Discussion of:
Balancing covariates in randomized experiments with the Gram–Schmidt Walk design

Maximilian Kasy

Department of Economics, University of Oxford

June 3, 2021

Summary

- Treatment assignment based on covariates X, in a static experiment, for estimation of sample ATE, evaluated by MSE.
- Observations $1, \ldots, n$, treatment $Z_i \in \{-1, 1\}$, potential outcomes Y_i^{-1}, Y_i^1, $\mu_i = (Y_i^{-1} + Y_i^1)/2$.
- The ex-ante MSE, given potential outcomes, of the Horovitz-Thompson estimator, for any assignment $\boldsymbol{Z}$ with $E[\boldsymbol{Z}] = 0$, equals

$$\frac{4}{n^2}\boldsymbol{\mu}' \operatorname{Cov}(\boldsymbol{Z})\boldsymbol{\mu}.$$

- The paper proposes and characterizes an algorithm which aims to trade off two minimization objectives:

$$\sup_{\boldsymbol{\mu}:\|\boldsymbol{\mu}\|^2=1} \boldsymbol{\mu}' \operatorname{Cov}(\boldsymbol{Z})\boldsymbol{\mu} \quad (\text{“}\textit{robustness}\text{''})$$

$$\sup_{\boldsymbol{\mu}:\boldsymbol{\mu}=\boldsymbol{X}\theta,\ \|\theta\|^2=1} \boldsymbol{\mu}' \operatorname{Cov}(\boldsymbol{Z})\boldsymbol{\mu} \quad (\text{“}\textit{balance}\text{''})$$

- The paper provides an elegant treatment,
  an interesting algorithm,
  and a range of theoretical characterizations.

- In my discussion, I will reconsider the evaluation criteria used:
  - "Robustness" is used to justify randomization.
    Is this convincing?
  - Robustness is defined, effectively, over an L^2 ball for $\boldsymbol{\mu}$.
    Why not over some other set?
  - Balance is defined for the averages of covariates.
    Why not over the entire distribution?

Randomization and robustness

- Decision theory does not justify randomization:
  - Consider risk functions / Bayes risk / **ex-post** minimax risk for a randomized procedure.
  - Then risk equals the expectation (average) of risk for deterministic decisions, which is worse than the best possible deterministic decision. (e.g. Kasy 2016)
- Adversarial games can justify randomization:
  - "Robustness" formalized as **ex-ante** minimax here (and in the literature since the 1980s).
  - Game against adversarial nature, which chooses μ **after** the choice of assignment algorithm, but **before** the realization of randomness.

Is this a plausible justification of randomization?

- The same "robustness" argument applies to estimation, inference, and prediction:
  - Suppose, e.g., the data might be generated by one of k models.
  - Then "robustness" for binary prediction requires that you sometimes report a prediction from one of the models chosen at random.
- "Robustness" only holds ex-ante, not ex post:
  - It requires us to evaluate decisions not just based on what happened, but also based on what could have happened!
  - How much comfort should we derive from the possibility of a counterfactual world where different treatment assignments might have been chosen?
    E.g., when covariates are unbalanced in your data, but balanced in expectation?
- Also noteworthy:
  - Randomization was introduced to experiments in the 1920s.
  - The rationalization of randomization by ex-ante minimax arguments was only introduced more than 50 years later.

Minimax over what? Balance of what?

- The minimax objective (robustness) here is defined over an L^2 ball for $\boldsymbol{\mu}$,

$$\sup_{\boldsymbol{\mu}:\|\boldsymbol{\mu}\|^2=1} \boldsymbol{\mu}' \operatorname{Cov}(\boldsymbol{Z})\boldsymbol{\mu}.$$

  - Why not subject to $\boldsymbol{\mu}' A \boldsymbol{\mu} \leq 1$ for some other A?
  - Choice of set to minimax over $\Leftrightarrow$ choice of prior!
    $A \Leftrightarrow$ inverse prior covariance matrix.

- Balance here is defined over linear functions only.

$$\|\boldsymbol{X}' \operatorname{Cov}(\boldsymbol{Z})\boldsymbol{X}\| = 0.$$

  - Eliminates bias when $\boldsymbol{\mu} = X\theta$.
  - But what about non-linear relationships?
  - We should balance the entire distribution, not just means! (e.g. Kasy 2016)

Thank you!