A Distributional Approach Using Propensity Scores

Zhiqiang Tan

Department of Statistics
Rutgers University

http://www.stat.rutgers.edu/~ztan

April 3, 2009
Outline

• Introduction

• Framework

• Illustration

• Methodology
  – No-confounding case: estimation
  – Confounding case: sensitivity analysis

• Application
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?
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

- **$Y = (1 - T) Y_0 + T Y_1$:** observed outcome

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

- **Propensity score:**
  \[
  \pi(X) = P(T = 1 \mid X)
  \]

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

Day
Proportion of Surviving
RHC, Raw
No RHC, Raw

Density

Raw histogram of aps

Density

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

<table>
<thead>
<tr>
<th></th>
<th>80</th>
<th>20</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>100</td>
</tr>
</tbody>
</table>

- Weight each patient such that

\[
80w_1(1) = 1/2, \quad 40w_1(0) = 1/2,
\]

\[
20w_0(1) = 1/2, \quad 60w_0(0) = 1/2.
\]

- 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:

\[ P(\text{obs survival} \mid BP =*, RHC = 1) = 70.0\% , \]
\[ P(\text{obs survival} \mid BP =*, RHC = 0) = 58.3\% . \]

- Weight each patient such that

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

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.

\[ P(\text{!obs survival} \mid BP =*, RHC = 1) \uparrow 72.2\% , \]
\[ P(\text{!obs survival} \mid BP =*, RHC = 0) \downarrow 55.0\% . \]
No-confounding case

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

- Likelihood:

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

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 \((\hat{G}_0, \hat{G}_1)\):

\[
\hat{\mu}_1 = \int y_1 \, d\hat{G}_1(x, y_1), \\
\hat{\mu}_0 = \int y_0 \, d\hat{G}_0(x, y_0).
\]
Thirty-day survival curves

<table>
<thead>
<tr>
<th>Day</th>
<th>Proportion of Surviving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RHC, Raw</td>
</tr>
<tr>
<td></td>
<td>No RHC, Raw</td>
</tr>
</tbody>
</table>

- **Raw histogram of aps**
- **Raw histogram of meanbp**
- **Raw histogram of pafi**
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 $\hat{h}^c = (\hat{\pi}, 1 - \hat{\pi}, \hat{h}_1, ..., \hat{h}_m^c)$. For a value of $\Lambda$, find bounds for $\int y_t \lambda_t \ dH_t$ by linear programming:

$$\text{min or max } \int y_t \lambda_t \ d\hat{G}_t$$

subject to

$$\int \lambda_t \ d\hat{G}_t = 1,$$

$$\int \hat{\pi}(x) \lambda_t \ d\hat{G}_t = \int \hat{\pi}(x) \ d\hat{G}_t,$$

$$\int \hat{h}_j(x) \lambda_t \ d\hat{G}_t = \int \hat{h}_j(x) \ d\hat{G}_t, \ j = 1, ..., m^c,$$

and $\frac{1}{\Lambda} \leq \lambda_t \leq \Lambda$.

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

− The unknowns are the values of $\lambda_t$ on observed data: $\lambda_{1i} = \lambda_1(Y_{1i}; X_i), \ i = 1, ..., n_1$,

$\lambda_{0i} = \lambda_0(Y_{0i}; X_i), \ i = n_1 + 1, ..., n$.

• Comparisons of the distributions

$$\hat{G}_0 \rightarrow [Y_0|T = 0, X][X], \quad \hat{G}_1 \rightarrow [Y_1|T = 1, X][X]$$

$$\lambda_0 \ d\hat{G}_0 \rightarrow [Y_0|T = 1, X][X], \quad \lambda_1 \ d\hat{G}_1 \rightarrow [Y_1|T = 0, X][X]$$

indicate (i) balance on covariates, (ii) hidden bias, and (iii) causal effects.
Non-inferiority trial without placebo

Focus: How to use historical data to guide determination of NI margin

Here is a wild thought:

• Need to estimate \( P_0/C_0 \) from historical trials with placebo

• Need to account for variability from trial to trial in \( P_0/C_0 \)

• Meta-analysis ??

• Propensity score ??