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:

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

• How can we incorporate information from $X$?
  In late session, we have discussed bias due to non-collapsibility if applying to adjustment.
"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 ⇒ 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
  \[
  \hat{\mu}_1 = n_1^{-1} \sum_{i=1}^{n} \left( \frac{Z_i}{\pi(X_i)} (Y_{1i} - \mu(X_i, \hat{\gamma}_1)) + \mu(X_i, \hat{\gamma}_1) \right)
  \]
Parametric Model Review

• Consider the parametric model

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

where \( q \) and \( r \) are finite and parameter \( \beta \) is of interest while \( \eta \) is a nuisance parameter.

• Previously in 581?, we studied that RAL estimators of \( \beta_0 \) have influence functions satisfying the following properties

1. Any influence functions of \( \beta \) belong to the orthogonal complement of the nuisance tangent space, which is defined as

\[ \Lambda_{\eta} \equiv \{ B^{q \times r} \hat{l}_{\eta} : B^{q \times r} \text{ any fixed real matrix} \} \]

where

\[ \hat{l}_{\eta} = \frac{\partial \log p(X; \beta_0, \eta_0)}{\partial \eta} \]
2. The most efficient influence function of $\beta$ is

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

where $i^{\text{eff}}_\beta (X; \beta_0, \eta_0) = i_\beta (X; \beta_0, \eta_0) - \prod [i_\beta (X; \beta_0, \eta_0) | \Lambda_\eta]$
But can we make connections between what we know and the novel semi-parametric model?

- A semi-parametric model: \( \{ P_\theta : \theta \in \Theta \} \)
  \[ \theta = (\beta^T, \eta^T)^T : \beta \in \mathbb{R}^q < \infty, \eta \text{ is not restricted and thus we are allowing } \Theta \text{ to be infinite-dimensional.} \]
- 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 \( \Lambda_\beta^a \) orthogonal to the nuisance tangent space \( \Lambda_\eta^a \);
  3. By doing the above steps infinite times, the intersection of all \( \Lambda_\beta^a \)'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 $\beta_0$ 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

$$\sup_{P_s} \mathbb{E}_{\beta_0, \eta_0} \left[ i_{\beta}^{\text{eff}} (X; \beta_0, \eta_0) i_{\beta}^{\text{eff}} (X; \beta_0, \eta_0)^T \right]^{-1}$$
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: \( \eta \) and \( \psi \) separately, and thus we can find their tangent space: \( \Lambda_\eta \) and \( \lambda_\psi \) 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 \(\eta\) and \(\psi\) is

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

where \(S_\eta(Y, X, Z) \equiv \frac{\partial}{\partial \eta} \log \left( \log p_{Y|Z}(y, x|z; \beta_0, \eta_0, \psi_0) \right)\) and similarly for \(S_\psi(Y, X, Z)\).

Also define \(\Lambda^*_\eta \equiv \frac{\partial}{\partial \eta} \log \left( \log p_Y|Z(y|z; \beta_0, \eta_0) \right)\).
Working under the first assumption

In first restriction, taking derivative of $\eta$ after taking log of both sides of the first equation:

$$B_{1}^{q \times s_{1}} \frac{\partial}{\partial \eta} \log \int p_{Y, X|Z}(y, x|z; \beta_{0}, \eta_{0}, \psi_{0}) \, dx$$

$$= B_{1}^{q \times s_{1}} \frac{\int \frac{\partial}{\partial \eta} p_{Y, X|Z}(y, x|z; \beta_{0}, \eta_{0}, \psi_{0}) \, dx}{\int p_{Y, X|Z}(y, x|z; \beta_{0}, \eta_{0}, \psi_{0}) \, dx}$$

$$= B_{1}^{q \times s_{1}} \left( \frac{\partial}{\partial \eta} \log p_{Y, X|Z}(y, x|z; \beta_{0}, \eta_{0}, \psi_{0}) \right) p_{Y, X|Z}(y, x|z; \beta_{0}, \eta_{0}, \psi_{0}) \, dx$$

$$= B_{1}^{q \times s_{1}} \mathbb{E}(S_{\eta}(Y, X, Z)|Y = y, Z = z) = B_{1}^{q \times s_{1}} \frac{\partial}{\partial \eta} \log p_{Y|Z}(y|z; \beta_{0}, \eta_{0}) \in \Lambda_{\eta}^{*}$$

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

$$\mathbb{E}(h(Y, X, Z)|Y, Z) \in \Lambda_{\eta}$$  \hspace{1cm} (3)
Working under the second assumption

As above, with similar trick in taking derivative of $\eta$ and $\psi$ after taking log of both sides of the second equation:

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

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

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

$$\mathbb{E}(h(Y, X, Z)|X, Z) \in \Lambda_X$$  \hspace{1cm} (4)

where $\Lambda_X \equiv \{ h(X) : \mathbb{E}h(X) = 0 \}$
Orthogonal complement of semi-parametric nuisance tangent space

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

- \((\Lambda^* \perp \eta \cap \Lambda^\perp_1)\): exactly the 'primitive' estimating equation!
  - Original part.

- \((\Lambda^\perp x \cap \Lambda^\perp_2)\): Augmentation part.
  - \(\Lambda^\perp_2 = \{ h(X, Z) : \mathbb{E}h(X, Z) = 0 \}\);
  - \(\Lambda^\perp x = \{ h(X, Z) : \mathbb{E}\{h(X, Z)|X\} = 0 \}\);
  - 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 \(\beta_0\) is

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

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

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

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

in which $\beta$ is the parameter being estimated.

Data generating mechanism:

$$\text{logit} \left( \Pr(Y = 1|Z = g, X) \right) = \alpha_{0g} + \alpha_g^T X, \ g = 1, 2.$$  

- 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

<table>
<thead>
<tr>
<th>Method</th>
<th>$\beta_2$</th>
<th>MC Bias</th>
<th>MC SD</th>
<th>Ave. SE</th>
<th>Cov. Prob</th>
<th>Rel. Eff.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mild Association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted</td>
<td>-0.494</td>
<td>0.00044</td>
<td>0.1668</td>
<td>0.1661</td>
<td>95.0%</td>
<td>1.00</td>
</tr>
<tr>
<td>Aug.</td>
<td>-0.494</td>
<td>-0.00042</td>
<td>0.1545</td>
<td>0.1533</td>
<td>94.9%</td>
<td>1.16</td>
</tr>
<tr>
<td>Adjusted</td>
<td>-0.494</td>
<td>-0.091</td>
<td>0.1831</td>
<td>0.1822</td>
<td>92.6%</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Moderate Association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted</td>
<td>-0.490</td>
<td>-0.0025</td>
<td>0.1634</td>
<td>0.1650</td>
<td>95.5%</td>
<td>1.00</td>
</tr>
<tr>
<td>Aug.</td>
<td>-0.490</td>
<td>-0.0026</td>
<td>0.1390</td>
<td>0.1392</td>
<td>95.1%</td>
<td>1.39</td>
</tr>
<tr>
<td>Adjusted</td>
<td>-0.490</td>
<td>-0.2208</td>
<td>0.2015</td>
<td>0.2015</td>
<td>81.2%</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Strong Association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted</td>
<td>-0.460</td>
<td>-0.0026</td>
<td>0.1662</td>
<td>0.1655</td>
<td>95.2%</td>
<td>1.00</td>
</tr>
<tr>
<td>Aug.</td>
<td>-0.460</td>
<td>-0.0026</td>
<td>0.132</td>
<td>0.131</td>
<td>95.2%</td>
<td>1.55</td>
</tr>
<tr>
<td>Adjusted</td>
<td>-0.460</td>
<td>-0.3266</td>
<td>0.222</td>
<td>0.2210</td>
<td>68.8%</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Questions?