Efficient Semiparametric Estimators via Preliminary Estimators for Kullback-Leibler Minimizers

Eric Slud, Math Dept, Univ of Maryland

(Ongoing joint project with Ilia Vonta, Univ. of Cyprus.)

OUTLINE

I. MOTIVATION — Semiparametric Frailty Problems

II. APPROACH — Finite-Dimensional Version

III. BACKGROUND LITERATURE —
Profile Likelihood, Semiparametrics, Frailty Models

IV. Semiparametric Examples

V. Efficient Estimator in Frailty Case

Survival Models with 'Frailties'

Variables: T_i Survival times, Discrete Covariates Z_i
C_i Censoring Times, cond surv fcn $R_z(c)$ given (T_i, C_i)

DATA: iid triples $(\min(T_i, C_i),\ I_{[T_i \leq C_i]},\ Z_i)$

Observable processes:

$$N_z^i(t) \;=\; I_{[Z_i=z,\ T_i \leq \min(C_i, t)]} \quad , \qquad Y_z^i(t) \;=\; I_{[\min(T_i, C_i) \geq t]}$$

TRANSF. MODEL: $S_{T|Z}(t|z) \;=\; \exp(-G(e^{\beta'z}\,\Lambda))$

G known , $\beta \in \mathbf{R}^m$, Λ cumulative-hazard fcn

$\boxed{\textbf{PROBLEM:}\ \text{efficient estimation of } \beta.}$

Special Cases: (1) *Cox 1972:* $G(x) \equiv x$

(2) Frailty: unobserved random intercept $\beta_0 = \xi_i$, $G \equiv x$

$$\implies \quad G(x) \;=\; -\log \int_0^\infty e^{-sx} dF(s)$$

(3) *Clayton-Cuzick 1986:* $G(x) \equiv \frac{1}{b}\,\log(1+bx)$

Finite-dimensional Case

$$X_i, \quad i = 1, \ldots, n \quad iid \quad \sim \quad f(x, \beta, \lambda),$$
$$\beta \in \mathbf{R}^m, \quad \lambda \in \mathbf{R}^d \quad \text{unknown , with true values } (\beta_0, \lambda_0)$$

$$logLik(\beta, \lambda) \;\; = \;\; \sum_{i=1}^{n} \log f(X_i, \beta, \lambda)$$

$$\textit{Profile Likelihood} \;\; = \;\; logLik(\beta, \hat{\lambda}_\beta) \qquad \text{with}$$
$$\text{restricted MLE} \qquad \hat{\lambda}_\beta \;\; = \;\; \arg\max_\lambda \; logLik(\beta, \lambda)$$

Min Kullback-Leibler Modified Profile Approach
(Severini and Wong 1992)

$$\mathcal{K}(\beta, \lambda) \;\; \equiv \;\; E_{\beta_0, \lambda_0}(\log f(X_1, \beta, \lambda))$$
$$= \;\; \int \{\log f(x, \beta, \lambda)\} \; f(x, \beta_0, \lambda_0) \; dx$$

Define: $\qquad \lambda_\beta \;\; = \;\; \arg\max_\lambda \; \mathcal{K}(\beta, \lambda)$

Then: $\qquad \tilde{\lambda}_\beta \;\; \text{estimates curve} \;\; \lambda_\beta$

Candidate Estimator

$$\tilde{\beta} \;\; \equiv \;\; \arg\max_\beta \;\; logLik(\beta, \tilde{\lambda}_\beta)$$

Key mathematical features of this approach are

- the convenience of restricting attention to nuisance parameters such as hazards or density functions which satisfy smoothness restrictions;

- the replacement of operator-inversion within (blocks of) the generalized information operator by differentiation of the restricted Kullback-Leibler minimizer; and

- diminished need for high-order consistency of estimation, when consistent estimators of Kullback-Leibler minimizers *and their derivatives with respect to structural parameters* are available.

NB: Kullback-Leibler Functional $= -\mathcal{K}(\beta, \lambda)$

Sketch of (Finite-dimensional) Theory

Fix β_0, λ_0 and $f_0(x) = f(x, \beta_0, \lambda_0)$.

Information Matrix : $\quad \mathcal{I}(\beta, \lambda) = \begin{pmatrix} A_{\beta,\lambda} & B_{\beta,\lambda} \\ B^*_{\beta,\lambda} & C_{\beta,\lambda} \end{pmatrix}$

$$= - \int \begin{pmatrix} \nabla^{\otimes 2}_\beta \log f(x,\beta,\lambda) & \nabla^2_{\beta\lambda} \log f(x,\beta,\lambda) \\ \nabla^2_{\lambda\beta} \log f(x,\beta,\lambda) & \nabla^{\otimes 2}_\lambda \log f(x,\beta,\lambda) \end{pmatrix} f_0(x) dx$$

$\lambda_\beta = \arg\max_\lambda \mathcal{K}(\beta, \lambda)$ satisfies: $\quad \nabla_\lambda \mathcal{K}(\beta, \lambda_\beta) = 0$

Differentiate implicitly (total deriv) wrt β to find:

$$-\nabla^T_\beta [\nabla_\lambda \mathcal{K}(\beta, \lambda_\beta))] = B + C \nabla'_\beta \lambda_\beta = 0$$

This implies $\qquad \nabla'_\beta \lambda_\beta = -C^{-1} B$ and

(β, λ_β) is a **least-favorable** nuisance-parameterization

and usual Information about β for this model is

$$I^0_\beta = A_{\beta_0,\lambda_0} - (\nabla'_\beta \lambda_\beta)' C_{\beta_0,\lambda_0} (\nabla'_\beta \lambda_\beta)$$

$$= A_{\beta_0,\lambda_0} - B_{\beta_0,\lambda_0} C^{-1}_{\beta_0,\lambda_0} B^*_{\beta_0,\lambda_0}$$

Theory, continued

Information Ineq says: $\lambda_{\beta_0} = \lambda_0$, $\quad \nabla_\beta \mathcal{K}(\beta_0, \lambda_0) = 0$

Now assume $\tilde{\lambda}_\beta$, $\nabla_\beta^{\otimes 2} \tilde{\lambda}_\beta$ consistent for λ_β, $\nabla_\beta^{\otimes 2} \lambda_\beta$ unif on nbhd of β_0.

Can check successively:

(A) $(\nabla_\beta^T)^{\otimes 2} logLik(\beta, \tilde{\lambda}_\beta)$ neg-def unif on β_0 nbhd

(B) $n^{-1} logLik(\beta_0, \tilde{\lambda}_{\beta_0}) \xrightarrow{P} 0$

(C) $\tilde{\beta}$ unique loc sol'n of $\nabla_\beta^T logLik(\beta, \tilde{\lambda}_\beta) = 0$, consistent for β_0.

(D) $n^{-1}\nabla_\lambda logLik(\tilde{\beta}, \tilde{\lambda}_{\tilde{\beta}}) = n^{-1}\nabla_\lambda logLik(\beta_0, \lambda_0) + o_P(\tilde{\beta} - \beta_0)$ $\qquad$ [because $-C^{-1}B = \nabla_\beta \tilde{\lambda}_\beta$]

(E) $n^{-1}\nabla_\beta logLik(\tilde{\beta}, \tilde{\lambda}_{\tilde{\beta}}) = n^{-1}\nabla_\beta logLik(\beta_0, \lambda_0) - $

$\qquad I_\beta^0 \cdot (\tilde{\beta} - \beta_0) + o_P(\tilde{\beta} - \beta_0)$

(F) $\sqrt{n}\,(\tilde{\beta} - \beta_0) = (I_\beta^0)^{-1} \frac{1}{\sqrt{n}} (\nabla_\beta logLik(\beta_0, \lambda_0) + $

$\qquad \nabla_\lambda' logLik(\beta_0, \lambda_0) \nabla_\beta \lambda_{\beta_0})$

(G) $\sqrt{n}\,(\tilde{\beta} - \beta_0) \stackrel{\mathcal{D}}{\approx} \mathcal{N}(\mathbf{0}, (I_\beta^0)^{-1})$

LITERATURE

Profile Likelihood

Cox, D. R. & Reid, N. (1987) **JRSSB**

McCullagh, P. and Tibshirani, R. (1990) **JRSSB**

Severini, T. and Wong, W. (1992) *Ann Stat*

Semiparametrics

Bickel, Klaassen, Ritov, & Wellner: 1993 Book

Cox, D. R. (1972)   *'Cox-Model' paper* **JRSSB**

Owen, A. (1988) *Biometrika*: Empirical likelihood

Qin, J. & Lawless, J. (1994) *Ann Stat*: EL & GEE

Murphy & van der Vaart (2000) *JASA*

Transformation/Frailty Models

Cheng, Wei, & Ying (1995) *Biometrika*

Clayton and Cuzick (1986) *ISI Centenary Session*

Slud, E. and Vonta, I. (2002) *submitted*

Parner (1998) *Ann. Stat.*

∞-dim Examples & Applications

(I). *Mean estimation in the location model*

$$X_i \sim \lambda_0(x - \beta_0), \quad \beta_0 = E(X_1)$$

$\lambda_0 \in L^1(dx)$ compactly supported 0-mean density.

(Easy example, $\overline{X}$ efficient. Owen 1988, Qin & Lawless 1994 studied in connection with *empirical likelihood*.)

$$\text{For } \beta \neq \beta_0, \qquad \lambda_\beta(x) = \lambda_0(x + \beta)/(1 - \alpha x)$$

with $\quad \alpha$ solving $\int x \, \lambda_\beta(x) \, dx = 0$

Define $\tilde{\lambda}_\beta$ first at β_0 using density estimator

$$\tilde{\lambda}_{\beta_0}(x) = \frac{1}{n \, h_n} \sum_{i=1}^{n} \phi((x - X_i + \overline{X})/h_n), \quad h_n \searrow 0$$

$$\tilde{\lambda}_\beta(x) = \frac{\tilde{\lambda}_{\beta_0}(x + \beta)}{1 - \tilde{\alpha}x}, \qquad \int \frac{x \, \tilde{\lambda}_{\beta_0}(x + \beta)}{1 - \tilde{\alpha}x} \, dx = 0$$

(II). *Cox model.* For $q_z(t) = p_Z(z) R_z(t) \exp(-e^{z'\beta_0} \Lambda_0(t))$,

$$\Lambda_\beta(t) \equiv \int_0^t \lambda_\beta(s) \, ds = \frac{\Sigma_z \, q_z(x) \, e^{z'\beta_0} \, \lambda_0(x)}{\Sigma_z \, e^{z'\beta} \, q_z(x)}$$

8

Example (II), cont'd.

First let $\tilde{\lambda}_0$ be consistent density estimate of $\lambda_0(x) = \Lambda_0'(x)$ (eg by smoothing and differentiating the Kaplan-Meier cumulative-hazard estimator on data in a $z = 0$ data-stratum.) Then estimate elsewhere by

$$\tilde{\lambda}_\beta(t) \; = \; \sum_z e^{z'\beta_0} \, Y_z(t) \, \tilde{\lambda}_{\beta_0}(t) \, / \, \sum_z e^{z'\beta} \, Y_z(t)$$

NB. In this example, any $\tilde{\beta}$ estimator produced in this way collapses to the usual Cox Max Partial Likelihood Estimator !

(III). *Transformation/Frailty Models*

In the general G transformation model case, must assume for some finite time τ_0 with $\Lambda^0(tau_0) < \infty$ that all data are censored at τ_0.

In this model, Slud and Vonta (2002) characterize the $\mathcal{K}$-optimizing hazard intensity λ_β in its integrated form $L = \Lambda_\beta = \int_0^\cdot \lambda_\beta(x)\,dx$, through the second order ODE system:

Example (III), cont'd.

$$\frac{dL}{d\Lambda_0}(s) \;=\; \frac{\Sigma_z \; e^{z'\beta_0} \, q_z(s) \, G'(e^{z'\beta_0}\Lambda_0(s))}{\Sigma_z \; e^{z'\beta} \, q_z(s) \, G'(e^{z'\beta}L(s)) \;+\; Q(s)}$$

$$\frac{dQ}{d\Lambda_0}(s) \;=\; \sum_z e^{z'\beta} \, q_z(s) \, \frac{G''}{G'}\big|_{e^{z'\beta}L(s)} \cdot$$

$$(e^{z'\beta_0} \, G'(e^{z'\beta_0}\Lambda_0(s)) - (e^{z'\beta} \, G'((e^{z'\beta}L(s)) \frac{dL}{d\Lambda_0}(s))$$

subject to the initial/terminal conditions

$$L(0) \;=\; 0 \quad, \qquad Q(\tau_0) \;=\; 0$$

Slud and Vonta (2002) show that these ODE's have
unique solutions, smooth with respect to β and dif-
ferentiable in t, which (with $\lambda_\beta \equiv L'$) maximize the
functional $\mathcal{J}(\beta, \lambda_\beta)$ as desired.

Consistent preliminary estimators $\tilde{\lambda}_\beta$ can be devel-
oped by substitute for β_0, Λ_0 in those equations (smoothed
wrt t) consistent preliminary estimators.

PUNCHLINE: the new estimator $\tilde{\beta} \;=\; \arg\max_\beta logLik(\beta, \tilde{\lambda}_\beta)$
is efficient !