\lambda}$ and $\boldsymbol{w}$ in $\mathbb{R}_+^{3m}$.

Remark: The vector $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers and $\mathbf{w}$ is called a vector of "slack variables" for the constraints; see, e.g., [22], page 164. Defining the function $\phi_{\boldsymbol{\lambda}}$ by

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

we can write condition (5.2) in the form

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

Note that $\boldsymbol{g}(\hat{\mathbf{z}}) + \mathbf{w} = 0$, for $\mathbf{w} \in \mathbb{R}_+^{3n}$, implies $\boldsymbol{g}(\hat{\mathbf{z}}) \leq 0$, and that $\lambda_i > 0$ implies $w_i = 0$, by (5.4), and hence $g_i(\hat{\mathbf{z}}) = 0$, by (5.3). Thus:

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

The proof of Theorem 1 can be found in [16], 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 $\boldsymbol{\lambda}$ 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 (3.1) 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 $2m$ 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.

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

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

For $\boldsymbol{\lambda}$ and $\mathbf{w}$ we take as starting values $\boldsymbol{\lambda}_0 = \mathbf{w}_0 = 0.5 \cdot \mathbf{1}$, where $\mathbf{1}$ 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 $\boldsymbol{A}$; for example, $\boldsymbol{\Lambda}$ is the diagonal matrix with element λ_i as its ith diagonal element. The first (Newton) iteration step now solves the system of equations:

$$\begin{pmatrix} \nabla_{\mathbf{z}_0\mathbf{z}_0}\phi(\mathbf{z}_0) & \boldsymbol{G}' & 0 \\ \boldsymbol{G} & \mathbf{0} & \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{W}_0 & \boldsymbol{\Lambda}_0 \end{pmatrix} \begin{pmatrix} \mathbf{z} - \mathbf{z}_0 \\ \boldsymbol{\lambda} - \boldsymbol{\lambda}_0 \\ \mathbf{w} - \mathbf{w}_0 \end{pmatrix} = - \begin{pmatrix} \nabla\phi_{\boldsymbol{\lambda}_0}(\mathbf{z}_0) \\ g(\mathbf{z}_0) + \mathbf{w}_0 \\ (\boldsymbol{\Lambda}_0\boldsymbol{W}_0 - \sigma\boldsymbol{\mu}_0)\mathbf{1} \end{pmatrix} \tag{5.5}$$

in $(\mathbf{z}, \mathbf{w}, \boldsymbol{\lambda})$, 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 $\phi_{\boldsymbol{\lambda}}(\mathbf{z})$ w.r.t. $\mathbf{z}$, the so-called *Hessian* of the function $\phi_{\boldsymbol{\lambda}}$ with respect to $\mathbf{z}$. We now define, for fixed $\beta > 0$, the set

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

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

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

By taking $\boldsymbol{\lambda} = \mathbf{w} = 0.5\cdot\mathbf{1}$, we have made this parameter equal to 0.25 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

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

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

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

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

$$\begin{pmatrix} \nabla_{\mathbf{z}_1\mathbf{z}_1}\phi(\mathbf{z}_1) & \boldsymbol{G}' & \mathbf{0} \\ \boldsymbol{G} & \mathbf{0} & \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{W}_1 & \boldsymbol{\Lambda}_1 \end{pmatrix} \begin{pmatrix} \mathbf{z}-\mathbf{z}_1 \\ \boldsymbol{\lambda}-\boldsymbol{\lambda}_1 \\ \mathbf{w}-\mathbf{w}_1 \end{pmatrix} = -\begin{pmatrix} \nabla\phi_{\boldsymbol{\lambda}_1}(\mathbf{z}_1) \\ \boldsymbol{g}(\mathbf{z}_1)+\mathbf{w}_1 \\ (\boldsymbol{\Lambda}_1\boldsymbol{W}_1 - \sigma\boldsymbol{\mu}_1)\,\mathbf{1} \end{pmatrix} \tag{5.6}$$

and find the new $(\mathbf{z}(\alpha), \boldsymbol{\lambda}(\alpha), \mathbf{w}(\alpha))$, required to lie in $\mathcal{N}(\mu_1)$, for this system, denoted by $(\mathbf{z}_2, \boldsymbol{\lambda}_2, \mathbf{w}_2)$, 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

$$\begin{pmatrix} \nabla_{\mathbf{zz}}\phi(\mathbf{z}) & \boldsymbol{G}' & \mathbf{0} \\ \boldsymbol{G} & \mathbf{0} & I \\ \mathbf{0} & \boldsymbol{W} & \boldsymbol{\Lambda} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{z} \\ \Delta\boldsymbol{\lambda} \\ \Delta\mathbf{w} \end{pmatrix} = -\begin{pmatrix} \nabla\phi_{\boldsymbol{\lambda}}(\mathbf{z}) \\ g(\mathbf{z})+\mathbf{w} \\ (\boldsymbol{\Lambda}\boldsymbol{W} - \sigma\mu)\,\mathbf{1} \end{pmatrix} \tag{5.7}$$

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

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

if $(\mathbf{z}_k, \boldsymbol{\lambda}_k, \mathbf{w}_k)$ is the value at the start of the kth iteration, and we solve for $\Delta\mathbf{z}$, $\Delta\boldsymbol{\lambda}$ 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:

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

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

$$\begin{pmatrix} \nabla_{\mathbf{zz}}\phi(\mathbf{z}) & \boldsymbol{G}' \\ \boldsymbol{G} & -\boldsymbol{D}^{-1} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{z} \\ \Delta\boldsymbol{\lambda} \end{pmatrix} = -\begin{pmatrix} \nabla\phi_{\boldsymbol{\lambda}}(\mathbf{z}) \\ \boldsymbol{g}(\mathbf{z}) + \sigma\mu\boldsymbol{\Lambda}^{-1}\mathbf{1} \end{pmatrix}. \tag{5.8}$$

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

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

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

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

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

$$\nabla_{\mathbf{zz}}\phi(\mathbf{z}) + \boldsymbol{G}'\boldsymbol{D}\boldsymbol{G} \ , \tag{5.9}$$

and this matrix is a symmetric positive definite matrix at each step. The matrices $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ are diagonal matrices, so inversion of these is trivial. In our case, the matrix (5.9) 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.

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 Primal-Dual Interior Point algorithm for the estimation of the MLE of F_1 and F_2 was applied to the data set presented in Table 1. Figure 2 shows the estimates. It can be seen that the difference with Figure 1 is rather minor. An `R` script for computing the estimates is given in [10] which uses the `R` package `Rcpp` to connect to the original `C++` code (also given there) for the primal-dual interior point algorithm.

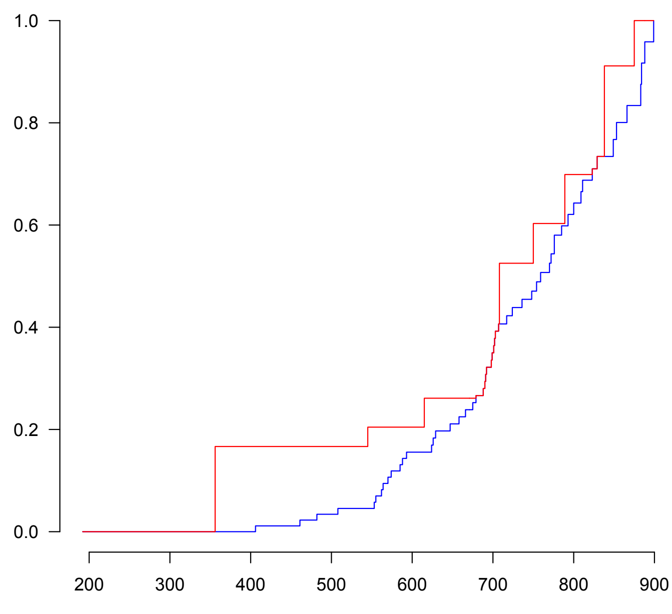


Fig 2: Joint MLE estimates of F_1 and F_2, computed by the primal-dual interior point algorithm.

5.2. EM Algorithm

We discuss here the implementation of the EM algorithm for the survival-sacrifice model in the case of ties. See also [19]. A 2-dimensional discrete distribution in the plane for (T, U) is considered, where T is time of onset of the disease and U is time of death due to the disease. Note that our description is a bit different from the discussion on p. 43 and 44 of [19] and closer to their discussion on p. 45 on the actual implementation.

There are three situations to consider.

(i) $T > Y$ and $U > Y$. In this case mass has to be distributed over points (t_k, u_k) such that $t_k > Y$ and $u_k > Y$.
(ii) $T \le Y$ and $U > Y$. In this case we should have mass points for $T \le Y$ and for U strictly to the right of Y. So mass should be distributed over points (t_k, u_k) with $t_k \le Y$ and $u_k > Y$.
(iii) Our observation is a time of death U. In this case mass should be distributed over points (t_k, u_k) such that $u_k = U$.

Our goal is again to have a distribution of mass on these points so that the likelihood maximized.

We consider the log likelihood

$$\ell(F_1, F_2, f_2) = \sum_{i=1}^{m} \{f_{i1} \log(1 - F_1(y_i)) + f_{i2} \log(F_1(y_i) - F_2(y_i)) \\ + f_{i3} \log\{F_2(y_i) - F_2(y_i-)\}\},$$

where m is the number of strictly different observation times, and where f_{i1} to f_{i3} are the frequencies of the three possibilities, mentioned above, in that order. According to (i) above, we have:

$$1 - F_1(y) = \sum_{k:\, t_k > y,\, u_k > y} p_k, \tag{5.10}$$

Similarly,

$$F_1(y) - F_2(y) = \sum_{k:\, t_k \le y,\, u_k > y} p_k. \tag{5.11}$$

and

$$F_2(y) - F_2(y-) = \sum_{k:\, u_k = y} p_k. \tag{5.12}$$

Since (with added outside points!) we have the side restriction that $\sum_k p_k = 1$, we add a Lagrange term to the log likelihood, by which the criterion to maximize becomes:

$$\ell_\lambda(F_1, F_2) \\ = \sum_{i=1}^{m} \{f_{i1} \log\{1 - F_1(y_i)\} + f_{i2} \log\{F_1(y_i) - F_2(y_i)) \\ + f_{i3} \log\{F_2(y_i) - F_2(y_i-)\}\} - \lambda \sum_k p_k.$$

Standard arguments show that $\lambda = N$, where N is the total number of observations (sum of the frequencies). Differentiation w.r.t. p_k then yields the so-called self-consistency equations. The maximizing density at the point (t_k, u_k) is therefore given by

$$p_k = p_k \frac{1}{n} \sum_{i=1}^{m} \left\{ \frac{\{t_k > y_i,\, u_k > y_i\} f_{i1}}{1 - F_1(y_i)} + \frac{\{t_k \le y_i,\, u_k > y_i\} f_{i2}}{F_1(y_i) - F_2(y_i)} + \frac{\{u_k = y_i\} f_{i3}}{F_2(y_i) - F_2(y_i-)} \right\}, \tag{5.13}$$

with the convention $0/0 = 0$, and the EM iteration will be:

$$p_k^{new} = p_k^{old}\frac{1}{n}\sum_{i=1}^{m}\left\{\frac{\{t_k > y_i,\, u_k > y_i\}f_{i1}}{1-F_1^{old}(y_i)} + \frac{\{t_k \le y_i,\, u_k > y_i\}f_{i2}}{F_1^{old}(y_i)-F_2^{old}(y_i)} + \frac{\{u_k = y_i\}f_{i3}}{\{F_2^{old}(y_i)-F_2^{old}(y_i-)\}}\right\}.$$

After having computed the values $p^{new}(t_k, u_k)$ we can update F_1, F_2 and f_2 via (5.10) to (5.12). This is done in the order:

1.

$$F_1^{new}(y) = 1 - \sum_{k:\, t_k > y,\, u_k > y} p_k,$$

2.

$$F_2^{new}(y) = F_1^{new}(y) - \sum_{k:\, t_k \le y,\, u_k > y} p_k$$

After this the iteration is repeated until some convergence criterion is reached.

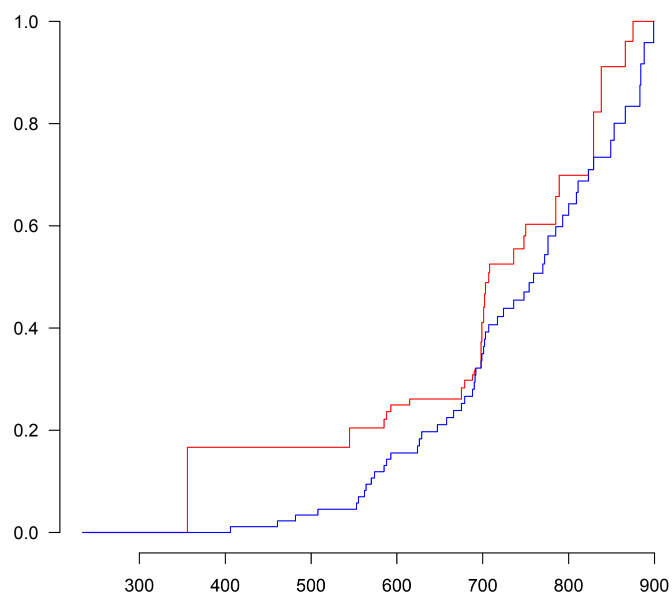


Fig 3: Joint MLE estimates of F_1 and F_2, computed by the EM algorithm

We ran the EM algorithm for the data set presented in Table 1, taken from [19], until the Fenchel duality criteria (i) and (ii) of Theorem 3.1 were satisfied at accuracy 10^{-10}, after having sorted the observations in strict order of magnitude, correcting for ties. This took 3139 iterations. After the correction for ties, there are 102 strictly different observation times (the same correction for ties was applied for the primal-dual algorithm) on a total of 109 observations. The Lagrange multipliers of Theorem 3.1, divided by sample size, were given by $\lambda_1 = 0.055214$ and $\lambda_2 = 0.220856$. As in the case of the primal-dual interior point algorithm, the implementation is given in [10].

5.3. Simulation results and computing times for the primal-dual interior point algorithm and the EM algorithm

We consider simulation of the example in Section 6.1 below:

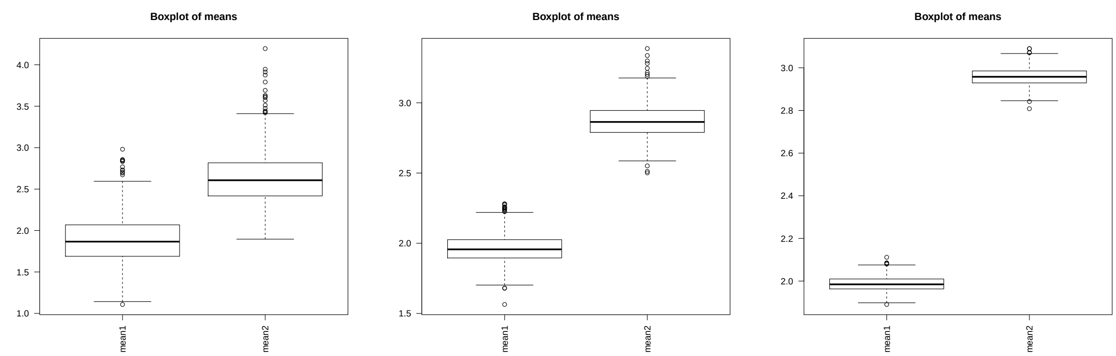
$$T_1 \sim \exp(0.5), \quad T_2 - T_1 \sim \exp(1), \quad C \sim \exp(0.4),$$

i.e.,

$$F_1(t) = 1 - e^{-t/2}, \quad F_2(t) = 1 - 2e^{-t/2} + e^{-t}, \quad G(t) = 1 - e^{-2t/5}, \quad t \ge 0 \tag{5.14}$$

This is the primary example considered by [20].

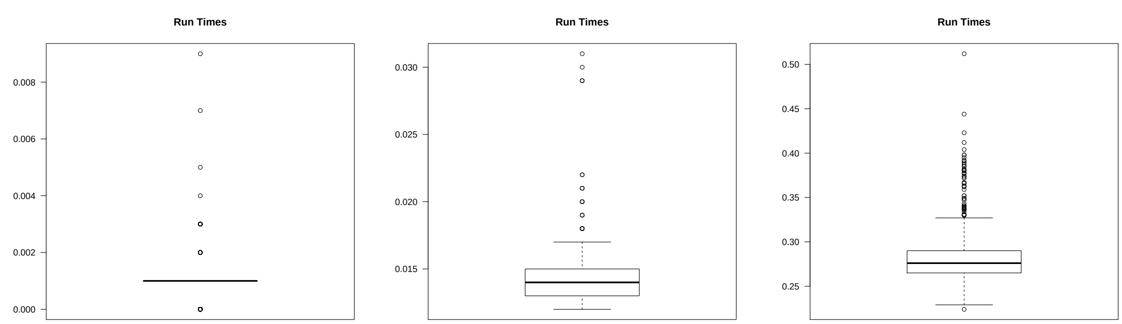
We use the implementation of the algorithms in [10]. Because of the speed of the primal-dual interior point algorithm, we can easily run simulations up to sample size $n = 10,000$. First we present the boxplots of the results of the primal-dual interior point algorithm for sample sizes $n = 100, 1000$ and $10,000$.



(a) n=100 (b) n=1000 (c) n=10,000

Fig 4: Boxplots for 1000 replications of the two means for sample sizes $100, 1000$ and $10,000$ in model (5.14).

Noting that the means for F_1 and F_2 are 2 and 3, respectively, it is seen that the estimates are still rather biased for sample size 100 (in particular for F_2), but become less biased for bigger sample sizes. The estimates of the means were defined as in (7.1) below.

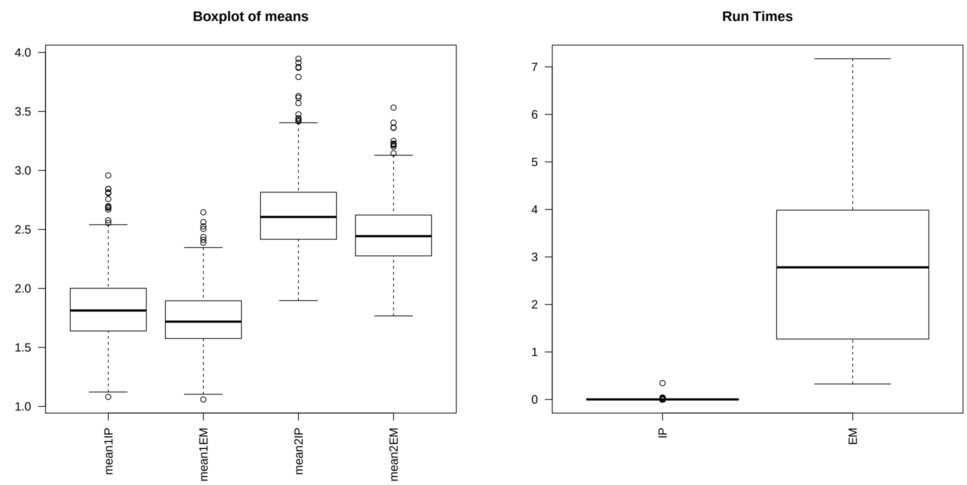


(a) n=100 (b) n=1000 (c) n=10,000

Fig 5: Boxplots for 1000 replications of the computing times of the MLE for sample sizes $100, 1000$ and $10,000$ in model (5.14).

The boxplots of the results for the corresponding computing times of the primal-dual interior point algorithm for sample sizes $n = 100, 1000$ and $10,000$ are given in Figure 5. The simulations were run on a MacBook Pro and time was measured in seconds.

We compare the estimates and run times in seconds for the primal-dual interior point method and the EM algorithm below for 1000 replications of sample size $n = 100$. The interior point method is for this sample size more than 1000 times faster than the EM algorithm and the discrepancy in computing time even grows for increasing sample size. We gave the EM algorithm an upper bound of $10,000$ iterations, but the (Fenchel duality) convergence criterion was then often still not reached, in contrast with the situation for the interior point method, which was run until the convergence criterion was reached. The estimates seem roughly similar, though, as can be seen from Figure 6, although the EM algorithm gives even more biased estimates of the means for F_2 (probably due to non complete convergence of the EM algorithm in computing the MLE).



(a) means for interior point and EM (b) run times for interior point and EM

Fig 6: Boxplots for 1000 replications of means and computing time in seconds of the MLE, computed by the primal-dual interior point algorithm and the EM algorithm, respectively, for sample size 100 in model (5.14).

All examples in this subsection can be checked by running the `R` scripts in [10].

6. Smooth functionals

It is somewhat surprising that results for smooth functionals of the distribution functions F_1 and F_2 can be derived without first proving simple facts for its local behavior like consistency. This is a consequence of the fact that local properties of the MLE are "washed away" in the global behavior of functionals of the MLE. We consider in particular moments as examples of smooth functionals. The theory for the computations, given in the Appendix, Section 9, is based on [3] and more recently Chapter 10 of [11]. We mainly give the computations for completeness and introduction of the notation, used in the examples.

6.1. Examples

Example 1. In the particular case studied by [20], we have

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

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.

Let R be the integral operator,

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

(see(9.14)). In this case we would get, if $\Psi(x) = x$ (the functional, corresponding to the first moment of F_1),

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

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

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, for b_2, defined by (9.5), we get:

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

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 (9.7)). 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

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

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:

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

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 (9.18) and (9.19). 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

$$b_2(t) = -\frac{2}{5}\int_0^t \frac{(8-u)e^{2u/5}}{2-e^{-u/2}}\,du + \frac{(8-t)e^{2t/5}}{2-e^{-t/2}} - 8. \tag{6.1}$$

Moreover,

$$b(t,1,1) = -\frac{2}{5}\int_0^t \frac{(8-u)e^{2u/5}}{2-e^{-u/2}}\,du - 8. \tag{6.2}$$

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

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

under the side condition $b(0,1,1) = -8$, which follows from (9.7), and by using (6.1). 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:

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

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 $10\int_0^{10} x^2 e^{-x/10}\,dx \approx 1606.03$. 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

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

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

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

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 (9.25) 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 (9.25) from the interval $[0,\tau_n]$ can be considerably smaller than I_κ^{-1}. The following table gives the values of

$$I_{\kappa,n}^{-1} \equiv \int_0^{\tau_n} \frac{[R_{F_2}(\Psi)]^2}{\overline{G}^2} \overline{F_2}\,\overline{G} d\boldsymbol{\lambda}_F = \int_0^{\tau_n} \frac{[R_{F_2}(\Psi)]^2}{\overline{G}} dF_2\,. \tag{6.3}$$

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.

7. Quantiles of F_1 and F_2

In order to have a proper view of the behavior of the MLE of F_1 and the Weighted Least Squares estimator proposed by [20] 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 6.1: $T_1 \sim \exp(0.5)$, $\;T_2 - T_1 \sim \exp(1)$, $\;C \sim \exp(0.4)$, i.e. $F_1(t) = 1 - e^{-t/2}$, $\;t \ge 0$, $F_2(t) = 1 - 2e^{-t/2} + e^{-t}$, $\;t \ge 0$, e $G(t) = 1 - e^{-2t/5}$, $\;t \ge 0$. It will make a comparison of our results with theirs possible. Here is our further example:

Example 3. In this example we have

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

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

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

$C \sim U(0,3)$.
In this case, the (marginal) distribution functions are

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

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

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 [20] 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

$$\widehat{Var}(\widehat{MSE}(t_j))$$

$$= \frac{1}{r}\left\{\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^4 - \frac{1}{r}\left(\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^2\right)^2\right.$$

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

$$\left.-\frac{2j}{5}\left[\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^3 - \left(\frac{1}{r}\sum_{i=1}^{r}\left[\hat{F}_{1,i}(t_j)\right]^2\right)\left(\frac{1}{r}\sum_{i=1}^{r}\hat{F}_{1,i}(t_j)\right)\right]\right\}.$$

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).

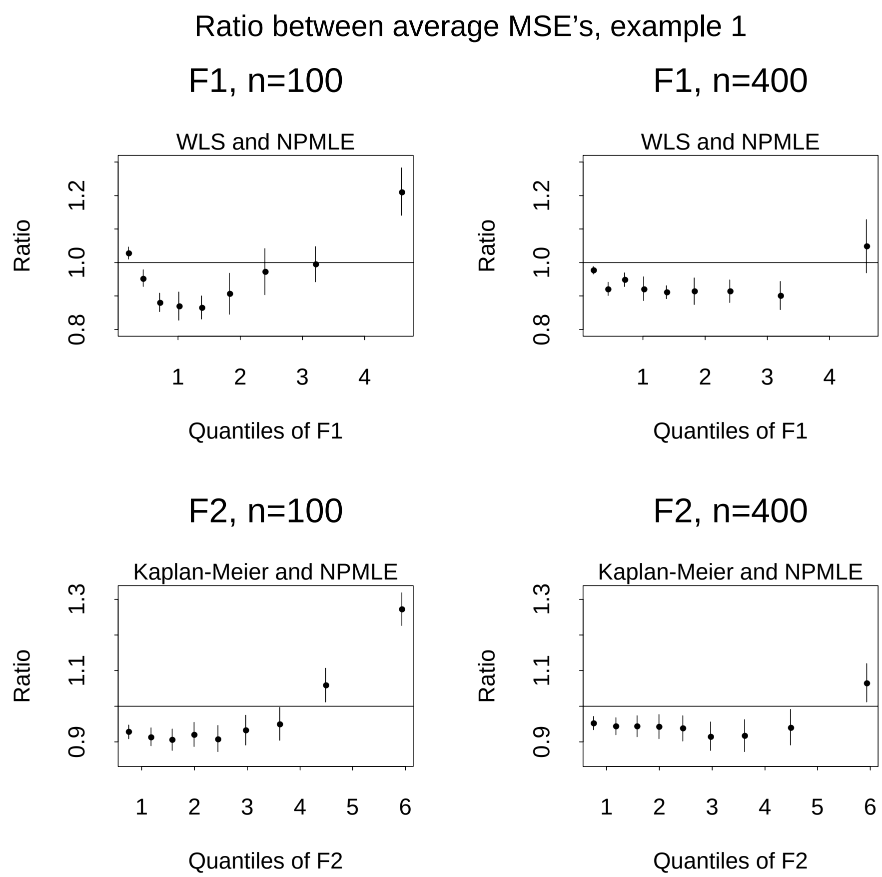


Fig 7: Ratio between average MSE's of WLSE and MLE of F_1 and between the Kaplan-Meier and MLE of F_2 (example 1).

Figures 7 to 8 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).

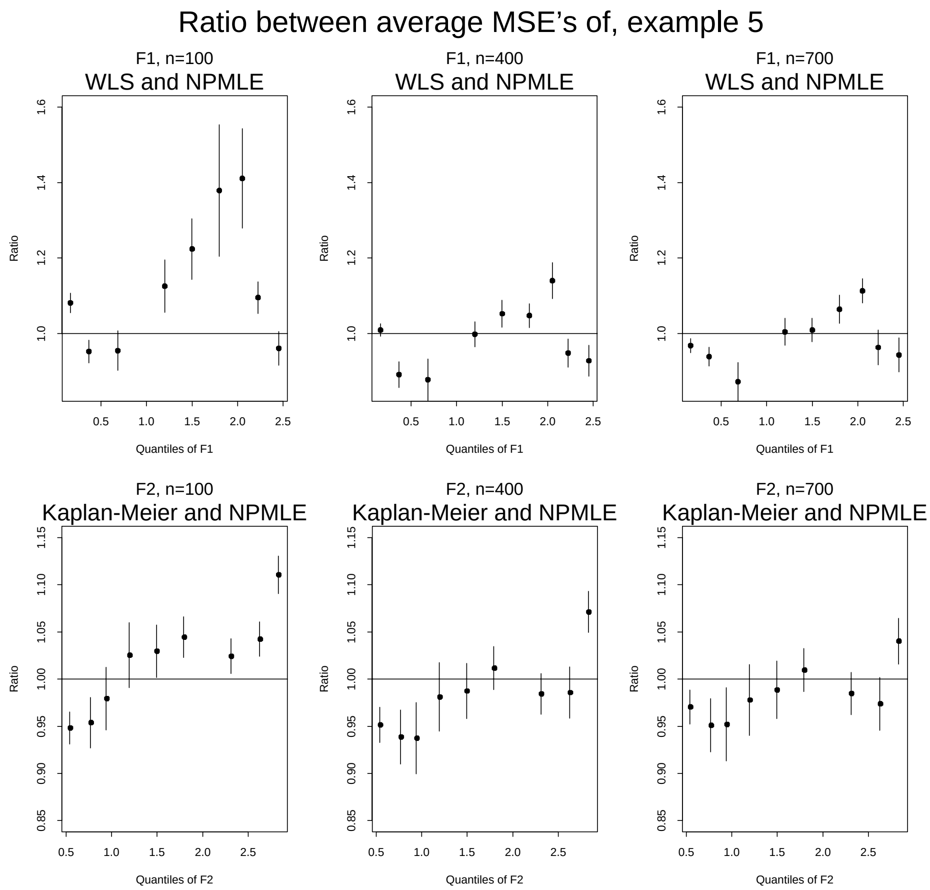


Fig 8: Ratio between average MSE's of WLSE and MLE of F_1 and between Kaplan-Meier and MLE of F_2 (example 3).

Figure 7 shows that our results for example 1, obtained using the Primal-Dual Interior Point algorithm, do not support the conclusions in [20] since the MLE 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 MLE 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 MLE 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.

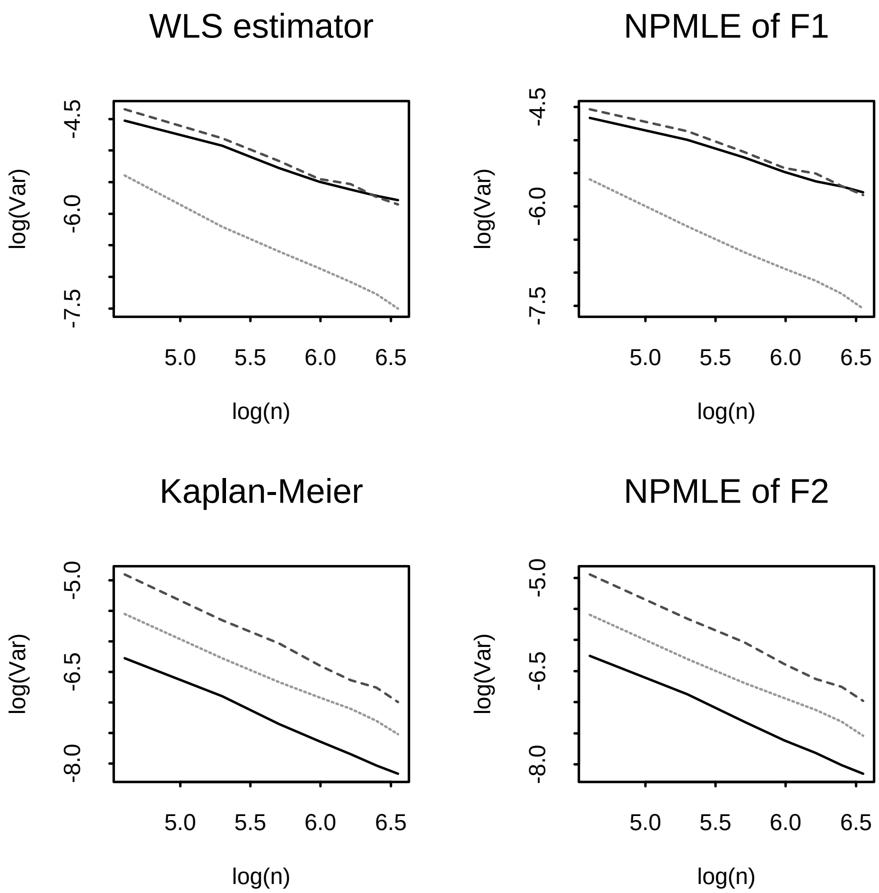


Fig 9: log-log plot of variance of the estimators versus sample size.

For Example 3, the WLS estimator is beaten by the MLE in the central quantiles for all sample sizes. These results indicate that the good performance of their estimator claimed by [20] tends to happen when F_1 and F_2 are far apart since it was generally beaten by the MLE in the tails for Example 1 or in the central quantiles in Example 3, where $F_1 = F_2$.

Figure 9 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

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

In each plot, the solid line refers to the second quantile, the dotted line refers to the fifth quantile, and the

TABLE 3
Slopes of the log-log plots at each quantile.

quantile	WLS	MLE of F1	Kaplan-Meier	MLE 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

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 9 and table 3, where all the lines have slopes close to -1. This 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 MLE, the pseudo MLE and the Weighted Least Squares estimator of F_1 proposed by [20] coincide. The same is true for the MLE, the pseudo MLE 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 MLE of $F \equiv F_1$ in the Interval Censoring, case 1 (see Groeneboom and Wellner (1992)), implying that the MLE of F_1, the pseudo MLE 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 MLE of F_1 and $\hat{F}_{1,WLS}$ is the estimator proposed by [20]. 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.

7.1. Estimation of moment functionals

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

$$\int_0^{Y_{(n)}} (1 - \hat{F}_n(x))\, dx \qquad \text{or} \qquad \int_0^{Y_{(n)}} 2x(1 - \hat{F}_n(x))\, dx \tag{7.1}$$

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

$$\int_0^{Y_{(n)}} x\, d\hat{F}_n(x) \qquad \text{or} \qquad \int_0^{Y_{(n)}} x^2\, d\hat{F}_n(x), \tag{7.2}$$

but putting $\hat{F}_n$ equal to 1 at $Y_{(n)}$. It is also possible to allow defective distribution functions $\hat{F}_n$ in (7.2), 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 (7.1).

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

TABLE 7
Summary results of the simulations for the WLS and MLE 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				MLE			
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 MLE 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				MLE			
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 MLE 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				MLE			
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

8. Concluding remarks

Theoretical properties of the real maximum likelihood estimator for the survival-sacrifice problem such as consistency were unknown and for this reason several pseudo maximum likelihood estimators have been proposed for this model in the past. The Weighted Least Squares of F_1, proposed in [20] (LJP (1997)) tends to be more efficient than the MLE 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 MLE of F_1 and F_2 by the Primal-Dual Interior Point algorithm, since the algorithm LJP (1997) applied for that purpose does not yield the true MLE estimator of F_1 and F_2, which invalidates their results about the relative efficiency of their estimator and the MLE of F_1. The magnitude of the variance of each estimator at the quantiles seems to explain the differences between the MLE and the Weighted Least Squares estimator of F_1, since the bias of both estimators are similar.

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 [20], 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.

We gave a characterization of the real MLE as an element of a cone in Section 3, and used this to derive a self-consistency equation, which in turn was used to prove consistency of the MLE in Section 4, turning the self-consistency equation into a Volterra integral equation. Further properties of the MLE are still unknown, but we believe that (under some natural conditions) the present approach will also lead to further distribution results, in particular rate $n^{1/2}$ results for the estimate of F_2 (the "Kaplan-Meier part") and rate $n^{1/3}$ results F_1 (the "current status part"). Proving this still remains a challenge at this point.

We introduced a primal-dual interior point method for computing the MLE, which was shown to be very much faster than the EM algorithm, introduced for this model in [19]. The results of both algorithms can be checked by running the `R` scripts in [10], where the implementation of both methods is given. The duality criteria of Theorem 3.1 give a natural stopping criterion for the EM algorithm, for which a stopping criterion, based on distances between successive iterations, is not appropriate because of the slow convergence of the algorithm.

9. Appendix

If we want to estimate a functional of the bivariate distribution H of (T_1, T_2), like the first moment of F_1, the score operator is given by:

$$\begin{aligned}[Aa](y,\delta_1,\delta_2) &= E[a(T_1,T_2)\,|\,T_2\wedge C=y,\Delta_1=\delta_1,\Delta_2=\delta_2]\\ &= \frac{(1-\delta_1)(1-\delta_2)\iint_{y<t_1\le t_2} a(t_1,t_2)\,dH(t_1,t_2)}{1-F_1(y)}\\ &\quad+\frac{\delta_1(1-\delta_2)\iint_{t_1\le y<t_2} a(t_1,t_2)\,dH(t_1,t_2)}{F_1(y)-F_2(y)}\\ &\quad+\frac{\delta_1\delta_2\int_{t_1<y} a(t_1,y)\,dH(t_1,y)}{f_2(y)}\ \text{ a.e. } [P_{H,G}]. \qquad (9.1)\end{aligned}$$

where G is the distribution of the censoring time C. This operator may be defined on $L_2(H)$, with range in $L_2(P_{H,G})$. Its range is contained in $L_2^0(P_{H,G})$. The adjoint of A on $L_2^0(P_{H,G})$ can be written as $[A^*b](t_1,t_2) = E[b(Y,\Delta_1,\Delta_2)\,\big|\,(T_1,T_2)=(t_1,t_2)]$ and we get

$$\begin{aligned}&[A^*b](t_1,t_2) = E\left[b(Y,\Delta_1,\Delta_2)\,\big|\,(T_1,T_2)=(t_1,t_2)\right]\\ &= \int_0^{t_1} b(u,0,0)\,dG(u) + \int_{t_1}^{t_2} b(u,1,0)\,dG(u) + \ b(t_2,1,1)\ \{1-G(t_2)\}\ \text{ a.e. } [H]. \qquad (9.2)\end{aligned}$$

We have pathwise differentiability of a functional κ of F_i, $i=1,2$, with canonical gradient $\tilde\kappa = \tilde\kappa_{F_i}$ 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 $\overline{\mathcal{R}(A)} \subset L_2^0(P_{H,G})$ satisfying

$$A^*\tilde l_\kappa = \tilde\kappa. \qquad (9.3)$$

See [21], and [3], 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 and can be therefore estimated at rate $\sqrt{n}$. We assume smoothness of the distribution functions and the function Ψ, allowing us to differentiate in the relations we get below. We also assume $\Psi(0)=0$, as is true for the moment functionals, considered in the examples. We have:

$$\tilde\kappa_{F_1}(t_1) = \Psi(t_1) - E_{F_1}\Psi(T_1),$$

and we have to solve the equation

$$\begin{aligned}&\Psi(t_1) - E_{F_1}\Psi(X)\\ &= \int_0^{t_1} b(u,0,0)\,g(u)\,du + \int_{t_1}^{t_2} b(u,1,0)\,dG(u) + b(t_2,1,1)\ \{1-G(t_2)\}\,. \qquad (9.4)\end{aligned}$$

Differentiating with respect to t_1 we obtain

$$b(t_1,0,0)g(t_1) - b(t_1,1,0)g(t_1) = \psi(t_1).$$

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

$$b_2(t_1) = b(t_1,1,0),\ t_1 \ge 0, \qquad (9.5)$$

we obtain

$$b(t_1,0,0) = \frac{\psi(t_1) + b_2(t_1)\,g(t_1)}{g(t_1)}\,,\ t_1 \ge 0. \qquad (9.6)$$

By letting $t_1 \downarrow 0$ in (9.4), we obtain

$$b(t_2,1,1) = -\frac{\int_0^{t_2} b_2(u)\,dG(u) + E_{F_1}\Psi(X)}{1-G(t_2)}\,,\ y \ge 0. \tag{9.7}$$

We note here, that, if

$$\lim_{t\to\infty} b(t,1,1)\{1-G(t)\} = 0, \tag{9.8}$$

we get

$$\int_0^\infty b_2(u)\,dG(u) = -E_{F_1}\Psi(X)\,. \tag{9.9}$$

But we will see an example below of a smooth functional for which (9.8) is not satisfied and therefore also (9.9) does not hold.

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

$$\begin{aligned}[A^*\phi](t_1,t_2) &= E\left[\phi(Y,\Delta_1,\Delta_2)\,\middle|\,(T_1,T_2)=(t_1,t_2)\right]\\ &= \int_0^{t_1}\phi(u,0,0)\,g(u)\,du + \int_{t_1}^{t_2}\phi(u,1,0)\,g(u)\,du + \phi(t_2,1,1)\,\{1-G(t_2)\} = 0,\ \text{ a.e. } [H].\end{aligned}$$

By taking the derivative with respect to t_1 we obtain:

$$\phi(t_1,0,0) = \phi(t_1,1,0),\ \text{ a.e. } [F_1]. \tag{9.10}$$

Letting $t_1 \downarrow 0$ yields:

$$\phi(t_2,1,1) = \frac{-\int_0^{t_2}\phi(u,1,0)\,dG(u)}{1-G(t_2)}\,,\ \text{ a.e. } [F_2]. \tag{9.11}$$

Note that if ϕ has compact support, we have:

$$\lim_{t_2\to\infty}\phi(t_2,1,1)\{1-G(t_2)\} = 0, \tag{9.12}$$

and hence, in that case,

$$\phi(t_2,1,1) = \frac{-\int_0^{t_2}\phi(u,1,0)\,dG(u)}{1-G(t_2)} = \frac{\int_{t_2}^{\infty}\phi(u,1,0)\,dG(u)}{1-G(t_2)}\,,\ \text{ a.e. } [F_2]. \tag{9.13}$$

We have the following lemma.

Lemma 9.1. *$\mathcal{N}(A^*) = \{\phi \in L_2(P_{H,G}) : $ (9.10), and (9.11) hold$\}$.*

Proof. It is clear that $\phi \in \mathcal{N}(A^*)$ implies (9.10) and (9.11). Conversely, if (9.10) and (9.11) hold, then

$$[A^*\phi]\,(t_1,t_2) = \int_0^{t_1}\phi(u,1,0)dG(u) + (1-G(t_2))\phi(t_2,1,1) = 0,\ \text{ a.e. } [H],$$

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

Remark. If $\phi \in L_2(P_{H,G})$ satisfies (9.10), and (9.11), then also $\phi \in L_2^0(P_{H,G})$, since

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

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

We now have:

$$\begin{aligned}
&E\tilde{l}_\kappa(Y,\Delta_1,\Delta_2)\phi(Y,\Delta_1,\Delta_2)\\
&= Eb(C,0,0)\phi(C,0,0)1_{\{C<T_1\}} + Eb(C,1,0)\,\phi(C,1,0)1_{\{T_1\le C<T_2\}}\\
&\qquad\qquad\qquad\qquad + Eb(T_2,1,1)\,\phi(T_2,1,0)1_{\{T_2<C\}}\\
&= \int_0^\infty b(y,0,0)\,\phi(y,0,0)\,g(t)\{1-F_1(y)\}\,dy\\
&\qquad + \int_0^\infty b(y,1,0)\phi(y,1,0)\,g(y)\{F_1(y)-F_2(y)\}\,dy\\
&\qquad\qquad\qquad + \int_0^\infty b(y,1,1)\,\phi(y,1,1)\{1-G(y)\}\,dF_2(y).
\end{aligned}$$

By (9.10), this can be written

$$\begin{aligned}
&E\tilde{l}_\kappa(Y,\Delta_1,\Delta_2)\phi(Y,\Delta_1,\Delta_2)\\
= &\int_0^\infty \{b(y,0,0)-b(y,1,0)\}\,\phi(y,0,0)\,\{1-F_1(y)\}\,dG(y)\\
&+\int_0^\infty b(y,1,0)\,\phi(y,0,0)\,\{1-F_2(y)\}\,dG(y)\\
&+\int_0^\infty b(y,1,1)\,\phi(y,1,1)\{1-G(y)\}\,dF_2(y)
\end{aligned}$$

Inserting the values of the function b yields:

$$\begin{aligned}
&E\tilde{l}_\kappa(Y,\Delta_1,\Delta_2)\phi(Y,\Delta_1,\Delta_2)\\
= &\int_0^\infty \psi(y)\phi(y,0,0)\,\{1-F_1(y)\}\,dG(y)\\
&+\int_0^\infty b_2(y)\,\phi(y,0,0)\,g(y)\{1-F_2(y)\}\,dG(y)\\
&-\int_0^\infty \left\{E_{F_1}\Psi(X)+\int_0^y b_2(u)\,dF_2(u)\right\}\phi(y,1,1)\,dF_2(y)\\
= &\ 0.
\end{aligned}$$

Now define the R-operator, corresponding to the distribution G, by

$$Rh(y) = h(y) - \frac{\int_y^\infty h(u)\,dG(u)}{1-G(y)}\,. \tag{9.14}$$

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

$$\alpha(y) = \phi(y,0,0) = \phi(y,1,0), \qquad y\ge 0. \tag{9.15}$$

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

$$\begin{aligned}
&\int_0^\infty \alpha(y)\psi(y)\,\{1-F_1(y)\}\,dy + \int_0^\infty b_2(y)\,\alpha(y)\,g(y)\{1-F_2(y)\}\,dy\\
&\qquad\qquad + \int_0^\infty \frac{E_{F_1}\Psi(X)+\int_0^y b_2 dG}{1-G(y)}\int_0^y \alpha(u)\,dG(u)\,dF_2(y)\\
= &\int_0^\infty \alpha(y)\psi(y)\,\{1-F_1(y)\}\,dy\\
&\quad + \int_0^\infty \left\{b_2(y) + \frac{E_{F_1}\Psi(X)+\int_0^y b_2(x)\,dG(x)}{1-G(y)}\right\}[R\alpha](y)\,\{1-F_2(y)\}\,dG(y)\\
= &\ 0,
\end{aligned}$$

where we use (9.13). So we obtain the relation

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

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

$$\{1-F_2(y)\}\left\{b_2(y)+\frac{E_{F_1}\Psi(X)+\int_0^y b_2(x)\,dG(x)}{1-G(y)}\right\} = -\left[R\left(\frac{\psi\{1-F_1\}}{g}\right)\right](y)\,, \tag{9.16}$$

since $R\circ R^* = I$; see, e.g., Proposition 8, Part B, p. 421, in [3].

So b_2 can be determined from (9.16). In fact, dividing by $1-F_2(y)$ yields:

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

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

$$\begin{aligned}&\{1-G(y)\}b_2(y)+\int_0^y b_2(x)\,dG(y)+E_{F_1}\Psi(X)\\ &= -\frac{1-G(y)}{1-F_2(y)}\left[R\left(\frac{\psi\{1-F_1\}}{g}\right)\right](y),\end{aligned} \tag{9.17}$$

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

$$b_2'(y) = -\{1-G(y)\}^{-1}\frac{d}{dy}\left\{\frac{1-G(y)}{1-F_2(y)}\left[R\left(\frac{\psi\{1-F_1\}}{g}\right)\right](y)\right\}, \tag{9.18}$$

and

$$b_2(0) = -\frac{\psi(0)}{g(0)} - E_{F_1}\Psi(X) + \int_0^\infty \psi(y)\{1-F_1(y)\}\,dy\,. \tag{9.19}$$

Note that when $\Psi(0)=0$, (9.19) reduces to

$$b_2(0) = -\frac{\psi(0)}{g(0)}\,. \tag{9.20}$$

The complete solution b can now be found by first getting $b_2 = b(\cdot,1,0)$ from (9.17) (or (9.18) and (9.19)), and next getting $b(\cdot,0,0)$ and $b(\cdot,1,1)$ from relations (9.6) and (9.7).

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

$$\tilde{l}_\kappa(t,\Delta_1,\Delta_2) = \begin{cases} (\psi(t)+b_2(t)g(t))/g(t)\,, & \text{if } (\Delta_1,\Delta_2)=(0,0),\\ b_2(t)\,, & \text{if } (\Delta_1,\Delta_2)=(1,0),\\ -\dfrac{E_{F_1}\Psi(X)+\int_0^t b_2(x)\,dG(x)}{1-G(t)}\,, & \text{if } (\Delta_1,\Delta_2)=(1,1),\end{cases} \tag{9.21}$$

where b_2 is determined by (9.17) (or (9.18) and (9.19)). The information bound is just

$$\begin{aligned} I_\kappa^{-1} &= E(\tilde{l}_\kappa(Y,\Delta_1,\Delta_2)^2) \\ &= \int_0^\infty \left(\frac{\psi(y)+b_2(y)g(y)}{g(y)}\right)^2 g(y)(1-F_1(y))dy \\ &\quad + \int_0^\infty b_2^2(y)g(y)(F_1(y)-F_2(y))dy \\ &\quad + \int_0^\infty \frac{\left\{E_{F_1}\Psi(X)+\int_0^y b_2 dG\right\}^2}{(1-G(y))^2}(1-G(y))dF_2(y). \end{aligned} \tag{9.22}$$

Remark: Note that when the d.f. F_2 is concentrated at $+\infty$ and $\Psi(x)=x$, then (9.17) 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. [3], page 209; [9], page 115, and [13], 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

$$\begin{aligned} \Psi(y) - E_{F_2}\Psi(Y) &= \int_0^{t_1} b(u,0,0)\,g(u)\,du + \int_{t_1}^{t_2} b(u,1,0)\,g(u)\,du \\ &\qquad + b(t_2,1,1)\,\{1-G(t_2)\}\,. \end{aligned} \tag{9.23}$$

By differentiating with respect to t_1 we get:

$$b(t_1,0,0) = b(t_1,1,0), \text{ a.e. } [G]\,. \tag{9.24}$$

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. [3], pages 272 - 280, and especially page 276. Hence the efficient influence function $\tilde{l}_\kappa$ for functionals of this type is given by

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

where

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

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

$$I_\kappa^{-1} = \int_0^\infty \frac{[R_{F_2}(\Psi)]^2}{\overline{G}^2}\overline{F_2}\,\overline{G}d\boldsymbol{\lambda}_F = \int_0^\infty \frac{[R_{F_2}(\Psi)]^2}{\overline{G}}\,dF_2\,. \tag{9.25}$$

This is in agreement with the results of [5], [18], and [1].

References

[1] Michael G. Akritas. The central limit theorem under censoring. *Bernoulli*, 6(6):1109–1120, 2000. ISSN 1350-7265. . URL https://doi.org/10.2307/3318473.

[2] R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk. *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.

[3] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semiparametric models.* Springer-Verlag, New York, 1998. ISBN 0-387-98473-9. Reprint of the 1993 original.

[4] G.E. Dinse and S.W. Lagakos. Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. *Biometrics*, 38:921–932, 1982.

[5] Richard Gill. Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, 11(1):49–58, 1983. ISSN 0090-5364. . URL https://doi.org/10.1214/aos/1176346055.

[6] A. E. Gomes. Consistency of the non-parametric maximum pseudo-likelihood estimator of the disease onset distribution function for a survival-sacrifice model. *J. Nonparametr. Stat.*, 20(1):39–46, 2008. ISSN 1048-5252. . URL https://doi.org/10.1080/10485250701830121.

[7] Antonio Eduardo Gomes. Asymptotics for a weighted least squares estimator of the disease onset distribution function for a survival-sacrifice model. *Ann. Inst. Statist. Math.*, 56(4):683–700, 2004. ISSN 0020-3157. . URL https://doi.org/10.1007/BF02506483.

[8] P. Groeneboom. Lectures on inverse problems. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 67–164. Springer, Berlin, 1996. . URL http://dx.doi.org/10.1007/BFb0095675.

[9] P. Groeneboom and J.A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1992. ISBN 3-7643-2794-4.

[10] Piet Groeneboom. R scripts for the survival-sacrifice model. https://github.com/pietg/survival-sacrifice-model, 2018.

[11] Piet Groeneboom and Geurt Jongbloed. *Nonparametric estimation under shape constraints*, volume 38 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2014. ISBN 978-0-521-86401-5. . URL https://doi.org/10.1017/CBO9781139020893. Estimators, algorithms and asymptotics.

[12] M. G. Gu and C.-H. Zhang. Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, 21(2):611–624, 1993. ISSN 0090-5364. . URL https://doi.org/10.1214/aos/1176349140.

[13] J. Huang and J.A. Wellner. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Statist. Neerlandica*, 49:153–163, 1995. ISSN 0039-0402. . URL http://dx.doi.org/10.1111/j.1467-9574.1995.tb01462.x.

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

[15] Johan Lim, Seung Jean Kim, and Xinlei Wang. Estimating stochastically ordered survival functions via geometric programming. *J. Comput. Graph. Statist.*, 18(4):978–994, 2009. ISSN 1061-8600. . URL https://doi.org/10.1198/jcgs.2009.06140.

[16] David G. Luenberger. *Optimization by vector space methods.* John Wiley & Sons, Inc., New York-London-Sydney, 1969.

[17] Lu Mao. Nonparametric identification and estimation of current status data in the presence of death. *Statistica Neerlandica*, 73:to appear, 2019.

[18] Winfried Stute. The central limit theorem under random censorship. *Ann. Statist.*, 23(2):422–439, 1995. ISSN 0090-5364. . URL https://doi.org/10.1214/aos/1176324528.

[19] B.W. Turnbull and T.J. Mitchell. Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments. *Biometrics*, 40(1):41–50, 1984. URL https://www.jstor.org/stable/2530742.

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

[21] A.W. van der Vaart. On differentiable functionals. *Ann. Statist.*, 19:178–204, 1991. ISSN 0090-5364. . URL http://dx.doi.org/10.1214/aos/1176347976.

[22] S.J. Wright. *Primal-dual interior-point methods*, volume 54. Siam, 1997.