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

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

A. E. Gomes, P. Groeneboom, J. A. Wellner

Relatório Técnico
RTP-07/2000

Relatório Técnico
Série Pesquisa

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

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

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

Jon A. Wellner
Department of Statistics - University of Washington
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 $\Delta_2 = 1_{\{T_2 \le C\}}$, T_1 and T_2 have a joint distribution function F such that $P(T_1 \le T_2) = 1$, and are independent of C. Some authors call this structure "survival/sacrifice model".

van der Laan, Jewell, and Peterson (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 NPMLE (nonparametric maximum likelihood estimator) of F_1, what is not always true since the algorithm applied by them for the calculation of the joint NPMLE of F_1 and F_2 does not always yield the true NPMLE.

The Primal-Dual Interior Point algorithm was applied in our simulations to obtain the true NPMLE of F_1. The results showed a better performance of the NPMLE of F_1 over the Weighted Least Squares estimator for the points where $F_1 \cong F_2$, while the opposite seems to occur wen F_1 and F_2 are far apart.

Contents

1 Introduction

In experiments for the study of onset and mortality from undetectable irreversible diseases (occult tumors, e.g.) a possible data structure consists of the time of death, whether or not the disease of interest was present at death, and if present, whether the disease was a probable cause of death or not. This data structure is related to moderately lethal incurable diseases when the cause of death is known. Defining the non-negative variables T_1 (time of disease onset), T_2 (time of death from the disease) and C (time of death from an unrelated cause), we observe, for the ith individual, $(C_i \wedge T_{2,i}, \Delta_{1,i}, \Delta_{2,i})$, where $\Delta_{1,i} = 1_{\{T_{1,i} \leq C_i\}}$, $\Delta_{2,i} = 1_{\{T_{2,i} \leq C_i\}}$, $C_i \wedge T_{2,i} = \min\{C_i, T_{2,i}\}$, $T_{1,i}$ and $T_{2,i}$ have a joint distribution function F such that $P(T_{1,i} \leq T_{2,i}) = 1$, C_i has distribution function G and is independent of $(T_{1,i}, T_{2,i})$. When ordering the data by observation time $(C_i \wedge T_{2,i})$ it is assumed that in case of ties the observations with $(\Delta_{1,i}, \Delta_{2,i}) = (1,1)$ occurs first, followed by the ones with $(\Delta_{1,i}, \Delta_{2,i}) = (1,0)$ and finally by the ones with $(\Delta_{1,i}, \Delta_{2,i}) = (0,0)$. The interval censoring, case 1 (see, e.g., Groeneboom and Wellner (1992)), can be seen as a particular case of this data structure when the disease is non-lethal, i.e., $\Delta_{2,i} = 0, i = 1, \ldots, n$. The right censoring problem can also be considered as a special case of data with the structure above when a lethal disease is always present at the moment of death, i.e., $\Delta_{1,i} = 1, i = 1, \ldots n$. Some authors call it *survival sacrifice model.*

Our interest here is to study nonparametric estimation of the distribution functions F_1 and F_2 of T_1 and T_2, respectively (or equivalently the disease resistance function S_1 and the survival function S_2). The parameter space can be defined as

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

where $F_1 <_s F_2$ means that $F_1(x) \geq F_2(x)$ for every $x \in I\!R$ and $F_1(x) > F_2(x)$ for some $x \in I\!R$. Some estimation approaches for this problem are based on the 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 $K(g, G)$ is a term involving only the distribution G of C.
Kodell, Shaw and Johnson (1982) also studied nonparametric estimation of S_1 and S_2, but their work is restricted to the case where $R(t) = S_1(t)/S_2(t)$ is non-increasing, an assumption that may not be reasonable, for example, for progressive diseases whose incidence is concentrated in the early or middle part of the life span.
Turnbull and Mitchell (1984) proposed an EM algorithm for the joint estimation of F_1 and F_2 which converges very slowly to the NPMLE of (F_1, F_2) (provided the support of the initial estimator contains the support of the NPMLE). It should be noticed that the two-dimensional nature of their method enables us to avoid the use of Lagrange multipliers.
Another possible way of estimating F_1 is plugging in the Kaplan-Meier estimator of F_2 and and calculating the pseudo NPMLE of F_1. The part of the log-likelihood involving F_1 is

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

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) can be written as

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

with $f = F_1$, $\phi = d\Phi/df$, $g = 1 - (1 - \hat{F}_{2,KM})(1 - \Delta_1)$, $w = (1 - \Delta_2)/(1 - \hat{F}_{2,KM})$ and $\Phi(y) = (y - F_2)\log(y - F_2) + (1 - y)\log(1 - y)$, Dinse and Lagakos (1982) concluded that the values of $F_1(Y_{(i)}), i = 1, \ldots, n$, maximizing the log-likelihood (1) could be obtained applying theorem 1.10 in Barlow, Bartholomew, Bremner, and Brunk (1972), i.e., the pseudo NPMLE of F_1 would be given by the isotonic regression g^* of $g(Y_{(i)})$ with weights $w(Y_{(i)}), i = 1, \ldots, n$. However, theorem 1.10 in Barlow *et al.* (1972) is applicable to a real convex function Φ defined on $I\!R$ while in the application above the function Φ is in fact defined on $I\!R^2$ since the value of F_2 is not supposed to be constant. It should be mentioned here that, although the Kaplan-Meier estimator $\hat{F}_{2,KM}$ is uniquely defined, except possibly at times exceeding the largest observation, the pseudo NPMLE $\hat{F}_{1,n}$ is uniquely defined only over certain data-determined intervals. Specifically, $\hat{F}_{1,n}$ is always uniquely defined at the observed C_i's, i.e., the observations for which $\Delta_{2,i} = 0$.

2 A Weighted Least Squares Estimator

Another possibility to estimate F_1 is by calculating a weighted least squares estimator as suggested by van der Laan, Jewell, and Peterson (1997). Making $S_1 = 1 - F_1$ and $S_2 = 1 - F_2$, in terms of populations, $R(c) = S_1(c)/S_2(c)$ is the proportion of subjects in the population alive at time c who are disease free (i.e., $1 - R(c)$ is the prevalence function at time c), and it can be written as

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

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

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

under the constraint that S_1 is nonincreasing. This minimization problem can be solved by using results from the theory of isotonic regression (see Barlow *et al.* (1972)) and is given by

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

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

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

for $m=1,\ldots,n$. If we use $w_j=(1-\Delta_{2(j)})/\hat{S}^2_{2,KM}(Y_{(j)})$ instead, we have an estimator with a closed form that can be calculated as the left derivative of the least concave majorant of the cumulative sum diagram $(0,0),(W_1,G_1),\ldots,(W_n,G_n)$, where $W_i=\sum_{j=1}^{i} w_j$ and

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

$\hat{S}_1(t)$ is the slope of the Least Concave Majorant at W_i for $t\in\left(Y_{(i-1)},Y_{(i)}\right]$.
van der Laan, Jewell and Peterson (1997) claim that the estimator above is more efficient than the NPMLE of F_1 which would be calculated, according to the 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

$$F_{2n}(t) = \frac{\sum_{i=1}^{n}\frac{\int_{C_i}^{t} dF_{2n}}{(S_{2n}-S_{1n})(C_i)}1_{\{\delta_i=(1,0)\}} + 1_{T_{2,i}\le t}1_{\{\delta_i=(1,1)\}}}{\sum_{i=1}^{n} 1_{\{\delta_i=(1,1)\}} + \frac{S_{2n}(C_i)}{(S_{2n}-S_{1n})(C_i)}1_{\{\delta_i=(1,0)\}}}, \tag{3}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{1n}(C_i)} - \frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)-S_{1n}(C_i)}\right\}=0, \tag{4}$$

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

$$S_{1n}(t)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{\Delta_{1,i}=0,\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S_{2n}(C_i)R_n(C_i)(1-R_n(C_i))}\Big/\sum_{i=1}^{n}\left\{\frac{\{\Delta_{2,i}=0,C_i\in(t_j,t_{j+1}]\}}{S^2_{2n}(C_i)R_n(C_i)(1-R_n(C_i))}\right\}\right., \tag{5}$$

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 $F_{1n}(t)\ge F_{2n}(t)$ are active (these are precisely the points where we will get a Langrange multiplier $\lambda_i>0$ in the primal-dual interior point algorithm, described in section 3), the expression above is not defined, since we get zeros in the denominators!
As an example, in one of our simulations studies we found the following values for the NPMLE of the pair (F_1,F_2) between points $t_j=1.141807$ and $t_{j+1}=1.567906$:

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

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

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

on the same interval. So F_{2n} becomes equal to F_{1n} at the upper part of this interval. Exactly at the point where this happens, the constraint $F_{1n} \geq F_{2n}$ becomes active, and this means that the equation (4) is *not* satisfied. In fact, the sum on the left-hand side of (4) was precisely the value of the Lagrange multiplier divided by n, missing in (4)! And, as noted above, the expression on the right-hand side of (5) is undefined in this situation, because of the fact that the denominators become zero if the constraint is active.
The conclusion is that the NPMLE is *not* the reweighted least squares estimator as introduced by van der Laan, Jewell and Peterson (1997), because that estimator is not well-defined, their (not iteratively defined) weighted least squares estimator, claimed to be superior to the NPMLE, actually often *coincides* with the NPMLE.
One might perhaps hope that maximizing the likelihood, ignoring the constraint $F_1 \geq F_2$ (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 NPMLE, even if we start the procedure with positive masses at each point, as recommended in van der Laan, Jewell and Peterson (1997). Use of Lagrange multipliers or a similar technique is the only way out here.

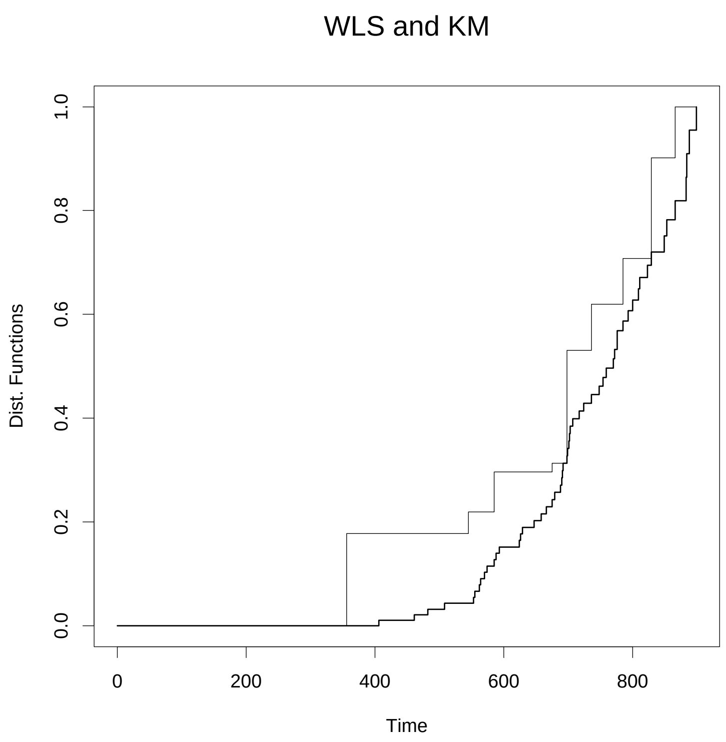

Figure 1: **Weighted Least Square estimate of F_1 and Kaplan-Meier estimate of F_2.**

We close this section by showing the the Kaplan-Meier estimator and the (non-iterative) weighted least

squares estimator introduced in van der Laan, Jewell and Peterson (1997) for a real data set studied by Dinse and Lagakos (1982) and Turnbull and Mitchell (1984) representing the ages at death (in days) of 109 female RFM mice (Table 1). The disease of interest is reticulum cell sarcoma (RCS). These mice formed the control group in a survival experiment to study the effects of prepubertal ovariectomy in mice given 300 R of X-rays. The smoother picture for the estimate of F_2 in Figure 1 is a consequence of the fact that the estimators of F_1 have a $n^{-1/3}$ rate of convergence.
The value of the log-likelihood for these data set at the estimates obtained through the algorithm proposed by van der Laan *et al.* (1997) above is smaller than that obtained at the true NPMLE of F_1 and F_2 obtained through the Primal-Dual Interior Point algorithm (-262.7964 and -262.5468, respectively).

3 Primal-Dual Interior Point Algorithm

Our problem here is to maximize the log-likelihood function

$$\begin{aligned} \mathcal{L}(\mathbf{x},\mathbf{y}) &= \sum_{i=1}^{n} \{(1-\delta_{1,(i)})(1-\delta_{2,(i)})\log(1-x_i) \\ &\quad + \delta_{1,(i)}(1-\delta_{2,(i)})\log(x_i-y_i) \\ &\quad + \delta_{1,(i)}\delta_{2,(i)}\log(y_i-y_{i-1})\}\,, \end{aligned} \tag{6}$$

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

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

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

$$\begin{aligned} g_i(\mathbf{z}) &= x_i - x_{i+1},\ i = 1, \ldots, n-1, \\ g_n(\mathbf{z}) &= x_n - 1, \\ g_{n+1}(\mathbf{z}) &= -y_1, \\ g_{n+i}(\mathbf{z}) &= y_{i-1} - y_i,\ i = 2, \ldots, n, \\ g_{2n+i}(\mathbf{z}) &= y_i - x_i,\ i = 1, \ldots, n, \end{aligned}$$

and the matrix G by

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

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

$$g(\mathbf{z}) \le 0,\ \mathbf{z} \in \mathbb{R}^{2n}\ . \tag{7}$$

Theorem 1: Let $\hat{\mathbf{z}} = (\hat{x}_1, \ldots, \hat{x}_n, \hat{y}_1, \ldots, \hat{y}_n)'$ be a vector in $I\!R^{2n}$ such that $\phi(\hat{\mathbf{z}}) < \infty$. Then $\hat{\mathbf{z}}$ minimizes $\phi(\hat{\mathbf{z}})$ over the set of vectors $\mathbf{z}$, satisfying (7), if and only if the following conditions are satisfied:

$$\begin{aligned} \nabla\phi(\hat{\mathbf{z}}) + G'\lambda &= 0 && (8)\\ g(\hat{\mathbf{z}}) + \mathbf{w} &= 0 && (9)\\ \langle \lambda, \mathbf{w}\rangle &= 0, && (10) \end{aligned}$$

for vectors λ and $\mathbf{w}$ in $I\!R_+^{3n}$.

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

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

we can write condition (8) in the form

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

Note that $g(\hat{\mathbf{z}}) + \mathbf{w} = 0$, for $\mathbf{w} \in I\!R_+^{3n}$, implies $g(\hat{\mathbf{z}}) \le 0$, and that $\lambda_i > 0$ implies $w_i = 0$, by (10), and hence $g_i(\hat{\mathbf{z}}) = 0$, by (9). Thus:

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

The proof of theorem 1 can be found in Groeneboom (1998).

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 the dual problem for the vectors λ and $\mathbf{w}$).

A peculiar difficulty is that not all variables appear in the object function that we want to maximize (the log-likelihood). For example, if $\delta_{1,(i)} = \delta_{2,(i)} = 0$, then only x_i Figures in the (6) and not y_i or y_{i-1} (y_{i-1} could appear in a preceding term, though). For this reason the log-likelihood will never have $2n$ arguments, unless only terms $\log(x_i - y_i)$ occur. Although we could in principle work with the full set of $2n$ variables, we prefer to make a preliminary reduction, where we only keep the arguments x_i and y_j that really are arguments of the log-likelihood function. For these arguments we get a unique solution in contrast with the situation for the full set of arguments, for which this is generally not true.

Suppose that $x_1 \le \ldots \le x_{n_1}$ and $y_1 \le \ldots \le y_{n_2}$ are the arguments of the log-likelihood corresponding to values of F_1 and F_2, respectively. We then define the vector $\mathbf{z}$ by

$$\mathbf{z} = (z_1, \ldots, z_{n_3})' = (x_1 \le \ldots \le x_{n_1}, y_1 \le \ldots \le y_{n_2})',$$

where $n_3 = n_1 + n_2$. Defining $z_0 = 0$, we get that the vector $\mathbf{z}$ will have to satisfy the following $m = n_3 + n_2 + 1$ constraints:

$$\begin{aligned} g_i(\mathbf{z}) &= z_{i-1} - z_i \le 0,\ i = 1, \ldots, n_1,\\ g_{n_1+1}(\mathbf{z}) &= z_{n_1} - 1 \le 0,\\ g_{n_1+2}(\mathbf{z}) &= -z_{n_1+1} \le 0,\\ g_{n_1+i+1}(\mathbf{z}) &= z_{n_1+i-1} - z_{n_1+i} \le 0,\ i = 2, \ldots, n_2,\\ g_{n_3+i+1}(\mathbf{z}) &= z_{n_1+i} - z_{k_i} \le 0,\ i = 1, \ldots, n_2, \end{aligned}$$

where z_{k_i} is the smallest component of the set $z_1, \ldots, z_{n_1}$ for which we know that $z_{k_i} \ge z_{n_1+i}$ has to be satisfied; if there is no such component (this happens if the biggest observation has $\delta_{1,(i)} = \delta_{2,(i)} = 1$), we put $z_{k_i} = 1$. As before, we will denote this vector of constraints by $g(\mathbf{z})$ and the corresponding matrix by G. Note that the constraints are not linear but affine; we can write them in the form

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

or

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

where $\mathbf{a}$ is a vector having zero components, except for the $(n_1+1)th$ component, which is 1, and possibly $(n_3+i+1)th$ components, corresponding to elements z_{n_1+i} for which no "majorizing" component in the set $z_1, \ldots, z_{n_1}$ need exist (see above); in such a case the $(n_3+i+1)th$ component of $\mathbf{a}$ is also equal to 1.

We now start the computation of the NPMLE with a vector $\mathbf{z}_0 \in I\!R^{n_3}$, strictly satisfying all constraints, i.e., $g(\mathbf{z}_0) < 0$. For λ and $\mathbf{w}$ we take as starting values $\lambda_0 = \mathbf{w}_0 = e$, where e denotes the vector in $I\!R^m, m = n_3 + n_2 + 1$, with all components equal to 1. For a vector $\mathbf{a}$ we denote the diagonal matrix with component a_i as its ith diagonal element by A; for example, Λ is the diagonal matrix with element λ_i as its ith diagonal element. The first (Newton) iteration step now solves the system of equations:

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

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

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

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

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

By taking $\lambda = \mathbf{w} = 0.5e$, we have made this parameter equal to 0.5 at the start of the iterations.
We now take a final parameter $\gamma \in (0,1)$, and take α_1 as the first number in the sequence

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

such that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\nabla_{\mathbf{zz}}\phi(\mathbf{z}) + G'DG\ , \tag{11}$$

and this matrix is a symmetric positive definite matrix at each step. The matrices W and Λ are diagonal matrices, so inversion of these is trivial. In our case, the matrix (11) is also "sparse", since most off-diagonal elements are zero. This fact can be used for fast and efficient inversion methods, where we only have to reserve computer memory for the elements that can be non-zero. Figure 2 illustrates the application of the primal-dual interior point algorithm for the estimation of the NPMLE of F_1 and F_2 for simulated data following the distribution functions of example 1, with several sample sizes. The Primal-Dual Interior Point algorithm for the estimation of the NPMLE of F_1 and F_2 was also applied to the data set presented in Table 1. Figure 3 shows the estimates, and we can notice the larger number of jumps that the estimate of F_2 has compared to the estimate of F_1, a fact related to the difference in the rate of convergence of the estimators of each distribution function.

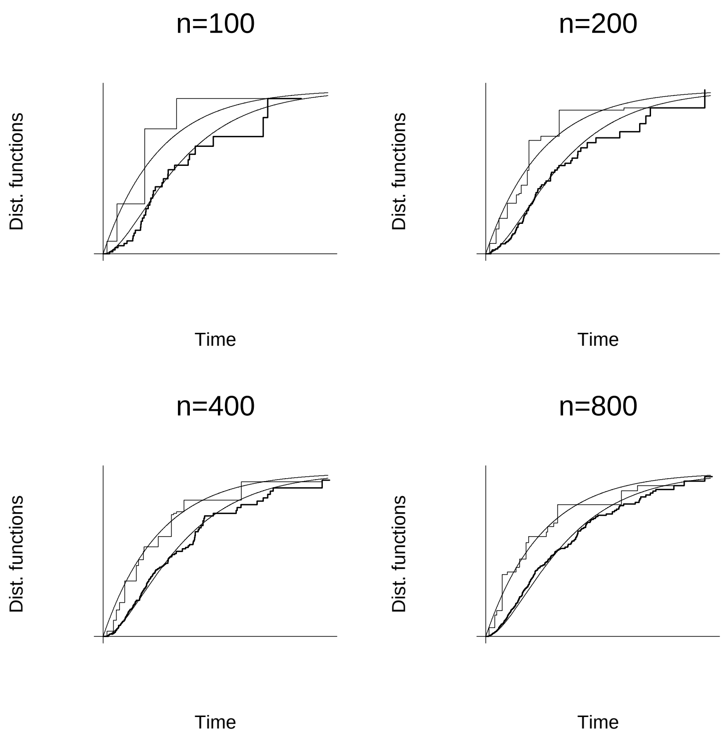

Figure 2: **Joint NPML estimates of F_1 and F_2 (simulated data for example 1).**

4 Simulation Studies

In order to have a proper view of the behavior of the NPMLE of F_1 and the Weighted Least Squares estimator proposed by van der Laan *et al.* (1997) to solve the problem of estimating F_1, two examples of distributions for T_1, T_2 and C were considered. In example 1, $T_1 \sim \exp(0.5)$, $\quad T_2 - T_1 \sim \exp(1)$, $\quad C \sim \exp(0.4)$, i.e., $F_1(t) = 1 - e^{-t/2}, \quad t \geq 0$, $F_2(t) = 1 - 2e^{-t/2} + e^{-t}, \quad t \geq 0$, e $G(t) = 1 - e^{-2t/5}, \quad t \geq 0$. These are the distributions van der Laan *et al.* (1997) considered in their simulations and will make a comparison of our results with theirs possible. In example 2, we have

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

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

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

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

In this case, the distribution functions are

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

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

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

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

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

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

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

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

In Tables 3, 4 and 5, the bias and variance of the estimators at each quantile are presented as well as the standard error of the Mean Squared Error, estimated by the square root of

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

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

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

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

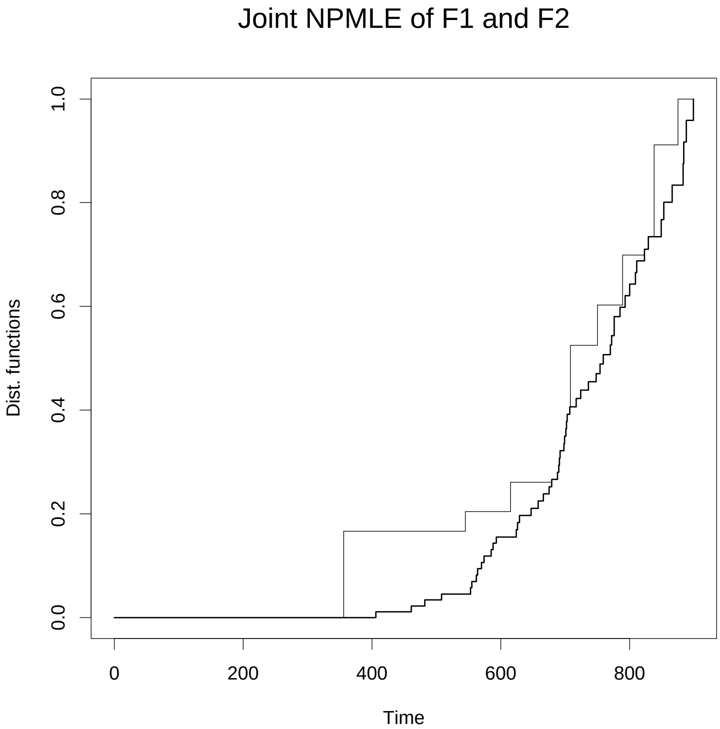

Figure 3: **Joint NPML estimates of F_1 and F_2.**

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

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

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

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

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

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

Figure 4 shows that our results for example 1, obtained using the Primal-Dual Interior Point algorithm, do not support the conclusions in van der Laan *et al.* (1997) since the NPMLE seems to estimate F_1 better in the right-hand tail for both sample sizes (efficiency > 1), while in their results they claim to have observed a relative efficiency of 0.92 for both sample sizes in the last quantile considered. The difference in the results is caused by the way they calculated the NPMLE of F_1, which may not yield the correct estimate. Looking at column (2) for both estimators in Table 3 we see that the lower variance of the NPMLE of F_1 at the last quantile is the reason for its superior performance there, and the variance seems to explain also the better performance of the WLS estimator at the other quantiles since the variance is the biggest component of the MSE for both estimators.

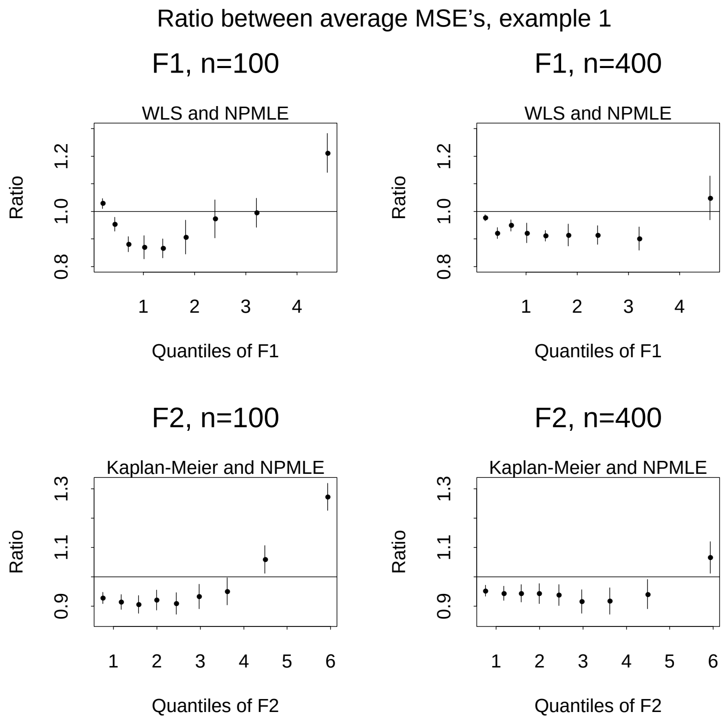

Figure 4: **Ratio between average MSE's of of WLSE and NPMLE of F_1 and between the Kaplan-Meier and NPMLE of F_2 (example 1).**

For example 2 (Figure 5), the WLS estimator is beaten by the NPMLE in the central quantiles for all sample sizes. These results indicate that the good performance of their estimator claimed by van der Laan *et al.* (1997) tends to happen when F_1 and F_2 are far apart since it was generally beaten by the NPMLE in the tails for example 1 or in the central quantiles in example 2, where $F_1 = F_2$.

Table 2: *Slopes of the log-log plots at each quantile.*

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

Figure 6 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 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 2 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 $F_1 = F_2$

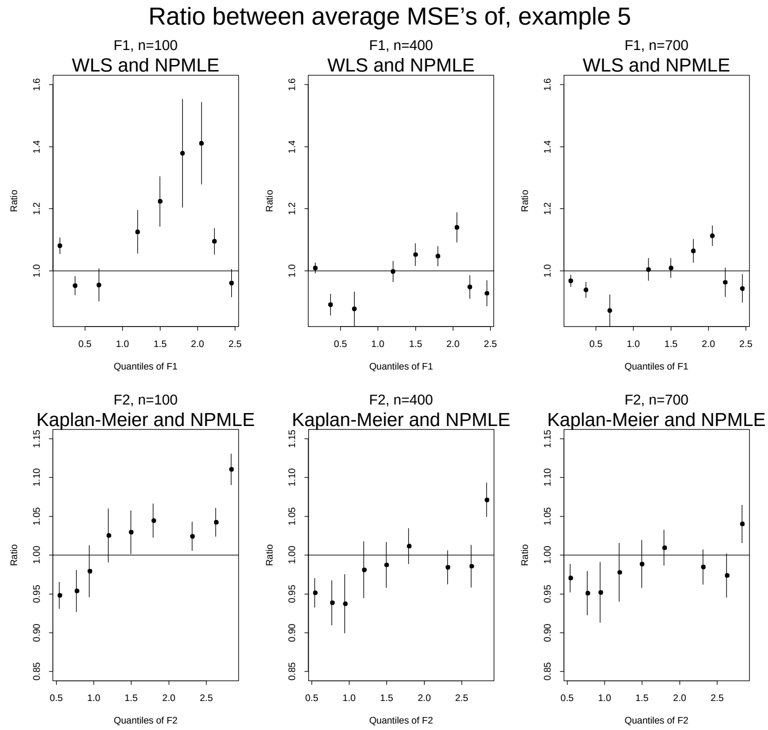

Figure 5: **Ratio between average MSE's of WLSE and NPMLE of F_1 and between Kaplan-Meier and NPMLE of F_2 (example 5).**

we probably have $n^{-1/2}$ as the rate of convergence of the estimators of F_1, a property that the estimators of F_2 have for the whole real line, as we can see in the plots for the Kaplan-Meier and NPML estimators of F_2 in Figure 6 and Table 2, 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 NPMLE, the pseudo NPMLE and the Weighted Least Squares estimator of F_1 proposed by van der Laan *et al.* (1997) coincide. The same is true for the NPMLE, the pseudo NPMLE and the Kaplan-Meier estimator for F_2. That happens because $\| F_1 - F_2 \|_\infty = \sup_{t\in I\!R} \mid F_1 - F_2 \mid = 1$ implies that the ordered observed death times $U_i = C_i \wedge T_{2,i}$ are such that there are two blocks of observations, the first one with $\Delta_{2,i} = 0$, and the second one with $\Delta_{2,i} = 1$. That makes $\hat{F}_{2,KM}(U_{(i)}) = \hat{F}_{2,n}(U_{(i)}) = 0$ where $\hat{F}_{2,KM}$ is the Kaplan-Meier estimator and $\hat{F}_{2,n}$ is the NPML estimator of F_2 for the observations in the first block. Also, in that block the weights w_i become equal to 1 and the cumulative sum diagram that will be used to calculate the WLS estimator is similar to the one used to calculate the NPMLE of $F \equiv F_1$ in the Interval Censoring, case 1 (see Groeneboom and Wellner (1992)), implying that the NPMLE of F_1, the pseudo NPMLE of F_1 and the weighted least squares estimators will coincide. On the other hand, in the second block of observations we have $\Delta_{1,(i)} = 1$. A look at the expression of the log-likelihood shows that it will be maximized making $\hat{F}_{1,n} = 1$ for all the observations in the second block, and hence the log-likelihood coincides with the log-likelihood for the right censoring problem, which is maximized in F_2 by the Kaplan-Meier estimator. Also, the cumulative sum diagram for the calculation of the weighted least squares estimator has $G_i = 0$ for those observations, making $\hat{F}_{1,WLS}(U_{(i)}) = 1$, where $\hat{F}_{1,n}$ is the NPMLE of F_1 and $\hat{F}_{1,WLS}$ is the estimator proposed by van der Laan *et al.* (1997). Putting

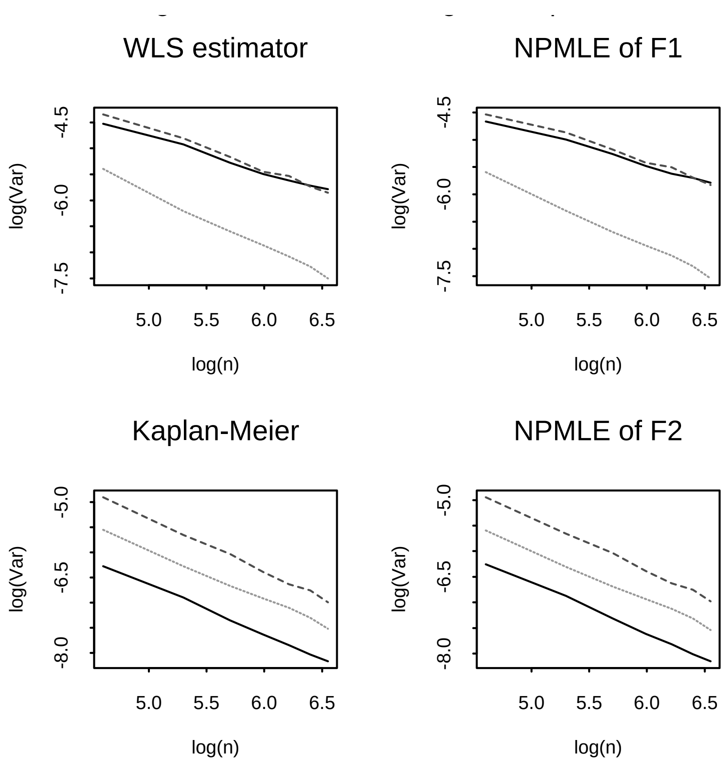

Figure 6: **log-log plot of variance of the estimators versus sample size.**

all these facts together we see that in the case where $\| F_1 - F_2 \|_\infty = 1$, all the estimators of F_1 and F_2 considered here coincide.

5 Conclusions

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

The magnitude of the variance of each estimator at the quantiles seems to explain the differences between the NPMLE and the Weighted Least Squares estimator of F_1, since the bias of both estimators are similar.

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

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

Table 3: *Summary results of the simulations for the WLS and NPML estimators of F_1 (example1, n=100, 400).*

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

Table 4: *Summary results of the simulations for the WLS and NPML estimators of F_1 (example 5, n=100, 400).*

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

Table 5: *Summary results of the simulations for the WLS and NPML estimators of F_1 (example 5, n=700).*

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

6 References

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M. AND BRUNK, H.D. (1972), *Statistical Inference under Order Restrictions*, John Wiley and Sons, New York.

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

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

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

HOLLAND, J.M., MITCHELL, T.J. AND WALBURG, H.E. (1977), Effects of prepubertal ovariectomy on survival and specific diseases in female RFM mice given 300 R of X rays, *Radiation Research*, **69**, 317-327.

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

KODELL, R., SHAW, G. AND JOHNSON, A. (1982), Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments, *Biometrics*, **38**, 43-58.

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

ROBERTSON,T., WRIGHT,F.T. AND DYKSTRA,R.L. (1988), *Order Restricted Statistical Inference*, John Wiley and Sons, New York.

TURNBULL, B.W. AND MITCHELL, T.J. (1984), Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments, *Biometrics* **40**, 41-50.

VAN DER LAAN, M.J., JEWELL, N.P. AND PETERSON, D. (1997), Efficient estimation of the lifetime and disease onset distribution, *Biometrika*, **84**, 539-554.

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