Estimation of

(iii)the marginal mean of response curve in the presence of dependent censoring by time varying factors

James Robins, Eric Tchetgen,
Lingling Li, Aad van der vaart.

$$Obs\ Data : (Y, X, R)$$
$$Y = \text{ binary } response$$
$$X = high \text{ dim } vect \text{ of cont baseline confounders}$$
$$R = binary\ treatment\ indicator$$

Parameter of Interest:

$$Average\ Treatment\ Effect = E[Y(1)] - E[Y(0)]$$

,

Identifiable Under Assumtion of No Unmeasured Confounders

$$(Y(1), Y(0)) \text{ ind } R|X$$
$$p[R = 1|X, Y(1)]$$

as Functional

$$\psi_1 - \psi_0$$

of Obserevd data Distribution

$$One\ Occassion$$

$$O = (RY, X, R)$$

$$L = (Y, X, R)$$

$$Y = \text{ binary } response$$

$$X = high \text{ dim } vect\text{or of always observed cov}$$

$$R = binary\ mis \sin gness\ indicator$$

$$\psi = E[Y]$$

$$pr[R = 1|Y(1), X] = pr[R = 1|, X], \text{ MAR}$$

$$\psi = E[b(X)] = E\left[\frac{R}{\omega(X)} Y\right]$$

$$b(X) = E[Y|X, R = 1]$$

$$\omega(X) = pr[R = 1|X] = E[R|X]$$

In desined missiness, $\omega(X)$ known by design so HT estimator of ψ

$$n^{-1} \sum_i \frac{I(R_i = r)}{\omega_r(X_i)} Y_i$$

is unbiased (not efficient) , AN, $n^{1/2} - consistent.$

Allows for asymptotically 1-α Wald confidence for the ATE ψ

In observational studies $\omega(X)$ is unknown.

How do we proceed?

Longitudinal Studies:

Dependent Right Censoring -Monotone Missingness:

$$TemporaX\ Obs\ Data : (X_0, R_0, R_0 X_1, R_1, R_0 R_1 Y)$$

$$Y = \text{ binary } response$$

$$\psi = E[Y]$$

CAR

$$E[R_0 = 1 | X_0, (Y, X_1)] = E[R_0 = 1 | X_0]$$

$$E[R_1 | X_0, X_1, R_0 = 1, Y] = E[R_1 | X_0, X_1, R_0 = 1]$$

$$b_1(X_0, X_1) = E[Y | X_0, X_1]$$

$$= E[Y | X_0, X_1, R_1 = 1, R_0 = 1]$$

$$b_0(X_0) = E[Y | X_0] =$$

$$E[b_1(X_0, X_1) | X_0, R_0 = 1]$$

$$\omega_1(X_0, X_1) = E[R_1 | X_0, X_1, R_0 = 1]$$

$$\omega_0(X_0) = E[R_0 | X_0]$$

$$
\begin{aligned}
&\psi \\
&= \int b_1(X_0, X_1) f(X_1 | R_0 = 1, X_0) f(X_0) \, dX_0 dX_1 \\
&= E\left[\frac{R_1 R_0}{\omega_1(X_0, X_1)\, \omega_0(X_0)} Y\right] \\
&= E\left[b_0(X_0)\right]
\end{aligned}
$$

Sequential known design $\omega_1(X_0, X_1), \omega_0(X_0)$ known

-

$$
n^{-1} \sum_i \left[\frac{R_{1i} R_{0i}}{\omega_1(X_{0i}, X_{1i})\, \omega_0(X_{0i})} Y_i\right]
$$

is unbiased AN, $n^{1/2} - consistent$

Back to Point Exposure Study: All models in book with van der laan on complex longitudinal data

Our goal is to construct confidence intervals fin the non-parametric model subject only to smoothness restrictions.

Specifically we will use the following smoothness restrictions.

Def: Define $h(X)$ to be Holder β with radius c if

(i)all partial derivatives up to order int[β] exist and

(ii) all int[β] partial derivatives $m(X)$ satisfy

$$|m(x) - m(z)| < c\,||x - z||^{\beta - int[\beta]}$$

for some known constant c.

Our goal is to construct asymptotic 1-α confidence intervals (length shrinking to 0 fastest rate) given model that $b(X\ ), \omega(X), f(X)$ lie in the Holder balls

$$(\beta_b, c_b)\ (\beta_\omega, c_\omega), \left(\beta_f, c_f\ \right)$$

.

Not possible if no bounds on smothness given. Robins and Ritov (1997) no uniform consistent estimates of ψ

without restrictions on either (β_b, c_b) or (β_ω, c_ω).

Uniform consistency required for CI

In practice sensitivity analysis as adaption not possible for CI.

Training on Prior Smoothness . Criteria Other than Derivatives

Approx: If a $h(X)$ is a Holder β^*, optimal approximation bias based on k terms is

$$E\left[\left\{h(X) - \overline{h}_k(X)\right\}^2\right] = O\left(k^{-2\beta^*/d}\right)$$

if

$$\overline{Z}_k = \overline{z}_k(X) = (z_1(X) \ldots . z_k(X))$$

is polynomial, spline or compact wavelet basis with appropriate no. of vanishing moments. .

$$\overline{h}_k(X) = \eta_{PLS}^T \overline{z}_k(X)$$

Est: If $h\left(X\right)$ is a regression function or density , the optimal rate of estimation i.e. $\left|\widehat{h}\left(X\right)-h\left(X\right)\right|$ is $n^{\frac{-\beta^{*}}{2\beta^{*}+d}}$ in L_p norms

Because Holder β^{*} functions are dense in L_2, we have nonparametric model.

We will use the theory of higher order influence functions to construct our intervals.

Obtain adaptive estimate $\widehat{\theta} = \left(\widehat{b}, \widehat{\omega}, \widehat{f}\right)$,

Conider plug in

$$\psi\left(\widehat{\theta}\right) = \int \widehat{f}(x)\,\widehat{b}(x)\,dx.$$

Bias can be first order.

Usual Solution Use: $\widehat{\psi}_1 = \psi\left(\widehat{\theta}\right) + IF_1\left(\widehat{\theta}\right)$

$$IF_1\left(\widehat{\theta}\right) = n^{-1}\sum_i \frac{R_i}{\widehat{\omega}(X_i)}\left(Y_1 - \widehat{b}(X_i)\right) + \widehat{b}(X_i) - \psi\left(\widehat{\theta}\right)$$

$$\widehat{\psi}_1 = n^{-1}\sum_i \frac{R_i}{\widehat{\omega}(X_i)}\left(Y_1 - \widehat{b}(X_i)\right) + \widehat{b}(X_i)$$

$$E\left[\widehat{\psi}_1 - \psi(\theta)\,|\widehat{\theta}\right]$$

$$= E\left[\left\{b(X_i) - \widehat{b}(X_i)\right\}\{\omega(X_i) - \widehat{\omega}(X_i)\}/\widehat{\omega}(X_i)\right]$$

Doubly robust.

But Confidence intervals $\widehat{\psi}_1$

$$\widehat{\psi}_1 \pm var\left[IF_1\left(\widehat{\theta}\right)\right]^{1/2} z_\alpha$$

$$Width = O\left(n^{-1/2}\right)$$

fail unless

$$\left\{b\left(X_i\right) - \widehat{b}\left(X_i\right)\right\}\left\{\omega\left(X_i\right) - \widehat{\omega}\left(X_i\right)\right\} \leq O_p\left(n^{-1/2}\right)$$

$$\left\{b\left(X_i\right)-\widehat{b}\left(X_i\right)\right\}\left\{\omega\left(X_i\right)-\widehat{\omega}\left(X_i\right)\right\} \leq O_p\left(n^{-1/2}\right)$$

$$\Leftrightarrow$$

$$n^{\frac{-\beta_\omega^*}{2\beta_\omega^*+d}} n^{\frac{-\beta_b^*}{2\beta_b^*+d}} \leq n^{-1/2}$$

This requires

$$\beta_b^* = \beta_\omega^* \geq .5d, \text{ then}$$
$$n^{\frac{-\beta^*}{2\beta^*+d}} = n^{\frac{-.5d}{2(.5)d+d}} = n^{-1/4}$$

.

If this fails what then????

Use $\widehat{\psi}_m = \psi\left(\widehat{\theta}\right) + IF_m\left(\widehat{\theta}\right)$ where $IF_m\left(\widehat{\theta}\right)$ higher order IF and we sample split

$$\widehat{\psi}_m \pm var\left[IF_m\left(\widehat{\theta}\right)\right]^{1/2} z_\alpha$$

$$IF_2\left(\widehat{\theta}\right) = IF_1\left(\widehat{\theta}\right) + IF_{2,2}\left(\widehat{\theta}\right)$$

$$IF_{2,2}\left(\widehat{\theta}\right) = \left[n\left(n-1\right)\right]^{-1} \sum_{i \neq j} h_2\left(O_i, O_j; \widehat{\theta}\right)$$

$$h_2\left(O_i, O_j; \widehat{\theta}\right) = \frac{R_i}{\widehat{\omega}\left(X_i\right)}\left(Y_i - \widehat{b}\left(X_i\right)\right) \overline{Z}_{ki}^T \overline{Z}_{kj} \times \left(\frac{R_j}{\widehat{\omega}\left(X_j\right)} - 1\right)$$

Must choose $k = k_2(n)$ such that

$$var\left\{IF_2\left(\widehat{\theta}\right)\right\}$$
$$= O\left(max\left\{\frac{1}{n}\frac{k}{n}, \frac{1}{n}\right\}\right)$$
$$=$$
$$E\left[IF_{2,2}\left(\widehat{\theta}\right) - \psi\right]^2$$
$$= \max\left\{\left[n^{-\frac{-2\beta_f^*}{2\beta_f^*+d}} + n^{\frac{-2\beta_\omega^*}{2\beta_\omega^*+d}}\right] n^{\frac{-2\beta_\omega^*}{2\beta_\omega^*+d}} n^{\frac{-2\beta_b^*}{2\beta_b^*+d}}, k^{-2\left(\beta_b^*+\beta_\omega^*\right)/d}\right\}$$
$$= \max\left\{EB^2, TB^2\right\}$$
$$EB\ .$$
$$= E\left[\left\{b\left(X_i\right) - \widehat{b}\left(X_i\right)\right\}^2 \left\{\omega\left(X_i\right) - \widehat{\omega}\left(X_i\right)\right\}^4 / \widehat{\omega}\left(X_i\right)^4\right]$$
$$+ E\left[\left[\frac{\left\{f\left(X_i\right) - \widehat{f}\left(X_i\right)\right\}}{\widehat{f}\left(X_i\right)}\right]^2 \left\{b\left(X_i\right) - \widehat{b}\left(X_i\right)\right\}^2 \left\{\frac{\omega\left(X_i\right) - \widehat{\omega}}{\widehat{\omega}\left(X_i\right)}\right.$$

$If\ \beta_b^*/d = \beta_\omega^*/d < 1/4 \Rightarrow k > n$

CI does not shrink at $n^{-1/2}$

Price of valid intervals

$If\ 1/2 > \beta_b^*/d = \beta_\omega^*/d > 1/4 \Rightarrow k < n$ if β_f^* is large

CI does shrink at $n^{-1/2}$ for some $m, m = 2$ if $\beta_f^*/d > .$

$Need$ higher order IF to get first order effciiency

$$IF_3\left(\widehat{\theta}\right) = IF_2\left(\widehat{\theta}\right) + IF_{3,3}\left(\widehat{\theta}\right)$$

$$IF_{3,3}\left(\widehat{\theta}\right) = [n(n-1)(n-2)]^{-1} \sum_{i \neq j \neq s} h_3\left(O_i, O_j, O_s; \widehat{\theta}\right)$$

$$h_3\left(O_i, O_j, O_s; \widehat{\theta}\right) = \frac{R_i}{\widehat{\omega}(X_i)}\left(Y_i - \widehat{b}(X_i)\right) \times \overline{Z}_{ki}^T \left\{\overline{Z}_{ks}\overline{Z}_{ks}^T - I\right\} \overline{Z}_{kj} \times \left(\frac{R_j}{\widehat{\omega}\left(X_j\right)} - 1\right)$$

Must choose $k = k_3(n)$ such that

$var\left\{IF_3\left(\widehat{\theta}\right)\right\}$

$= O\left(\frac{1}{n}\frac{k}{n}\frac{k}{n}\right)$

$=$

$E\left[IF_{3,3}\left(\widehat{\theta}\right) - \psi\right]^2$

$= \max\left\{\left[n^{-\frac{-4\beta_f^*}{4\beta_f^*+d}} + n^{\frac{-4\beta_\omega^*}{4\beta_\omega^*+d}}\right] n^{\frac{-2\beta_\omega^*}{2\beta_\omega^*+d}} n^{\frac{-2\beta_b^*}{2\beta_b^*+d}}, k^{-2\left(\beta_b^*+\beta_\omega^*\right)/d}\right\}$

EB .

$= E\left[\left\{b(X_i) - \widehat{b}(X_i)\right\}\{\omega(X_i) - \widehat{\omega}(X_i)\}^3 / \widehat{\omega}(X_i)^3\right]$

$+ E\left[\left[\frac{\left\{f(X_i) - \widehat{f}(X_i)\right\}}{\widehat{f}(X_i)}\right]^2 \left\{b(X_i) - \widehat{b}(X_i)\right\}\left\{\frac{\omega(X_i) - \widehat{\omega}(}{\widehat{\omega}(X_i)}\right.\right.$

The advantage of $IF_3\left(\widehat{\theta}\right)$ is that if $IF_2\left(\widehat{\theta}\right)$ has estimation bias dominate truncation bias

then $IF_3\left(\widehat{\theta}\right)$ with smaller EB

can improve rate of convergence.

Mapping from smoothness assumptions to optimal CI?

What smoothness or other size controlling assumptions.

Given a sufficiently smooth $p-$ dimensional parametric submodel $\widetilde{\theta}(\varsigma)$ mapping $\varsigma \in R^p$ injectively into Θ, define

$$\psi_{\backslash i_1 \ldots i_m}(\theta) = \left(\psi \circ \widetilde{\theta}\right)_{\backslash i_1 \ldots i_m}(\varsigma)\,|_{\varsigma = \widetilde{\theta}^{-1}(\theta)}$$

and

$$f_{\backslash i_1 \ldots i_m}(\mathbf{O}; \theta) = \left(f \circ \widetilde{\theta}\right)_{\backslash i_1 \ldots i_m}(\varsigma)\,|_{\varsigma = \widetilde{\theta}^{-1}(\theta)}$$

where each $i_s \in \{1, ..., p\}$

$$f(\mathbf{O}; \theta) \triangleq \prod_{i=1}^{n} f(O_i; \theta)$$

Canonical (Hoeffding) Representation of Order 1 and 2 Mean 0 U-stat-

$$U_1(\theta) = \sum_{i \neq j} u_1(O_i), E[u_1(O_i)] = 0 :$$
$$u(\cdot,\cdot) \ \textit{not} \text{ necc } \textit{sym}$$

$$U_2(\theta) = \sum_{i \neq j} u\left(O_i, O_j\right), E\left[u\left(O_i, O_j\right)\right] = 0 :$$
$$u(\cdot,\cdot) \ \textit{not} \text{ necc } \textit{sym}$$

$$U_2(\theta) = \sum_{i} d(O_i, \theta) + \sum_{i \neq j} m\left(O_i, O_j\right),$$
$$E_\theta[d(O_i, \theta)] = 0,$$
$$E\left[m\left(O_i, O_j\right) | O_i\right] = E\left[m\left(O_i, O_j\right) | O_j\right] = 0,$$
$$m(\cdot,\cdot) \ \textit{not} \text{ necc } \textit{sym}$$
$$\sum_{i} d(O_i, \theta) \text{ and } \sum_{i \neq j} m\left(O_i, O_j\right) \text{ uncorr}$$

Canonical Representation of Order 3 Mean 0 U-stat-

$$U_3(\theta) = U_2(\theta) + \sum_{i \neq j \neq X} t\left(O_i, O_j, O_X\right)$$

$$E\left[t\left(O_i, O_j, O_X\right) | O_i, O_j\right] = E\left[t\left(O_i, O_j, O_X\right) | O_i, O_X\right]$$

$$= E\left[t\left(O_i, O_j, O_X\right) | O_j, O_X\right] = 0,$$

$$t(\cdot, \cdot, \cdot) \ \textit{not} \text{ necc } \textit{sym}$$

$$\sum_{i \neq j \neq X} m\left(O_i, O_j, O_X\right) \text{ and } U_2(\theta) \text{ uncorr}$$

Formula for higher order scores associated with $\widetilde{\theta}(\varsigma)$

$$S_{i_1 \ldots i_m}(\theta) = f_{/i_1 \ldots i_m}(\mathbf{O};\theta) / f(\mathbf{O};\theta)$$
$$f(\mathbf{O};\theta) = \prod_{i=1}^{n} f(O_i;\theta)$$

of order m in terms of the subject specific scores (Waterman and Lindsay (1996)

$$S_{i_1 \ldots i_m, j}(\theta) = f_{/i_1 \ldots i_m, j}\left(O_j;\theta\right) / f_j\left(O_j;\theta\right), j = 1, \ldots, n$$

(Waterman and Lindsay (1996) .

$$S_{i_1} = \sum_j S_{i_1, j}$$

$$S_{i_1 i_2} = \sum_j S_{i_1 i_2, j} + \sum_{X \neq j} S_{i_1, j} S_{i_2, X}$$
$$S_{i_1 i_2, j}(\theta) = S_{i_1, j}(\theta) S_{i_2, j}(\theta) + \partial S_{i_1, j}(\theta(\varsigma)) / \partial \varsigma_{i_2}|_{\widetilde{\theta}(\varsigma) = \theta}$$

$$S_{i_1 i_2 i_3}$$
$$= \sum_{j} S_{i_1 i_2 i_3, j} + \sum_{X \neq j} S_{i_1 i_2, j} S_{i_3, X} + S_{i_3 i_2, j} S_{i_1, X} + S_{i_1 i_3, j} S_{i_2, X}$$
$$\sum_{X \neq j \neq t} S_{i_1, j} S_{i_2, X} S_{i_3, t}$$

Definition of a kth order influence function: A U-statistic $U_k(\theta) = u_k(\mathbf{O};\theta)$ of order k, dimension and finite variance is said to be an kth order influence function for $\psi(\theta)$ if (i)

$$E_\theta[U_k(\theta)] = 0, \theta \in \Theta$$

(ii) for $m = 1, 2, ..., k$, and every $\widetilde{\theta}(\varsigma)$, $p = 1, 2, ..$

$$\psi_{\backslash i_1 ... i_m}(\theta) = E_\theta\left[U_k(\theta) S_{i_1 ... i_m}(\theta)\right]$$

$p = k$ sufficient. We say that $\psi(\theta)$ is kth order pathwise differntiable

Theorem: If the model is nonparametric .then there is at most one mth order estimation influence function $IF_m^{est}(\theta)$, the efficient mth order IF.

Lemma: $IF_m^{est}(\theta) = IF_{m-1}^{est}(\theta) + IF_{mm}^{est}(\theta)$,

$IF_{mm}^{est}(\theta) = \sum_{\{i_1 \neq i_2 \neq ... \neq i_m; i_X \in \{1,2,...,n\}, X \in =1,...,m\}} d_m \Big(O_{i_1},$
where $d_m\left(O_{i_1}, O_{i_2}, ..., O_{i_m}\right)$ is canonical

Var$\left[IF_m^{est}(\theta)\right]$ increases with m

Var$\left[IF_m^{est}(\theta)\right]$ /Var$\left[IF_m^{est}\left(\widehat{\theta}\right)\right] = 1 + o(1)$

The following Extended Information Theorem is closely related to result in McLeish and Small (1994).

Theorem: Given $U_k(\theta)$, for all $\widetilde{\theta}(\varsigma)$ for $s \leq k$

$$\begin{aligned} &\partial^s E_\theta \left[U_k(\theta(\varsigma))\right] / \partial\varsigma_{i_1 \ldots} \partial\varsigma_{i_s} \\ &= -E_\theta \left[U_k(\theta)\, S_{i_1 \ldots i_s}(\theta)\right] \\ &= -\psi_{\backslash i_1 \ldots i_s}(\theta) \end{aligned}$$

$$E_\theta \left[U_k\left(\widehat{\theta}\right)\right] = -\left[\psi\left(\widehat{\theta}\right) - \psi(\theta)\right] + O_p\left(\left\|\widehat{\theta} - \theta\right\|^{k+1}\right)$$

since as functions of $\widehat{\theta}$, the functions $E_\theta\left[U_k\left(\widehat{\theta}\right)\right]$ and $-\left[\psi\left(\widehat{\theta}\right) - \psi(\theta)\right]$ have the same Taylor expansion around θ up to order k

$$\widehat{\psi}_m = \psi\left(\widehat{\theta}\right) + IF_m\left(\widehat{\theta}\right)$$

where $\widehat{\theta}$ is an initial estimator of θ.from a separate sample (no Donsker like needed).

But, by extended info equality

$$E_\theta\left[IF_m\left(\widehat{\theta}\right)\right] = -\left[\psi\left(\widehat{\theta}\right) - \psi(\theta)\right] + O_p\left(\left\|\widehat{\theta} - \theta\right\|^{m+1}\right)$$

so (conditional) bias of $\widehat{\psi}_m$ is

$$\left\{\psi\left(\widehat{\theta}\right) + E_\theta\left[IF_m\left(\widehat{\theta}\right)\right] - \psi(\theta)\right\} = O_p\left(\left\|\widehat{\theta} - \theta\right\|^{m+1}\right), \downarrow m$$

Var$\left[\widehat{\psi}_m\right]$ increases with m

$IF_m\left(\widehat{\theta}\right)$ and $\widehat{\psi}_m = \psi\left(\widehat{\theta}\right) + IF_m\left(\widehat{\theta}\right)$ are AN given $\widehat{\theta}$ often normal.

Shortest conservative uniform asymptotic confidence intervals based on

$$\widehat{\psi}_{m_{conf}} \pm var\left[IF_{m_{conf}}\left(\widehat{\theta}\right)|\widehat{\theta}\right]^{1/2} z_\alpha$$

where k_{conf} is the smallest k with $var\left[U_k\left(\theta\right)\right]$ higher order (or equal if constants dealt with) than the squared bias.

Example: Problem: If $IF_1(\theta)$ depends on θ through a nonparametric $\rho(\theta)$ where $\rho(\theta)$ infinte dimensional

$IF_m(\theta)$ does not exist for $m \geq 2$

Example:

$$IF_1\left(\widehat{\theta}\right) = n^{-1}\sum_i \frac{R_i}{\widehat{\omega}(X_i)}\left(Y_1 - \widehat{b}(X_i)\right) + \widehat{b}(X_i) - \psi\left(\widehat{\theta}\right)$$

Use sieves ie $k = k(n)$ dimensional submodels for $b(X)$, $\omega(X_i)$. Then IF_m exists for all m.

Bur then truncation bias

$$TB_k = E\left[\left\{b(X_i) - \overline{b}_k(X_i)\right\}\{\omega(X_i) - \overline{\omega}_k(X_i)\}/\overline{\omega}_k(X_i)\right]$$

added to estimation bias $\left|\left|\widehat{\theta} - \theta\right|\right|^{m+1}$

where $\overline{b}_k(X_i)$ is the limit of the model

Specifically $b\left(X\right) = b^{*}\left(Z_k^T B_k\right), \omega\left(X\right) = \omega^{*}\left(Z_k^T \alpha_k\right)$

where Z_k a basis in R^d as $k \longrightarrow \infty$