RIT talk 2 September, 2005

Approximating Semiparametric Information and Estimators by Parametric Ones

Eric Slud, Stat Program, UMCP

`OUTLINE`

I. Parametric Information as function of increasing nuisance-parameter dimension

II. Limiting Parametric Information versus Semiparametric Information Bound

III. Example: Regression with infinitely many predictors

IV. General Approach via *Modified Profile Likelihood*

Parametric Background

Data sample $X_1, \ldots, X_n \sim f(x, \beta, \lambda)$ *iid*

True parameter $\vartheta_0 \equiv (\beta_0, \lambda_0) \in \Theta$ open subset of $\mathbf{R}^d \times \mathbf{R}^k$

Regularity conditions, on integrability & dominatedness of ϑ-derivatives of $f(X_1, \vartheta)$, $\log f(X_1, \vartheta)$ under P_{ϑ_0}

Information Matrix $I_{\vartheta_0} = \begin{pmatrix} I_{\beta\beta} & I_{\beta\lambda} \\ I_{\lambda\beta} & I_{\lambda\lambda} \end{pmatrix} \equiv$

$$E_{\vartheta_0} \left\{ \begin{pmatrix} \nabla_\beta \\ \nabla_\lambda \end{pmatrix} \log f(X_1, \beta_0, \lambda_0) \begin{pmatrix} \nabla_\beta \\ \nabla_\lambda \end{pmatrix}^{tr} \log f(X_1, \beta_0, \lambda_0) \right\}$$

assumed positive-definite.

`Recall:` Information $\mathcal{J}^\beta(\vartheta_0)$ about β is

$$\Big(\mathrm{aVar}_{\vartheta_0}(\sqrt{n}\,\hat{\beta}_{ML})\Big)^{-1} = \Big((I_{\vartheta_0}{}^{-1})_{\beta\beta} \Big)^{-1} = I_{\beta\beta} - I_{\beta\lambda}\, I_{\lambda\lambda}{}^{-1}\, I_{\lambda\beta}$$

$$= \quad \mathrm{aVar}_{\vartheta_0}\big(\tfrac{1}{\sqrt{n}} \{\nabla_\beta - I_{\beta\lambda}\, I_{\lambda\lambda}{}^{-1}\, \nabla_\lambda\} \log f(X_1, \vartheta_0)\big)$$

aVar of **Orthog. Projection** of $\nabla_\beta \log f(X_1, \vartheta)$ on $(\mathrm{span}(\nabla_\lambda \log f(X_1, \vartheta)))^\perp$

Finite-dimensional Case

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

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

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

Neyman (1959: efficiency for test-statistic holds much more generally, with $\hat{\lambda}$ replaced by 'preliminary' estimator consistent for λ_0 at rate $o_P(n^{-1/4})$.

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: $\quad \lambda_\beta \quad = \quad \arg\min_\lambda \; \mathcal{K}(\beta, \lambda)$
Then: $\quad \tilde{\lambda}_\beta$ estimates curve λ_β

Candidate Estimator

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

This material is in many texts & papers introducing semiparametric ideas, including

- Charles Stein (1956) *3rd Berkeley Symposium*
- L. LeCam & G. Yang (1990) *book*
- Bickel, Klaassen, Ritov & Wellner (1993) *book*
- A. van der Vaart (1998) *Asymptotic Statistics* book

Consequences of these ideas:

(1) In problems with fixed P_{ϑ_0} parameterized with larger $\dim(\lambda)$, $\mathcal{J}^\beta(\vartheta_0)$ is smaller. Restrict attention only to alternative nuisance parameterizations $\lambda^{(k)}$ with values such that $P_{true} = P_{(\beta_0, \lambda_0^{(k)})}$.

(2) So in problems with ∞-dim nuisance parameter, can look at (decreasing) limit of $\mathcal{J}^\beta(\vartheta_0)$ over increasing $\dim(\lambda) \uparrow \infty$ as *limiting Fisher Information.*

(3) Alternatively define *semiparametric information bound* as $inf\, \mathcal{J}^\beta(\beta_0, \lambda_0^{(k)})$ over all finite-dim. parameterizations $\lambda^{(k)}$. Under regularity, *inf* attained at **least favorable submodel**.

Generally information bounds (2) < (3) **are not the same.** Wong (1986, Ann. Stat.) gave sufficient but restrictive conditions for them to be equal.

Infinite-Dimensional Regression

$Y_i \;=\; \Sigma_{j=1}^{\infty}\, Z_{ij}\, b_j \,+\, \epsilon_i \;\;, \quad (\{Z_{ij}\}_{j=1}^{\infty}, \epsilon_i)$ indep, *iid*

L^2 convergent expansion

For simplicity, $\epsilon_i \sim \mathcal{N}(0,1)$.

Consider submodels with $p+q$ column design matrix, $\{Z_{ij}\}_{j=1}^{\infty}$ Gaussian

$\mathbf{Z}_{p,q} \;=\; (\mathbf{Z}_p \,|\, \mathbf{Z}_q^*)\,, \quad \beta = (b_1, \ldots, b_p)\,, \;\; \lambda = (b_{p+1}, \ldots, b_{p+q})$

Information Matrix $\quad I_{\vartheta_0} \;=\; \begin{pmatrix} \mathbf{Z}_p^{tr}\mathbf{Z}_p & \mathbf{Z}_p^{tr}\mathbf{Z}_q^* \\ \mathbf{Z}_q^{*tr}\mathbf{Z}_p & \mathbf{Z}_q^{*tr}\mathbf{Z}_q^* \end{pmatrix}$

Information about β : $\mathbf{Z}_p^{tr}\mathbf{Z}_p - \mathbf{Z}_p^{tr}\underbrace{\mathbf{Z}_q^*\;(\mathbf{Z}_q^{*tr}\mathbf{Z}_q^*)^{-1}\,\mathbf{Z}_q^{*tr}}\mathbf{Z}_p$

The under-bracket indicates **projection** onto $\mathrm{col}(\mathbf{Z}_q^*)$.

So in this example, if V is the space spanned by $(Z_{1j}, \;\; j = p+1, \ldots, \infty,$ then the limiting Fisher information and semiparametric information bound are both given by: $E\left((Z_{1j})_{j=1}^{p}\;(I - \Pi_V)\;((Z_{1j})_{j=1}^{p})^{tr}\right)$

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 generalized information operator $\begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$ by differentiation of the KL minimizer, since $\nabla^*_\beta \lambda_\beta = -C^{-1} B$. Then the semiparametric Information about β is

$$\mathcal{J}(\beta_0 \,|\, \lambda_0) \;=\; A_{\beta_0,\lambda_0} \;-\; (\nabla^*_\beta \lambda_{\beta_0})^* \, C^{-1}_{\beta_0,\lambda_0} \, (\nabla^*_\beta \lambda_{\beta_0})$$

- 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

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

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

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

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*

Accelerated Failure / Censored Regression

Koul, Susarla & van Ryzin (1981) *Ann. Stat.*
Tsiatis (1990) *Ann. Stat.*

Partially Linear Models

Bhattacharyya, P.K. and Zhao (1997) Ann. Statist.

More on ∞-Order Regression

$$Y_i \;=\; \Sigma_{j=1}^{\infty}\, Z_{ij}\, b_j \;+\; \epsilon_i \;\;, \quad (\{Z_{ij}\}_{j=1}^{\infty}, \epsilon_i) \;\text{indep}, \;\; iid$$

<u>*Relation to Other Models*</u>

- Semiparametric regression: f_ϵ also unknown, but $\mathbf{b} = (\beta_1, \ldots, \beta_p, \mathbf{0})$.

- right-censored data version of semiparametric regression is treated as *accelerated failure time model*, usually with $Y_i = \log T_i$

- Partially linear regression: $Y_i - \beta' X_i - g(W_i) = \epsilon_i$, Now $(b_1, \ldots, b_p) = (\beta_1, \ldots, \beta_p)$, and X_i is the vector of first p components $(Z_{i1}, \ldots, Z_{ip})$.

  Other variables $Z_{i,p+k}$ can be orthogonal basis of functions $h_k(W_i)$ so that

$$g(w) \;=\; \sum_{k=1}^{\infty} b_{p+k} h_k(w)$$

  within suitable function space.

- Model of increasing-complexity covariates with sample size can be treated as ∞-order case where we fit p_n coeff's from finite increasing set of predictors.

Solving the KL-minimizing Eq'n

`Normal-data partially linear case`:

$$l(b, \sigma^2, \lambda) = -\frac{n}{2} \log \sigma^2 \; - \; \frac{1}{2\sigma^2} \sum_{i=1}^{n} \|Y_i - b' W_i - \lambda' Z_i\|^2$$

and per-observation Kullback-Leibler function is

$$\mathcal{K}((b, \sigma^2, \lambda); (b_0, \sigma_0^2, \lambda_0)) \; = \; \log \sigma^2 - \log \sigma_0^2$$

$$+ \, E \left\{ \sigma^{-2} \|Y_1 - b' W_1 - \lambda' Z_1\|^2 - \sigma_0^{-2} \|Y_1 - b_0' W_1 - \lambda_0' Z_1\|^2 \right\}$$

$$= \; \log \sigma^2 - \log \sigma_0^2 + (\frac{\sigma_0^2}{\sigma^2} - 1) + \frac{1}{\sigma^2} \, E \, \|b' W_1 + \lambda' Z_1 - b_0' W_1 - \lambda_0' Z_1\|^2$$

minimized at

$$\lambda_b \; = \; \lambda_0 \; - \; \Sigma_Z^{-1} \Sigma_{ZW} (b - b_0)$$

Interpretation of modified profile likelihood result: if we can produce a preliminary estimator for $\lambda_0, b_0,$ can use this to provide functional estimator for $\tilde{\lambda}_b$ and $\tilde{\beta}$ maximizing

$$logLik(b, \; \tilde{\lambda}_b)$$

`Semiparametric linear regression case`

In this case, take β and $\lambda_0 \equiv \lambda_\epsilon$ as unknown parameters. Put $q_z(t) = p_Z(z)\,\exp(-\Lambda_0(t - z'\beta^0))$ in case of discrete Z. Then

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

and preliminary estimator $\tilde{\lambda}_0$ immediately generates **efficient** estimator for β maximizing

$$logLik(\beta,\, \tilde{\lambda}_\beta)$$