Introduction to Empirical Processes and Semiparametric Inference Lecture 05: Overview Continued

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Example: Partly Linear Logistic Regression

The observed data are n i.i.d. realizations of (Y, Z, U), where

- $Z \in \mathbb{R}^p$ and $U \in \mathbb{R}$ are covariates which are not linearly dependent,
- Z is restricted to a bounded set, $U \in [0, 1]$,
- Y is a dichotomous outcome with conditional expectation $\nu[\beta' Z + \eta(U)]$,
- $\beta \in \mathbb{R}^p$ ($p = 1$ hereafter for simplicity),
- $\nu(t) = e^t/(1 + e^t)$, and where
- η is an unknown smooth function.

We assume, for some integer $k \geq 1$, that

$$J^2(\eta) = \int_0^1 \left[\eta^{(k)}(t)\right]^2 dt < \infty.$$

We estimate β and η based on the following penalized log-likelihood:

$$\tilde{L}_n(\beta, \eta) = n^{-1} \sum_{i=1}^{n} \log p_{\beta,\eta}(X_i) - \hat{\lambda}_n^2 J^2(\eta),$$

where

$$p_{\beta,\eta}(x) = \{\nu\left[\beta z + \eta(u)\right]\}^y \{1 - \nu\left[\beta z + \eta(u)\right]\}^{1-y}$$

and $\hat{\lambda}_n$ satisfies $\hat{\lambda}_n = o_P(n^{-1/4})$ and $\hat{\lambda}_n^{-1} = O_P(n^{k/(2k+1)})$.

Denote

- $\hat{\beta}_n$ and $\hat{\eta}_n$ to be the maximizers of $\tilde{L}_n(\beta, \eta)$,
- with $P_{\beta,\eta}$ the expectation under the model, and
- β_0 and η_0 the true values of the parameters.

Consistency of $\hat{\beta}_n$ and $\hat{\eta}_n$ and efficiency of $\hat{\beta}_n$ are established for partly linear generalized linear models in Mammen and van de Geer (1997).

We now

- derive the efficient score for β
- then sketch a verification that $\hat{\beta}_n$ is asymptotically linear
- with influence function equal to the efficient influence function.

Several difficult steps will be reserved for Chapter 15.

Our approach diverges only slightly from that used by Mammen and van de Geer (1997).

Let $\mathcal{H}$ be the linear space of functions $h : [0, 1] \mapsto \mathbb{R}$ with $J(h) < \infty$.

For $t \in [0, \epsilon)$ and ϵ sufficiently small, let

$$\beta_t = \beta + tv \text{ and } \eta_t = \eta + th$$

for $v \in \mathbb{R}$ and $h \in \mathcal{H}$.

If we differentiate the non-penalized log-likelihood, we deduce that the score for β and η, in the direction (v, h), is

$$(vZ + h(U))(Y - \mu_{\beta,\eta}(Z, U)),$$

where $\mu_{\beta,\eta}(Z, U) = \nu[\beta Z + \eta(U)]$.

Thus the usual score for β is

$$\dot{\ell}_{\beta,\eta}(X) = vZ(Y - \mu_{\beta,\eta}(Z,U)),$$

and the tangent space for η is

$$\dot{\mathcal{P}}^{(\eta)}_{P_{\beta,\eta}} = \left\{h(U)(Y - \mu_{\beta,\eta}(Z,U)) : \; h \in L_2(U)\right\}.$$

We need to find the projection of $\dot{\ell}_{\beta,\eta}(X)$ onto $\dot{\mathcal{P}}^{(\eta)}_{P_{\beta,\eta}}$, and subtract it from $\dot{\ell}_{\beta,\eta}$ to obtain the efficient score $\tilde{\ell}_{\beta,\eta}$.

Now let

$$h_1(u) = \frac{E\{ZV_{\beta,\eta}(Z,U)|U=u\}}{E\{V_{\beta,\eta}(Z,U)|U=u\}},$$

where $V_{\beta,\eta}(Z,U) = \mu_{\beta,\eta}(U,Z)(1-\mu_{\beta,\eta}(U,Z))$, and assume that h_1 is bounded (and thus $h_1 \in \mathcal{H}$).

Note first that $q_{\beta,\eta}(X) \equiv h_1(U)(Y - \mu_{\beta,\eta}(X)) \in \dot{\mathcal{P}}^{(\eta)}_{P_{\beta,\eta}}$ by assumption on h_1.

If we can also verify that $(Z - h_1(U))(Y - \mu_{\beta,\eta}(X))$ is uncorrelated with $h(U)(Y - \mu_{\beta,\eta}(X))$ for all $h \in L_2(U)$, then $q_{\beta,\eta}$ is the desired projection onto $\dot{\mathcal{P}}^{(\eta)}_{P_{\beta,\eta}}$.

Note that

$$
\begin{aligned}
& P\left[q_{\beta,\eta}(X)h(U)(Y-\mu_{\beta,\eta}(X))\right] \\
&= P\left[P\left\{(Z-h_1(U))h(U)\left(Y-\mu_{\beta,\eta}(X)\right)^2\Big| U\right\}\right] \\
&= P\left[P\left\{h_1(U)h(U)P[V_{\beta,\eta}(Z,U)|U]-h_1(U)h(U)P[V_{\beta,\eta}(Z,U)|U]\right\}\right] \\
&= 0,
\end{aligned}
$$

and thus we have the desired uncorrelation.

Hence $\tilde{\ell}_{\beta,\eta}(X)=(Z-h_1(U))(Y-\mu_{\beta,\eta}(X))$ is the efficient score for β.

Hence the efficient information for β is

$$\begin{aligned} \tilde{I}_{\beta,\eta} &= P\left[\tilde{\ell}_{\beta,\eta}\tilde{\ell}'_{\beta,\eta}\right] \\ &= P_{\beta,\eta}\left[(Z - h_1(U))^2 V_{\beta,\eta}(Z,U)\right] \end{aligned}$$

and the efficient influence function is

$$\tilde{\psi}_{\beta,\eta} = \tilde{I}^{-1}_{\beta,\eta}\tilde{\ell}_{\beta,\eta},$$

provided $\tilde{I}_{\beta,\eta} > 0$, which we assume hereafter for $\beta = \beta_0$ and $\eta = \eta_0$.

In order to prove asymptotic linearity of $\hat{\beta}_n$, we need to slightly strengthen our assumption that Z and U are not linearly dependent to require that

$$P_{\beta_0,\eta_0}\left[Z - \tilde{h}_1(U)\right]^2 > 0,$$

where $\tilde{h}_1(u) = E\{Z|U = u\}$.

We prove in Chapter 15 that $\hat{\beta}_n$ and $\hat{\eta}_n$ are both uniformly consistent for β_0 and η_0, respectively, and that

$$\mathbb{P}_n\left[(\hat{\beta}_n - \beta_0)Z + \hat{\eta}_n(U) - \eta_0(U)\right]^2 \quad = \quad o_P(n^{-1/2}), \qquad (1)$$

which we will for now take as given.

Let $\hat{\beta}_{ns} = \hat{\beta}_n + s$ and $\hat{\eta}_{ns}(u) = \hat{\eta}_n(u) - sh_1(u)$.

If we now differentiate $\tilde{L}_n(\hat{\beta}_{ns}, \hat{\eta}_{ns})$ and evaluate at $s = 0$, we obtain

$$\begin{aligned} 0 &= \mathbb{P}_n \left[(Y - \mu_{\beta_0,\eta_0})(Z - h_1(U)) \right] \\ &\quad - \mathbb{P}_n \left[(\mu_{\hat{\beta}_n,\hat{\eta}_n} - \mu_{\beta_0,\eta_0})(Z - h_1(U)) \right] \\ &\quad - \lambda_n^2 \left\{ \partial J^2(\hat{\eta}_{ns}) / (\partial s)|_{s=0} \right\} \\ &= A_n - B_n - C_n, \end{aligned}$$

since $\tilde{L}_n(\beta, \eta)$ is maximized at $\hat{\beta}_n$ and $\hat{\eta}_n$ by definition.

Using (1), we obtain by Taylor expansion

$$\begin{aligned} B_n &= \mathbb{P}_n \left[V_{\beta_0,\eta_0}(Z,U) \left\{ (\hat{\beta}_n - \beta_0)Z + \hat{\eta}_n(U) - \eta_0(U) \right\} (Z - h_1(U)) \right] \\ &\quad + \mathbb{P}_n \left[\frac{\dot{V}_{\beta^*,\eta^*}(Z,U)}{2} \left\{ (\hat{\beta}_n - \beta_0)Z + \hat{\eta}_n(U) - \eta_0(U) \right\}^2 \right. \\ &\quad \left. \times (Z - h_1(U)) \right] \\ &= Q_n + E_n, \end{aligned}$$

where (β^*, η^*) is on the line segment between $(\hat{\beta}_n, \hat{\eta}_n)$ and (β_0, η_0).

Since $\dot{V}_{\beta^*,\eta^*}(X)$, Z and $h_1(U)$ are all bounded, we obtain from (1) that $E_n = o_P(n^{-1/2})$.

By definition of h_1, we have

$$P_{\beta_0,\eta_0}\left[V_{\beta_0,\eta_0}(Z,U)(\hat{\eta}_n(U)-\eta_0(U))(Z-h_1(U))\right]=0.$$

If we can also show that

$$\sqrt{n}(\mathbb{P}_n-P)V_{\beta_0,\eta_0}(Z,U)(\hat{\eta}_n(U)-\eta_0(U))(Z-h_1(U)) \quad = \quad o_P(1), \tag{2}$$

then we know that

$$B_n \quad = \quad (\hat{\beta}_n-\beta_0)\mathbb{P}_n\left[V_{\beta_0,\eta_0}(Z,U)(Z-h_1(U))\right]+o_P(n^{-1/2}). \tag{3}$$

To verify (2), we need to show that for each $\tau > 0$, $\hat{\eta}_n(U) - \eta_0(U)$ lies in a bounded P_{β_0,η_0}-Donsker class with probability $> (1-\tau)$ for all $n \geq 1$ large enough and all $\tau > 0$.

Then, since products of bounded Donsker classes are themselves Donsker, we will have that $V_{\beta_0,\eta_0}(X,U)(\hat{\eta}_n(U) - \eta_0(U))(Z - h_1(U))$ also lies in a Donsker class with increasingly high probability.

Since also

$$P_{\beta_0,\eta_0}\left[V_{\beta_0,\eta_0}(Z,U)(\hat{\eta}_n(U) - \eta_0(U))(Z - h_1(U))\right]^2 \xrightarrow{\mathsf{P}} 0,$$

the desired conclusion (2) will follow.

This means that (3) will follow, and thus

$$\begin{aligned} B_n &= (\hat{\beta}_n - \beta_0) P_{\beta_0,\eta_0} \left[V_{\beta_0,\eta_0}(Z,U)(Z - h_1(U))^2 \right] + o_P(n^{-1/2}), \\ &= (\hat{\beta}_n - \beta_0) \tilde{I}_{\beta_0,\eta_0} + o_P(n^{-1/2}), \end{aligned} \qquad (4)$$

since

$$P_{\beta_0,\eta_0} \left[V_{\beta_0,\eta_0}(Z,U) Z (Z - h_1(U)) \right] = P_{\beta_0,\eta_0} \left[V_{\beta_0,\eta_0}(Z,U)(Z - h_1(U))^2 \right]$$

by definition of h_1.

If we can show that $C_n = o_P(n^{-1/2})$, then since $A_n = \mathbb{P}_n \tilde{\ell}_{\beta_0,\eta_0}$, we have

$$\mathbb{P}_n \tilde{\ell}_{\beta_0,\eta_0} - (\hat{\beta}_n - \beta_n)\tilde{I}_{\beta_0,\eta_0} = o_P(n^{-1}),$$

and thus $\sqrt{n}(\hat{\beta}_n - \beta_0) = \sqrt{n}\mathbb{P}_n \tilde{\psi}_{\beta_0,\eta_0} + o_P(1)$, where

$$\tilde{\psi}_{\beta_0,\eta_0} = \tilde{I}^{-1}_{\beta_0,\eta_0} \tilde{\ell}_{\beta_0,\eta_0}$$

is the efficient influence function.

Thus $\hat{\beta}_n$ is regular and semiparametric efficient!

We are now left with the tasks of verifying that

1. $\hat{\eta}_n(U) - \eta_0(U)$ lies in a Donsker class with increasingly high probability and
2. $C_n = o_P(n^{-1/2})$.

For 1, let $\mathcal{H}_c$ be the subset of $\mathcal{H}$ with functions h satisfying $J(h) \leq c$ and $\|h\|_\infty \leq c$.

We will show in Chapter 15 that $\{h(U) : h \in \mathcal{H}_c\}$ is indeed Donsker for each $c < \infty$ and also that $J(\hat{\eta}_n) = O_P(1)$.

This yields the desired Donsker property, and 1 follows.

Since

$$\frac{\partial}{\partial s}\int_0^1 [\hat{\eta}_n(u) - s\eta_0(u)]^2\,du\bigg|_{s=0} = -2\int_0^1 \hat{\eta}_n(u)\eta_0(u)du$$

and $\lambda_n = o_P(n^{-1/4})$ by assumption,

$$C_n \le 2\lambda_n^2 J(\hat{\eta}_n)J(\eta_0) = o_P(n^{-1/2}),$$

and 2 follows.

In this example, the steps for solving the problem are essentially:

- Obtain consistency of $\hat{\beta}_n$ and $\hat{\eta}_n$ (uniform).
- Show that $J(\hat{\eta}_n) = O_P(1)$.
- Derive the efficient score $\tilde{\ell}_{\beta_0,\eta_0}$, efficient information $\tilde{I}_{\beta_0,\eta_0}$ and efficient influence function $\tilde{\psi}_{\beta_0,\eta_0}$.
- Show that $\sqrt{n}(\hat{\beta}_n - \beta_0) = \sqrt{n}\mathbb{P}_n\tilde{\psi}_{\beta_0,\eta_0} + o_P(n^{-1/2})$.

With three additional/intermediate steps including:

- Show that $P\left[\hat{\eta}_n(U) - \eta_0(U)\right]^2 = o_P(1)$.
- Show that $\mathbb{P}_n\left[(\hat{\beta}_n - \beta_0)Z + \hat{\eta}_n(U) - \eta_0(U)\right]^2 = o_P(n^{-1/2})$.
- Show that $\|\hat{\eta}_n(U) - \eta_0(U)\|_{P,2} = o_P(n^{-k/(2k+1)})$.