Brazilian Journal of Probability and Statistics (2001), 15, pp. 135–145.

CHARACTERIZATION OF THE NPMPLE OF THE DISEASE ONSET DISTRIBUTION FUNCTION FOR A SURVIVAL-SACRIFICE MODEL

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

Summary
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 $Y = \min\{T_2, C\}$, $\Delta_1 = 1_{\{T_1 \leq C\}}$, and $\Delta_2 = 1_{\{T_2 \leq C\}}$, T_1 and T_2 have a joint distribution function F such that $P(T_1 \leq T_2) = 1$, and are independent of C. Some authors call this model a "survival-sacrifice model". The interest here is to estimate the marginal distribution functions F_1 and F_2 of T_1 and T_2, respectively (since F is not identifiable). One possible way of doing that is by using a consistent estimator $\hat{F}_2$ for F_2 (Kaplan-Meier, for example) and then plugging it in the loglikelihood to obtain $\hat{F}_1$, the nonparametric maximum pseudo likelihood estimator (NPMPLE) of F_1. A characterization theorem of $\hat{F}_1$ is stated here and an algorithm for its calculation is presented.

Key words: Maximum likelihood estimation, nonparametric estimation; survival-sacrifice.

1 Introduction

In experiments for the study of onset and mortality from undetectable irreversible diseases (occult tumors, e.g.) a possible data structure consists of the time of death, whether the disease of interest was present at death, and if present, whether the disease was a probable cause of death. This data structure is related to moderately lethal incurable diseases when the cause of death is known. Defining the non-negative variables T_1 (time of disease onset), T_2 (time of death from the disease) and C (time of death from an unrelated cause), we observe, for the ith individual, $(Y_i, \Delta_{1,i}, \Delta_{2,i})$, where $\Delta_{1,i} = 1_{\{T_{1,i} \leq C_i\}}$, $\Delta_{2,i} = 1_{\{T_{2,i} \leq C_i\}}$, $Y_i = C_i \wedge T_{2,i} = \min\{C_i, T_{2,i}\}$, $T_{1,i}$ 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 $(T_{1,i}, T_{2,i})$. Some authors call that model a *survival-sacrifice model*.

We will suppose, without loss of generality, that $Y_1 \leq Y_2 \leq \ldots \leq Y_n$. In case of ties it is assumed that 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 case 1 of interval censoring model, also called "current status data" (see, e.g., Groeneboom and Wellner (1992)), can be seen as a particular case of this model when the disease is nonlethal, 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., $\Delta_{1,i} = 1, i = 1, \ldots n$.

For this survival-sacrifice model, 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 $x \in I\!R$, a consequence of $P(T_1 \leq T_2) = 1$. The loglikelihood function for this data structure is

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

where $K(g, G)$ is a term involving only the distribution function G and the probability density function g of variable C.

Kodell, Shaw and Johnson (1982) also studied the nonparametric estimation of $S_1 = 1 - F_1$ and $S_2 = 1 - F_2$, but their work is restricted to the case where $R(t) = S_1(t)/S_2(t)$ is 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 nonparametric maximum likelihood estimator (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.

Gomes, Groeneboom and Wellner (2001) used the Primal-Dual Interior Point algorithm to calculate the joint NPMLE of F_1 and F_2.

Van der Laan, Jewell, and Peterson (1997) proposed a weighted least squares estimator of F_1 making $F_2 = \hat{F}_{2,KM}$ (its Kaplan-Meier estimate).

Another possible way of estimating F_1 is by plugging in the Kaplan-Meier estimator $\hat{F}_{2,KM}$ (or another consistent estimator) of F_2 and calculating the nonparametric maximum pseudo likelihood estimator (NPMPLE) of F_1. The part of the loglikelihood involving F_1 is

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

where $x_i = F_1(Y_i)$. Notice that (1.1) can be written as

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

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 $\Phi(y) = (y-\hat{F}_{2,KM})\log(y-\hat{F}_{2,KM})+(1-y)\log(1-y), 0 < y < 1$. Using this representation, Dinse and Lagakos (1982) concluded that the values of $F_1(Y_i), i = 1,\ldots,n$, maximizing the pseudo loglikelihood (1.1) could be obtained by applying Theorem 1.10 in Barlow et al. (1972), i.e., the NPMPLE of F_1 would be given by the isotonic regression g^* of $g(Y_i)$ with weights $w(Y_i), i = 1,\ldots,n$. However, that Theorem is applicable to a real convex function Φ defined on $\mathbb{R}$ while the function Φ defined above is, in fact, defined on $\mathbb{R}^2$ since the value of $\hat{F}_{2,KM}$ is not supposed to be constant.

It should be mentioned here that, although the Kaplan-Meier estimator $\hat{F}_2$ is uniquely defined, except possibly at times exceeding the largest observation, the NPMPLE $\hat{F}_1$ is uniquely defined only over certain data-determined intervals. Specifically, $\hat{F}_1$ is always uniquely defined at the observed C_i's, i.e., the observations for which $\Delta_{2,i} = 0$.

In Section 2, we review the characterization of the NPMLE of the distribution function of the time of disease onset for the case 1 of interval censoring model (current status data) and present the characterization of the NPMPLE of F_1 for the survival-sacrifice model under study. In Section 3, we introduce an Iterative Convex Minorant algorithm based on the characterization of the NPMPLE of F_1. In Section 4, we present an example of a data set with the structure studied here and calculate the NPMPLE $\hat{F}_1$ and $\hat{F}_{2,KM}$.

2 Characterization of the NPMPLE of F_1

First we restate a characterization theorem for the NPMLE of the distribution function of the time of disease onset for current status data (see Groeneboom and Wellner (1992)). Defining independent positive variables X and T, we observe (T, δ) where $\delta = 1_{\{X \le T\}}$. Here X is the (completely censored) time of disease onset and T is the time of occurance of an examination (possibly an autopsy). The loglikelihood for F (the d.f. of X) is

$$\mathcal{L}(F) = \sum_{i=1}^{n} \left\{ \delta_i \log(F(T_i)) + (1-\delta_i) \log(1 - F(T_i)) \right\}. \tag{2.2}$$

We will assume, without loss of generality, that $\delta_1 = 1$ and $\delta_n = 0$ since we could maximize (2.2) for the first observations with $\delta_i = 0$ by making $F(T_i) = 0$ at those points. Similarly, we could maximize (2.2) for the last observations with $\delta_i = 1$ by making $F(T_i) = 1$ at those points.

Theorem 2.1 characterizes the NPMLE of F in terms of the Fenchel conditions (2.3) and (2.4).

Theorem 2.1 *Let $\delta_1 = 1$ and $\delta_n = 0$, and $x_i = F(T_i)$, $i = 1, \ldots, n$. A vector $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ maximizes (2.2) if and only if*

$$\sum_{j=i}^{n} \left\{ \frac{\delta_j}{x_j^*} - \frac{1-\delta_j}{1-x_j^*} \right\} \ \leq \ 0, \ i = 1, \ldots, n, \tag{2.3}$$

and

$$\sum_{i=1}^{n} \left\{ \frac{\delta_i}{x_i^*} - \frac{1-\delta_i}{1-x_i^*} \right\} x_i^* \ = \ 0. \tag{2.4}$$

Moreover, $\mathbf{x}^*$ is uniquely determined by (2.3) and (2.4).

We will now present and demonstrate an equivalent result for the survival-sacrifice model under study. Consider the problem of minimizing

$$\phi(\mathbf{x}) = -\sum_{i=1}^{n} \Big(1 - \Delta_{2(i)}\Big) \Big\{ \Delta_{1(i)} \log(x_i - k_i) + \Big(1 - \Delta_{1(i)}\Big) \log(1 - x_i) \Big\} \tag{2.5}$$

over $\mathcal{K}$ where

$$\mathcal{K} = \{ \mathbf{x} \in \mathbb{R}^n : 0 \leq x_1 \leq \ldots \leq x_n \leq 1 \}$$

subject to $x_i \geq k_i, i = 1, \ldots, n$, where $x_i = F_1(Y_i)$, $k_i = \hat{F}_{2,KM}(Y_i)$ and the vector $\mathbf{k} = (k_1, \ldots, k_n) \in \mathcal{K}$.

In other words we want to minimize $\phi(x)$ over $\mathcal{K} \cap \mathcal{L}$ where

$$\mathcal{L} = \{ \mathbf{x} \in \mathbb{R}^n : x_i \geq k_i, i = 1, \ldots, n \}.$$

Since ϕ is a convex function on $\mathcal{K}$ (a convex set of a linear vector space) and $G(\mathbf{x}) = -(\mathbf{x} - \mathbf{k})$ is a convex mapping from $\mathcal{K}$ into a normed space, by Theorem 1, page 217 in Luenberger (1969) (restated in the Appendix) there exists a vector $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ with $\hat{\lambda}_i \geq 0$, $i = 1, \ldots, n$, such that

$$\inf_{\mathbf{x} \in \mathcal{K}} \left\{ \phi(\mathbf{x}) - \sum_{i=1}^{n} \hat{\lambda}_i (x_i - k_i) \right\} = \inf_{\mathbf{x} \in \mathcal{K} \cap \mathcal{L}} \phi(\mathbf{x}).$$

So, in order to characterize the solution of the minimization problem above we introduce a vector of Lagrange multipliers $\hat{\lambda} \in I\!R^n_+$ and define

$$\psi(\mathbf{x}, \hat{\lambda}) \equiv \phi(\mathbf{x}) - \sum_{i=1}^{n} \hat{\lambda}_i \left(x_i - k_i\right).$$

Notice that we can take $\hat{\lambda}_i = 0$ if $\Delta_{1(i)} = 1$ since then the $\log{(x_i - k_i)}$ term in ϕ forces $x_i > k_i$. We may also reduce the problem to involving just those x_i's with $\Delta_{2,i} = 0$, since those with $\Delta_{2,i} = 1$ do not contribute to the function ϕ. Thus, we may take the $\hat{\lambda}_i$'s to be

$$\hat{\lambda}_i = (1 - \Delta_{2,i})(1 - \Delta_{1,i})\gamma_i$$

where we want $\gamma_i > 0$ in the cases when $\Delta_{1,i} = \Delta_{2,i} = 0$ and the solution $\mathbf{x}$ has $x_i = k_i$.

The vector of gradients of ψ with respect to $\mathbf{x}$ is given by

$$\begin{aligned} (\nabla_x \psi)_i &= -\left(1 - \Delta_{2,i}\right) \left(\frac{\Delta_{1,i}}{x_i - k_i} - \frac{1 - \Delta_{1,i}}{1 - x_i} \right) - \left(1 - \Delta_{2,i}\right) \left(1 - \Delta_{1,i}\right) \gamma_i \\ &= \left(1 - \Delta_{2,i}\right) \left\{ \left(1 - \Delta_{1,i}\right) \left(\frac{1}{1 - x_i} - \gamma_i \right) - \frac{\Delta_{1,i}}{x_i - k_i} \right\}, i = 1, \ldots, n, \end{aligned}$$

and the vector of second partial derivatives of ψ has ith coordinate

$$\frac{\partial^2}{\partial x_i^2} \psi = \left(1 - \Delta_{2,i}\right) \left\{ \frac{\Delta_{1,i}}{\left(x_i - k_i\right)^2} + \frac{1 - \Delta_{1,i}}{\left(1 - x_i\right)^2} \right\},\ i = 1, \ldots, n.$$

Thus the Fenchel conditions for minimizing ψ over $\mathcal{K}$ are given by

$$\begin{aligned} 0 &= \langle \hat{\mathbf{x}}, \nabla_x \psi(\hat{\mathbf{x}}, \hat{\lambda}) \rangle \\ &= -\sum_{i=1}^{n} \left(1 - \Delta_{2,i}\right) \hat{x}_i \left\{ \frac{\Delta_{1,i}}{\hat{x}_i - k_i} - \frac{1 - \Delta_{1,i}}{1 - \hat{x}_i} \right\} - \sum_{i=1}^{n} \hat{x}_i \hat{\lambda}_i \\ &= \sum_{i=1}^{n} \left(1 - \Delta_{2,i}\right) \hat{x}_i \left\{ \frac{1 - \Delta_{1,i}}{1 - \hat{x}_i} - \frac{\Delta_{1,i}}{\hat{x}_i - k_i} \right\} - \sum_{i=1}^{n} \hat{x}_i \hat{\lambda}_i \qquad (2.6) \end{aligned}$$

and, with $\mathbf{1}_i$ defined to be the vector with 0 in the first $i-1$ coordinates and 1 in the coordinates i through n,

$$\begin{aligned} 0 &\leq \langle \mathbf{1}_i, \nabla_x \psi(\hat{\mathbf{x}}, \hat{\lambda}) \rangle \\ &= \sum_{j=i}^{n} (1-\Delta_{2,j}) \left\{ \left(1-\Delta_{1(j)}\right) \left(\frac{1}{1-\hat{x}_j} - \gamma_j \right) - \frac{\Delta_{1(j)}}{(\hat{x}_j - k_j)} \right\}, \quad (2.7) \\ & \quad i = 1, \ldots, n. \end{aligned}$$

We take the Lagrange multipliers to be of the form

$$\hat{\lambda}_i = (1-\Delta_{2,i})(1-\Delta_{1,i})\gamma_i 1_{\{\hat{x}_i = k_i\}} \quad (2.8)$$

for some $\gamma_i > 0$. Then we have

$$\hat{x}_i = k_i, \qquad \text{if} \qquad \hat{\lambda}_i > 0 \quad \text{and} \quad \Delta_{1,i} = 0, \quad i = 1, \ldots, n; \quad (2.9)$$

and

$$\hat{\lambda}_i = 0, \qquad \text{otherwise.} \quad (2.10)$$

The Fenchel conditions characterize the arg min of a convex function on a convex cone (see Lemma 2.1, page 8 in Jongbloed (1995), for example, restated in the Appendix). In our case, condition (2.7) is equivalent to condition (ii) in the Lemma since the vectors $(1, \ldots, 1)$, $(0, 1, \ldots, 1)$, $\ldots$, $(0, \ldots, 0, 1)$ generate cone $\mathcal{K}$.

Theorem 2.2 *Suppose that (2.6) to (2.10) hold. Then $\hat{\mathbf{x}}$ minimizes ϕ over $\mathcal{K} \cap \mathcal{L}$.*

Proof: Since the function ϕ is continuous and $\mathcal{K} \cap \mathcal{L}$ is a compact subset of a normed linear space, there exists $\hat{\mathbf{x}}$ minimizing $\phi(\mathbf{x})$ over $\mathcal{K} \cap \mathcal{L}$. Since $\mathcal{K}$ is a convex subset of $\mathbb{R}^n$, and $G(\mathbf{x}) = -(\mathbf{x} - \mathbf{k})$ is a convex mapping from $\mathcal{K}$ into $\mathbb{R}^n$, we have, by Theorem 1, page 217 in Luenberger (1969), that there exists a vector $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ with $\hat{\lambda}_i \geq 0$ such that

$$\inf_{\mathbf{x} \in \mathcal{K}} \left\{ \phi(\mathbf{x}) - \sum_{i=1}^{n} \hat{\lambda}_i (x_i - k_i) \right\} = \inf_{\mathbf{x} \in \mathcal{K}} \phi(\mathbf{x})$$

subject to $x_i \geq k_i, i = 1, \ldots, n$.

Moreover, from Luenberger (1969), $\hat{\mathbf{x}}$ minimizes $\psi(\mathbf{x}, \hat{\lambda})$ on $\mathcal{K} \cap \mathcal{L}$ and $\sum_{i=1}^{n} \hat{\lambda}_i (\hat{x}_i - k_i) = 0$. Since $\hat{\lambda}_i > 0$ if and only if $\hat{x}_i = k_i$, we have

$$\phi(\hat{\mathbf{x}}) = \psi(\hat{\mathbf{x}}, \hat{\lambda}) \leq \psi(\mathbf{x}, \hat{\lambda}) \leq \phi(\mathbf{x}).$$

But (2.6) and (2.7) are the Fenchel conditions to minimize ψ over $\mathcal{K}$. So, $\hat{\mathbf{x}}$ obtained from those conditions will minimize $\psi(\mathbf{x}, \hat{\lambda})$ over $\mathcal{K}$, and hence minimize ϕ over $\mathcal{K} \cap \mathcal{L}$. $\square$

3 The iterative convex minorant algorithm

An algorithm can be developed based on Theorem 2.2. The iterative convex minorant algorithm is an adaptation of the ICM algorithm for the calculation of the NPMLE of the distribution function of the time of disease onset for the case 2 of interval censoring (see Groeneboom and Wellner (1992)).

(0) Take $x_i^{(0)} = (k_i + 1)/2, \lambda_i = 0, i = 1, \ldots, n$. Set $k = 0$.

(i) Form

$$V_i^{(k)} = \sum_{j=1}^{i} x_j^{(k)} \frac{\partial^2}{\partial x_j^2} \psi\left(\mathbf{x}^{(k)}\right) - \sum_{j=1}^{i} \left(\nabla_x \psi\right)_j \left(\mathbf{x}^{(k)}\right), \qquad i = 1, \ldots, n$$

$$G_i^{(k)} = \sum_{j=1}^{i} \frac{\partial^2}{\partial x_j^2} \psi\left(\mathbf{x}^{(k)}\right), \qquad i = 1, \ldots, n$$

(ii) Form the cumulative sum diagram $\left\{\left(G_i^{(k)}, V_i^{(k)}\right), i = 1, \ldots, n\right\}$, compute its greatest convex minorant $GCM^{(k)}$ and $x_i^{(k+1)}$ = left-derivative of $GCM^{(k)}$ at $G_i^{(k)}$.

(iii) If $x_i^{(k+1)} \leq k_i$, set $x_i^{(k+1)} = k_i$; set $\lambda_i^{(k+1)} = 0$ if $x_i^{(k+1)} > k_i$.

(iv) Verify whether the Fenchel conditions (2.6) and (2.7) are satisfied using the current values $\hat{\mathbf{x}}^{(k)}, \lambda^{(k)}$. If the conditions are satisfied, stop; otherwise replace $\hat{\mathbf{x}}^{(k)}$ by $\hat{\mathbf{x}}^{(k+1)}$, and continue.

(v) Find the remaining $\lambda_i^{(k+1)}$'s from points $\{i_m\}$ where equality between $GCM^{(k)}$ and the cusum diagram holds, i.e., points where

$$0 = \sum_{j=1}^{i_m} (1 - \Delta_{2,j}) \left\{ (1 - \Delta_{1,j}) \left(\frac{1}{1 - \hat{x}_j^{(k+1)}} - \gamma_j^{(k+1)} \right) - \frac{\Delta_{1,j}}{\left(\hat{x}_j^{(k+1)} - k_j\right)} \right\}.$$

Go to **(i)**.

In step (iv) above, we say that the Fenchel condition (2.6) is satisfied if the absolute value of the expression on the right-hand side of (2.6) is negligible. The Fenchel condition (2.7) is said to be satisfied if the value of the expression on the right-hand side of (2.7) is greater than a negligible negative value for $i = 1, \ldots, n$. This algorithm was used to calculate $\hat{F}_1$ for the example in the next section.

4 Example

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

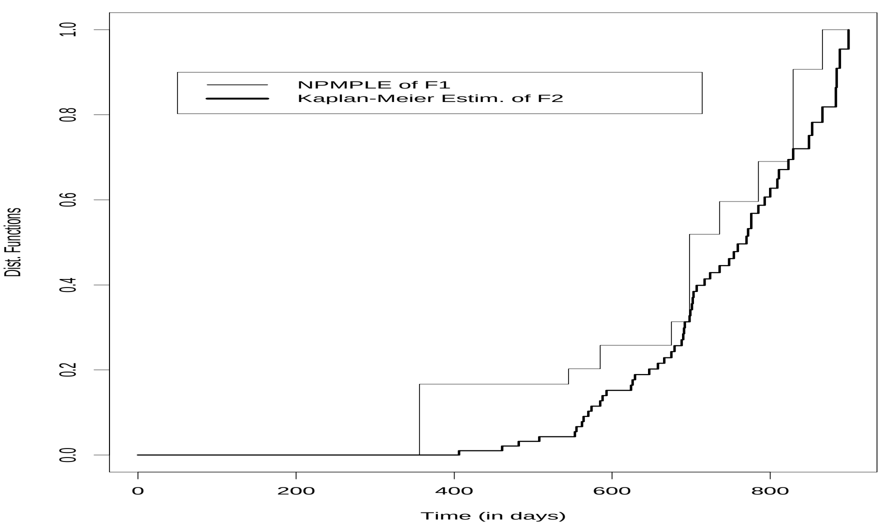

Figure 1
NPMPLE of F_1 and Kaplan-Meier estimator of F_2.

The data in Table 1 were studied by Dinse and Lagakos (1982) and Turnbull and Mitchell (1984) and represent the ages at death (in days) of 109 female RFM mice. The disease of interest is reticulum cell sarcoma (RCS). These mice formed the control group in a survival experiment to study the effects of prepubertal ovariectomy in mice given 300 R of X-rays.

Figure 1 shows the Kaplan-Meier estimate of F_2 and the NPMPLE of F_1. As mentioned in Section 2, we should have $\lambda_i > 0$ for those observations with $\hat{F}_1(Y_i) = \hat{F}_2(Y_i)$ and with $\Delta_{1,i} = 0$. In our example, two observations (the ones with Y equal to 694 and 828) are in that situation. For those observations, we have λ equal to 1.455938 and 3.276873, respectively (λ is equal to zero for all the other observations). The value of the expression on the right-hand side of (2.6) is 4.774×10^{-6}. The minimum value of the expression on the right-hand side of (2.7) is -1.973153×10^{-6}. These values are negligible, showing that both Fenchel conditions (2.6) and (2.7) are satisfied.

We can notice that the estimate of F_2 has more discontinuity points than that of F_1. Also, the magnitudes of the variations of $\hat{F}_2$ at its discontinuity points are smaller than those of $\hat{F}_1$. These facts are related to the different convergence rates of the estimators of F_1 and F_2 ($n^{-1/3}$ and $n^{-1/2}$, respectively).

5 Discussion

The importance of the characterization Theorem 2.2 is that it allows checking whether an estimate of F_1 is in fact the nonparametric maximum pseudo likelihood estimator (NPMPLE) $\hat{F}_1$, as seen for the example in Section 4. The characterization result,then, allows checking whether any algorithm proposed to calculate $\hat{F}_1$ actually converges to $\hat{F}_1$. That is the case of the Iterative Convex Minorant algorithm proposed in Section 3. The example in Section 4 illustrates that situation since both Fenchel conditions (2.6) and (2.7) are satisfied for the data in Table 1. The version of the ICM algorithm proposed here is an adaptation of the ICM algorithm used to estimate the nonparametric maximum likelihood estimator for the case 2 of interval censoring (see Groeneboom and Wellner (1992)). Jongbloed (1995) presents a general version of the algorithm. Another algorithm that may be used to calculate $\hat{F}_1$ is the Primal-Dual Interior Point algorithm (see Wright (1997) or Groeneboom (1998)). However, no performance comparison study has been carried out so far for those algorithms.

Appendix

(Lemma 2.1, page 8, Jongbloed (1995)) Let $\phi : \mathbb{R}^n \to (-\infty, \infty]$ be a continuous convex function such that ϕ is continuously differentiable on the set $\{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) < \infty\}$. Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex cone. Then

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{K}} \phi(\mathbf{x})$$

if and only if $\hat{\mathbf{x}} \in \mathcal{K}$ satisfying (the Fenchel conditions)

$$(i) \qquad \langle \hat{\mathbf{x}}, \nabla\phi(\hat{\mathbf{x}})\rangle = 0.$$

$$(ii) \qquad \langle \mathbf{x}, \nabla\phi(\hat{\mathbf{x}})\rangle \geq 0,\ \forall\ \mathbf{x} \in \mathcal{K}$$

(Theorem 1, page 217, Luenberger (1969)) Let $\mathcal{X}$ be a linear vector space, Λ a normed space, $\mathcal{K}$ a convex subset of $\mathcal{X}$, and C the positive cone in Λ. Assume that C contains an interior point. Let ϕ be a real-valued convex functional on $\mathcal{K}$ and G a convex mapping from $\mathcal{K}$ into Λ. Assume the existence of a point $\mathbf{x}_1 \in \mathcal{K}$ for which $G(\mathbf{x}_1) \leq \mathbf{0}$ (i.e., $G(\mathbf{x}_1)$ is an interior point of $N = -C$). Let

$$\mu_0 = \inf f(\mathbf{x}) \tag{A.1}$$

subject to $\mathbf{x} \in \mathcal{K}$, and $G(\mathbf{x}) \leq \mathbf{0}$, and assume μ_0 is finite. Then there is an element $\lambda_0^* \geq \mathbf{0}$ in Λ^* (the ”normal dual” of Λ, i.e., the space of all the bounded linear functionals on Λ) such that

$$\mu_0 = \inf_{\mathbf{x}\in\mathcal{K}} \left\{\phi(\mathbf{x}) + \langle G(\mathbf{x}), \lambda_0^*\rangle\right\}. \tag{A.2}$$

Furthermore, if the infimum is achieved in (A.1) by an $\mathbf{x}_0 \in \mathcal{K}$, with $G(\mathbf{x}_0) \leq \mathbf{0}$, it is achieved by $\mathbf{x}_0$ in (A.2) and $\langle G(\mathbf{x}_0), \lambda_0^*\rangle = 0$.

Acknowledgements

The author would like to thank Dr. Jon A. Wellner and Dr. Piet Groeneboom for the useful discussions during the preparation of this paper. The author is also thankful for the comments by the referee which helped improving substantially the final version of the paper.

References

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

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

Gomes, A.E., Groeneboom, P. and Wellner, J.A. (2001). Nonparametric estimation of the lifetime and disease onset distributions for a survival-sacrifice model. Paper submitted for publication.

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*. Basel: Birkhauser Verlag.

Jongbloed, G. (1995). Three Statistical Inverse Problems, estimators-algorithms-asymptotics. Ph.D. Thesis, Delft University of Technology, The Netherlands.

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*. New York: Wiley.

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

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*. Philadelphia: SIAM.