A Distributional Approach
Using Propensity Scores

Zhiqiang Tan

Department of Biostatistics
Johns Hopkins School of Public Health

http://www.biostat.jhsph.edu/~ztan

June 20, 2005
Outline

- Introduction
- Counterfactual framework
- Illustration
- Application
- No-confounding case
  - Known propensity score
  - Parametric propensity score
- Confounding case
Introduction

- Right heart catheterization (RHC) is performed daily in hospitals since 1970s.

- The benefit of RHC had NOT been demonstrated in a successful randomized clinical trial.

- Connors et al.’s (1996) observational study raised the concern that RHC might not benefit critically ill patients and might in fact cause harm.

- Data were collected on 5735 critically ill patients admitted to the ICUs of five medical centers:
  - Treatment: No-RHC or RHC
  - Outcome: 30-day survival
  - Covariates: 75 covariates

- HOW to evaluate the “effect” of RHC on survival?
Counterfactual framework

- **X**: covariates measured

- **T**: treatment variable taking value “0” or “1” if a patient actually receives No-RHC or RHC

- \((Y_0, Y_1)\): potential outcome that would be observed if a patient received No-RHC or RHC

- \(Y = (1 - T)Y_0 + TY_1\): observed outcome

- We are interested in “average causal effect”
  \[ E(Y_1 - Y_0) = E(Y_1) - E(Y_0) \]
  or \( P(\{Y_1\}) \) versus \( P(\{Y_0\}) \)

- Assignment mechanism
  - No-confounding: \( T \perp (Y_0, Y_1) \mid X \)
  - Confounding: \( T \not\perp (Y_0, Y_1) \mid X \)

- Propensity score:
  \[ \pi(X) = P(T = 1 \mid X) \]
Thirty-day survival curves

Day
Proportion of Surviving
RHC, Raw
No RHC, Raw

Raw histogram of aps

Density

RHC
No RHC

Raw histogram of meanbp

Density

Raw histogram of pafi

Density
Illustration

<table>
<thead>
<tr>
<th></th>
<th>RHC= 1</th>
<th>RHC= 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP= 1</td>
<td>(52, 28) 80</td>
<td>(11, 9) 20</td>
</tr>
<tr>
<td>BP= 0</td>
<td>(30, 10) 40</td>
<td>(37, 23) 60</td>
</tr>
<tr>
<td></td>
<td>82, 38 120</td>
<td>48, 32 80</td>
</tr>
</tbody>
</table>

• Patients get RHC at random

\[ P(\text{ survival } \mid \text{ RHC } = 1) = \frac{82}{120} = 68.3\% \]

\[ P(\text{ survival } \mid \text{ RHC } = 0) = \frac{48}{80} = 60.0\% \]

• Patients get RHC at random given blood pressure

- Weight each patient such that

\[ 80w_1(1) = \frac{1}{2}, \quad 40w_1(0) = \frac{1}{2}, \]

\[ 20w_0(1) = \frac{1}{2}, \quad 60w_0(0) = \frac{1}{2}. \]

- Compare the weighted probabilities

\[ 52w_1(1) + 30w_1(0) = 70.0\%, \]

\[ 11w_0(1) + 37w_0(0) = 58.3\%. \]
• WHAT IF patients are NOT equally likely to get RHC at each level of blood pressure?

  – Previous estimates:

    \[ P(\text{obs survival } | \ BP = *, \ RHC = 1 ) = 70.0\% , \]

    \[ P(\text{obs survival } | \ BP = *, \ RHC = 0 ) = 58.3\% . \]

  – Weight each patient such that

    \[
    \sum_{i=1}^{80} \lambda_{1i}w_1(1) = \frac{1}{2} , \quad \sum_{i=81}^{120} \lambda_{1i}w_1(0) = \frac{1}{2} , \quad \sum_{i=121}^{140} \lambda_{0i}w_0(1) = \frac{1}{2} , \quad \sum_{i=141}^{200} \lambda_{0i}w_0(0) = \frac{1}{2} ,
    \]

    where \( \Lambda^{-1} \leq \lambda_{1i}, \lambda_{0i} \leq \Lambda \) (\( \Lambda = 1.5 \)).

  – Bound the weighted probabilities

    \[
    \sum_{i=1}^{120} \lambda_{1i}w_1(X_i)Y_{1i} , \quad \sum_{i=121}^{200} \lambda_{0i}w_0(X_i)Y_{0i} ,
    \]

    subject to the foregoing constraints.

    \[ P(\text{!obs survival } | \ BP = *, \ RHC = 1 ) \geq 72.2\% , \]

    \[ P(\text{!obs survival } | \ BP = *, \ RHC = 0 ) \leq 55.0\% . \]
Thirty-day survival curves

Day
Proportion of Surviving
RHC, Observed
RHC, Counterfactual
No RHC, Observed
No RHC, Counterfactual

Weighted histogram of aps
Density

Weighted histogram of meanbp
Density

Weighted histogram of pafi
Density
No-confounding case

• Data: \((X_i, Y_{T_i}, T_i), \ i = 1, 2, ..., n\)

• Likelihood:

\[
L_1 \times L_2 = \prod_{i=1}^{n} \left[ (1 - \pi(X_i))^{1-T_i} \pi(X_i)^{T_i} \right] \\
\times \prod_{i=1}^{n} \left[ G_0(\{X_i, Y_{0i}\})^{1-T_i} G_1(\{X_i, Y_{1i}\})^{T_i} \right]
\]

where \(G_0\) is the joint distribution of \((X, Y_0)\) and \(G_1\) is the joint distribution of \((X, Y_1)\).

• \(G_0\) and \(G_1\) induce the same marginal distributions on the covariate space \(\mathcal{X}\). Equivalently,

\[
\int h(x) \, dG_0(x, y_0) = \int h(x) \, dG_1(x, y_1)
\]

for each bounded function \(h\) on \(\mathcal{X}\).

• Take finitely many constraints and find MLE \((\hat{G}_0, \hat{G}_1)\):

\[
\hat{\mu}_1 = \int y_1 \, d\hat{G}_1(x, y_1), \\
\hat{\mu}_0 = \int y_0 \, d\hat{G}_0(x, y_0).
\]
**Known propensity score** [Model S0: known $\pi^*$]

- Maximize the likelihood subject to the constraints

$$
\int \pi^*(x) \, dG_0 = \int \pi^*(x) \, dG_1, \\
\int h_j^*(x) \, dG_0 = \int h_j^*(x) \, dG_1, \quad j = 1, \ldots, m.
$$

- Let $h^* = (\pi^*, 1 - \pi^*, h_1^*, \ldots, h_m^*)$. Maximize

$$
\frac{1}{n} \sum_{i=1}^{n_1} \log(\lambda^\top h^*(X_i)) + \frac{1}{n} \sum_{i=n_1+1}^{n} \log(1 - \lambda^\top h^*(X_i)).
$$

Then

$$
\hat{G}_1\{(X_i, Y_{1i})\} = \frac{n^{-1}}{\lambda^\top h^*(X_i)}, \quad i = 1, \ldots, n_1,
$$

$$
\hat{G}_0\{(X_i, Y_{0i})\} = \frac{n^{-1}}{1 - \lambda^\top h^*(X_i)}, \quad i = n_1 + 1, \ldots, n.
$$

- First-order approximation:

$$
\bar{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{1i} T_i}{\pi^*(X_i)} - \beta_1^\top \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{h^*(X_i)}{1 - \pi^*(X_i)} \left( \frac{T_i}{\pi^*(X_i)} - 1 \right) \right],
$$

$$
\bar{\mu}_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{0i} (1 - T_i)}{1 - \pi^*(X_i)} - \beta_0^\top \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{h^*(X_i)}{\pi^*(X_i)} \left( \frac{1 - T_i}{1 - \pi^*(X_i)} - 1 \right) \right],
$$

where $\beta_1 = \bar{B}^{-1} \bar{C}_1$ and $\beta_0 = \bar{B}^{-1} \bar{C}_0$. 

• The method of control variates:

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{Y_{1i} T_i}{\pi^*(X_i)} - b_1^\top \left[ \frac{1}{n} \sum_{i=0}^{n} \frac{h^*(X_i)}{1 - \pi^*(X_i)} \left( \frac{T_i}{\pi^*(X_i)} - 1 \right) \right].
\]

The optimal choice of \( b_1 \) is \( \beta_1 = B^{-1}C_1 \).

• A more general class of estimators:

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{Y_{1i} T_i}{\pi^*(X_i)} - \frac{1}{n} \sum_{i=1}^{n} \phi_1(X_i) \left( \frac{T_i}{\pi^*(X_i)} - 1 \right).
\]

The optimal choice of \( \phi_1(x) \) is \( \mathbb{E}(Y_1|X = x) \).

\( \leftrightarrow \) achieves semiparametric efficiency under \( S0 \).

• Choose \( h^* \) such that

\( E(Y_1|X = x) \) is contained the linear span of \( \frac{h^*(x)}{1 - \pi^*(x)} \),

\( E(Y_0|X = x) \) is contained the linear span of \( \frac{h^*(x)}{\pi^*(x)} \).

• Outcome regression [Model R]

\[
E(Y_1|X) = \Psi(\alpha_1^\top g_1(X)),
\]

\[
E(Y_0|X) = \Psi(\alpha_0^\top g_0(X)).
\]

Choose \( h^* = (\pi^*, 1 - \pi^*, \pi^* \Psi(\hat{\alpha}_0^\top g_0), (1 - \pi^*) \Psi(\hat{\alpha}_1^\top g_1)) \).
Parametric propensity score [Model S: $\pi(\cdot; \gamma)$]

- Maximize the likelihood subject to the constraints
  \[
  \int \hat{\pi}(x) \, dG_1 = \int \hat{\pi}(x) \, dG_0, \\
  \int \hat{h}_j(x) \, dG_1 = \int \hat{h}_j(x) \, dG_0, \quad j = 1, \ldots, m.
  \]

- Let $\hat{h} = (\hat{\pi}, 1 - \hat{\pi}, \hat{h}_1, \ldots, \hat{h}_m)$. Maximize
  \[
  \frac{1}{n} \sum_{i=1}^{n_1} \log(\lambda^T \hat{h}(X_i)) + \frac{1}{n} \sum_{i=n_1+1}^{n} \log(1 - \lambda^T \hat{h}(X_i)).
  \]
  Then
  \[
  \hat{G}_1\{(X_i, Y_{1i})\} = \frac{n^{-1}}{\lambda^T \hat{h}(X_i)}, \quad i = 1, \ldots, n_1, \\
  \hat{G}_0\{(X_i, Y_{0i})\} = \frac{n^{-1}}{1 - \lambda^T \hat{h}(X_i)}, \quad i = n_1 + 1, \ldots, n.
  \]

- First-order approximation:
  \[
  \bar{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{1i} T_i}{\hat{\pi}(X_i)} - \tilde{\beta}_1^\top \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{h}(X_i)}{1 - \hat{\pi}(X_i)} \left( \frac{T_i}{\hat{\pi}(X_i)} - 1 \right) \right], \\
  \bar{\mu}_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{0i}(1 - T_i)}{1 - \hat{\pi}(X_i)} - \tilde{\beta}_0^\top \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{h}(X_i)}{\hat{\pi}(X_i)} \left( \frac{1 - T_i}{1 - \hat{\pi}(X_i)} - 1 \right) \right],
  \]
  where $\tilde{\beta}_1 = \tilde{B}^{-1} \tilde{C}_1$ and $\tilde{\beta}_0 = \tilde{B}^{-1} \tilde{C}_0$. 
• Our strategy is
  – To build and check propensity score models to ensure consistency
  – To use outcome regression models for variance and bias reduction

• Propensity score models can be checked with the following idea:
  – Pick up a collection of test functions $\hat{h}_j$'s on $\mathcal{X}$, for example, $(\hat{\pi}, 1 - \hat{\pi}, \hat{\pi}X, (1 - \hat{\pi})X)$.
  – Compute the sample average
    $$\bar{E}\left[\hat{h}_j(X)\left(\frac{T}{\hat{\pi}(X)} - \frac{1 - T}{1 - \hat{\pi}(X)}\right)\right]$$
    i.e. average difference in $\hat{h}_j(X)$ between the treated and control after propensity score weighting.
  – If model $S$ is correct, then the sample averages relative to standard errors, or $z$-ratios, should be statistically nonsignificant from zero.

Examination of $z$-ratios against the standard normal can reveal possible misspecification of model $S$. 
Confounding case

• Data: \((X_i, Y_{Ti}, T_i), i = 1, 2, ..., n\)

• Likelihood:
\[
L_1 \times L_2 = \prod_{i=1}^{n} \left[(1 - \pi(X_i))^{1-T_i} \pi(X_i)^{T_i}\right] \\
\times \prod_{i=1}^{n} \left[H_0(\{X_i, Y_{0i}\})^{1-T_i} H_1(\{X_i, Y_{1i}\})^{T_i}\right]
\]

where \(H_0\) is the distribution \(P(\{Y_0\}|T = 0, X)P(\{X\})\) and \(H_1\) is the distribution \(P(\{Y_1\}|T = 1, X)P(\{X\})\).

• \(H_0\) and \(H_1\) induce the same marginal distributions on the covariate space \(\mathcal{X}\). Equivalently,
\[
\int h(x) \, dH_0(x, y_0) = \int h(x) \, dH_1(x, y_1)
\]

for each bounded function \(h\) on \(\mathcal{X}\).

• Convergence of previous estimates:
\[
(\hat{G}_0, \hat{G}_1) \to (H_0, H_1) \\
\hat{\mu}_1, \hat{\mu}_1 \to E[E(Y_1|T = 1, X)] \\
\hat{\mu}_0, \hat{\mu}_0 \to E[E(Y_0|T = 0, X)]
\]
• Unmeasured confounding: gaps between
  
  \[ P(\{Y_0\}|T = 0, X) \text{ and } P(\{Y_0\}|T = 1, X) \]
  
  \[ P(\{Y_1\}|T = 0, X) \text{ and } P(\{Y_1\}|T = 1, X) \]
  
  i.e. systematic differences between the treated and untreated even if they received the same treatment.

• Define the Radon-Nikodym derivatives:

  \[ \lambda_0(Y_0; X) = \frac{P(dY_0|T = 1, X)}{P(dY_0|T = 0, X)} , \]
  
  \[ \lambda_1(Y_1; X) = \frac{P(dY_1|T = 0, X)}{P(dY_1|T = 1, X)} . \]

  The case \( \lambda_0 = \lambda_1 = 1 \) corresponds to “no confounding”, while deviations of \( \lambda_0 \) and \( \lambda_1 \) from 1 indicate unmeasured confounding.

• By Bayes’ rule, \( \lambda_0 \) and \( \lambda_1 \) can be seen as odds ratios:

  \[ \lambda_0(Y_0; X) = \frac{1 - \pi(X) P(T = 1|Y_0, X)}{\pi(X) P(T = 0|Y_0, X)} , \]
  
  \[ \lambda_1(Y_1; X) = \frac{\pi(X) P(T = 0|Y_1, X)}{1 - \pi(X) P(T = 1|Y_1, X)} . \]

• A sensitivity analysis model:

  \[ \lambda^{-1} \leq \lambda_0(Y_0; X), \lambda_1(Y_1; X) \leq \Lambda, \]

  where \( \Lambda \geq 1 \) indicates the degree of departure from “no confounding”.
• Let \( \hat{h}^c = (\hat{\pi}, 1 - \hat{\pi}, \hat{h}_1, \ldots, \hat{h}_{m^c}) \). For a value of \( \Lambda \), find bounds for \( \int y_t \lambda_t \, dH_t \) by linear programming:

\[
\begin{align*}
\text{min or max} & \quad \int y_t \lambda_t \, d\hat{G}_t \\
\text{subject to} & \quad \int \lambda_t \, d\hat{G}_t = 1, \\
& \quad \int \hat{\pi}(x) \lambda_t \, d\hat{G}_t = \int \hat{\pi}(x) \, d\hat{G}_t, \\
& \quad \int \hat{h}_j(x) \lambda_t \, d\hat{G}_t = \int \hat{h}_j(x) \, d\hat{G}_t, \quad j = 1, \ldots, m^c, \\
\text{and} \quad & \frac{1}{\Lambda} \leq \lambda_t \leq \Lambda.
\end{align*}
\]

– \( \hat{G}_1 \) is supported on \( \{(X_i, Y_{1i})\}_{i=1,\ldots,n_1} \) and \( \hat{G}_0 \) on \( \{(X_i, Y_{0i})\}_{i=n_1+1,\ldots,n} \). Integral is finite sum.

– The unknowns are the values of \( \lambda_t \) on observed data: \( \lambda_{1i} = \lambda_1(Y_{1i}; X_i), \quad i = 1, \ldots, n_1, \)
  \( \lambda_{0i} = \lambda_0(Y_{0i}; X_i), \quad i = n_1 + 1, \ldots, n. \)

• Comparisons of the distributions

\[
\begin{align*}
\hat{G}_0 \rightarrow [Y_0|T = 0, X][X], \quad & \quad \hat{G}_1 \rightarrow [Y_1|T = 1, X][X] \\
\lambda_0 \, d\hat{G}_0 \rightarrow [Y_0|T = 1, X][X], \quad & \quad \lambda_1 \, d\hat{G}_1 \rightarrow [Y_1|T = 0, X][X]
\end{align*}
\]

indicate (i) balance on covariates, (ii) hidden bias, and (iii) causal effects.