Statistical Analysis of Randomized Experiments with "Truncation by Death"

Kosuke Imai

Department of Politics
Princeton University

August 31, 2007

- Sharp Bounds on the Causal Effects in Randomized Experiments with "Truncation-by-Death". *Statistics & Probability Letters*, Forthcoming.
- Identification Analysis for Randomized Experiments with Noncompliance and "Truncation-by-Death".

What is “Truncation by Death”?

- Consider a randomized clinical trial
- Sample of very old and sick patients
- Treatment: a drug for hypertension
- Outcome of interest: blood pressure

- Problem: some patients died during the trial
- Blood pressure **not defined** for the dead!
- Analysis of the survivors leads to **post-treatment bias**
- Unless death occurs independent of the treatment

Motivating Example: Seguro Popular de Salud (SPS)

- Evaluation of the Mexican universal health insurance program (King *et al.*, 2007)
- Aim: “provide social protection in health to the 50 million uninsured Mexicans” (Frenk *et al.*, 2003)
- Matched-pair, cluster randomized trials (Imai *et al.*, 2007)
- Encouragement design: must affiliate to receive SPS services
- Noncompliance: always-takers and never-takers

- One outcome of interest: satisfaction with the received health care
- Satisfaction is undefined for those who have not been to clinics
- **Skip-pattern questions** in survey

Additional Examples

- An Internet-based survey experiment before 2004 Japanese Upper House election (Horiuchi *et al.*, 2007)
- Causal effect of policy information on voting behavior
- Political party websites about pension policies
- Encouragement design: never-takers

- Questions of interest:
  1. How many switched from LDP to DPJ (from DPJ to LDP)?
  2. How many switched from abstention to LDP (DPJ)?

Field	Treatment	Outcome	Truncation
Economics	job training	wages	unemployment
Education	teaching program	test scores	drop-out

The Analytical Approach and Contributions

- How to avoid post-treatment bias in **conditional inference**?

- Use of principal stratification (Frangakis and Rubin 2002)
- Focus on those who would survive under **both** treatment and control conditions (Zhang and Rubin 2003)
- Identification analysis: derive sharp (best possible) bounds

- Formalize the derivation of the bounds in the literature
- Simplify the expressions
- Derive the sharp bounds on quantile TE as well as ATE
- Extend the results to experiments with noncompliance

Framework for Standard Randomized Experiments

- Causal inference via potential outcomes:
  - Binary treatments: $T_i \in \{0, 1\}$
  - Potential "truncation" variable: $W_i(T_i)$
  - Observed "truncation" variable: $W_i = T_i W_i(1) + (1 - T_i) W_i(0)$
  - Potential outcomes: $Y_i(W_i, T_i)$
  - $Y_i(0, t)$ exists but $Y_i(1, t)$ does not
  - Observed outcome (defined only for $W_i = 0$): Y_i
- Randomized treatment:

$$(Y_i(0,0), Y_i(0,1), W_i(1), W_i(0)) \perp\!\!\!\perp T_i \quad \text{for all } i$$

- Estimands:
  - Average Treatment Effect (ATE):

$$\tau_{ATE} \equiv E[Y_i(0,1) - Y_i(0,0) \mid W_i(0) = 0, W_i(1) = 0]$$

  - Quantile Treatment Effect (QTE):

$$\tau_{QTE}(\alpha) = q_{00|1}(\alpha) - q_{00|0}(\alpha)$$

Identification Problem

- Identifiable distributions: for $t = 0, 1$

$$P_t \equiv P(y \mid W_i = 0, T_i = t) \quad \text{and} \quad p_t \equiv \Pr(W_i = 1 \mid T_i = t)$$

- (Unidentifiable) Distributions of interest: $P_{00|1}$ and $P_{00|0}$ where

$$\begin{aligned} P_{w0|1} &\equiv P(y(0,1) \mid W_i(0) = w, W_i(1) = 0, T_i = 1) \\ P_{0w|0} &\equiv P(y(0,0) \mid W_i(0) = 0, W_i(1) = w, T_i = 0) \end{aligned}$$

- What is the relationship?

$$\begin{aligned} P_0 &= \frac{\pi_{00}}{1 - p_0} P_{00|0} + \left(1 - \frac{\pi_{00}}{1 - p_0}\right) P_{01|0}, \\ P_1 &= \frac{\pi_{00}}{1 - p_1} P_{00|1} + \left(1 - \frac{\pi_{00}}{1 - p_1}\right) P_{10|1}, \end{aligned}$$

- The sharp bounds of $\pi_{00} \equiv \Pr(W_i(0) = 0, W_i(1) = 0)$:

$$(0, 1] \bigcap [1 - p_0 - p_1,\ \min(1 - p_0, 1 - p_1)]$$

Theoretical Results

- Sharp bounds on τ_{ATE} and τ_{QTE} without assumptions
- Tighter (sharp) bounds under additional assumptions.
- **Stochastic Dominance**: for all $y \in \Omega$

$$\begin{aligned} P_{00|0}[-\infty, y] &\leq P_{01|0}[-\infty, y], \\ P_{00|1}[-\infty, y] &\leq P_{10|1}[-\infty, y]. \end{aligned}$$

  Those who always survive are healthier than those who sometimes die
- **Monotonicity**: $W_i \leq W_i(0)$.
  Treatment never kills people
- Bounds are in closed form if Monotonicity holds
- Under both assumptions, the naïve estimate is the lower bound

Extension to 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 "truncation" variable: $W_i(T_i)$
  - Observed "truncation" variable: $W_i = T_i W_i(1) + (1 - T_i) W_i(0)$
  - Potential outcomes: $Y_i(W_i, T_i)$
  - Observed outcome (defined only for $W_i = 0$): Y_i
- Randomization of encouragement:

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

- Intention-To-Treat (ITT) Effect:

$$\tau_{ITT} \equiv E[Y_i(1) - Y_i(0) \mid W_i(0) = 0, W_i(1) = 0]$$

- Complier Average Causal Effect (CACE):

$$\tau_{CATE} \equiv E[Y_i(1) - Y_i(0) \mid W_i(0) = 0, W_i(1) = 0, C_i = c]$$

Identification Problem

- Monotonicity (No Defier): Angrist *et al.* (1996)
- **Exclusion restriction** for noncompliers: $W_i(1) = W_i(0)$ for $C_i \in \{a, n\}$
- Identifiable distributions:

$$\begin{aligned} P_{tz} &\equiv P(y \mid W_i = 0, T_i = t, Z_i = z) \\ p_{jtz} &\equiv \Pr(W_i = 1 \mid T_i = t, Z_i = z). \end{aligned}$$

- (Unidentifiable) Distributions of interest: $P_{c00|1}$ and $P_{c00|0}$ where

$$P_{sjk|z} = p(Y_i(j,k) \mid C_i = s, W_i(1) = j, W_i(0) = k, Z_i = z)$$

- What is the relationship?

$$\begin{aligned} \frac{p_{000}P_{00} - p_{001}P_{10}}{p_{000} - p_{001}} &= \frac{\pi_{c00}}{p_{000} - p_{001}} P_{c00|0} + \left(1 - \frac{\pi_{c00}}{p_{000} - p_{001}}\right) P_{c01|0}, \\ \frac{p_{011}P_{11} - p_{010}P_{01}}{p_{011} - p_{010}} &= \frac{\pi_{c00}}{p_{011} - p_{010}} P_{c00|1} + \left(1 - \frac{\pi_{c00}}{p_{011} - p_{010}}\right) P_{c10|1}. \end{aligned}$$

Identification Analysis

Observed Strata (W_i, T_i, Z_i)	Principal Strata $(C_i, W_i(0), W_i(1))$
$(0,0,0)$	$(n,0,0)$, $(n,0,1)$, $(c,0,0)$, $(c,0,1)$
$(0,1,0)$	$(a,0,0)$, $(a,0,1)$
$(1,0,0)$	$(n,1,0)$, $(n,1,1)$, $(c,1,0)$, $(c,1,1)$
$(1,1,0)$	$(a,1,0)$, $(a,1,1)$
$(0,0,1)$	$(n,0,0)$, $(n,1,0)$
$(0,1,1)$	$(a,0,0)$, $(a,1,0)$, $(c,0,0)$, $(c,1,0)$
$(1,0,1)$	$(n,0,1)$, $(n,1,1)$
$(1,1,1)$	$(a,0,1)$, $(a,1,1)$, $(c,0,1)$, $(c,1,1)$

- Identification of π_{c00} as a **linear programming** problem
- Enumerate all vertices of the implied polyhedron
- Additional assumptions to point-identify $\pi_c 00$:
  1. Monotonicity: $W_i(1) \leq W_i(0)$ for all i
  2. Stochastic Dominance: $P_{c00|z}[-\infty, y] \leq P_{c01|z}[-\infty, y]$

Concluding Remarks

- “Truncation by death” frequently occurs even in non-medical experiments
- Naïve analysis would lead to post-treatment bias
- Can’t simply “control” for observed post-treatment variables
- Causal effects are not identifiable
- How much can we learn from the observed data?
- Propose analytical techniques to derive sharp bounds
- Various assumptions to tighten the bounds
- Ongoing project: measurement error in causal inference