The NPMLE for Doubly Censored Current Status Data

Mark J. van der Laan and Nicholas P. Jewell

Division of Biostatistics

University of California, Berkeley

May 15, 2001

Abstract

In biostatistical applications interest often focuses on the estimation of the distribution of time T between two consecutive events. If the initial event time is observed and the subsequent event time is only known to be larger or smaller than an observed point in time, then the data is described by the well understood singly censored current status model, also known as interval censored data, case I. Jewell, Malani and Vittinghoff (1994) extended this current status model by allowing the initial time to be unobserved, but with its distribution over an observed interval $[A, B]$ known to be uniformly distributed; the data is referred to as doubly censored current status data. These authors used this model to handle applications in AIDS partner studies focussing on the NPMLE of the distribution G of T. The model is a submodel of the current status model, but the distribution G is essentially the derivative of the distribution of interest F in the current status model.

In this paper we establish that the NPMLE of G is uniformly consistent and that the resulting estimators for the $n^{1/2}$-estimable parameters are efficient. We propose an iterative weighted Pool-Adjacent-Violator-Algorithm to compute the estimator. It is also shown that, without smoothness assumptions, the NPMLE of F converges at rate $n^{-2/5}$ in L^2-norm while the NPMLE of F in the nonparametric current status data model converges at rate $n^{-1/3}$

in L^2-norm, which shows that there is a substantial gain in using the submodel information.

Key words: doubly censored current status data, nonparametric maximum likelihood estimator, efficiency, AIDS partner studies

1 Introduction

In many longitudinal studies, interest focuses on the distribution of the length of the interval between two events. This situation is particularly common in epidemiological investigations of the natural history of a disease. Jewell, Malani and Vittinghoff (1994) give two examples that arise from studies of Human Immunodeficiency Virus (HIV) disease.

Notationally, let I and J be the chronological times of the two events, respectively, with $I \leq J$ always. Consider two observed times A and B, and assume that individual observations are randomly sampled conditional on $A \leq I \leq B$. Suppose that, for a given individual, we further observe whether the second event has occurred at the *single point in time*, B. That is, in addition to A and B, we observe $\Delta \equiv I(J \leq B)$. To avoid unidentifiability, we assume that the conditional distribution of I on $[A, B]$ is known, and, following Jewell, Malani and Vittinghoff (1994), we take this distribution to be the uniform distribution on $[A, B]$. We are concerned with nonparametric estimation of the distribution G of $T = J - I$, based on n i.i.d. observations $Y_i = (A_i, B_i, \Delta_i \equiv I(J_i \leq B_i)$, under the assumption that (i) T is independent of (A, B) and (ii) T is independent of I, given (A, B).

We refer to this data structure as *doubly censored current status data* since observation of both I and J are interval censored. If I is observed, then the data structure is commonly referred to as (singly censored) *current status data*. The qualifier "singly censored" serves to differentiate this structure from the data structure given above. Previous work and examples of current status data can be found in Diamond, McDonald and Shah (1986), Jewell and Shiboski

(1990), Diamond and McDonald (1991), and Keiding (1991). In that case, one observes n i.i.d. observations $Y_i = (C_i, \Delta_i)$, where $C_i \equiv B_i - I_i$ and $\Delta_i \equiv I(J_i \leq B_i) = I(T_i \leq C_i)$. We have

$$P(\Delta = 1 \mid C = c) = G(c) \text{ and } P(\Delta = 0 \mid C = c) = 1 - G(c).$$

Hence, the density of the distribution P_G of (C, Δ) w.r.t. the distribution H of C is given by

$$p_G(c, \Delta) \equiv \frac{dP_G}{dH}(c, \Delta) = G(c)\Delta + (1 - G(c))(1 - \Delta). \tag{1}$$

Here H is unspecified, but it should be noticed that the model for H does not affect maximum likelihood estimation for G and it does not affect the information bound for smooth functionals of G either. If G is assumed to be completely unknown, then this model is also referred to as interval censoring, case I, by Groeneboom (1991). Uniform consistency, pointwise convergence to a limit distibution at rate $n^{-1/3}$ and efficiency of smooth functionals of the nonparametric maximum likelihood estimator (NPMLE) G_n of G, is established in Groeneboom and Wellner (1992). Alternative shorter proofs of efficiency of the NPMLE of smooth functionals of G (e.g. its moments) are given by van de Geer (1994) and Huang and Wellner (1995). In van de Geer (1993) it is also proved that the NPMLE G_n converges at rate $n^{-1/3}$ in L^2-norm.

Returning to the doubly censored situation, if we define $C = B - A$, then by assumptions (i) and (ii) we have

$$P(\Delta = 1 \mid A, B) = P(\Delta = 1 \mid C) = F_G(C) \equiv \frac{1}{C} \int_0^C G(t)dt. \tag{2}$$

Thus, for inference on G, we can reduce the data to $Y = (C, \Delta)$. Let H be the unknown sampling cumulative distribution of $C = B - A$. The density of the data $Y = (C, \Delta)$ w.r.t. H is

$$p_G(c, \Delta) = F_G(c)\Delta + (1 - F_G(c))(1 - \Delta). \tag{3}$$

Because F_G is a distribution function this is just a *submodel* of (singly censored) current status data.

There is a simple direct relation between F_G and G since

$$G(c) = \frac{d}{dc}\{cF_G(c)\} = F_G(c) + cf_G(c) \text{ with } f_G \equiv \tfrac{dF_G}{dx}. \tag{4}$$

From current status data, estimation of F_G corresponds with estimation of a monotone density and hence this relation shows that, for doubly censored current status data, estimation of G is comparable with estimation of a derivative of a density.

This doubly censored current status data model was introduced in Jewell, Malani and Vittinghoff (1994) and the NPMLE was applied to two applications in AIDS-research, but no asymptotic results were established. We suggest that readers consult this paper for a description and analysis of the specific examples, including illustrations of the NPMLE. In this paper it is shown that the NPMLE G_n of G is uniformly consistent and that, in spite of the fact that G_n converges at most at rate $n^{-1/5}$, smooth functionals $\int r(x)(1-G_n)(x)dx$ are $n^{1/2}$-consistent, asymptotically normal and asymptotically efficient. It will also be shown that F_{G_n} converges at rate $n^{-2/5}$ in $L^2(H)$-norm.

Based on the above results, we make the following observations regarding the gain from using the extra structure of the doubly censored current status data model – in (3) it is required that $F_G = F$ for which $F(c) + cf(c)$ is a distribution– relative to the nonparametric current status data model where F is unspecified. For estimation of F_G and G this extra structure helps in the sense that F_G is estimated at rate $n^{-2/5}$ (compare with the rate $n^{-1/3}$ for the nonparametric current status data NPMLE) and that its derivative (i.e. G) is uniformly consistent (while the density of the current status data NPMLE which is a step function does not even exist). For smooth functionals the picture is different. Because of the linearity of $G \rightarrow F_G$ one can write $\mu = \int r(1-G)dx$ as $\mu = \int r_1(1-F_G)dx$ for some r_1 (see section 4). In spite of the extra structure, the efficiency bound for estimation of μ in the nonparametric current status data model equals the information bound for estimation of μ in the doubly censored current status data model (van der Laan, Bickel and Jewell, 1997; Rabinowitz and Jewell, 1996; and section 4). Therefore $\int r_1(1-F_n)dx$ is an efficient estimate of μ where F_n is the nonparametric current status data NPMLE. However, our efficiency proof for the true NPMLE $\int r_1(1-F_{G_n})dx$ of μ in this paper establishes a rate $n^{-4/5}$ for the second order term while the second order term for the estimator $\int r_1(1-F_n)dx$ is of the order $n^{-2/3}$ (Huang and Wellner, 1995). Thus the extra structure appears to improve the second order efficiency for estimation of smooth functionals of F_G and thus of G and therefore we propose use of the true NPMLE for estimation of smooth functionals.

The organisation of the paper is as follows. In section 2, relying on results in Jongbloed (1995), we discuss a general iterative weighted-Pool-Adjacent-Violator-Algorithm (IWPAVA), for maximizing a concave function under monotonicity constraints and apply it to our loglikelihood. In section 3 we prove that if $\int_0^\infty dH(x)/x < \infty$, then the NPMLE F_{G_n} converges in $L^2(H)$ at rate $n^{-2/5}$ to F_G. We also show that, if G is continuous, then G_n is uniformly consistent on the support of H. In section 4 we establish asymptotic linearity and efficiency of smooth functionals of G. We conclude with a discussion of the issue of the convergence rate for the NPMLE G_n.

2 The iterative weighted Pool-Adjacent-Violator-Algorithm for the NPMLE

From Barlow, Bartholomew, Bremner and Brunk (1972), Groeneboom and Wellner (1992) and Jongbloed (1995), the maximizer of a concave function (e.g. loglikelihood) over the set of all cumulative distribution functions can be computed using a (possibly modified) iterative weighted Pool-Adjacent-Violator algorithm or equivalently with a (possibly modified) iterative convex minorant algorithm. In the next subsections we will state this general result and apply it to our loglikelihood.

2.1 The modified iterative weighted Pool-Adjacent-Violator-Algorithm to maximize a concave function under order restrictions.

The proof of proposition 1.1 in Groeneboom and Wellner (1992) (see also Jongbloed, 1995) establishes the following result.

Theorem 2.1 *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function and let $A = \{\vec{x} : 0 < x_1 \leq x_2 \leq \ldots \leq x_n < 1\}$. Let $t_1 < \ldots < t_n$ be a given set of time-points. For a given cdf G we identify G with $G(t_1), \ldots, G(t_n)$ so that $\phi(G) = \phi(G(t_1), \ldots, G(t_n))$. Let $\phi_j(G) = \frac{d}{dx_j}\phi(\vec{x})\Big|_{\vec{x}=G}$ be the j-th partial derivative of ϕ at G.*

Then $G_n = (G_n(t_1), \ldots, G_n(t_n))$ minimizes ϕ over A if and only if

$$\sum_{j \geq i} \phi_j(G_n) \geq 0 \text{ for } i = 1, \ldots, n, \text{ and} \sum_{j=1}^{n} \phi_j(G_n) G_n(t_j) = 0. \tag{5}$$

If ϕ is strictly convex on A, then ϕ has a unique minimum G_n.

An immediate corollary of Theorem 2.1 is the following.

Corollary 2.1 *We have that $G_n = (G_n(t_1), \ldots, G_n(t_n))$ minimizes ϕ over A if and only if G_n minimizes $G \rightarrow \Psi(G \mid G_n)$ over A where Ψ is defined by*

$$G \rightarrow \sum_{j=1}^{n} \{G(t_j) - Y_{j,G_n}\}^2 w_{j,G_n},$$

$$Y_{G_n,j} = \frac{\phi_j(G_n)}{w_{j,G_n}} + G_n(t_j)$$

and $\{w_{j,G_n} : j = 1, \ldots, n\}$, is an arbitrary sequence of positive numbers.

This corollary is proved by applying Theorem 2.1 to ψ and noting that it gives the same characterizing inequality and equality constraints (5).

For a given estimator G_n^0 the isotonic regression problem, i.e. minimizing $\Psi(\cdot \mid G_n^0)$ over A, is solved with the simple and very fast weighted Pool-Adjacent-Violator-Algorithm (Barlow, Bartholomew, Bremner and Brunk, 1972). Therefore Corollary 2.1 suggests the following iterative weighted Pool-Adjacent-Violator-Algorithm: Set G_n^0 equal to an initial estimator and iterate $G_n^{k+1} = \min_A^{-1} \Psi(\cdot \mid G_n^k)$, $k = 0, 1, \ldots$. This algorithm corresponds with the iterative convex minorant algorithm proposed in Groeneboom and Wellner (1992) who highlight its fast practical performance. They suggest setting $w_{j,G}$ to be equal to the diagonal elements of the matrix of second order partial derivatives of ϕ to accelerate convergence of the algorithm. In this way the algorithm can be viewed as an approximation of the Newton-Raphson-algorithm since each step corresponds with minimizing the second order Taylor approximation of ϕ at G_n^k, but where all the off diagonal elements of the matrix of second order derivatives of Φ are set equal to zero (Jongbloed 1995, section 2.3).

Jongbloed (1995) considers the convergence properties of the iterative convex minorant algorithm. He shows that if the above algorithm converges, then it converges to the minimum. However, he also points out that there is no guarantee that the algorithm converges. He shows that the following modification of the algorithm guarantees the wished convergence.

Step 1: Set G_n^0 equal to an initial estimator and let $\epsilon \in (0, 0.5)$ be given.

Step 2: $G_n^{k+1,*} = \min_A^{-1} \Psi(\cdot \mid G_n^k)$.

Step 3: If $\Phi(G_n^{k+1,*}) < \Phi(G_n^k) + (1-\epsilon)\langle d\Phi(G_n^k)(G_n^{k+1,*} - G_n^k)\rangle$, then set $G_n^{k+1} = G_n^{k+1,*}$.

Else find λ such that $Y \equiv \lambda G_n^{k+1,*} + (1-\lambda)G_n^k$ satisfies:

$$(1-\epsilon)d\Phi(G_n^k)(Y - G_n^k) \leq \Phi(Y) - \Phi(G_n^k) \leq \epsilon d\Phi(G_n^k)(Y - G_n^k). \tag{6}$$

Set $G_n^{k+1} = Y$.

The latter Y can be found with the following binary search: Set $\text{low} = 0, \lambda = 1/2, \text{high} = 1$. If the first equality fails, then set $\text{low} = 1/2$, $\lambda = 3/4$ and $\text{high} = 1$. If the second equality fails, then set low $= 0$, $\lambda = 1/4$ and high $= 1/2$. Keep updating low, λ and high in this manner untill (6) holds.

Step 4 Repeat step 2 and 3 till convergence.

2.2 Application to the doubly censored current status data model

For the sake of presentation we will assume that C is continuous so that we have no ties, but the formulas below are trivially generalized to discrete data. It will be assumed that the observations are ordered w.r.t. to the monitoring times so that $C_1 < C_2 < \ldots < C_n$. We will represent this ordered set of monitoring times with $t_1 < t_2 < \ldots < t_n$. It will be assumed that $\Delta_1 = 1$ and $\Delta_n = 0$ so that the loglikelihood is maximized at a G with $G(t_1) > 0$ and $G(t_n) < 1$; otherwise just delete elements of the tails of $t_1 < t_2, \ldots < t_n$ till it holds. Since the loglikelihood only depends on $F_G(t_i)$, $i = 1, \ldots, n$, it is maximized at a step function (piecewise constant) G with jumps at $t_1 < \ldots < t_n$. The loglikelihood as a function of $G = (G(t_1), \ldots, G(t_n))$, with G piecewise constant, multiplied by -1, is given by:

$$\phi(G) = -\sum_{i=1}^{n} \log(F_G(C_i))\Delta_i + \log(1 - F_G(C_i))(1 - \Delta_i).$$

Note that ϕ is strictly convex on A. We have

$$\phi_j(G) = \sum_{i=1}^{n} \frac{\Delta_i - F_G(t_i)}{F_G(t_i)\bar{F}_G(t_i)} \times \frac{d}{dx_j} F_G(t_i), \text{ where } x_j = G(t_j),$$

and

$$F_G(t_i) = \frac{1}{t_i} \sum_{k=1}^{i} x_{k-1}(t_k - t_{k-1}).$$

Therefore

$$\frac{d}{dx_j} F_G(t_i) = I(j \leq i - 1)\frac{1}{t_i}(t_{j+1} - t_j)$$

and

$$\phi_j(G) = -\sum_{i=1}^{n} \frac{\Delta_i - F_G(t_i)}{F_G(t_i)\bar{F}_G(t_i)} \times \left\{ I(j \le i-1) \frac{1}{t_i}(t_{j+1} - t_j) \right\}. \tag{7}$$

Following the advice of Groeneboom and Wellner (1992) we set the weights w_j of the iterative WPAVA equal to the diagonal elements of the matrix of second order derivatives.

Straightforward calculation shows:

$$\begin{aligned} w_{j,G} &\equiv \left. \frac{d^2}{dx_j^2}\phi(\vec{x}) \right|_{\vec{x}=G} \qquad (8) \\ &= \sum_{k=1}^{n} \left\{ I(j \le k-1) \frac{t_{j+1} - t_j}{t_k} \right\}^2 \left\{ \frac{1}{F_G(t_k)(1 - F_G(t_k))} + \frac{(\Delta_k - F_G(t_k))(1 - 2F_G(t_k))}{F_G^2(t_k)(1 - F_G)^2(t_k)} \right\}. \end{aligned}$$

The NPMLE G_n can now be computed with the modified iterative WPAVA given in the preceding subsection with

$$Y_{G_n^k, j} = \frac{\phi_j(G_n^k)}{w_{j,G_n^k}} + G_n^k(t_j)$$

and $\phi_j(G_n^k)$, w_{j,G_n^k} are defined in (7) and (8), respectively.

3 L2-consistency and uniform consistency

The following theorem establishes a rate result for the NPMLE F_{G_n}.

Theorem 3.1 *Assume that H has compact support and $\int_0^\infty \frac{dH(x)}{x} < \infty$. Let G_n be the NPMLE of G. Then*

$$\|F_{G_n} - F_G\|_H \equiv \sqrt{\int (F_{G_n} - F_G)^2(x)dH(x)} = O_P(n^{-2/5}).$$

Proof. We will follow closely the proof of van de Geer for proving the Hellinger-consistency rate $n^{-1/3}$ for the nonparametric current status data NPMLE (see example 4.8(a), page 35, van de Geer, 1993). We first change our notation to her setting. Parametrize the distributions in our model with $\theta = K_G = \int_0^{\cdot} G(y)dy$ so that the parameter space is given by $\Theta = \{K_G : G \text{ cdf}\}$.

Then the model is given by $\{P_\theta : \theta \in \Theta\}$, where the monitoring density h of C is considered fixed. Let $\mu = \mu_1 \times \nu$ with $d\mu_1(x) = \frac{h(x)dx}{x}$ and ν be the counting measure on $\{0, 1\}$. We note that, by assumption, μ is a finite dominating measure for this model. For $\theta \in \Theta$ we define $p_\theta \equiv dP_\theta/d\mu$ as the density w.r.t. μ:

$$p_\theta(c, \delta) = \theta(c)^\delta (c - \theta(c))^{1-\delta}.$$

The NPMLE of the true θ_0 is defined by $\theta_n = \max^{-1}_{\theta\in\Theta} \int \log(p_\theta(c, \delta))dP_n(c, \delta)$, where P_n is the empirical distribution function. Let

$$H^2(P_{\theta_n}, P_\theta) \equiv 1/2 \int \left(\sqrt{\frac{dP_{\theta_n}}{d\mu}} - \sqrt{\frac{dP_\theta}{d\mu}} \right)^2 d\mu$$

be the squared Hellinger-norm between P_{θ_n} and P_θ. Suppose that

$$H(P_{\theta_n}, P_\theta) = O_P(n^{-2/5}). \tag{9}$$

Since $\int(\sqrt{\theta_n} - \sqrt{\theta})^2 dH(x)/x \leq H^2(P_{\theta_n}, P_\theta)$ this implies that $\int(\sqrt{\theta_n} - \sqrt{\theta})^2 dH(x)/x = O_P(n^{-4/5})$. Now, note that $\int(\sqrt{F_{G_n}} - \sqrt{F_G})^2 dH(x) = \int(\sqrt{\theta_n} - \sqrt{\theta})^2 dH(x)/x = O_P(n^{-4/5})$. Finally,

$$\begin{aligned} \int (F_{G_n} - F_G)^2 dH(x) &= \int (\sqrt{F_{G_n}} - \sqrt{F_G})^2 (\sqrt{F_{G_n}} + \sqrt{F_G})^2 dH(x) \\ &\leq 4 \int (\sqrt{F_{G_n}} - \sqrt{F_G})^2 dH(x) = O_P(n^{-4/5}). \end{aligned}$$

Thus for proving the theorem it suffices to prove (9).

We can now apply the proof of van de Geer (1993, example 4.8a) to our model and we refer to van de Geer for the definition of the entropy used below. First, the map $\theta \to p_\theta$ is linear and Θ is convex. Since Θ is contained in the set of convex functions with derivative bounded by 1 we have the following entropy bound:

$$\mathcal{H}^B(\delta, \{\sqrt{p_\theta} : \theta \in \Theta\}, \|\cdot\|_P) \leq \frac{const}{\sqrt{\delta}}$$

uniformly in P, and obviously also uniformly in any measure bounded by a fixed constant (Birman and Solomak, 1967). Since μ is finite, we thus have this bound for the entropy of $\{p_\theta : \theta \in \Theta\}$ endowed with Hellinger metric to our disposal:

$$\mathcal{H}^B(\delta, \{\sqrt{p_\theta} : \theta \in \Theta\}, \|\cdot\|_\mu) \leq \frac{const}{\sqrt{\delta}}.$$

Furthermore, again because μ is finite (i.e. $1/\sqrt{p_{\theta_0}}$ is P_0-square integrable) the same bound is valid for $\mathcal{G}_u = \{\sqrt{p_\theta / p_{u,\theta}} : \theta \in \Theta, p_{u,\theta} = up_\theta + (1-u)p_{\theta_0}\}$, equipped with the empirical metric $\|\cdot\|_{P_n}$ (see van de Geer, 1993, for the definition). Now, application of Theorem 4.5 of van de Geer (1993) yields $H(P_{\theta_n}, P_\theta) = O_P(n^{-2/5})$. □

This result can now also be used to prove uniform consistency of G_n.

Corollary 3.1 *Assume* $\int_0^\infty \frac{h(x)}{x} dx < \infty$ *and that* G *is continuous. If* $h > 0$ *on* $[0, \tau)$*, where we allow* $\tau = \infty$*, then* $\sup_{x \in [0,\tau)} \mid G_n(x) - G(x) \mid \to 0$ *almost surely.*

Proof. By Helly's selection theorem, G_n has a subsequence G_{n_k} for which $G_{n_k}(x) \to G_\infty(x)$ for some non-decreasing function between 0 and 1 at every continuity point x of G_∞. This implies that $f_n(x) \equiv \left(\sqrt{F_{G_{n_k}}(x)} - \sqrt{F_{G_\infty}(x)}\right)^2 \to 0$ for all x. Now $f_n(x) \leq 2$ where the constant function 2 is integrable w.r.t. μ: $\int 2 d\mu(x) < \infty$. By the dominated convergence theorem this implies $\int \left(\sqrt{F_{G_{n_k}}} - \sqrt{F_{G_\infty}}\right)^2 dH(x) \to 0$. By Theorem 3.1 $\int \left(\sqrt{F_{G_{n_k}}} - \sqrt{F_G}\right)^2 dH(x) \to 0$. This proves that $\int (\sqrt{F_{G_\infty}} - \sqrt{F_G})^2 (x) dH(x) = 0$. Thus, if $h > 0$ on $[0, \tau)$, $\int_0^x G_\infty(y) dy = \int_0^x G(y) dy$ for $x \in (0, \tau)$. This proves $G_\infty = G$ Lebesgue a.e. on $[0, \tau)$. Since G is continuous this is only possible if G_∞ is continuous. As a consequence, we have that $G_\infty(x) = G(x)$ for each $x \in [0, \tau)$. Thus $G_{n_k}(x) \to G(x)$ for each $x \in [0, \tau)$. Since G is continuous and G_{n_k} is monotone this proves that G_{n_k} converges uniformly to G on $[0, \tau)$. Finally, since we could also have applied this proof to any subsequence of G_n (instead of the original sequence G_n) each subsequence of G_n has a uniformly convergent subsequence. □

4 Efficiency of Smooth functionals

We have the following simple relation between smooth functionals of G and smooth functionals of F_G.

Lemma 4.1 *Let r have a Lebesgue derivative r' and satisfy $\lim_{t\to\infty} tR(t) = 0$, where $R(t) = \int_0^t r(s)ds$. Then*

$$\int R(t)dG(t) = \int r(s)(1-G)(s)ds = \int r_1(t)(1-F_G)(t)dt \text{ where } r_1(t) = -tr'(t).$$

Proof. By Fubini's theorem we have

$$\begin{aligned} \int r_1(1-F_G)dx &= \int r_1(x)\int_x^\infty \frac{1}{y^2}\int_0^y tdG(t)dydx \\ &= \int \left\{\int \frac{I(y\geq t)}{y^2}\left(\int r_1(x)I(Y>x)dx\right)dy\right\} tdG(t) \\ &= \int \left\{\int_t^\infty \frac{R_1(y)}{y^2}dy\right\} tdG(t). \end{aligned}$$

Now note that $R_1(t) = \int_0^t r_1(s)ds = R(t) - tr(t)$ so that $\int_t^\infty \frac{R_1(y)}{y^2}dy = R(t)/t$, using that $R(\infty)\infty = 0$ by assumption. This proves that $\int r_1(1-F_G)dx = \int\{R(t)/t\}tdG(t) = \int R(t)dG(t)$.

□

From van der Laan, Bickel and Jewell (1997) we have the following result:

Result 4.1 *Consider the parameter $\mu = \int r_1(x)(1-F_G(x))dx$ for a given function r_1. If G is absolutely continuous w.r.t. the Lebesgue measure, H is absolutely continuous w.r.t. Lebesgue measure with density h, and $\int \frac{F_G(x)(1-F_G(x))}{h(x)}r_1^2(x)dx < \infty$, then the efficient influence curve or canonical gradient for μ is given by:*

$$\ell^*(F_G, H, r_1) = \frac{r_1(C)}{h(C)}(\Delta - F_G(C)).$$

We note that this equals the efficient influence curve for $\mu = \int r_1(1-F)dx$ in the nonparametric current status data model where F is unrestricted. In this section we are concerned with proving that the NPMLE $\mu_n = \int r_1(x)(1 - F_{G_n}(x))dx$ is asymptotically efficient:

$$\mu_n - \mu = (P_n - P_G)\ell^*(F_G, H, r_1) + o_P(n^{-1/2}),$$

where we used for a given function f of Y and a probability measure P the notation $Pf \equiv \int f(y)dP(y)$. By lemma 4.1 this also yields the efficiency of $\int R(t)dG_n(t) = \int r(1 - G_n)dt$ by setting $r_1(t) = -tr'(t)$.

We now give the general setup of our proof. The score operator $A_G : L_0^2(G) \rightarrow L_0^2(P_G)$ at G is given by:

$$A_G(h) = \frac{\int_0^C \phi(h)(x)dF_G(x)}{F_G(x)}\Delta + \frac{\int_C^\infty \phi(h)(x)dF_G(x)}{1 - F_G(x)}(1 - \Delta),$$

where $\phi(h) = \int_0^x th(t)dG(t)/\int_0^x tdG(t)$. Using that $\int hdG = 0$ we can also represent this score operator as:

$$A_G(h) = \int_0^C \phi(h)(x)dF_G(x)\left[\frac{\Delta - F_G(C)}{F_G(C)(1 - F_G(C))}\right].$$

Consider the one-dimensional submodel $G_{n,\epsilon,h}$ defined by $\epsilon \rightarrow \int_0^{\cdot}(1 + \epsilon h(x))dG_n(x)$ through G_n at $\epsilon = 0$, where h has finite supremum norm and satisfies $\int h(x)dG_n(x) = 0$. Since G_n is an NPMLE we have that $\epsilon \rightarrow \int \log(p_{G_{n,\epsilon,h}})dP_n$ is maximized at $\epsilon = 0$ which proves that $P_n A_{G_n}(h) = 0$ for all $h \in L_0^2(G_n)$ with finite supremum norm.

For notational convenience, we will set $F_n \equiv F_{G_n}$ in the rest of this proof. We note that the range of the score operator A_{G_n} can be represented as:

$$\mathcal{S} \equiv \left\{ g\frac{\Delta - F_n}{F_n(1 - F_n)} : g \text{ piecewise linear on support } G_n \right\}.$$

Thus we have that

$$P_n\left\{ g\frac{\Delta - F_n}{F_n(1 - F_n)} \right\} = 0 \text{ for all } g \in \mathcal{S}. \tag{10}$$

Define the function $m_n = (r_1/h)F_n(1-F_n)$ and $m = (r_1/h)F(1-F)$. Let $\tilde{m}_n, \tilde{m} \in \mathcal{S}$ be the piecewise linear approximations of m_n and m on the support of G_n, respectively. Since $r_1/(hm_n) = 1/(F_n(1-F_n))$ application of (10) with $g = \tilde{m}_n$ and $g = \tilde{m}$ yields:

$$P_n \left\{ \frac{r_1}{h} \frac{\tilde{m}_n}{m_n} (\Delta - F_n) \right\} = 0 = P_n \left\{ \frac{r_1}{h} \frac{\tilde{m}}{m_n} (\Delta - F_n) \right\}.$$

We also note that the general identity in convex linear models (van der Laan, 1998) holds if r_1/h is uniformly bounded:

$$P_G \ell^*(G, H, r_1) = P_G \left\{ \frac{r_1}{h} (\Delta - F_n) \right\} = -(\mu_n - \mu).$$

This identity can be explicitly verified. Combining the last two equations and noting that

$$P_G \left\{ \frac{r_1}{h} (\Delta - F_n) \right\} - P_G \left\{ \frac{r_1}{h} \frac{\tilde{m}_n}{m_n} (\Delta - F_n) \right\} = \int \frac{r_1(c)}{h(c)} \left(\frac{\tilde{m}_n}{m_n}(c) - 1 \right) (F_n - F)(c) dH(c)$$

yields

$$\mu_n - \mu = (P_n - P_G) \left\{ \frac{r_1}{h} \frac{\tilde{m}_n}{m_n} (\Delta - F_n) \right\} + \int_0^\tau \frac{r_1(c)}{h(c)} \left(\frac{\tilde{m}_n(c)}{m_n(c)} - 1 \right) (F_n - F)(c) dH(c). \quad (11)$$

Similarly, by using that $0 = P_n \left\{ \frac{r_1}{h} \frac{\tilde{m}}{m_n} (\Delta - F_n) \right\}$ we obtain the same equation (11) but with $\tilde{m}_n$ replaced by $\tilde{m}$. The first and second term on the right-hand side will be referred to as the empirical process term and second order term, respectively.

In order to avoid the implicit singularity $F_n(x)$ at $x = 0$ we restrict ourselves to the case where $r = 0$ on $[0, \epsilon)$ for some $\epsilon > 0$ with $F_G(\epsilon) > 0$. We do not expect that this is a necessary restriction to the class of smooth functionals, but, because of the implicitness of F_n, we have not been able to extend our result.

Theorem 4.1 *Let* $\mu_n = \int r_1(1 - F_{G_n})dx$ *be the NPMLE of* $\mu = \int r_1(1 - F_G)dx$. *If* $\int_0^\infty h(x)dx/x < \infty$, *$G$ is absolutely continuous w.r.t. the Lebesgue measure, the derivative of* r_1/h *is uniformly bounded,* $r_1(x) = 0$ *for* $x \le \epsilon$ *and* $x > \tau$ *for some* $\epsilon > 0$ *with* $F_G(\epsilon) > 0$ *and* $\tau < \infty$, $\int \frac{F_G(x)(1-F_G(x))}{h(x)} r_1^2(x) dx < \infty$, *then*

$$\mu_n - \mu = (P_n - P_G)\ell^*(F_G, H, r_1) + O_P(n^{-4/5}).$$

Proof. We chose (11) with $\tilde{m}_n$ replaced by $\tilde{m}$ as starting point. We first prove that the second order term is $O_P(n^{-4/5})$. By the uniform consistency of F_n to F_G there exists a $\delta > 0$ so that with probability tending to 1 $m_n > \delta > 0$ on (ϵ, τ). By $r_1/h < M < \infty$ and the Cauchy-Schwarz inequality, the second order term is, with probability tending to 1, bounded by:

$$\frac{M}{\delta}\|\tilde{m} - m_n\|_H \|F_n - F\|_H.$$

By Theorem 3.1 we have that $\|F_n - F\|_H = O_P(n^{-2/5})$. Furthermore, we have

$$\|\tilde{m} - m_n\|_H \leq \|\tilde{m} - m\|_H + \|m_n - m\|_H.$$

The term $\|m_n - m\|_H$ can be directly bounded by $\|F_n - F\|_H$ (times a constant). The first term is the approximation error in $L^2(H)$-norm of the piecewise linear approximation of m. Consider the linear approximation on $(t_j, t_{j+1}]$. It is easily verified that if for some $\delta > 0$ and $M < \infty$ $F' > \delta > 0$ and $\mid m' \mid < M < \infty$ on $(t_j, t_{j+1}]$, then the area between the linear approximation $\tilde{m}$ of m and m itself can be bounded by the area between the linear approximation $\tilde{F}$ of F and F itself. We indeed have that $F'(x) = 1/x^2 \int_0^x tdG(t) > 1/\tau^2 \int_0^\epsilon tdG(t) > 0$ on $[\epsilon, \tau]$. Furthermore, since the derivative of r_1/h and $F'(x) < 1/\epsilon^2 \int_0^\tau tdG(t) < \infty$ are bounded on $[\epsilon, \tau]$ we can also bound $\mid m' \mid$ on $[\epsilon, \tau]$. This proves that $\|\tilde{m} - m\|_H \leq \text{const}\|\tilde{F} - F\|_H = O(n^{-2/5})$. We conclude that the second order term is $O_P(n^{-4/5})$.

We now analyze the empirical process term. We first show that for each $\delta \in \{0, 1\}$ the univariate real valued function

$$f_n(\cdot, \delta) \equiv \frac{r_1}{h}\frac{\tilde{m}}{m_n}(\delta - F_n) = \frac{\tilde{m}}{F_n(1 - F_n)}(\Delta - F_n)$$

has a variation norm $\|f_n(\cdot, \delta)\|_{\mathrm{v}}$ bounded by a constant $M < \infty$, with probability tending to 1. By our assumptions we have that $r_1/hF(1 - F)$ has a uniformly bounded derivative on $[\epsilon, \tau]$ and thus $\tilde{m}$ has a variation norm bounded by the variation norm of $r_1/hF(1 - F)$. In addition, $F_n(1 - F_n)$ has a uniformly in n bounded variation norm since it is a product of two bounded monotone functions. Moreover, the denominator $\inf_{x\in[\epsilon,\tau]} F_n(1-F_n) > \delta > 0$, with probability tending to one, for some $\delta > 0$. We conclude that $\|f_n(\cdot,\delta)\|_{\mathrm{v}} < M < \infty$, with probability tending to 1. This proves that, with probability tending to one, f_n falls in the Donsker class of functions with a uniform bound on the variation norm. Let $f \equiv \ell^*(G, H, r_1)$. By the tightness of an empirical process indexed by a Donsker class we have that $\sqrt{n}(P_n - P_G)(f_n - f) = o_P(1)$ if $\int (f_n - f)^2 dP_G \to 0$ in probability (see e.g. van der Vaart and Wellner, 1996). We already established that $\|\tilde{m} - m\|_H \to 0$. The $L^2(P_G)$-convergence follows immediately from the fact that F_n converges uniformly to F, $\|\tilde{m} - m\|_H \to 0$ and that $F_n(1-F_n)$ is bounded away from $\delta > 0$ with probability tending to 1. This completes the proof. □

5 Discussion

We remark that the rate of G_n (the natural conjecture is $n^{-1/5}$) and its limit distribution are open problems. The current literature on related problems only provide conjectures, but no formal results yet. A related problem has been studied by Jongbloed (1995, chapter 6) who studies the NPMLE for the model consisting of decreasing convex densities. He establishes a minimax lower bound of $n^{-1/5}$ for the derivative of this density and he conjectures a limit distribution for the NPMLE as well. We note that $cF_G(c) = \int_0^c G(s)ds$ is an increasing convex bounded function so that our density $\Delta \int_0^c G(s)ds + (1-\Delta)(c - \int_0^c G(s)ds)$ of the data w.r.t. $1/xh(x)$ is a simple function of an increasing convex bounded function. Therefore one would expect that his work can be used to establish the same minimax lower bound $n^{-1/5}$ for estimation of the derivative G of the increasing convex function $\int_0^c G(s)ds$. His conjectured limit distribution will not be applicable, though.

We also note the following identity:

$$E(C\Delta \mid C = c) = \int_0^c G(s)ds.$$

Therefore G is the derivative of a regression curve which is known to be convex. Mammen (1991) considers the least squares estimator for this regression problem and establishes a rate result for the derivative (i.e. G) of the regression curve given by $n^{-1/5}$. Thus his results show that the least squares estimator of G will converge at rate $n^{-1/5}$ giving a strong indication that also the NPMLE G_n will converge at rate $n^{-1/5}$. In addition, Mammen (1991) also conjectures a limit distribution for the least-squares estimator.

In van der Laan, Andrews (1999) the NPMLE for a class of doubly censored current status data models is considered in which the uniform distribution of I over $[A, B]$ is replaced by a mixture of a uniform distribution and a pointmass which naturally arises in AIDS-partner studies. In van der Laan, Andrews (1999) the modified-IWPAVA has been implemented for this class of models and used to analyze the AIDS-partner study. In these models, due to the pointmass, the NPMLE of G actually converges at rate $n^{-1/3}$. These models also differ from the uniform model by the facts that they cannot be viewed as a submodel of the nonparametric current status data model and do not allow explicit limit distributions for smooth functionals of the NPMLE.

References

Barlow, R. E., Bartholomew, D.J., Bremner, J.M. & Brunk, H.D. (1972). *Statistical inference under order restrictions*, New York, John Wiley.

Bickel, P.J. & Klaassen, C.A.J. & Ritov, Y. & Wellner, J.A. (1993), *Efficient and adaptive inference in semi-parametric models*, Johns Hopkins University Press, Baltimore.

Birman, M.S. & Solomak, M.Z. (1967). Piecewise-polynomial approximations of functions of the classes W_p^α. *Mat. Sb.*, **73**, 295–317.

Diamond, I.D., McDonald, J.W. & Shah, I.H. (1986). Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan. *Demography* **23**, 607–620.

Diamond, I.D. & McDonald, J.W. (1991), The analysis of current status data. *Demographic applications of event history analysis*, J. Trussell, R. Hankinson, & J. Tilton (eds.), Oxford: Oxford University Press.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14–44.

van de Geer, S. (1994). Asymptotic normality in mixture models, preprint University of Leiden, the Netherlands.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution, Technical report nr. **378**, Department of Statistics, Stanford University, California.

Groeneboom, P. & Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, Boston, Birkhäuser.

Huang, J. & Wellner, J.A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. *Statistica Neerlandica* **49**, 153–163.

Jewell, N.P. & Shiboski, S.C. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics* **46** 1133–1150.

Jewell, N.P., Malani, H.M. & Vittinghoff, E. (1994). Nonparametric estimation for a form of doubly censored data with application to two problems in aids. *J. Amer. Statist. Assoc.* **89** 7–18.

Jongbloed, G. (1995), *Three statistical inverse problems*, Ph.D. dissertation, Delft University of Technology, Delft.

Keiding, N. (1991). Age-specific incidence and prevalence (with discussion). *J. R. Statist. Soc.*

A **154**, 371–412.

van der Laan, M.J. (1998), Identity for NPMLE in Censored Data Models, *Lifetime Data Models* **4**, 83–102.

van der Laan, M.J., Andrews, C. (2000), *The NPMLE in a class of doubly censored current status data models with applications to AIDS Partner studies*, *Biometrika* **87**, Vol. 1, pp 61–71.

van der Laan, M.J., Bickel, P.J. & Jewell, N.P. (1997), Singly and doubly censored current status data; estimation, asymptotics and regression. *Scand. J. Statist.* **24**, 289–308.

Mammen, E. (1991), Nonparametric regression under qualitative smoothness assumptions, *Ann. Statist.* **19** 741–759.

Rabinowitz, D. and Jewell, N.P. (1996), "Regression with doubly censored current status data," *J. r. Statist. Soc. B* **58**, 541–550.

van der Vaart, A.W. & Wellner, J.A. (1996). *Weak convergence and empirical processes*, Springer Verlag.

Mark J. van der Laan

Division of Biostatistics

School of Public Health

140 Warren Hall

University of California, Berkeley

Berkeley, CA 94720-7360