NIH Talk, September '03

Efficient Semiparametric Estimators via Modified Profile Likelihood in Frailty & Accelerated-Failure Models

Eric Slud, Math Dept, Univ of Maryland

Ongoing joint project with Ilia Vonta, Univ. of Cyprus.
The talk is based on joint papers:

(i) about NPMLE *Scan. Jour. Stat.* to appear 2003.

(ii) about Modified Profile Likelihood, JSPI 2004?

TALK OUTLINE

I. MOTIVATION — Transformation Model Problems

II. APPROACH — Finite-Dimensional Version

III. BACKGROUND LITERATURE — Profile Likelihood, Frailty & Accelerated Failure Models

IV. FRAILTY CASE (info bounds)

V. ACCELERATED-FAILURE CASE (more details)

Transformation Survival Models

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

DATA: iid triples $(\min(T_i, C_i),\ I[T_i \le C_i],\ Z_i)$

Observable processes:

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

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

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

PROBLEM: efficient estimation of β.

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

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

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

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

Cox-Model Case , $G(x) = x$

$$S_{T|Z}(t|z) \;=\; \exp(-e^{\beta' z}\Lambda(t)) \;\;, \quad h_{T|Z}(t|z) \;=\; e^{\beta' z}\lambda(t)$$

is the *Proportional* or *Multiplicative Hazards* model.

Frailty

If $\beta' Z$ covariate has added to it an unobservable random-effect intercept $\log \xi$ called *frailty*,

$$P(T > t \,|\, Z = z) \;=\; E_\xi(\,\exp(-\xi\, e^{\beta' z}\Lambda(t)) \;\equiv\; \exp(-G(e^{\beta' z}\Lambda(t)))$$

In famous **Clayton-Cuzick (1986) frailty model** example, take $\xi \sim Gamma(b^{-1}, b^{-1})$, leading to

$$S_{T|Z}(t|z) \;=\; (1{+}be^{\beta' z}\Lambda(t))^{-1/b} \;, \quad h_{T|Z}(t|z) \;=\; \frac{e^{\beta' z}\lambda(t)}{1 + be^{\beta' z}\Lambda(t)}$$

Why are these 'Transformation Models' ?

(*Recall cum. haz. at failure is Expon(1) variate V.*)

$$G(\,e^{\beta' Z}\,\Lambda(T)\,) \;=\; V$$

or for G known, but β, Λ unknown,

$$\log \Lambda(T) \;=\; \log G^{-1}(V) \;-\; \beta' Z$$

‘Accelerated-Failure’ Models

These are also Transformation Models. Covariates now have an additive effect on transformed time-variable:

T_0 is ‘neutral’ individual’s failure-time
g a (known) ‘measurement scale’
survival fcn of $g(T_0)$ unknown and equal to $K(e^t)$.

Now suppose

$$g(T) \;=\; \beta' Z \;+\; g(T_0)$$

Then $\qquad S_{T|Z}(t|z) \;=\; P(g(T) > g(t)\,|\,Z = z)$

$$=\; P(g(T) > g(t) + \beta' z) \;=\; K(e^{\beta' z + g(t)})$$

has transformation-model form, for K unknown, g known (often equal to log).

Note: this is the same as the transformation model

$$\log \Lambda(T) \;=\; \log G^{-1}(V) \;-\; \beta' Z$$

for frailty if Λ were known and G unknown !

Finite-dimensional Case

$X_i, \;\; i = 1, \ldots, n \quad iid \quad \sim \quad f(x, \beta, \lambda),$
$\beta \in \mathbf{R}^m, \;\; \lambda \in \mathbf{R}^d$ unknown , with true values (β_0, λ_0)

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

Profile Likelihood $= \; logLik(\beta, \; \hat{\lambda}_\beta)$ with

restricted MLE $\quad \hat{\lambda}_\beta \; = \; \arg\max_\lambda \; logLik(\beta, \lambda)$

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

Then: $\quad \tilde{\lambda}_\beta$ estimates curve λ_β

Candidate Estimator

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

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

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

Note that by definition of $\mathcal{K}$ and implicit (total) differentiation

$$-\nabla_\beta^T \left[\nabla'_\lambda \mathcal{K}(\beta,\lambda_\beta))\right] \;=\; B \,+\, C\, \nabla'_\beta \lambda_\beta \;=\; 0$$

The usual Information about β for this model, defined (as in the Cramer-Rao Inequality) as inverse of the minimum variance matrix for unbiased estimators of β, is

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

Equivalently, to test $\beta = \beta_0$, denoting ‘restricted MLE’ $\hat{\lambda}_r$ as maximizer of $logLik(\beta_0,\lambda)$, efficient test-statistic is

$$\frac{1}{\sqrt{n}} \left[\nabla_\beta logLik(\beta_0,\hat{\lambda}_r) \;-\; B^*_{\beta_0,\hat{\lambda}_r}\, (C^*_{\beta_0,\hat{\lambda}_r})^{-1}\, \nabla_\lambda\, logLik(\beta_0,\hat{\lambda}_r)\right]$$

Neyman (1959) indicated that the same efficiency for test-statistic can be obtained much more generally, with $\hat{\lambda}_r$ replaced by ‘preliminary’ estimator consistent for λ_0 at rate $o_P(n^{-1/4})$.

Key mathematical features of the **modified profile** approach via $\tilde{\beta}$, $\tilde{\lambda}_\beta$ are:

- the technical convenience of restricting attention to nuisance parameters such as hazards or density functions which satisfy smoothness restrictions;
- replacement of operator-inversion within (blocks of) the generalized information operator by differentiation of the Kullback-Leibler minimizer, since $\nabla'_\beta \lambda_\beta \; = \; -C^{-1} B$ and the semiparametric Information about β is

$$\mathcal{J}(\beta_0, \lambda_0) \; = \; A_{\beta_0,\lambda_0} \, - \, (\nabla^*_\beta \lambda_{\beta_0})^* \, C_{\beta_0,\lambda_0} \, (\nabla^*_\beta \lambda_{\beta_0})$$

  and
- there is no need for high-order consistency of estimation of λ, when consistent estimators of K-L minimizers λ_β *and their derivatives with respect to structural parameters* are available.

Whether dimension of nuisance parameter is finite or infinite, under regularity conditions:

$$\sqrt{n}\,(\tilde{\beta} \, - \, \beta_0) \; \overset{\mathcal{D}}{\approx} \; \mathcal{N}(\mathbf{0}, \; (\mathcal{J}(\beta_0, \lambda_0))^{-1})$$

Semiparametric Case (λ ∞-dim)

Kullback-Leibler Functional

$$\mathcal{K}(\beta, \lambda) = - \int (\log f(x, \beta, \lambda))\, f(x, \beta_0, \lambda_0)\, dx$$

Define

$$\lambda_\beta = \arg\max{}_\lambda \; \mathcal{K}(\beta, \lambda)$$

to satisfy: $\quad \nabla_\lambda \, \mathcal{K}(\beta, \lambda_\beta) = 0$

Under minimal regularity conditions:

$(\beta,\, \lambda_\beta)$ is a **least-favorable** nuisance-parameterization

Substitute preliminary estimators $\tilde{\beta}_0$, $\tilde{\lambda}_0$ (usually involves density-estimator for λ_0), into λ_β formula to get estimator $\tilde{\lambda}_\beta$.

Then **maximize** over β within

$$\textit{logLik}(\beta, \tilde{\lambda}_\beta) = \sum_{i=1}^{n} \log f_X(X_i, \beta, \tilde{\lambda}_\beta)$$

LITERATURE

Profile Likelihood

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

McCullagh, P. & 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**

Murphy & van der Vaart (2000) *JASA*

Transformation/Frailty Models

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

Clayton & Cuzick (1986) *ISI Centenary Session*

Slud, E. & Vonta, I. (2003) *Scand. Jour. Stat*

Parner (1998) *Ann. Stat.*

Accelerated Failure Time Models

Koul, Susarla & van Ryzin (1981) *Ann. Stat.*

Tsiatis (1990) *Ann. Stat.*

Ritov (1990) *Ann. Stat.*

∞-dim Examples

(I). *Cox model.*

For $q_z(t) = p_Z(z) R_z(t) \exp(-e^{z'\beta_0}\Lambda_0(t))$, can solve uniquely for:

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

Let $\tilde{\lambda}_0$ be a consistent density estimate of $\lambda_0(x)$ (eg by smoothing and differentiating the Nelson-Aalen estimator on data in a $z = 0$ data-stratum.) Estimate $q_z(t)$ by kernel-smoothing the *at-risk process* $Y_z(t)/n$,

$$\tilde{\lambda}_\beta(t) \;=\; \sum_{i=1}^n e^{Z_i'\beta_0}\, A(\frac{t - T_i}{b_n})\, \tilde{\lambda}_{\beta_0}(T_i) \,/\, \sum_{i=1}^n e^{Z_i'\beta_0}\, A(\frac{t - T_i}{b_n})$$

where A is a smooth cdf (kernel) and b_n a bandwidth parameter decreasing slowly to 0 as $n \to \infty$.

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

(II). *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 & Vonta (2003) 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:

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

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

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

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 differentiable in t, which (with $\lambda_\beta \equiv L'$) minimize the functional $\mathcal{K}(\beta,\, \lambda_\beta)$ as desired.

Consistent preliminary estimators $\tilde{\lambda}_\beta$ can be developed by substituting for β_0, Λ_0 in those equations (smoothed with respect to t) consistent preliminary estimators.

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

Software for these estimators so far 'not ready for prime time' because of need for general-purpose two-point boundary value problem for ODE, but has been used to generate formulas for Semiparametric Information.

That is technically easier because it only involves the **adjoint ODE system** obtained by differentiating the one above *at the true values* (β_0, λ_0) with respect to a parameter $Q(0)$.

Censored Linear Regression

Censored linear regression model (usually, for log-survival times) assumes ϵ_i independent of (Z_i, C_i) in

$$X_i \;=\; \beta^{tr} Z_i + \epsilon_i \quad , \qquad Z_i \text{ and } \epsilon_i \quad \text{independent}$$

Data: *iid* triples $(T_i,\, Z_i,\, \Delta_i)$,

$$T_i \;=\; \min(X_i,\, C_i) \quad , \quad \Delta_i \;=\; I_{[X_i \le C_i]}$$

Unknown parameters: β , $\lambda(u) \equiv F'_\epsilon(u)/(1 - F_\epsilon(u))$

Step 1. Preliminary estimator $\tilde{\beta}_0$ as in Koul-Susarla-van Ryzin (1981) by regression

$$\Delta_i\, T_i \,/\, \hat{S}_{C|Z}(T_i \,|\, Z_i) \qquad \text{on} \qquad Z_i$$

Generally, $\hat{S}_{C|Z}$ a kernel-based nonparametric regression estimator (Cheng, 1989). If Z_i, C_i independent, use Kaplan-Meier $\hat{S}^{KM}_C(T_i)$.

Step 2. $\tilde{\lambda}^0$ estimated by kernel-density variant of Nelson-Aalen estimator (Ramlau-Hansen 1983), with kernel cdf $A(\cdot)$, bandwidth $b_n \nearrow \infty$ slowly enough:

$$\tilde{\lambda}_0(w) \;=\; \frac{1}{b_n} \int \; A'(\,\frac{w-u}{b_n}) \, \frac{\Sigma_i \; dN_i(u + Z_i'\tilde{\beta}_0)}{\Sigma_i \; I_{[T_i \geq u + Z_i'\tilde{\beta}_0]}}$$

$$= \; \sum_{i=1}^{n} \; \Delta_i \, A'(\,\frac{w - T_i + Z_i'\tilde{\beta}_0)}{b_n}) \,/\, \sum_{j=1}^{n} \; I_{[T_j \geq T_i + (Z_j - Z_i)'\tilde{\beta}_0]}$$

Step 3. Next use K-L functional to find:

$$\lambda_\beta(t) \;\equiv\; \sum_z \; q_z(t + z'\beta) \, \lambda_0(t + z'(\beta - \beta_0)) \,/\, \sum_z \; q_z(t + z'\beta)$$

Step 4. Then define $\tilde{\lambda}_\beta$ by substituting $\tilde{\lambda}_0$ into

$$\sum_{i=1}^{n} \; A(\frac{T_i - t - Z_i'\beta}{b_n}) \, \tilde{\lambda}_0(t + Z_i'(\beta - \tilde{\beta}_0)) \,/\, \sum_{i=1}^{n} \; A(\frac{T_i - t - Z_i'\beta}{b_n})$$

and $\tilde{\Lambda}_\beta$ by numerical integral of $\tilde{\lambda}_\beta$ over $[0, t]$.

Step 5. Finally substitute these expressions into

$$logLik \;=\; \sum_{i=1}^{n} \; \{\Delta_j \; \log \tilde{\lambda}_\beta(T_j - Z_j'\beta) \;-\; \tilde{\Lambda}_\beta(T_j - Z_j'\beta)\}$$

and maximize numerically.