Improving Efficiency of Inferences in Randomized Clinical Trials Using Auxiliary Covariates

Min Zhang, Anastasios A. Tsiatis and Marie Davidian

(2008, Biometrics)

Presented by Rui Zhang

May 8, 2012

A motivating example

- Suppose in a randomized clinical trial (RCT), whether or not getting a certain disease after treatment (treated by drug or placebo, denoted by $Z = 1, 0$) is recorded in every patient as Y, and also baseline covariates, such as demographic information, is also recorded *before treatment assigned* as X, then how do we compare the treatment effect?
- One quick answer:

$$logit\{\mathbb{E}(Y|Z)\} = \beta_0 + \beta_1(Z = 1)$$

- How can we incorporate information from X?
  In late session, we have discussed bias due to non-collapsibility if applying to adjustment.

A quote

"Every clinical trial is a problem of missing data."

– Professor Scott Emerson

Missing-ness problem?

- Suppose we want to estimate average of outcome in treatment group: $Y_1's$ and in control group: $Y_0's$;
- Since $Y_1's$ are missing in control group and $Y_0's$ in treatment group $\Rightarrow$ **Missing completely at random**

Missing data strategy

- Complete case analysis: This is our quick answer.
  Not using any information in auxiliary covariates X.
- Inverse weight method:
  We utilize information of X by estimate the probability of non-missed given some X:

$$\pi(x) \equiv Pr(Z = 1 | X = x)$$

$$\hat{\mu}_1 = n_1^{-1} \sum_{i=1}^{n} \frac{Z_i Y_{1i}}{\pi(X_i)}$$

- Double Robust Estimator:
  Two models are specified for $\mathbb{E}(Y_1 | X = x) = \mu(x, \gamma_1)$ and $Pr(Z = 1 | X = x) = \pi(x, \gamma_2)$ and the estimator is

$$\hat{\mu}_1 = n_1^{-1} \sum_{i=1}^{n} \left(\frac{Z_i}{\pi(X_i)} \left(Y_{1i} - \mu(X_i, \hat{\gamma}_1) \right) + \mu(X_i, \hat{\gamma}_1) \right)$$

Parametric Model Review

- Consider the parametric model

$$\mathscr{P} = \{p(x;\beta,\eta) : \beta \in \Gamma \subset \mathbb{R}^q, \eta \in \Lambda \subset \mathbb{R}^r\}$$

where q and r are finite and parameter β is of interest while η is a nuisance parameter.

- Previously in 581?, we studied that RAL estimators of β_0 have influence functions satisfying the following properties
  1. Any influence functions of β belong to the **orthogonal complement of the nuisance tangent space**, which is defined as

$$\Lambda_\eta \equiv \{B^{q\times r}\dot{l}_\eta : B^{q\times r}\text{any fixed real matrix}\}$$

where $\dot{l}_\eta = \frac{\partial logp(X;\beta_0,\eta_0)}{\partial \eta}$

Parametric Model Review continued

2. The most efficient influence function of β is

$$\phi^{eff} = \mathbb{E}_{\beta_0,\eta_0}[\dot{l}_\beta^{eff}(X;\beta_0,\eta_0)\dot{l}_\beta^{eff}(X;\beta_0,\eta_0)^T]^{-1}\dot{l}_\beta^{eff}(X;\beta_0,\eta_0)$$

where $\dot{l}_\beta^{eff}(X;\beta_0,\eta_0) = \dot{l}_\beta(X;\beta_0,\eta_0) - \prod[\dot{l}_\beta(X;\beta_0,\eta_0)|\Lambda_\eta]$

But can we make connections between what we know and the novel semi-parametric model?

- A semi-parametric model: $\{P_\theta : \theta \in \Theta\}$
  $\theta = (\beta^T, \eta^T)^T$: $\beta \in \mathbb{R}^q < \infty$, η is not restricted and thus we are allowing Θ to be infinite-dimensional.
- Working procedure
  1. Define a parametric sub-model a as being contained in the semi-parametric model and containing the truth: $\beta = \beta_0 \in \mathbb{R}^q$ and $\eta = \eta_0 \in \mathbb{R}^r$;
  2. Working with the above defined semi-parametric model a, finding the space Λ_β^a orthogonal to the nuisance tangent space Λ_η^a;
  3. By doing the above steps infinite times, the intersection of all Λ_β^a's should serve as the space in the true data-generating model and we can the intersection.

Semi-parametric efficiency bound: "sup" of C-R bounds

Consequently,

1. The influence function of β_0 from semi-parametric model should be orthogonal to nuisance tangent spaces from all parametric sub-models.
2. The variance of any RAL semi-parametric must be greater than or equal to any C-R bound from parametric sub-models. So it can be written as

$$\underset{\mathscr{P}_s}{sup}\mathbb{E}_{\beta_0,\eta_0}[\dot{l}_{\beta}^{eff}(X;\beta_0,\eta_0)\dot{l}_{\beta}^{eff}(X;\beta_0,\eta_0)^T]^{-1}$$

Critical Assumption of this paper

$$\int p_{Y,X|Z}(y,x|z;\beta,\eta,\psi)dx = p_{Y|Z}(y|z;\beta,\eta) \tag{1}$$

$$\int p_{Y,X|Z}(y,x|z;\beta,\eta,\psi)dy = p_X(x) \tag{2}$$

Comment: though look quite trivial, these assumptions tell us that we can write the two nuisance parameter vectors: η and ψ separately, and thus we can find their tangent space: Λ_η and λ_ψ separately and form the final space by arbitrary linear combinations.

Proposing a parametric sub-model

Denote the true parameters as $(\beta_0, \eta_0, \psi_0)$: $\beta_0 \in \mathbb{R}^q$, $\eta_0 \in \mathbb{R}^{s_1}$ and $\psi_0 \in \mathbb{R}^{s_2}$. Then the parametric sub-model nuisance tangent space of η and ψ is

$$\{B_1^{q\times s_1} S_\eta(Y, X, Z) + B_2^{q\times s_2} S_\psi(Y, X, Z)\}$$

where $S_\eta(Y, X, Z) \equiv \frac{\partial}{\partial\eta} log(logp_{Y,X|Z}(y, x|z; \beta_0, \eta_0, \psi_0))$ and similarly for $S_\psi(Y, X, Z)$.
Also define $\Lambda_\eta^* \equiv \frac{\partial}{\partial\eta} log(logp_{Y|Z}(y|z; \beta_0, \eta_0))$.

Working under the first assumption

In first restriction, taking derivative of η after taking log of both sides of the first equation:

$$
\begin{aligned}
& B_1^{q\times s_1}\frac{\partial}{\partial\eta}log\int p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dx \\
&= B_1^{q\times s_1}\frac{\int \frac{\partial}{\partial\eta}p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dx}{\int p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dx} \\
&= B_1^{q\times s_1}\frac{\int\left(\frac{\partial}{\partial\eta}logp_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)\right)p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dx}{p_{Y|Z}(y|z;\beta_0,\eta_0)} \\
&= B_1^{q\times s_1}\mathbb{E}(S_\eta(Y,X,Z)|Y=y,Z=z)=B_1^{q\times s_1}\frac{\partial}{\partial\eta}logp_{Y|Z}(y|z;\beta_0,\eta_0)\in\Lambda_\eta^*
\end{aligned}
$$

Any element from parametric sub-model nuisance tangent space: $h(Y,X,Z)=B_1^{q\times s_1}S_\eta(Y,X,Z)+B_2^{q\times s_2}S_\psi(Y,X,Z)$ must satisfy the condition

$$\mathbb{E}(h(Y,X,Z)|Y,Z)\in\Lambda_\eta \tag{3}$$

Working under the second assumption

As above, with similar trick in taking derivative of η and ψ after taking log of both sides of the second equation:

$$B_1^{q \times s_1} \frac{\partial}{\partial \eta} log \int p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dy = B_1^{q \times s_1} \mathbb{E}(S_\eta(Y,X,Z)|X=x,Z=z)$$

$$B_2^{q \times s_2} \frac{\partial}{\partial \psi} log \int p_{Y,X|Z}(y,x|z;\beta_0,\eta_0,\psi_0)dy = B_2^{q \times s_1} \mathbb{E}(S_\psi(Y,X,Z)|X=x,Z=z)$$

Any element from parametric sub-model nuisance tangent space: $h(Y,X,Z) = B_1^{q \times s_1} S_\eta(Y,X,Z) + B_2^{q \times s_2} S_\psi(Y,X,Z)$ must also satisfy the condition

$$\mathbb{E}(h(Y,X,Z)|X,Z) \in \Lambda_x \qquad (4)$$

where $\Lambda_x \equiv \{h(X) : \mathbb{E}h(X) = 0\}$

Orthogonal complement of semi-paramtric nuisance tangent space

- Functions satisfying (3) is $\Lambda_\eta^* + \Lambda_1$ where $\Lambda_1 \equiv \{h_1(Y, X, Z) : \quad \mathbb{E}\{h_1(Y, X, Z)|Y, Z\} = 0\}$
- Functions satisfying (4) is $\Lambda_x + \Lambda_2$ where $\Lambda_2 \equiv \{h_2(Y, X, Z) : \quad \mathbb{E}\{h_2(Y, X, Z)|X, Z\} = 0\}$
- The nuisance tangent space from the above parametric sub-model *a* can be written out as $\Lambda_{\eta,\psi}^a = (\Lambda_\eta^* + \Lambda_1) \bigcap (\Lambda_x + \Lambda_2)$
- It can be shown that this space works just fine for our semi-parametric model!
- Then orthogonal complement: $\Lambda^\perp = (\Lambda_\eta^{*\perp} \bigcap \Lambda_1^\perp) + (\Lambda_x^\perp \bigcap \Lambda_2^\perp)$

Semi-parametric estimation equation deriving

- $(\Lambda_\eta^{*\perp} \bigcap \Lambda_1^\perp)$: exactly the 'primitive' estimating equation!: Original part.
- $(\Lambda_x^\perp \bigcap \Lambda_2^\perp)$: Augmentation part.
  - $\Lambda_2^\perp = \{h(X,Z) : \mathbb{E}h(X,Z) = 0\}$;
  - $\Lambda_x^\perp = \{h(X,Z) : \mathbb{E}\{h(X,Z)|X\} = 0\}$;
  - Taking intersection, we have $\{h(X,Z) : \mathbb{E}\{h(X,Z)|X\} = 0\}$.
- By simple projections we can show that the estimating equation for β_0 is

$$m(Y,Z;\beta) + \sum_{g=1}^{k} (I(Z=g) - \pi_g)\, \mathbb{E}(m(Y,Z;\beta)|X, Z=g) = 0$$

Simulation: binary outcome Y

A binary outcome from a two-arm RCT of 600 subjects, with $5,000$ Monte Carlo datasets:

$$logit\left(\mathbb{E}(Y|Z)\right) = \beta_1 + \beta_2 I(Z = 2)$$

in which β is the parameter being estimated.
Data generating mechanism:$logit\left(Pr(Y = 1|Z = g, X)\right) = \alpha_{0g} + \alpha_g^T X$, $g = 1, 2$.

- Mild association: $(\alpha_{01}, \alpha_{02}) = (0.025, -0.8)$, $\alpha_1 = (0.8, 0.5, 0, 0, 0, 0, 0, 0)$ and $\alpha_2 = (0.3, 0.7, 0.3, 0.8, 0, 0, 0, 0)$
- Moderate association: $(\alpha_{01}, \alpha_{02}) = (0.38, -0.8)$, $\alpha_1 = (1.2, 1.0, 0, 0, 0, 0, 0, 0)$ and $\alpha_2 = (0.5, 1.3, 0.5, 1.5, 0, 0, 0, 0)$
- Strong association:$\alpha_{02}) = (0.8, -0.8)$, $\alpha_1 = (1.5, 1.8, 0, 0, 0, 0, 0, 0)$ and $\alpha_2 = (1.0, 1.3, 0.8, 2.5, 0, 0, 0, 0)$.

On estimating $\mathbb{E}(m(Y, Z; \beta)|X, Z = g)$, they only used X used to generate the data to do OLS.
They used the same $X's$ in estimating $\mathbb{E}(m(Y, Z; \beta)|X, Z = g)$ to run the adjusted case.

Simulation results

Method	β_2	MC Bias	MC SD	Ave. SE	Cov. Prob	Rel. Eff.
Mild Association						
Unadjusted	-0.494	0.00044	0.1668	0.1661	95.0%	1.00
Aug.	-0.494	-0.00042	0.1545	0.1533	94.9%	1.16
Adjusted	-0.494	-0.091	0.1831	0.1822	92.6%	0.66
Moderate Association						
Unadjusted	-0.490	-0.0025	0.1634	0.1650	95.5%	1.00
Aug.	-0.490	-0.0026	0.1390	0.1392	95.1%	1.39
Adjusted	-0.490	-0.2208	0.2015	0.2015	81.2%	0.31
Strong Association						
Unadjusted	-0.460	-0.0026	0.1662	0.1655	95.2%	1.00
Aug.	-0.460	-0.0026	0.132	0.131	95.2%	1.55
Adjusted	-0.460	-0.3266	0.222	0.2210	68.8%	0.18

Questions?