Statistical Analysis of Randomized Experiments with Nonignorable Missing Binary Outcomes

Kosuke Imai

Department of Politics
Princeton University

July 31, 2007

Overview

- Missing outcomes in randomized experiments.
- A growing literature on the topic:
  - Method of bounds (e.g., Horowitz and Manksi, 2000).
  - Semiparametric methods (e.g., Scharfstein et al. 1999).
  - Ignorability (e.g., Yau and Little, 2001).
  - Latent ignorability (e.g., Frangakis and Rubin, 1999).
- Nonignorable missing outcomes:
  - Political science: self-reported voting behavior.
  - Economics: self-reported income.
  - Medicine: self-reported health status.
- The paper offers (with and without noncompliance):
  1. Alternative identification and estimation strategies.
  2. New sensitivity analyses.
  3. Applications in political science, psychology, and public health.

Framework for Standard Randomized Experiments

- Causal inference via potential outcomes (e.g., Holland 1986).
  - Experimental unit: $i = 1, 2, \ldots, n$.
  - Binary treatments: $T_i \in \{0, 1\}$.
  - Potential outcomes: $Y_i(T_i)$.
  - Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$.
  - Potential response indicators: $R_i(T_i)$.
  - Observed response indicator: $R_i = T_i R_i(1) + (1 - T_i) R_i(0)$.
  - Pre-treatment covariates: X_i.
- No interference among units (Cox 1958; Rubin 1990).
- Randomized treatment: $(Y_i(1), Y_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp T_i$ for all i.
- Estimands:
  - Average Treatment Effect (ATE):
    $\tau_{ATE} \equiv E[Y_i(1) - Y_i(0)] = E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0]$.
  - Conditional Average Treatment Effect (CATE):
    $\tau_{CATE} \equiv \frac{1}{n} \sum_{i=1}^{n} E[Y_i(1) - Y_i(0) \mid X_i]$.

Identification Problem in the Binary Case

- Assume $Y_i(0), Y_i(1) \in \{0, 1\}$.
- Define,

$$\begin{aligned} p_{jk} &\equiv \Pr(Y_i = 1 \mid T_i = j, R_i = k), \\ \pi_{jk} &\equiv \Pr(T_i = j, R_i = k), \end{aligned}$$

- Then, the ATE can be written as,

$$\tau_{ATE} = \frac{p_{10}\pi_{10} + p_{11}\pi_{11}}{\pi_{10} + \pi_{11}} - \frac{p_{00}\pi_{00} + p_{01}\pi_{01}}{\pi_{00} + \pi_{01}},$$

  where p_{00} and p_{10} are not identifiable from the data.
- Since $p_{j0} \in [0, 1]$, the sharp bounds (Horowitz & Manski, 2000) are given by,

$$\tau_{ATE} \in \left[\frac{p_{11}\pi_{11}(\pi_{00} + \pi_{01}) - (\pi_{00} + p_{01}\pi_{01})(\pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00} + \pi_{01})}, \right. \\ \left. \frac{(\pi_{10} + p_{11}\pi_{11})(\pi_{00} + \pi_{01}) - p_{01}\pi_{01}(\pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00} + \pi_{01})} \right].$$

Identification Strategies

- **Ignorability Assumption** (Little & Rubin, 1987): For $j \in \{0,1\}$,

$$\begin{aligned} & \Pr(R_i(j)=1 \mid T_i=j, Y_i(j)=1, X_i=x) \\ = & \Pr(R_i(j)=1 \mid T_i=j, Y_i(j)=0, X_i=x), \end{aligned}$$

- **Nonignorability (NI) Assumption**: For $k \in \{0,1\}$ and $x \in \mathcal{X}$,

$$\begin{aligned} & \Pr(R_i(j)=1 \mid T_i=0, Y_i(0)=k, X_i=x) \\ = & \Pr(R_i(j)=1 \mid T_i=1, Y_i(1)=k, X_i=x). \end{aligned}$$

- Missing-data mechanism directly depends on the realized value of the outcome variable itself, but is conditionally independent of the treatment status.
- Identification of the ATE is established via Bayes rule (PROPOSITION 1).

Inference under the Nonignorability Assumption

1. Without covariates (or within strata defined by covariates): the ML estimator is in a closed form (PROPOSITION 2).
2. With covariates:
   - Modeling approach (e.g., logistic regression):

$$\begin{aligned} q_j(x) &= \Pr(Y_i=1 \mid T_i=j, X_i=x), \\ r_{jk}(x) &= \Pr(R_i=1 \mid T_i=j, Y_i=k, X_i=x), \end{aligned}$$

   - Complete-data likelihood function:

$$\begin{aligned} \prod_{i=1}^{n} & \left[r_{\cdot 1}(X_i)^{R_i}\{1-r_{\cdot 1}(X_i)\}^{1-R_i}\right]^{Y_i} \left[r_{\cdot 0}(X_i)^{R_i}\{1-r_{\cdot 0}(X_i)\}^{1-R_i}\right]^{1-Y_i} \\ \times & \left[q_1(X_i)^{Y_i}\{1-q_1(X_i)\}^{1-Y_i}\right]^{T_i} \left[q_0(X_i)^{Y_i}\{1-q_0(X_i)\}^{1-Y_i}\right]^{1-T_i}, \end{aligned}$$

     where $r_{\cdot k}(x) = r_{1k}(x) = r_{0k}(x)$ for $x \in \mathcal{X}$ under the NI assumption.
   - Computation: *EM* algorithm, Gibbs sampler with prior distributions.

Sensitivity Analysis

- Neither MAR nor NI assumptions are testable.
- Sensitivity analysis based on the following parameter,

$$\theta_k^{NI} \equiv \frac{\Pr(R_i(1) = 1 \mid T_i = 1, Y_i(1) = k)}{\Pr(R_i(0) = 1 \mid T_i = 0, Y_i(0) = k)},$$

for $k = 0, 1$ where the range of the parameter is given by,

$$\frac{(1 - p_{11})\pi_{11}}{(1 - p_{11})\pi_{11} + \pi_{10}} \leq \theta_0^{NI} \leq \frac{(1 - p_{01})\pi_{01} + \pi_{00}}{(1 - p_{01})\pi_{01}},$$
$$\frac{p_{11}\pi_{11}}{p_{11}\pi_{11} + \pi_{10}} \leq \theta_1^{NI} \leq \frac{p_{01}\pi_{01} + \pi_{00}}{p_{01}\pi_{01}}.$$

- τ_{ATE} is now a function of θ_k^{NI} and identifiable parameters.
- See how τ_{ATE} varies along with the value of θ_k.

Randomized Experiments with Noncompliance

- Randomized "encouragement" design:
  - Binary encouragement: $Z_i \in \{0, 1\}$.
  - Potential binary treatments: $T_i(Z_i) \in \{0, 1\}$.
  - Observed treatment: $T_i = Z_i T_i(1) + (1 - Z_i) T_i(0)$.
  - Potential outcomes: $Y_i(Z_i)$.
  - Observed outcome: $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$.
  - Potential response indicators: $R_i(Z_i)$.
  - Observed response indicator: $R_i = Z_i R_i(1) + (1 - Z_i) R_i(0)$.
- Randomization of encouragement:

$$(Y_i(1), Y_i(0), T_i(1), T_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp Z_i,$$

- Intention-To-Treat (ITT) effect: $\tau_{ITT} \equiv E[Y_i(T_i(1), 1) - Y_i(T_i(0), 0)]$.

Instrumental Variables (Angrist, Imbens & Rubin, 1996)

- Noncompliance
  - Complier: $T_i(1) = 1$ and $T_i(0) = 0$.
  - Noncomplier:
    1. Always-taker ($C_i = c$): $T_i(1) = T_i(0) = 1$.
    2. Never-taker ($C_i = n$): $T_i(1) = T_i(0) = 0$.
    3. Defier ($C_i = d$): $T_i(1) = 0$ and $T_i(0) = 1$.
- Assumptions:
  1. Monotonicity (no defier): $T_i(1) \geq T_i(0)$.
  2. Exclusion restriction for noncompliers: $Y_i(1) = Y_i(0)$ for $C_i = a, n$ (i.e., zero ITT effect for always-takers and never-takers).
- Complier Average Causal Effect (IV estimand):

$$\tau_{CACE} \equiv E[Y_i(1) - Y_i(0) \mid C_i = c] = \frac{E[Y_i(1) - Y_i(0)]}{E[T_i(1) - T_i(0)]}.$$

Identification Strategies

- **Ignorability** (Yau & Little, 2001): For $j = 0, 1$ and $l = 0, 1$,

$$\begin{aligned} & \Pr(R_i(l) = 1 \mid Y_i(l) = 1, T_i(l) = j, Z_i = l, X_i = x) \\ = \; & \Pr(R_i(l) = 1 \mid Y_i(l) = 0, T_i(l) = j, Z_i = l, X_i = x). \end{aligned}$$

- **Latent Ignorability** (Frangakis & Rubin, 1999):
  1. Latent ignorability: For $l = 0, 1$ and $t \in \{c, n, a\}$,

$$\begin{aligned} & \Pr(R_i(l) = 1 \mid Y_i(l) = 1, Z_i = l, C_i = t, X_i = x) \\ = \; & \Pr(R_i(l) = 1 \mid Y_i(l) = 0, Z_i = l, C_i = t, X_i = x). \end{aligned}$$

  2. Compound exclusion restriction for noncompliers: $Y_i(0) = Y_i(1)$, and $R_i(1) = R_i(0)$, for $C_i = n, a$.
- **Nonignorability**: For $j = 0, 1$, and $k = 0, 1$,

$$\begin{aligned} & \Pr(R_i(1) = 1 \mid T_i(1) = j, Y_i(1) = k, Z_i = 1, X_i = x) \\ = \; & \Pr(R_i(0) = 1 \mid T_i(0) = j, Y_i(0) = k, Z_i = 0, X_i = x). \end{aligned}$$

Theoretical Results in the Binary Case

- Apply the same analytical strategy as before.
- Define,

$$\begin{aligned} p_{jkl} &\equiv \Pr(Y_i = 1 \mid T_i = j, R_i = k, Z_i = l), \\ \pi_{jkl} &\equiv \Pr(T_i = j, R_i = k, Z_i = l). \end{aligned}$$

- Rewrite the ITT effect as,

$$\tau_{ITT} = \frac{\sum_{j=0}^{1}\sum_{k=0}^{1} p_{jk1}\pi_{jk1}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1}\sum_{k=0}^{1} p_{jk0}\pi_{jk0}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk0}},$$

  where π_{jkl} and p_{j1l} are identifiable, but p_{j0l} is not.
- Thus, the identification of τ_{ITT} requires four constraints (PROPOSITION 3).

Inference and Sensitivity Analysis

- With no covariate:
  - ML estimator and its asymptotic variance are in a closed-form.
  - Sensitivity analysis parameters:

$$\psi_{jk}^{NI} \equiv \frac{\Pr(R_i(1) = 1 \mid T_i(1) = j, Y_i(1) = k, Z_i = 1)}{\Pr(R_i(0) = 1 \mid T_i(0) = j, Y_i(0) = k, Z_i = 0)},$$

- Modeling approach:

$$\begin{aligned} p_{jl}(x) &\equiv \Pr(Y_i = 1 \mid T_i = j, Z_i = l, X_i = x), \\ q_l(x) &\equiv \Pr(T_i = 1 \mid Z_i = l, X_i = x), \\ r_{jk}(x) &\equiv \Pr(R_i = 1 \mid T_i = j, Y_i = k, X_i = x). \end{aligned}$$

  $\tau_{ITT}(x) = [p_{11}(x)q_1(x) + p_{01}(x)\{1 - q_1(x)\}] - [p_{10}(x)q_0(x) + p_{00}(x)\{1 - q_0(x)\}]$

Concluding Remarks

- Nonignorable missing data in randomized experiments.
- Identification and estimation strategies for randomized experiments with and without noncompliance.
- Sensitivity analyses to examine robustness of conclusiosns.