Statistica Sinica **10**(2000), 265-280

COMPUTING THE EFFICIENT SCORE IN SEMI-PARAMETRIC PROBLEMS

Daniel Rabinowitz

Columbia University

Abstract: An approach to organizing the calculation of the efficient score for a finite dimensional parameter of interest in the presence of an infinite dimensional nuisance parameter is presented. The approach involves choosing a set of submodels for the nuisance parameter. The condition that the scores for the submodels be orthogonal to the efficient score takes the form of an equation whose solution appears in a representation of the efficient score. The approach is illustrated with several examples.

Key words and phrases: Case-control, censoring, interval censored data, missing covariates, projection, tangent space.

1. Introduction

Computing the efficient score in a semiparametric problem can allow comparison of the variance of available estimators to the semi-parametric variance bound, aid in the development of efficient or improved estimating equations, and aid in the calculation of standard errors for estimators based on estimating equations. Bickel, Klaassen, Ritov and Wellner (1993) describe three approaches to finding the efficient score in the presence of an infinite dimensional nuisance parameter. They are via guessing a candidate estimator, via solving the orthogonality equations, and via inversion of the information operator. See also Newey (1990), Pfanzagl (1990) or Stein (1956), for example, or the references given there. When a candidate estimator is not readily guessed, it may not be immediately obvious how to approach the orthogonality equations or the information operator, and the efficient score is arrived at through a combination of ingenuity and intuition. The purpose of this paper is to present a systematic approach that can be helpful in organizing the derivation of the orthogonality equations.

Begin with some notation and definitions. Consider a model described by a parameter of interest, β, and nuisance parameters, α. Let Y denote observed data and let f denote its density, so that the parameterization may be expressed as

$$(\alpha, \beta) \longrightarrow f_{\alpha,\beta}(Y). \qquad (1)$$

Let S_β denote the score for the parameter of interest, the derivative of the log-likelihood with respect to β, $S_\beta = \frac{d}{d\beta}\log f_{\alpha,\beta}(Y)$. A submodel for the nuisance parameter is a mapping of the form $\eta \longrightarrow \alpha_\eta$. Scores for submodels of the nuisance parameters take the form $\frac{d}{d\eta}\log f_{\alpha_\eta,\beta}(Y)\Big|_{\alpha_\eta=\alpha}$. The efficient score for β is the difference between the score for β and the score for a particular submodel of the nuisance parameters. The particular submodel is the one with the property that the difference is uncorrelated with the scores for all submodels of the nuisance parameters. That is, the efficient score is

$$S_\beta - \frac{d}{d\eta}\log f_{\alpha_\eta,\beta}(Y)\Big|_{\alpha_\eta=\alpha},$$

for a submodel with the property that

$$E\Big\{\Big(S_\beta - \frac{d}{d\eta}\log f_{\alpha_\eta,\beta}(Y)\Big|_{\alpha_\eta=\alpha}\Big)S\Big\} = 0 \tag{2}$$

for all scores S for submodels of the nuisance parameters. The equations (2) are the orthogonality equations.

To motivate the approach developed here, it is helpful to first consider the orthogonality equations for the case of finite dimensional nuisance parameters. Let S_α denote the gradient of the log likelihood with respect to α, $S_\alpha = \nabla_\alpha \log f_{\alpha,\beta}(Y)$. Then, scores for submodels of the nuisance parameters are linear combinations of the components of S_α, $\frac{d}{d\eta}\log f_{\alpha_\eta,\beta}(Y)\Big|_{\alpha_\eta=\alpha} = \varphi^T S_\alpha$, where the row vector of coefficients, φ^T, is $\varphi = \frac{d\alpha}{d\eta}\Big|_{\alpha_\eta=\alpha}$. It follows that the efficient score for β is the difference between S_β and the particular linear combination of the components of S_α with the the property that the difference is uncorrelated with S_α. That is, the efficient score is given by

$$S_\beta - \varphi^{\star T} S_\alpha. \tag{3}$$

where the coefficients of the linear combination, $\varphi^\star$, are defined implicitly by the orthogonality equations,

$$E\left\{\left(S_\beta - \varphi^{\star T} S_\alpha\right) S_\alpha^T\right\} = 0. \tag{4}$$

These equations take the form of a matrix equation in $\varphi^\star$ with as many dimensions as there are in the nuisance parameter,

$$\left(ES_\alpha S_\alpha^T\right)\varphi^\star = ES_\alpha S_\beta. \tag{5}$$

With finite dimensional nuisance parameters, using the components of S_α as a basis for the space of scores for submodels for the nuisance parameters leads to the tractable formulation of the orthogonality equations, (4). When the nuisance parameter is not finite dimensional, a tractable formulation of the orthogonality equations is not always immediately obvious. Here, an infinite dimensional analogue to using the components of S_α as a basis is developed in order to arrive at a tractable approach to organizing the derivation of the orthogonality equations when α is infinite dimensional.

In the next section, the approach to formulating the orthogonality equations when the nuisance parameter is infinite dimensional is presented. The presentation is introduced through a simple example. In the third section, the approach is applied to some examples. Efficiency calculations for some of the examples are well known, and so allow comparisons between the heuristic presented here and other approaches to computing the efficient score.

2. The Orthogonality Equations

The approach presented here is to start with parameterization in the same form as (1) and then form an analogue of the tractable representation of the orthogonality equations, (4). To form the analogue, the components of S_α in (4) are replaced by scores for a subset of submodels of the nuisance parameter. The submodels in the set are chosen so that their scores span the space of all scores for the nuisance parameter. The resulting formulation of the orthogonality equations leads to an integral equation analogue of (5). Correspondingly, the efficient score is not expressed as the difference between S_β and a finite dimensional linear combination of the components of S_α. Instead, the finite dimensional linear combination is replaced by an infinite dimensional linear combination of the scores for the set of submodels.

To motivate the approach to forming the analogue to (4), it is useful to consider a simple example of a semi-parametric regression problem. Suppose that the data are a sample of triplets, (X_i, T_i, W_i), $i = 1, \ldots, n$, where the W_i are outcomes and the (X_i, T_i) are covariates. Suppose that the outcomes may be expressed in terms of the covariates and independent and identically distributed error terms ϵ_i by $W_i = \alpha(T_i) + X_i\beta + \epsilon_i$. Suppose that the joint distribution of the (X_i, T_i) pairs are unspecified, and suppose that the error terms are independent of the (X_i, T_i) pairs. For convenience, suppose that the distribution of the ϵ_i are specified as standard normal. The score for β is

$$\begin{aligned} S_\beta &= \frac{d}{d\beta} \log \prod_{i=1}^{n} (2\pi)^{-1/2} \exp(-(W_i - X_i\beta - \alpha(T_i))^2/2) \\ &= \sum_{i=1}^{n} (W_i - X_i\beta - \alpha(T_i))\, X_i. \end{aligned}$$

(Here, the joint distribution of the (X_i, T_i) pairs is conditioned out of the likelihood, as they are ancillary for α and β.) Similarly, scores for submodels of the nuisance parameter α take the form

$$\frac{d}{d\eta}\log\prod_{i=1}^{n}(2\pi)^{-1/2}\exp(-\frac{1}{2}(W_i - X_i\beta - \alpha_\eta(T_i))^2/2)$$
$$=\sum_{i=1}^{n}(W_i - X_i\beta - \alpha(T_i))\,\varphi(T_i),$$

where φ corresponds to the derivative of α_η with respect to η, $\varphi(t)=\frac{d}{d\eta}\alpha_\eta(t)\Big|_{\alpha_\eta=\alpha}$. The efficient score is therefore given by $S_\beta - \sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\varphi^\star(T_i)$, where $\varphi^\star$ solves the orthogonality equations:

$$\text{for all }\varphi,\quad E\Big\{\Big(S_\beta - \sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\varphi^\star(T_i)\Big)\sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\varphi(T_i)\Big\}=0.$$

These equations are similar to the equations (4) derived for finite dimensional nuisance parameters: the function $\varphi^\star$ and the functions φ in these equations play the role of the vectors $\varphi^{\star T}$ and φ in equation (4). In these equations, however, $\varphi^\star$ and φ are indexed by the range of the nuisance function α, while in (4), $\phi^\star$ and φ are indexed by the dimensions of the finite dimensional vector of nuisance parameters α. In the infinite dimensional case, as in the finite dimensional case, to compute the efficient score it suffices to find $\varphi^\star$.

The correspondence between the dimensions of the vector α in the finite dimensional case and the range of the function α in this semi-parametric regression example suggests an approach to organizing the calculation of $\varphi^\star$. In the finite dimensional case, the components of S_α were used as a basis for the space of scores for the nuisance parameters. The components of S_α are the scores for the submodels in which all but one component of the nuisance parameters are fixed. Analogues to these submodels in the semi-parametric regression example are submodels in which α is fixed except at one value of its range. That is, the analogues are submodels of the form

$$\eta \longrightarrow \alpha + \eta\delta^{t^\star}(t), \tag{6}$$

where $\delta^{t^\star}(t) = 1_{\{t\in(t^\star,t^\star+dt)\}}/dt$ is the Dirac δ function at $t^\star$. The score for such a submodel is $\sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\delta^{t^\star}(t)$. Substituting these scores into the orthogonality equations results in an analogue to (4),

$$E\Big\{\Big(S_\beta - \sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\varphi^\star(T_i)\Big)\sum_{i=1}^{n}[W_i - X_i\beta - \alpha(T_i)]\,\delta^{t^\star}(t).\Big\} = 0. \tag{7}$$

Formally evaluating the left hand side of (7) results in

$$\begin{aligned}
&\sum_{i=1}^{n} \int f_T(t)dtE\left\{ [W_i - X_i\beta - \alpha(T_i)]^2 (X_i - \varphi^\star(t))\delta^{t^\star}(t). \middle| T_i = t \right\} \\
&= \sum_{i=1}^{n} \int f_T(t)dt\delta^{t^\star}(t). \left(E\{X_i|T_i = t^\star\} - \varphi^\star(t^\star)\right) \\
&= nf_T(t^\star)\left(E\{X_i|T_i = t^\star\} - \varphi^\star(t^\star)\right),
\end{aligned}$$

where f_T is the marginal density of the T_i. Setting the left hand side to zero reveals that $\varphi^\star(t) = E\{X_i|T_i = t\}$. Strategies for achieving this bound, and discussions of smoothness conditions may be found, for example, in Chen (1988), Chen (1995) and Schick (1993).

Finding the efficient score in the example was made tractable by using the scores for the subset of submodels, (6), when formulating the orthogonality equations. A generalization of the approach is now described.

Suppose that the possible values of the nuisance parameter may be identified with a space of functions. Let $\mathcal{X}$ denote the range of the functions and let Ψ denote the mapping that takes functions to values of the nuisance parameter. The mapping Ψ need not be one-to-one. It is convenient, and there is no loss of generality, to suppose that the function that is identically zero is mapped to the true value of α. For each element, ξ, in the range of the function, $\mathcal{X}$, denote the Dirac function at ξ by $\delta^\xi(x) = 1_{\{x\in(\xi,\xi+dx)\}}/dx$. Let S^ξ denote the score for the submodel traced by the image under Ψ of $\eta\delta^\xi$, as η varies,

$$S^\xi = \frac{d}{d\eta}\log f_{\Psi(\eta\delta^\xi),\beta}(Y)\bigg|_{\eta=0}.$$

In a variety of settings, the S^ξ span the space of scores for submodels of the nuisance parameters in the sense that for any submodel, $\eta \longrightarrow \Psi(g_\eta)$, the score for the submodel can be expressed as a linear combination of the S^ξ,

$$\frac{d}{d\eta}\log f_{\Psi(g_\eta),\beta}(Y) = \int_{\mathcal{X}} \frac{d}{d\eta} g_\eta(\xi) S^\xi d\xi \bigg|_{g_\eta=0}.$$

In such settings, the efficient score for β is

$$S_\beta - \int_{\mathcal{X}} \varphi^\star(x)S^x dx, \tag{8}$$

where $\varphi^\star$ is defined by the orthogonality equations,

$$\text{for all } \xi,\quad E\left\{ \left(S_\beta - \int_{\mathcal{X}} \varphi^\star(x)S^x dx \right) S^\xi \right\} = 0. \tag{9}$$

The orthogonality equations, (9), will generally take the form of an integral equation for $\varphi^\star$.

In some situations, natural choices for Ψ and $\mathcal{X}$ will not be evident. This may especially be the case for models involving constraints. In such cases, the complexities in the structure of the tangent space induced by the constraints are faced through the choice of the parameterization. In some settings, natural choices for Ψ and $\mathcal{X}$ will lead to over-parameterization of the nuisances. In these situations, the integral equation will not have a unique solution. Nevertheless, all the solutions to the integral equation when substituted into (8) will result in the same efficient score. In many situations, different choices of Ψ and $\mathcal{X}$, corresponding to different parameterizations for the nuisance parameter, are available. A particular choice of parameterization may lead to simpler computations.

The general approach presented here parallels the calculations in settings where the nuisance is finite dimensional in the same way as in the case of the example. The set $\mathcal{X}$ is an infinite dimensional analogue of the indices of components of a finite dimensional vector of parameters, α. The submodels $\eta \longrightarrow f_{\Psi(\eta\delta^x),\beta}$ correspond to the submodels in finite dimensional settings that vary one component of α, while holding the other components fixed, and the scores S^ξ correspond to the components of the gradient of the log likelihood with respect to α. The function $x \longrightarrow \varphi^\star(x)$ in (8) is an analogue of the finite dimensional vector $\varphi^\star$ in (3). Solving the integral equation that results from evaluating (9) corresponds to inverting the matrix in $ES_\alpha S_\alpha^T$ that results from evaluating (4).

This section concludes with a brief description of parallels between the approach developed here and the approach based on information operators. The mapping

$$\dot{\mathrm{l}} \,:\, \frac{d}{d\eta}g_\eta \;\longrightarrow\; \int_{\mathcal{X}} \frac{d}{d\eta}g_\eta(\xi)S^\xi d\xi$$

is a score operator. The equation (9) is a representation of the normal equations in terms of the basis for the nuisance tangent space formed by the images of the Dirac functions. That is, the normal equations are that for $\xi \in \mathcal{X}$, $\langle S_\beta - \dot{\mathrm{l}}\varphi^\star, \dot{\mathrm{l}}\delta^\xi\rangle = 0$, instead of the more familiar form, $\dot{\mathrm{l}}^T S_\beta = \dot{\mathrm{l}}^T\dot{\mathrm{l}}\varphi^\star$. Solving the integral equation that results from evaluating (9) corresponds to inverting the information operator $\dot{\mathrm{l}}^T\dot{\mathrm{l}}$. See, for example, Begun, Hall, Huang and Wellner (1983), Van der Vaart (1991), Groeneboom and Wellner (1992, p.27), or Bickel, Klaassen, Ritov and Wellner (1993, p.79).

3. Examples

In this section, several examples are outlined. In each example, a representation of the parameter space in the form of (1) is proposed, and a choice of $\mathcal{X}$ and

Φ is presented. In many cases, the parameter space is represented in terms of an exponential tilt of the true distribution (or a tilt modified to respect constraints). Then, efficient score in the form of (8) and the orthogonality equations in the form (9) are found.

3.1. Median of a symmetric distribution

Suppose that T_i, $i = 1, \ldots, n$, are independent and identically distributed random variables with symmetric density g. Let β denote the parameter of interest, the median of g, and let $\alpha(t)$ denote the nuisance, $g(t-\beta)$. The likelihood is $\prod_{i=1}^n \alpha(T_i - \beta)$.

Let $\mathcal{X}$ be $[0, \infty)$ and define Ψ by

$$\Psi(\zeta)(t) = \alpha(t)\exp\{\zeta(|t|)\} \Big/ 2\int_0^\infty \alpha(u)\exp\{\zeta(|u|)\}du \ .$$

Then,

$$S^x = \sum_{i=1}^n \left[\frac{1_{\{|T_i-\beta|\in(x,x+dx)\}}}{dx} - 2\alpha(x)\right],$$

so that the efficient score, (8), is

$$\sum_{i=1}^n \left[\frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)} - \left(\varphi^\star(|T_i-\beta|) - 2\int_0^\infty \varphi^\star(u)\alpha(u)du\right)\right];$$

the orthogonality equations, (9), are

$$\text{for all } x,\ \ n2\alpha(x)\left\{\varphi^\star(x) - 2\int_0^\infty \varphi^\star(u)\alpha(u)du\right\} = 0,$$

from which it follows that $\varphi^\star$ is constant so that the efficient score for β is

$$\sum_{i=1}^n \frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)}.$$

This example is also treated in, for example, Stein (1956), Ibragimov and Has'minskii (1981) and Bickel, Klaassen, Ritov and Wellner (1993).

3.2. Mean of an arbitrary distribution

Suppose that T_i, $i = 1, \ldots, n$, are independent and identically distributed random variables with density g. Let β denote the parameter of interest, the expectation of the T_i, and let $\alpha(t)$ denote the nuisance, $g(t+\beta)$. The likelihood is $\prod_{i=1}^n \alpha(T_i - \beta)$.

Let $\mathcal{X}$ be the reals and define Ψ by $\Psi(h)(t) = \alpha(t+\mu(h))\exp\{h(t+\mu(h)) - \psi(h)\}$, where ψ is defined by $\exp(\psi(h)) = \log\int \alpha(s)\exp\{h(s)\}\,ds$ and μ is defined by $\mu(h) = \int \alpha(s) \exp\{h(s) - \psi(h)\} sds$. The term $\mu(h)$ is necessary so that the expectation of T under $f_{\beta,\Psi(h)}$ is β. Then

$$S^x = \sum_{i=1}^{n} \left\{ \frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)} \alpha(x)x + \frac{1_{\{T_i-\beta \in (x,x+dx)\}}}{dx} - \alpha(x) \right\},$$

so that the efficient score, (8), is

$$\sum_{i=1}^{n} \left\{ \frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)} \left(-1 - E\left[(T_i-\beta)\varphi^\star(T_i-\beta)\right]\right) - \varphi^\star(T_i-\beta) + E\left[\varphi^\star(T_i-\beta)\right] \right\}$$

the orthogonality equations (9) are: for all x,

$$n\alpha(x)x \left\{ E\left[\frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)}\right]^2 \left(-1 - E\left[\varphi^\star(T_i-\beta)(T_i-\beta)\right]\right) - E\left[\frac{\alpha'(T_i-\beta)}{\alpha(T_i-\beta)}\varphi^\star(T_i-\beta)\right] \right\}$$
$$+ n\alpha(x) \left\{ \left[\frac{\alpha'(x)}{\alpha(x)}\right] \left(-1 - E\left[\varphi^\star(T_i-\beta)(T_i-\beta)\right]\right) - \varphi^\star(x) + E\left[\varphi^\star(T_i-\beta)\right] \right\} = 0,$$

from which it follows that

$$\varphi^\star(t) = \frac{\alpha'(t)}{\alpha(t)} - \frac{t-\beta}{\mathrm{Var}(T_i)} + c,$$

for arbitrary c, so that the efficient score for β is, with finite $\mathrm{Var}(T_i)$, $\sum_{i=1}^{n} \frac{T_i-\beta}{\mathrm{Var}(T_i)}$. This example is also treated in, for example, Newey (1990) and Ibragimov and Has'minskii (1981).

3.3. Cumulative hazard from censored data

Suppose that (T_i, C_i), $i = 1, \ldots, n$, are independent identically distributed pairs of independent random variables. Let λ denote the hazard function of the T_i. Suppose that the distribution of the C_i is completely unspecified. Let β denote the parameter of interest, the cumulative hazard, $\beta = \int_{-\infty}^{t} \lambda(s)ds$, and define the nuisance parameter α by $\alpha(s) = \lambda(s)\beta^{-1_{\{s \le t\}}}$. Then, $\lambda(s) = \alpha(s)\beta$ for $s \le t$ and is $\alpha(s)$ for $s \ge t$, where $\alpha(s)$ is arbitrary except that $\int_{-\infty}^{t} \alpha(s) = 1$. The T_i are failure times, the C_i are censoring indicators, the Z_i are covariates, and the failure times given the covariates are assumed to follow a proportional hazards model. The C_i are ancillary and the conditional likelihood given the C_i is

$$\prod_{i=1}^{n} \exp\left(-\int_{-\infty}^{\min(T_i,C_i)} \alpha(s)ds\beta^{1_{\{s\le t\}}} \right) \left(\alpha(\min(T_i,C_i))\beta^{1_{\{s\le t\}}} \right)^{1_{\{T_i \le C_i\}}}.$$

Let $\mathcal{X}$ be the reals and define Ψ by $\Psi(g)(s) = \lambda(s)\exp(g(s) - \psi(g)1_{\{s\leq t\}})$, where ψ is defined by $\exp\{\psi(g)\} = \int_{-\infty}^{t} \lambda(s)\exp(g(s))ds$. Suppose that only the $\min(C_i, T_i)$ and $1_{\{T_i \leq C_i\}}$ are available. Then,

$$S^x = \sum_{i=1}^{n} \int_{-\infty}^{\infty} Y_i(s)(dN_i(s) - \lambda(s)ds)\Big(\frac{1_{\{s\in(x,x+dx)\}}}{dx} - 1_{\{s\leq t\}}1_{\{x\leq t\}}\frac{\lambda(x)}{\beta}\Big)$$

where $Y_i(s) = 1_{\{\min(T_i,C_i)\geq s\}}$ and $N_i(s) = (1 - Y_i(s))\,1_{\{T_i\leq C_i\}}$, so that the efficient score, (8), is

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} Y_i(s)\,(dN_i(s) - \lambda(s)ds)\,\Big(-\frac{1_{\{s\leq t\}}}{\beta} - \varphi^\star(s) + 1_{\{s\leq t\}}\frac{\int_{-\infty}^{t}\varphi^\star(u)\lambda(u)du}{\beta}\Big);$$

the orthogonality equations, (9), are

$$\begin{aligned}&\text{for all } x,\ n\lambda(x)EY_i(x)\Big(\frac{-1_{\{x\leq t\}}}{\beta} - \varphi^\star(x) + 1_{\{x\leq t\}}\frac{\int_{-\infty}^{t}\varphi^\star(u)\lambda(u)du}{\beta}\Big)\\ &\quad -n\lambda(x)\int_{-\infty}^{\infty} EY_i(s)\lambda(s)ds\Big(-\frac{1_{\{s\leq t\}}}{\beta} - \varphi^\star(s) + 1_{\{s\leq t\}}\frac{\int_{-\infty}^{t}\varphi^\star(u)\lambda(u)du}{\beta}\Big) = 0,\end{aligned}$$

from which it follows that

$$\varphi^\star(x) = \Big(\frac{\tau}{EY_i(x)} + c\Big)1_{\{s\leq t\}}, \qquad \text{where} \quad \tau = \beta^{-1}\int_{=\infty}^{t} \frac{\lambda(x)}{EY_i(x)}dx$$

for arbitrary c, so that the efficient score for β is

$$\sum_{i=1}^{n} \int_{-\infty}^{t} Y_i(s)\,(dN_i(s) - \lambda(s)ds)\,\Big(-\frac{1}{\beta} - \Big\{\frac{\tau}{EY_i(s)} - 1\Big\}\Big).$$

Derivations of the this and similar results may be found in, for example, Breslow and Crowley (1974), Gill (1983) and Andersen, Borgan, Gill and Keiding (1993).

3.4. Proportional hazards

Suppose that (T_i, C_i, Z_i), $i = 1, \ldots, n$, are independent identically distributed triplets of random variables. Suppose that the T_i are conditionally independent of the C_i given the Z_i. Suppose that the joint distribution of the (C_i, Z_i) pairs is completely unspecified and that only the $\min(C_i, T_i)$, the $1_{\{T_i\leq C_i\}}$ and the Z_i are available. Suppose that there is an unspecified non-negative function α and a parameter of interest β such that the hazard function of the T_i at t, given that $Z_i = z$, is of the form $\alpha(t)\exp(\beta z)$. The conditional likelihood given the censoring variables is

$$\prod_{i=1}^{n} \exp\left(-\int_{-\infty}^{\min(T_i,C_i)} \alpha(s)ds\exp(\beta Z_i)\right)(\alpha(\min(T_i, C_i))\exp(\beta Z_i))^{1_{\{T_i\leq C_i\}}}.$$

Let $\mathcal{X}$ be the reals and let Ψ be defined by $\Psi(g)(t) = \alpha(t)\text{exp}\left\{g(t)\right\}$. Then

$$S^x = \sum_{i=1}^{n} \int_{-\infty}^{\infty} Y_i(s)\left(dN_i(s) - \alpha(s)\text{exp}(Z_i\beta)ds\right) \frac{1_{\{s\in(x,x+dx)\}}}{dx}$$

where the $Y_i(s)$ and the $N_i(s)$ are as in the preceding example, so that the efficient score, (8), is

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} Y_i(s)\left(dN_i(s) - \alpha(s)\exp(Z_i\beta)ds\right)\left(Z_i - \varphi^\star(s)\right),$$

and the orthogonality equations, (9), are

$$\text{for all } x, \quad nEY_i(x)\alpha(x)\text{exp}(Z_i\beta)(Z_i - \varphi^\star(x)) = 0.$$

It follows that

$$\varphi^\star(x) = \frac{EY_i(x)\text{exp}(Z_i\beta)Z_i}{EY_i(x)\text{exp}(Z_i\beta)}.$$

Similar derivations may be found in Bickel, Klaassen, Ritov and Wellner (1993) and Andersen and Gill (1982). See also, for example, Tsiatis (1981) and Breslow (1974).

3.5. Missing covariates

Let $\mathcal{D} = \{1,\ldots,D\}$ denote a set of stratum indicators and suppose that $(X_{i,d}, Y_{i,d})$, $i = 1,\ldots,n_d$, $d \in \mathcal{D}$, are independent pairs of covariates and outcome pairs. Denote the marginal density of $X_{i,d}$, the nuisance parameter, by $\alpha(x,d)$ and denote the conditional density of $Y_{i,d}$ given $X_{i,d}$ by $f_\beta(y|x)$, assumed known up to the finite dimensional parameter of interest β. Let $Y_{j,d}$, $j = n_d+1, n_d+2,\ldots,n_d+m_d$, $d \in \mathcal{D}$, be independent outcomes with density $\int \alpha(x,d)f_\beta(y|x)dx$. The $Y_{i,d}$, $i = n_d+1,\ldots n_d+m_d$, have missing covariates. The likelihood is

$$\prod_{d\in\mathcal{D}} \Big(\prod_{i=1}^{n_d} \alpha(X_i,d)f_\beta(Y_i|X_i) \prod_{i=n_d+1}^{m_d} \int \alpha(x,d)f_\beta(Y_i|x)dx\Big).$$

Let $\mathcal{X}$ be the cross product of the support of the $X_{i,d}$ and $\mathcal{D}$, and define Ψ by $\Psi(g)(s,d) = \alpha(s,d)\text{exp}\left(g(x,d) - \psi(g,d)\right)$, where

$$\text{exp}\left(\psi(g,d)\right) = \int \alpha(x,d)\text{exp}\left(g(x,d)\right)dx.$$

Then

$$S^{x,d} = \sum_{i=1}^{n_d} \Big(\frac{1_{\{X_{i,d}\in(x,x+dx)\}}}{dx} - \alpha(x,d)\Big) + \sum_{j=n_d+1}^{n_d+m_d} \left(\frac{\alpha(x,d)f_\beta(Y_{j,d}|x)}{\int \alpha(s,d)f_\beta(Y_{j,d}|s)ds} - \alpha(x,d)\right)$$

so that the efficient score, (8), is

$$\sum_{d\in\mathcal{D}}\sum_{i=1}^{n_d}\left(s_\beta(Y_{i,d}|X_{i,d})-\varphi^\star(X_{i,d},d)-\int\alpha(x,d)\varphi^\star(x,d)dx\right)$$
$$+\sum_{d\in\mathcal{D}}\sum_{j=n_d+1}^{m_d}\frac{\int\alpha(x,d)f_\beta(Y_{j,d}|x)\left(s_\beta(Y_{j,d}|x)-\varphi^\star(x,d)-\int\alpha(s,d)\varphi^\star(s,d)ds\right)dx}{\int\alpha(u,d)f_\beta(Y_{j,d}|u)du},$$

where $s_\beta(y|x)=\frac{d}{d\beta}\log f_\beta(y|x)$; the orthogonality equations, (9), are

$$\text{for all } x,d,\quad \alpha(x,d)\left(n_d\varphi^\star(x)-m_d\int K(x,s,d)\varphi^\star(s)ds\right)=\nu(x,d),\qquad (10)$$

where

$$K(x,s,d)=\int\alpha(u,d)\int f_\beta(y|u)\frac{f(y|s)\alpha(s,d)f(y|x)ds}{\left(\int\alpha(v,d)f_\beta(y|v)dv\right)^2}dydu$$

and

$$\nu(x,d)=\alpha(x,d)\int\alpha(u,d)\int f_\beta(y|u)\frac{\int f(y|s)\alpha(s,d)f(y|x)s(y|s)ds}{\left(\int\alpha(v,d)f_\beta(y|v)dv\right)^2}dydu.$$

The equation (10) takes the form of a Fredholm type II integral equation for $\varphi^\star$. In general, there is not an explicit solution to the integral equation, but approximate solutions may be found. Derivations of the efficient score for this and related problems may be found in, for example, Robins, Rotnitzky and Zhao (1995), Cosslett (1981) and Chamberlain (1987).

3.6. Regression with interval censored data

Let failure times T_i, $i=1,\ldots,n$, satisfy $T_i=Z_i\beta+\epsilon_i$, where the ϵ_i are independent and independent of the Z_i with unspecified marginal density α, the Z_i are real valued, and β is a real valued parameter of interest. Let n_i denote numbers of examination times and let $X_i=(X_{i,1},\ldots,X_{i,n_i})$ denote the ordered sequence of examination times. For convenience, define $X_{i,0}=-\infty$, and $X_{i,n_i+1}=\infty$. Let $X_{i,L}$ be the last of the ith subjects examination times preceding T_i, and let $X_{i,U}$ be the first examination time following T_i. Suppose that only the $X_{i,L}$, $X_{i,U}$, Z_i and triplets are available. The failure times T_i, given the covariates Z_i follow an accelerated failure time model. The failure times are not observed, however, as they are only known to lie between the immediately preceding and following examination times. The likelihood is

$$\prod_{i=1}^{n}\left(F(X_{i,U}-Z_i\beta)-F(X_{i,L}-Z_i\beta)\right),$$

where F is the distribution function corresponding to α, $F(x) = \int_{-\infty}^{x} \alpha(u)du$.

Let $\mathcal{X}$ be the reals, and define Ψ by $\Psi(g)(e) = \alpha(e)\exp(g(e) - \psi(g))$ where ψ is defined by $\exp(\psi(g)) = \int \alpha(e)\exp(g(e))de$. Then,

$$S^x = \sum_{i=1}^{n} \frac{1_{\{x\in(X_{i,U}-Z_i\beta, X_{i,L}-Z_i\beta)\}}}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)},$$

so that (8) is

$$\sum_{i=1}^{n} \frac{(\alpha(X_{i,U} - Z_i\beta) - \alpha(X_{i,L} - Z_i\beta))\, Z_i}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)} - \frac{\int_{X_{i,U}-Z_i\beta}^{X_{i,L}-Z_i\beta} \alpha(u)\varphi^{\star}(u)du}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)}.$$

In this expression, the appearance of $\varphi^{\star}$ in the integral suggests that it may be convenient to parameterize the nuisance and to define the score operator in terms of the function $\mu(t) = \int_{-\infty}^{t} \alpha(u)\varphi^{\star}(u)du$, so that

$$S^x = \sum_{i=1}^{n} \frac{1_{\{x=X_{i,U}-Z_i\beta\}}/dx - 1_{\{x=X_{i,L}-Z_i\beta)\}}/dx}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)}$$

and the efficient score, (8), is

$$\sum_{i=1}^{n} \frac{(\alpha(X_{i,U} - Z_i\beta) - \alpha(X_{i,L} - Z_i\beta))\, Z_i}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)} - \frac{\mu(X_{i,U} - Z_i\beta) - \mu(X_{i,L} - Z_i\beta)}{F(X_{i,U} - Z_i\beta) - F(X_{i,L} - Z_i\beta)}.$$

The orthogonality equations, (9), take the form of the type II Fredholm equation for μ,

$$n\frac{P\{X_{i,U}-Z_i\beta=x\}}{dx} E\left\{\frac{\big(\alpha(x)-\alpha(X_{i,L}-Z_i\beta)\big)Z_i}{\big(F(x)-F(X_{i,L}-Z_i\beta)\big)^2} - \frac{\mu(x)-\mu(X_{i,L}-Z_i\beta)}{\big(F(x)-F(X_{i,L}-Z_i\beta)\big)^2}\,\middle|\, X_{i,U}-Z_i\beta=x\right\}$$

$$-n\frac{P\{\ X_{i,L}-Z_i\beta=x\}}{dx} E\left\{\frac{\big(\alpha(X_{i,U}-Z_i\beta)-\alpha(x)\big)Z_i}{\big(F(X_{i,U}-Z_i\beta)-F(x)\big)^2} - \frac{\mu(X_{i,U}-Z_i\beta)-\mu(x)}{\big(F(X_{i,U}-Z_i\beta)-F(x)\big)^2}\,\middle|\, X_{i,L}-Z_i\beta=x\right\}.$$

This does not have an explicit solution but approximate solutions may be found. Similar calculations appear in Rabinowitz, Tsiatis and Aragon (1995).

3.7. Logistic regression from case-control data with independent covariates

Let X_1 and X_2 be independent random variables with unspecified marginal densities α_1 and α_2 respectively, and let Y be Bernoulli with conditional expectation given X_1 and X_2 equal to $\pi(X_1, X_2; \beta)$, where β is a finite dimensional parameter of interest and π is known. Let $(X_{1,i}, X_{2,i}, Y_i)$, $i = 1, \ldots, n$, be drawn from the conditional distribution of (X_1, X_2, Y) given $Y = 1$, and let

$(X_{1,i}, X_{2,i}, Y_i)$, $i = n+1, \ldots, n+m$, be drawn from the conditional distribution of (X_1, X_2, Y) given $Y = 0$. The likelihood is

$$\prod_{i=1}^{n} \frac{\alpha_1(X_{1,i}\alpha_2(X_{2,i}\pi(X_{1,i}X_{2,i},1)}{\int\int \alpha_1(x_1)\alpha_2(x_2)\pi(x_1,x_2,1)} \prod_{i=1}^{n+m} \frac{\alpha_1(X_{1,i}\alpha_2(X_{2,i}\pi(X_{1,i}X_{2,i},0)}{\int\int \alpha_1(x_1)\alpha_2(x_2)\pi(x_1,x_2,0)}.$$

Let $\mathcal{X}$ be the cross product of the ranges of X_1 and X_2, assumed disjoint without loss of generality, and define Ψ by

$$\Psi_1(\zeta_1,\zeta_2)(x_1) = \frac{\alpha_1(t)\exp\{\zeta_1(x_1)\}}{\int \alpha_1(u_1)\exp\{\zeta_1(u_1)\}\,du_1},$$

$$\Psi_2(\zeta_1,\zeta_2)(x_2) = \frac{\alpha_2(t)\exp\{\zeta_2(x_2)\}}{\int \alpha_2(u_2)\exp\{\zeta_2(u_2)\}\,du_2}.$$

Then, for x_1 in the range of X_1,

$$S^{x_1} = \sum_{i=1}^{n+m}\left[\frac{1_{\{X_{1,i}\in(x_1,x_1+dx_1)\}}}{dx_1} - \frac{\int \alpha_1(x_1)\alpha_2(x_2)\pi(x_1,x_2;\beta)^{Y_i}\{1-\pi(x_1,x_2;\beta))\}^{1-Y_i}dx_2}{\int\int \alpha_1(u_1)\alpha_2(u_2)\pi(u_1,u_2;\beta)^{Y_i}\{1-\pi(u_1,u_2;\beta)\}^{1-Y_i}du_1du_2}\right],$$

and for x_2 in the range of X_2,

$$S^{x_2} = \sum_{i=1}^{n+m}\left[\frac{1_{\{X_{2,i}\in(x_2,x_2+dx_2)\}}}{dx_2} - \frac{\int \alpha_1(x_1)\alpha_2(x_2)\pi(x_1,x_2;\beta)^{Y_i}\{1-\pi(x_1,x_2;\beta)\}^{1-Y_i}dx_1}{\int\int \alpha_1(u_1)\alpha_2(u_2)\pi(u_1,u_2;\beta)^{Y_i}\{1-\pi(u_1,u_2;\beta)\}^{1-Y_i}du_1du_2}\right].$$

Then the residual from the projection of the jth component of the score for β, is

$$\sum_{i=1}^{n+m}\Bigg(\frac{\{Y_i-\pi(X_{1,i},X_{2,i};\beta)\}\frac{d}{d\beta_j}\pi(X_{1,i},X_{2,i};\beta)}{\pi(X_{1,i},X_{2,i};\beta)-\pi^2(X_{1,i},X_{2,i};\beta)} - \varphi^\star{}_1(X_{1,i}) - \varphi^\star{}_2(X_{2,i})$$
$$-\int\int\left[\frac{\{Y_i-\pi(x_1,x_2;\beta)\}\frac{d}{d\beta_j}\pi(x_1,x_2;\beta)}{\pi(x_1,x_2;\beta)-\pi^2(x_1,x_2;\beta)} - \varphi^\star{}_1(x_1) - \varphi^\star{}_2(x_2)\right]\mu(x_1,x_2,Y_i)dx_1dx_2\Bigg),$$

where

$$\mu(x_1,x_2,y) = \frac{\alpha_1(x_1)\alpha_2(x_2)\pi(x_1,x_2;\beta)^y\{1-\pi(x_1,x_2;\beta)\}^{1-y}}{\int\int \alpha_1(u_1)\alpha_2(u_2)\pi(u_1,u_2;\beta)^y\{1-\pi(u_1,u_2;\beta)\}^{1-y}du_1du_2};$$

the orthogonality equations, (9), may be written as the pair of Fredholm type II integral equations, for all x_1

$$m\int\left[\frac{\{0-\pi(x_1,x_2;\beta)\}\frac{d}{d\beta_j}\pi(x_1,x_2;\beta)}{\pi(x_1,x_2;\beta)-\pi^2(x_1,x_2;\beta)} - \varphi^\star{}_1(x_1) - \varphi^\star{}_2(x_2)\right]\mu(x_1,x_2,0)dx_2$$
$$+\ n\int\left[\frac{\{1-\pi(x_1,x_2;\beta)\}\frac{d}{d\beta_j}\pi(x_1,x_2;\beta)}{\pi(x_1,x_2;\beta)-\pi^2(x_1,x_2;\beta)} - \varphi^\star{}_1(x_1) - \varphi^\star{}_2(x_2)\right]\mu(x_1,x_2,1)dx_2,$$

and for all x_2

$$m\int\left[\frac{\{0-\pi(x_1,x_2;\beta)\}\frac{d}{d\beta_j}\pi(x_1,x_2;\beta)}{\pi(x_1,x_2;\beta)-\pi^2(x_1,x_2;\beta)}-\varphi^\star{}_1(x_1)-\varphi^\star{}_2(x_2)\right]\mu(x_1,x_2,0)dx_1$$
$$+\ n\int\left[\frac{\{1-\pi(x_1,x_2;\beta)\}\frac{d}{d\beta_j}\pi(x_1,x_2;\beta)}{\pi(x_1,x_2;\beta)-\pi^2(x_1,x_2;\beta)}-\varphi^\star{}_1(x_1)-\varphi^\star{}_2(x_2)\right]\mu(x_1,x_2,1)dx_1.$$

It may be of interest to note that, when π is a nontrivial function of both x_1 and x_2, the efficient score is not equivalent to the score that is obtained through the prospective likelihood described in Prentice and Pyke (1979).

3.8. Current status and complete data

Let T denote a failure time and let C denote a censoring time. Suppose that C and T are independent. Denote the distribution function of T by F, supposed unspecified. Denote the distribution function of C by G. Let Y be the indicator that T is less than C. Let β denote the parameter of interest, $F(t)$, and define the nuisance parameter α by

$$\alpha(s)ds=\begin{cases} dF(s)/\beta & \text{if } s\le\ t,\\ dF(s)/(1-\beta) & \text{if } s>\ t.\end{cases}$$

Let T_i, $i=1,\ldots,n$, be independent and identically distributed according to the distribution of T, and let (C_j,Y_j), $j=1,\ldots,m$, be independent and identically distributed according to the joint distribution of C and Y. The likelihood is

$$\prod_{i=1}^n\alpha(T_i)(1/\beta)^{1_{\{T_i\le t\}}}(1/(1-\beta))^{1_{\{T_i>t\}}}\prod_{j=1}^m\Big(\alpha(s)(1/\beta)^{1_{\{s\le t\}}}(1/(1-\beta))^{1_{\{s>t\}}}\Big)^{Y_j}$$
$$\times\ \Big(1-\alpha(s)(1/\beta)^{1_{\{s\le t\}}}(1/(1-\beta))^{1_{\{s>t\}}}\Big)^{1-Y_j}.$$

Let $\mathcal{X}$ be $[0,\infty)$ and define Ψ by

$$\Psi(\zeta)(s)=\begin{cases}\alpha(s)\exp\{\zeta(s)\}/\int_0^t\alpha(u)\exp\{\zeta(u)\}du. & \text{if } s\ \le\ t,\\ \\ \alpha(s)\exp\{\zeta(s)\}/\int_t^\infty\alpha(u)\exp\{\zeta(u)\}du. & \text{if } s\ >\ t.\end{cases}$$

Then,

$$S^x=\begin{cases}\sum_{i=1}^n\left[\frac{1_{\{T_i\in(x,x+dx)\}}}{dx}-\alpha(x)1_{\{T_i\le t\}}\right] \\ +\sum_{j=1}^m\frac{Y_j-F(A_j)}{F(A_j)-F^2(A_j)}\alpha(x)\left(F(t)1_{\{x\le A_j\le t\}}-F(A_j)1_{\{A_j\le t\}}\right) & \text{if } x\le t,\\ \\ \sum_{i=1}^n\left[\frac{1_{\{T_i\in(x,x+dx)\}}}{dx}-\alpha(x)1_{\{T_i\le t\}}\right] \\ +\sum_{j=1}^m\frac{Y_j-F(A_j)}{F(A_j)-F^2(A_j)}\alpha(x)\left(\{F(A_j)-F(t)\}1_{\{t\le A_j\le x\}}-\{1-F(A_j)\}1_{\{A_j>x\}}\right) & \text{if } x>t,\end{cases}$$

so that the efficient score, (8), is

$$\sum_{i=1}^{n} \frac{1_{\{T_i \le t\}} - \beta}{\beta(1-\beta)} + \sum_{j=1}^{m} \frac{Y_j - F(A_j)}{F(A_j)\{1-F(A_j)\}} \frac{F(A_j \wedge t)\{1 - F(A_j \vee t)\}}{\beta(1-\beta)}$$

$$+ \sum_{i=1}^{n} \varphi^{\star\star}(T_i) + \sum_{j=1}^{m} \frac{Y_j - F(A_j)}{F(A_j)\{1-F(A_j)\}} \int_{-\infty}^{A_j} \varphi^{\star\star}(s) f(s) ds,$$

where

$$\varphi^{\star\star}(s) = \varphi^{\star}(s) - 1_{\{s \le t\}} \int_{-\infty}^{t} \varphi^{\star}(u)\alpha(u)du - 1_{\{s>t\}} \int_{t}^{\infty} \varphi^{\star}(u)\alpha(u)du.$$

The orthogonality equations (9) are

$$f(x)b(x) = f(x)\varphi^{\star\star}(x) - f(x) \int_{-\infty}^{\infty} \varphi^{\star\star}(s) K(x, ds),$$

where

$$b(x) = \begin{cases} \int_{-\infty}^{t} dG(a) \left(\frac{1_{\{x \le a\}} - \frac{F(a)}{F(t)}}{F(t)\{1-F(a)\}} \right) & \text{if } x \le t, \\ \\ -\int_{t}^{\infty} dG(a) \left(\frac{1_{\{x > a\}} - \frac{1-F(a)}{1-F(t)}}{F(a)\{1-F(t)\}} \right) & \text{if } x > t, \end{cases}$$

and

$$K(x, ds) = \begin{cases} f(s)ds 1_{\{s \le t\}} \int_{s}^{t} dG(a) \left(\frac{1_{\{x \le a\}} - \frac{F(a)}{F(t)}}{F(a)\{1-F(a)\}} \right) & \text{if } x \le t, \\ \\ f(s)ds 1_{\{s > t\}} \int_{t}^{s} dG(a) \left(\frac{1_{\{x > a\}} - \frac{1-F(a)}{1-F(t)}}{F(a)\{1-F(a)\}} \right) & \text{if } x > t. \end{cases}$$

This problem is treated in Bagiella (1997).

References

Andersen, P. K., Borgan, O., Gill, R. and Keiding, N. (1993). Statistical models based on counting processes. Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Bagiella, E. (1997). Estimating a survival distribution from case-control family data. Columbia University Department of Biostatistics Doctoral Dissertation.

Begun, J., Hall, W., Huang, W. and Wellner, J. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432-452.

Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. (1993). Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press, Baltimore.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrika* **30**, 89-99.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437-453.

Chen, H. (1988). Convergence rates for the parametric component in a partly linear model. *Ann. Statist.* **16**, 136-146.

Chen, H. (1995). Asymptotically efficient estimation in semiparametric generalized linear models. *Ann. Statist.* **23**, 1102-1129.

Cosslett, S. R. (1981). Efficient estimation of discrete choice models. In *Structural Analysis of Discrete Data With Econometric Applications* (Edited by C. F. Manski and D. McFadden). MIT Press, Cambridge, MA.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34**, 305-324.

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

Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood. DMV Seminar, Band 19. Birkhauser Verlag, Basel.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation - Asymptotic Theory.* Springer-Verlag, New York.

Newey, W. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5**, 99-135.

Pfanszagl, J. (1990). Estimation in semi-parametric models: some recent developments. Lecture Notes in Statistics. Vol. 63. Springer-Verlag, New York.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403-411.

Rabinowitz, D., Tsiatis, A. and Aragon, J. (1995). Regression with interval censored data. *Biometrika* **82**, 501-514.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Estimation of regression coefficients when regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21**, 1486-1521.

Stein, C. (1956). Efficient non-parametric testing and estimation. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 187-195. University of California Press, Berkeley.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108.

Van der Vaart, A. (1991). On differentiable functionals. *Ann. Statist.* **19**, 178-204.

Department of Statistics, Columbia University, New York, NY 10027, U.S.A.

E-mail: dan@stat.columbia.edu

(Received July 1997; accepted February 1999)