11/17/2011

To do: 1) talk bout TMLE being more efficient than a given estimator
2) analyzing case control data.

Example: Given $O=(W,A,Y) \sim P_0$

We want to est. $\Psi(P) = E_P[E_P(Y|A=1,W) - E_P(Y|A=0,W)]$

If we look at set of possible gradients in $M(g_0)$, where g_0 known it is given by $\left(\frac{I(A=1)}{g} - \frac{I(A=0)}{g}\right)Y - \psi + \bar{Q}(1,W)\left(\frac{I(A=1)}{g} - 1\right) - \bar{Q}(0,W)\left(\frac{I(A=0)}{g} - 1\right)$

Wait!

↳ On second thought, lets make scenario simpler.

$\Psi(P) = E(E(Y|A=1,W))$

then $D^*(Q,g_0) = \left(\frac{I(A=1)}{g_0}Y - \psi - \bar{Q}(1,W)\left(\frac{I(A=1)}{g_0} - 1\right)\right)$, $\bar{Q}$

Let $\bar{Q}_\beta$ be a working model for $\bar{Q}_0$

We can define $\beta_n = \arg\min_\beta \frac{1}{n}\sum_{i=1}^n \left[\frac{I(A=1)}{g_0}Y_i - \psi_n^0 - \bar{Q}_\beta(1,W)\left(\frac{I(A=1)}{g_0} - 1\right)\right]^2$

$(*)$ $\psi_n = \frac{1}{n}\sum_{i=1}^n \frac{I(A_i=1)}{g_0(1|W)}Y_i - \bar{Q}_{\beta_n}(1,W_i)\left(\frac{I(A=1)}{g_0} - 1\right)$

chooses the β that will minimize the variance of the IC.

variance of estimator (missed name of this estimator, but recorded)

*In RCT, parametric models will be same as TMLE.

↳ if we run TMLE update, the $\hat{\epsilon}$ estimated will be 0.

We can define loss function as:

$L_{g_0,\psi_0}(\bar{Q}) = [D^*(\bar{Q},g_0,\psi_0)]^2$

we calculate $\bar{Q}^0 = \arg\min_{\bar{Q}} E_0[L_{g_0,\psi_0}(\bar{Q})]$

for candidate ests, $\hat{\bar{Q}}_k$ for $k=1,\ldots,K$

$k_n = \arg\min_k E_{B_n}[P_{n,B_n} L_{g_0,\psi_n}(\hat{\bar{Q}}_k(P^0_{n,B}))]$

Dissimilarity, $d(\bar{Q},\bar{Q}_0) = P_0 L_{g_0\psi_0}(\bar{Q}) - L_{g_0\psi_0}(\bar{Q}_0)$

$= var(D^*(\bar{Q},g_0,\psi_0)) - var(D^*(\bar{Q}_0,g_0,\psi_0))$

$= \int (D^*(\bar{Q},g_0,\psi_0) - D^*(\bar{Q}_0,g_0,\psi_0))\, dP_0(O)$

Going back to (*), how do we verify that TMLE is at least as good?

↳ $D^0(\bar{Q}_{\beta_0^*}, g_0, \psi_0)$ is the IC to beat

↑ β_0^* is the limit of β_n

$= D_{IPTW}(g_0, \psi) - \underbrace{D_{CAR}(\bar{Q}_{\beta_0^*}, g_0)}_{\text{we want to solve for this}}$

Note: general result for EIC is $D^*(\bar{Q}, g_0) = D_{IPTW}(g_0, \psi_0) - D_{CAR}(\bar{Q}, g_0)$

for TMLE, ~~set~~ calc $\text{Logit}(\bar{Q}_n^0(\varepsilon_1)) = \text{Logit}(\bar{Q}_n^0) + \varepsilon_1 \frac{I(A=1)}{g_n^0(1|W)}$

and $\text{Logit}(\bar{g}_n(\varepsilon)) = \text{Logit}(\bar{g}_n^0) + \varepsilon_{21} \frac{\bar{Q}_n^0(1,W)}{g_n^0(1|W)} + \varepsilon_{22} \underbrace{\frac{\bar{Q}_{\beta_n^*}(1,W)}{g_n^0(1|W)}}_{\bar{Q}_{\beta_n}(1,W) \text{ isn't updated}}$

* Notice how updated Q_n^0 is dep on g_n.

↳ keeps iterating until converges.

⇓

$P_n D^*(Q_n^*, g_n^*) = 0 = P_n D_{IPTW}(g_n^*, \psi(P_n^*)) - \underbrace{D_{CAR}(\bar{Q}_n^*, g_n^*)}_{=0}$

$P_n D_{CAR}(Q_n^*, g_n^*) = 0$

$P_n D_{CAR}(\bar{Q}_{\beta_n^*}, g_n^*) = 0$

$\Rightarrow P_n D_{IPTW}(g_n^*, \psi(P_n^*)) = 0$

$\rightarrow P_n D_{IPTW}(g_n^*, \underset{\uparrow \text{TMLE}}{\psi_n^*}) + D_{CAR}(g_n^*, \underset{\uparrow \text{External estimator}}{\bar{Q}_{\beta_n^*}}) = 0$

So, TMLE is at least as efficient as IPTW (b/c it is IPTW)

and is _____?