A Framework for Generalization and Transportation of Causal Estimates under Covariate Shift

Apoorva Lal, Wenjing Zheng, Simon Ejdemyr

February 15, 2023

Introduction

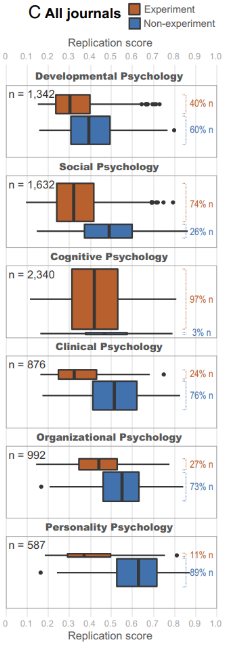
- Scientists, policy-makers, and practitioners care about different kinds of study validity
  - **Internal** validity: findings are informative about the *population under study*
    - estimand: Sample Average Treatment Effect (SATE)
  - **External** validity: findings are informative about the *population of interest*
    - estimand: Target Average Treatment Effect (TATE) ; Population ATE (PATE)
- Experiments are a corner solution that prioritizes *Internal* over *External* validity (Egami and Hartman 2022)

Introduction

- Scientists, policy-makers, and practitioners care about different kinds of study validity
  - **Internal** validity: findings are informative about the *population under study*
    - estimand: Sample Average Treatment Effect (SATE)
  - **External** validity: findings are informative about the *population of interest*
    - estimand: Target Average Treatment Effect (TATE) ; Population ATE (PATE)
- Experiments are a corner solution that prioritizes *Internal* over *External* validity (Egami and Hartman 2022)
- Design based solutions: design to balance in target population (Phan et al. 2021) or balanced sampling (Cytrynbaum 2021)
  - Limited feasibility in regime with many concurrent experiments

Introduction

- Scientists, policy-makers, and practitioners care about different kinds of study validity
  - **Internal** validity: findings are informative about the *population under study*
    - estimand: Sample Average Treatment Effect (SATE)
  - **External** validity: findings are informative about the *population of interest*
    - estimand: Target Average Treatment Effect (TATE) ; Population ATE (PATE)
- Experiments are a corner solution that prioritizes *Internal* over *External* validity (Egami and Hartman 2022)
- Design based solutions: design to balance in target population (Phan et al. 2021) or balanced sampling (Cytrynbaum 2021)
  - Limited feasibility in regime with many concurrent experiments
- **This project**: Framework and package for model-based solutions to bridging (generalizing or transporting) causal estimates to new populations



Perspective | Published: 22 July 2021

Behavioural science is unlikely to change the world without a heterogeneity revolution

Christopher J. Bryan, Elizabeth Tipton & David S. Yeager

Nature Human Behaviour **5**, 980–989 (2021) | Cite this article

Abstract

In the past decade, behavioural science has gained influence in policymaking but suffered a crisis of confidence in the replicability of its findings. Here, we describe a nascent heterogeneity revolution that we believe these twin historical trends have triggered. This revolution will be defined by the recognition that most treatment effects are heterogeneous, so the variation in effect estimates across studies that defines the replication crisis is to be expected as long as heterogeneous effects are studied without a systematic approach to sampling and moderation. When studied

Data and Estimands

- Data: $\mathcal{D}_i = (\mathbf{X}_i, S_i, S_i A_i, S_i Y_i)_{i=1}^N$ where
  - covariates $\mathbf{X}_i \in \mathbb{R}^p$,
  - treatment $A_i \in \mathcal{A} := \{0, \ldots, K\}$,
  - outcome $Y_i \in \mathbb{R}$,
  - selection indicator $S_i \in \{0, 1\}$

Data and Estimands

- Data: $\mathcal{D}_i = (\mathbf{X}_i, S_i, S_iA_i, S_iY_i)_{i=1}^N$ where
  - covariates $\mathbf{X}_i \in \mathbb{R}^p$,
  - treatment $A_i \in \mathcal{A} := \{0, \ldots, K\}$,
  - outcome $Y_i \in \mathbb{R}$,
  - selection indicator $S_i \in \{0, 1\}$
- Observe $(\mathbf{X}_i, A_i, Y_i)_{i=1}^{N_1}$ for observations with $S_i = 1$ (henceforth the *study* sample $\mathcal{S}_1$)
- Observe $(\mathbf{X}_i)_{i=N_1+1}^{N}$ for observations with $S_i = 0$ (henceforth the *target* sample $\mathcal{S}_0$).
- The *overall* sample is $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_0$.
- Two kinds of missing data

Data and Estimands

- Data: $\mathcal{D}_i = (\mathbf{X}_i, S_i, S_i A_i, S_i Y_i)_{i=1}^N$ where
  - covariates $\mathbf{X}_i \in \mathbb{R}^p$,
  - treatment $A_i \in \mathcal{A} := \{0, \ldots, K\}$,
  - outcome $Y_i \in \mathbb{R}$,
  - selection indicator $S_i \in \{0, 1\}$
- Observe $(\mathbf{X}_i, A_i, Y_i)_{i=1}^{N_1}$ for observations with $S_i = 1$ (henceforth the *study* sample $\mathcal{S}_1$)
- Observe $(\mathbf{X}_i)_{i=N_1+1}^{N}$ for observations with $S_i = 0$ (henceforth the *target* sample $\mathcal{S}_0$).
- The *overall* sample is $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_0$.
- Two kinds of missing data

Estimands

Generalizability

$$\mathbb{E}\left[Y^{a,S=1}\right]$$

Transportability

$$\mathbb{E}\left[Y^{a} | S = 0\right]$$

Lets us construct causal contrasts for any pair $a, a' \in \mathcal{A}$

Reasoning about extrapolation bias

- Let $\mathcal{X}_b, p_b(\mathbf{x}), b \in \{s, t\}$ denote Support and Distribution of covariates in study and target. Bias from naive extrapolation is

$$
\begin{aligned}
\text{TATE - SATE} &= \sum_{x \in \mathcal{X}_t} p_t(\mathbf{x})\tau_t(\mathbf{x}) - p_s(\mathbf{x})\tau_s(\mathbf{x}) \\
&= \sum_{x \in \mathcal{X}_t} \left(p_t(\mathbf{x}) - p_s(\mathbf{x})\right) \tau(\mathbf{x}) \\
&= \sum_{x \in \mathcal{X}_t} \underbrace{p_s(\mathbf{x})}_{\text{Strata Size}} \underbrace{\left(\frac{p_t(\mathbf{x})}{p_s(\mathbf{x})} - 1\right)}_{\text{Imbalance}} \underbrace{\tau(\mathbf{x})}_{\text{Heterogeneity}}
\end{aligned}
$$

Reasoning about extrapolation bias

- Let $\mathcal{X}_b, p_b(\mathbf{x}), b \in \{s, t\}$ denote Support and Distribution of covariates in study and target. Bias from naive extrapolation is

$$
\begin{aligned}
\text{TATE - SATE} &= \sum_{x \in \mathcal{X}_t} p_t(\mathbf{x})\tau_t(\mathbf{x}) - p_s(\mathbf{x})\tau_s(\mathbf{x}) \\
&= \sum_{x \in \mathcal{X}_t} \left(p_t(\mathbf{x}) - p_s(\mathbf{x})\right) \tau(\mathbf{x}) \\
&= \sum_{x \in \mathcal{X}_t} \underbrace{p_s(\mathbf{x})}_{\text{Strata Size}} \underbrace{\left(\frac{p_t(\mathbf{x})}{p_s(\mathbf{x})} - 1\right)}_{\text{Imbalance}} \underbrace{\tau(\mathbf{x})}_{\text{Heterogeneity}}
\end{aligned}
$$

Bias contributions

- Imbalance in effect-modifying strata $\mathbf{x}$ s.t. $\tau(\mathbf{x}) > 0$
- Failure of overlap: $p_s(\mathbf{x}) = 0$ but $p_s(\mathbf{x}) > 0$
- Heterogeneity model instability $\tau_s(\mathbf{x}) \neq \tau_t(\mathbf{x})$ DRO problem (Sahoo, Lei, and Wager 2022; Jeong and Namkoong 2020)

Identification Assumptions

1. Consistency / SUTVA : $Y_i = \sum_{a \in \mathcal{A}} \mathbb{1}_{A_i = a} Y_i^a$

Identification Assumptions

1. Consistency / SUTVA : $Y_i = \sum_{a \in \mathcal{A}} \mathbb{1}_{A_i = a} Y_i^a$
2. Ignorability of Treatment: $Y^0, \ldots, Y^a \perp\!\!\!\perp A | \mathbf{X} = \mathbf{x}, S = 1$

Identification Assumptions

1. Consistency / SUTVA : $Y_i = \sum_{a \in \mathcal{A}} \mathbb{1}_{A_i = a} Y_i^a$
2. Ignorability of Treatment: $Y^0, \ldots, Y^a \perp\!\!\!\perp A | \mathbf{X} = \mathbf{x}, S = 1$
3. Overlap
    3.1 Treatment overlap: $0 < \mathbf{Pr}\left(A = a | \mathbf{X} = \mathbf{x}, S = 1\right) < 1$
    3.2 Selection overlap: $0 < \mathbf{Pr}\left(S = 1 | \mathbf{X} = \mathbf{x}\right) < 1$

Identification Assumptions

1. Consistency / SUTVA : $Y_i = \sum_{a \in \mathcal{A}} \mathbb{1}_{A_i = a} Y_i^a$
2. Ignorability of Treatment: $Y^0, \ldots, Y^a \perp\!\!\!\perp A | \mathbf{X} = \mathbf{x}, S = 1$
3. Overlap
   3.1 Treatment overlap: $0 < \mathbf{Pr}\left(A = a | \mathbf{X} = \mathbf{x}, S = 1\right) < 1$
   3.2 Selection overlap: $0 < \mathbf{Pr}\left(S = 1 | \mathbf{X} = \mathbf{x}\right) < 1$
4. Selection
   4.1 $Y^0, \ldots, Y^a \perp\!\!\!\perp S | \mathbf{X} = \mathbf{x}$. Ignorability of Selection.
   4.2 $\mathbb{E}\left[Y | A, \mathbf{X}, S = 1\right] = \mathbb{E}\left[Y | A, \mathbf{X}, S = 0\right]$. The outcome model is stable across S strata.

Identification Assumptions

1. Consistency / SUTVA : $Y_i = \sum_{a \in \mathcal{A}} \mathbb{1}_{A_i = a} Y_i^a$
2. Ignorability of Treatment: $Y^0, \dots, Y^a \perp\!\!\!\perp A | \mathbf{X} = \mathbf{x}, S = 1$
3. Overlap
   3.1 Treatment overlap: $0 < \mathbf{Pr}\left(A = a | \mathbf{X} = \mathbf{x}, S = 1\right) < 1$
   3.2 Selection overlap: $0 < \mathbf{Pr}\left(S = 1 | \mathbf{X} = \mathbf{x}\right) < 1$
4. Selection
   4.1 $Y^0, \dots, Y^a \perp\!\!\!\perp S | \mathbf{X} = \mathbf{x}$. Ignorability of Selection.
   4.2 $\mathbb{E}\left[Y | A, \mathbf{X}, S = 1\right] = \mathbb{E}\left[Y | A, \mathbf{X}, S = 0\right]$. The outcome model is stable across S strata.

Under A1,2,3,4.1, the generalization effect is identified. (Dahabreh, Robertson, Tchetgen, et al. 2019; Bia, Huber, and Lafférs 2020)
Under A1,2,3,4.2, the transportation effect is identified. (Dahabreh, Robertson, Steingrimsson, et al. 2020; Josey et al. 2021)

Estimator Structure

Augmented Augmented IPW influence function (uncentered)

$$\psi^a = \overbrace{\underbrace{\mu^{(a)}(\mathbf{X}_i)}_{\text{Outcome Model}}}^{\text{Imputation}} + \overbrace{\underbrace{\gamma_i}_{\substack{S_i/\rho(\mathbf{X}_i) \\ \text{Sel Wt}}} \underbrace{\omega_i}_{\substack{\mathbb{1}_{A_i=a}/\pi^a(\mathbf{X}_i) \\ \text{Prop Wt}}} \left(Y_i - \mu^{(a)}(\mathbf{X}_i)\right)}^{\text{Reweighted Residuals}}$$

Estimator Structure

Augmented Augmented IPW influence function (uncentered)

$$\psi^a = \overbrace{\underbrace{\mu^{(a)}(\mathbf{X}_i)}_{\text{Outcome Model}}}^{\text{Imputation}} + \overbrace{\underbrace{\gamma_i}_{\substack{S_i/\rho(\mathbf{X}_i) \\ \text{Sel Wt}}} \quad \underbrace{\omega_i}_{\substack{\mathbb{1}_{A_i=a}/\pi^a(\mathbf{X}_i) \\ \text{Prop Wt}}} \quad \left(Y_i - \mu^{(a)}(\mathbf{X}_i)\right)}^{\text{Reweighted Residuals}}$$

- Difference between ψ^a and $\psi^{a'}$ estimates causal contrasts (Robins et al 1994, Newey 1994, Hahn 1998, Cattaneo 2010)
- Estimation: *put hats on* - Nuisance functions estimated by flexible nonparametric regression (L1/L2 reg, random forest, boosting) using cross-fitting
- Construct observation level influence function (=: doubly robust score)
- Average over target sample for point estimate, standard deviation for confidence interval

Estimators

	Generalization $\sum_x \mathbb{E}\left[Y \mid A=a, S=1, \mathbf{X}\right] P(\mathbf{X})$	**Transportation** $\sum_x \mathbb{E}\left[Y \mid A=a, S=1, \mathbf{X}\right] P(\mathbf{X} \mid S=0)$
OM	$\frac{1}{n}\sum_i \widehat{\mu}^a(\mathbf{X}_i)$	$\frac{1}{\lvert\mathcal{S}_0\rvert}\sum_i (1-S_i)\widehat{\mu}^a(\mathbf{X}_i)$
ISW	$\frac{1}{n}\sum_i \frac{S_i}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}Y_i$	$\frac{1}{n}\sum_i \frac{1}{\widehat{\mathbb{E}}[S_i=0]}\frac{S_i(1-\widehat{\rho}(\mathbf{X}_i))}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}Y_i$
AISW	$\frac{1}{n}\sum_i \widehat{\mu}^a(\mathbf{X}_i) + \frac{S_i}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}(Y_i - \widehat{\mu}^a(\mathbf{X}_i))$	$\frac{1}{n}\sum_i \frac{1}{\widehat{\mathbb{E}}[S_i=0]}\left((1-S_i)\widehat{\mu}^a(\mathbf{X}_i) + \frac{S_i(1-\widehat{\rho}(\mathbf{X}_i))}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}(Y_i - \widehat{\mu}^a(\mathbf{X}_i))\right)$

Estimators

	Generalization $\sum_x \mathbb{E}\left[Y \mid A=a, S=1, \mathbf{X}\right] P(\mathbf{X})$	Transportation $\sum_x \mathbb{E}\left[Y \mid A=a, S=1, \mathbf{X}\right] P(\mathbf{X} \mid S=0)$
OM	$\frac{1}{n}\sum_i \widehat{\mu}^a(\mathbf{X}_i)$	$\frac{1}{\lvert\mathcal{S}_0\rvert}\sum_i (1-S_i)\widehat{\mu}^a(\mathbf{X}_i)$
ISW	$\frac{1}{n}\sum_i \frac{S_i}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}Y_i$	$\frac{1}{n}\sum_i \frac{1}{\widehat{\mathbb{E}}[S_i=0]}\frac{S_i(1-\widehat{\rho}(\mathbf{X}_i))}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}Y_i$
AISW	$\frac{1}{n}\sum_i \widehat{\mu}^a(\mathbf{X}_i) + \frac{S_i}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}(Y_i - \widehat{\mu}^a(\mathbf{X}_i))$	$\frac{1}{n}\sum_i \frac{1}{\widehat{\mathbb{E}}[S_i=0]}\Big((1-S_i)\widehat{\mu}^a(\mathbf{X}_i) + \frac{S_i(1-\widehat{\rho}(\mathbf{X}_i))}{\widehat{\rho}(\mathbf{X}_i)}\frac{\mathbb{1}_{A=a}}{\widehat{\pi}^a(\mathbf{X}_i)}(Y_i - \widehat{\mu}^a(\mathbf{X}_i))\Big)$

- Outcome modelling (OM), Inverse Selection Weighting (ISW), and Augmented Inverse Selection Weighting (AISW)
- Target population
- Implemented in `ateGT`

Balancing weights

- goal for selection weights: **balance covariates across study and target**
- Inverse propensity weighting is indirect: fit $\mathbf{Pr}\left(S=1|\mathbf{X}\right)$, then invert
  - This inversion dramatically inflates errors when selection weights are small
  - Requires *individual level covariates* $\mathbf{X}_i$ for target population

Balancing weights

- goal for selection weights: **balance covariates across study and target**
- Inverse propensity weighting is indirect: fit $\mathbf{Pr}\left(S=1|\mathbf{X}\right)$, then invert
  - This inversion dramatically inflates errors when selection weights are small
  - Requires *individual level covariates* $\mathbf{X}_i$ for target population
- Alternatively calibrate a set of weights that balances covariate distributions

$$\sum_{i\in\mathcal{S}_1} \gamma_i c_{ij}(X_{ij}) \approx \sum_{i\in\text{target}} \underbrace{c_{ij}(X_{ij})}_{\text{Target moments}}$$

Balancing weights

- goal for selection weights: **balance covariates across study and target**
- Inverse propensity weighting is indirect: fit $\mathbf{Pr}(S = 1|\mathbf{X})$, then invert
  - This inversion dramatically inflates errors when selection weights are small
  - Requires *individual level covariates* $\mathbf{X}_i$ for target population
- Alternatively calibrate a set of weights that balances covariate distributions

$$\sum_{i \in \mathcal{S}_1} \gamma_i c_{ij}(X_{ij}) \approx \sum_{i \in \text{target}} \underbrace{c_{ij}(X_{ij})}_{\text{Target moments}}$$

$$\min_{\boldsymbol{\gamma}} = \sum_{i \in \mathcal{S}_1} f(\gamma_i) \text{ s.t.}$$

$$\text{Balance} \sum_{i \in \mathcal{S}_1} \gamma_i c_{ij}(X_{ij}) = \sum_{i \in \text{target}} c_{ij}(X_{ij})$$

$$\text{Simplex} \sum_{i \in \mathcal{S}_1} \gamma_i = 1 \text{ and } \gamma_i \geq 0 \; \forall \{i : i \in \mathcal{S}_1\}$$

Balancing weights

- goal for selection weights: **balance covariates across study and target**
- Inverse propensity weighting is indirect: fit $\mathbf{Pr}(S = 1|\mathbf{X})$, then invert
  - This inversion dramatically inflates errors when selection weights are small
  - Requires *individual level covariates* $\mathbf{X}_i$ for target population
- Alternatively calibrate a set of weights that balances covariate distributions

$$\sum_{i \in \mathcal{S}_1} \gamma_i c_{ij}(X_{ij}) \approx \sum_{i \in \text{target}} \underbrace{c_{ij}(X_{ij})}_{\text{Target moments}}$$

$$\min_{\boldsymbol{\gamma}} = \sum_{i \in \mathcal{S}_1} f(\gamma_i) \text{ s.t.}$$

$$\text{Balance} \sum_{i \in \mathcal{S}_1} \gamma_i c_{ij}(X_{ij}) = \sum_{i \in \text{target}} c_{ij}(X_{ij})$$

$$\text{Simplex} \sum_{i \in \mathcal{S}_1} \gamma_i = 1 \text{ and } \gamma_i \geq 0 \; \forall \{i : i \in \mathcal{S}_1\}$$

Dual is easy to solve for certain fs (L2, entropy) as regularized propensity score (Wang and Zubizarreta 2019). Implemented in `ateCAL`

Incorporating Surrogate Outcomes

- Suppose further that we observe a short-run outcome Z_i for both $\mathcal{S}_1, \mathcal{S}_0$, but outcome of interest Y_i only for $\mathcal{S}_1$
  - This could be because units arrive sequentially and 'mature' at some T period; $\mathcal{S}_1$ are early adopters

Incorporating Surrogate Outcomes

- Suppose further that we observe a short-run outcome Z_i for both $\mathcal{S}_1, \mathcal{S}_0$, but outcome of interest Y_i only for $\mathcal{S}_1$
  - This could be because units arrive sequentially and 'mature' at some T period; $\mathcal{S}_1$ are early adopters
- Intermediate outcome Z is said to be a 'surrogate' for the long-term outcome
- Literature on estimation of long-term treatment effects typically relies on variations of strong surrogacy assumption $Y \perp\!\!\!\perp A | Z$ (Athey et al. 2016; Chen and Ritzwoller 2021)

Incorporating Surrogate Outcomes

- Suppose further that we observe a short-run outcome Z_i for both $\mathcal{S}_1, \mathcal{S}_0$, but outcome of interest Y_i only for $\mathcal{S}_1$
  - This could be because units arrive sequentially and 'mature' at some T period; $\mathcal{S}_1$ are early adopters
- Intermediate outcome Z is said to be a 'surrogate' for the long-term outcome
- Literature on estimation of long-term treatment effects typically relies on variations of strong surrogacy assumption $Y \perp\!\!\!\perp A | Z$ (Athey et al. 2016; Chen and Ritzwoller 2021)

- Under analogues of A1-4, Kallus and Mao (2022) derive an influence function for the generalization effect with surrogates

$$
\begin{aligned}
\psi =& \mu^{(a)}(\mathbf{X}_i) + \\
& \underbrace{\overbrace{S_i/\rho(\mathbf{X}_i)}^{\gamma_i} \quad \overbrace{\mathbb{1}_{A_i=a}/\pi^a(\mathbf{X}_i)}^{\omega_i}}_{} \left(Y_i - \mu^{(a)}(\mathbf{X}_i)\right) + \\
& \omega_i \left(\nu^{(a)}(Z_i, \mathbf{X}_i) - \mu^{(a)}(\mathbf{X}_i)\right) - \tau
\end{aligned}
$$

- Second residual is information gained from incorporating surrogate outcome in prediction of Y

Characterizing Sensitivity via OVB

- Identification hinges on A4.1 (selection ignorability).
- Motivates our estimation strategy via reweighting γ_i
- Suppose there is an omitted variable U that makes A4.1 hold.
- True $(\mu, (\pi, \gamma) =: \alpha)$ and feasible $(\mu_s, (\pi, \gamma_s) =: \alpha_s)$ nuisance functions (Chernozhukov et al. 2022) where former includes U as covariate

Characterizing Sensitivity via OVB

- Identification hinges on A4.1 (selection ignorability).
- Motivates our estimation strategy via reweighting γ_i
- Suppose there is an omitted variable U that makes A4.1 hold.
- True $(\mu, (\pi, \gamma) =: \alpha)$ and feasible $(\mu_s, (\pi, \gamma_s) =: \alpha_s)$ nuisance functions (Chernozhukov et al. 2022) where former includes U as covariate
- Omitted Variables Bias = Covariance between regression error $(\mu - \mu_s)$ and Riesz Representer Error $(\alpha - \alpha_s)$

Characterizing Sensitivity via OVB

- Identification hinges on A4.1 (selection ignorability).
- Motivates our estimation strategy via reweighting γ_i
- Suppose there is an omitted variable U that makes A4.1 hold.
- True $(\mu, (\pi, \gamma) =: \alpha)$ and feasible $(\mu_s, (\pi, \gamma_s) =: \alpha_s)$ nuisance functions (Chernozhukov et al. 2022) where former includes U as covariate
- Omitted Variables Bias = Covariance between regression error $(\mu - \mu_s)$ and Riesz Representer Error $(\alpha - \alpha_s)$

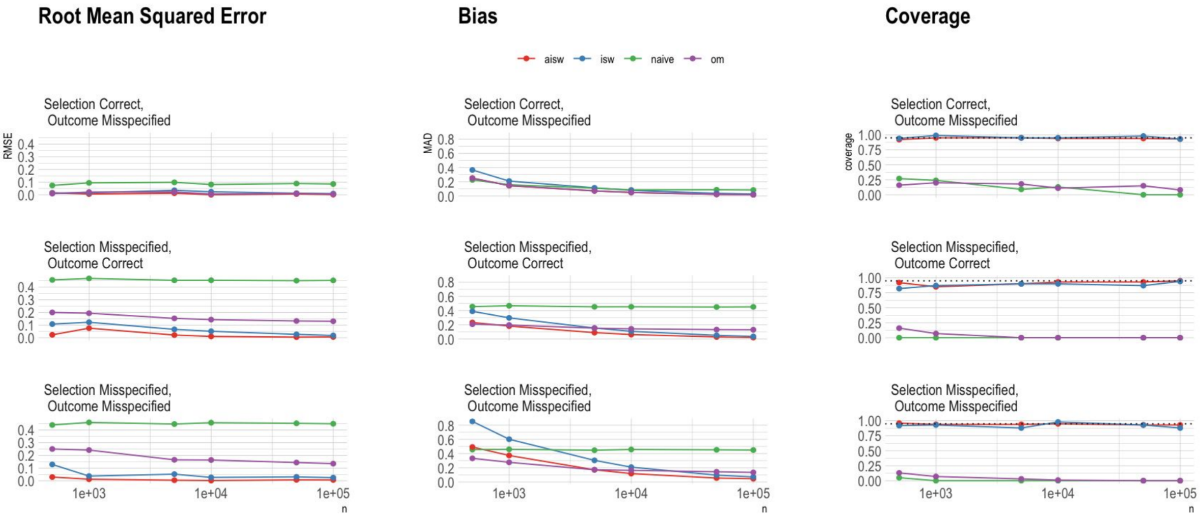
$$\alpha_s(\mathcal{D}) = \frac{S_i}{\rho(\mathbf{X}_i)} \left(\frac{\mathbb{1}_{A=a}}{\pi^a(\mathbf{X}_i)} - \frac{\mathbb{1}_{A=a'}}{\pi^{a'}(\mathbf{X}_i)} \right)$$

$$\alpha(\mathcal{D}, U) = \frac{S_i}{\rho(\mathbf{X}_i, U_i)} \left(\frac{\mathbb{1}_{A=a}}{\pi^a(\mathbf{X}_i)} - \frac{\mathbb{1}_{A=a'}}{\pi^{a'}(\mathbf{X}_i)} \right)$$

- Squared Bias can be bounded as $B^2 = S^2 C_Y^2 C_A^2$ where
- $S^2 := \mathbb{E}(Y - \mu_s)^2 \mathbb{E}\alpha_s^2$ (identifiable)
- $C_Y^2 = R^2_{Y-\mu_s \sim \mu - \mu_s}$: conjectured proportion of residual variance in outcome explained by confounders
- $C_A^2 = (1 - R^2_{\alpha \sim \alpha_s}) / R^2_{\alpha \sim \alpha_s}$: conjectured proportion of residual variance in long RR explained by confounders

Simulation Study

- **Covariates** $X_1, \ldots, X_10 \sim \text{U}[-1, 1]$
- Data Generating Process
  - $A_i \sim \text{Bernoulli}(0.5)$
  - $Y_i^* = Y^0(\mathbf{X}_i) + A_i \tau(\mathbf{X}_i)$
  - $Y_i = Y_i^*$ w.p. $\rho(\mathbf{X})$, else missing
    - For non-trivial functions $\rho(\mathbf{X})$, there is selection bias and SATE is biased for $\mathbb{E}[\tau(\mathbf{X})]$
- Vary functional form of
  - $Y^0(\mathbf{X})$ (Outcome Model)
  - $\rho(\mathbf{X})$ (Selection Model)



Root Mean Squared Error
Bias
Coverage
aisw
isw
naive
om
Selection Correct,
Outcome Misspecified
Selection Misspecified,
Outcome Correct
Selection Misspecified,
Outcome Misspecified
RMSE
MAD
coverage
0.0
0.1
0.2
0.3
0.4
0.6
0.8
0.00
0.25
0.50
0.75
1.00
1e+03
1e+04
1e+05
n

Implementation in causalTransportR

- Main function: `ateGT` (ATE Generalization or Transportation)
- Input vectors `y, a, s` of outcome, treatment, and selection where `y, a` are missing for `s = 0`
- Matrix of `X` of covariates (no missings)
- Additional arguments for nuisance function estimation

```r
ateGT(
  y,
  a,
  X,
  s = NULL,
  treatProb = NULL,
  Z = NULL,
  nuisMod = c("rlm", "rf"),
  target = c("generalize", "transport", "insample"),
  estimator = c("AISW", "ISW", "OM", "CW", "ACW"),
  hajekize = FALSE,
  separateMus = TRUE,
  glmnet_lamchoice = "lambda.min",
  glmnet_alpha = 1,
  glmnet_rho_family = "binomial",
  glmnet_pi_family = "binomial",
  glmnet_mu_family = "gaussian",
  glmnet_parl = FALSE,
  grf_tuneRf = "none",
  noi = FALSE
)
```

Conclusion

- Proposes a multi-treatment framework for causal generalization and transportation
- provide a performant computational implementation for it in `causalTransportR::ateGT`
  - calibrated generalization for when only summary statistics are available for target (`ateCAL`)
  - poststratification weights using fast fixed effects regressions (`ateGTreg`)
- Future work: more work on sensitivity analyses
  - Partial identification for generalization using marginal sensitivity model (Nie, Imbens, and Wager 2021) or Proportion with confounding (Kennedy and Bonvini 2021)
  - Estimate how different the population needs to be from the experimental sample to explain away the effect (Devaux and Egami 2022)

Thanks!

software: `https://github.com/apoorvalal/causalTransportR`
email: apoorval@stanford.edu
website: apoorvalal.github.io

[1] Susan Athey et al. “Estimating Treatment Effects using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index”. In: (Mar. 2016). arXiv: 1603.09326 [stat.ME]. URL: http://arxiv.org/abs/1603.09326.

[2] Michela Bia, Martin Huber, and Lukáš Lafférs. “Double machine learning for sample selection models”. In: (Nov. 2020). arXiv: 2012.00745 [econ.EM]. URL: http://arxiv.org/abs/2012.00745.

[3] Jiafeng Chen and David M Ritzwoller. “Semiparametric Estimation of Long-Term Treatment Effects”. In: (July 2021). arXiv: 2107.14405 [econ.EM]. URL: http://arxiv.org/abs/2107.14405.

[4] Victor Chernozhukov et al. *Long story short: Omitted variable bias in causal machine learning*. Tech. rep. National Bureau of Economic Research, 2022.

[5] Max Cytrynbaum. “Designing Representative and Balanced Experiments by Local Randomization”. In: *arXiv preprint arXiv:2111.08157* (2021).

[6] Issa J Dahabreh, Sarah E Robertson, Jon A Steingrimsson, et al. “Extending inferences from a randomized trial to a new target population”. en. In: *Statistics in medicine* 39.14 (June 2020), pp. 1999–2014.

[7] Issa J Dahabreh, Sarah E Robertson, Eric J Tchetgen, et al. “Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals”. en. In: *Biometrics* 75.2 (June 2019), pp. 685–694.

[8] Martin Devaux and Naoki Egami. “Quantifying Robustness to External Validity Bias”. In: (2022).

[9] Naoki Egami and Erin Hartman. “Elements of external validity: Framework, design, and analysis”. en. In: *American Political Science Review* (2022). URL: https://naokiegami.com/paper/external_full.pdf.

[10] Sookyo Jeong and Hongseok Namkoong. "Robust causal inference under covariate shift via worst-case subpopulation treatment effects". In: *Proceedings of Thirty Third Conference on Learning Theory*. Ed. by Jacob Abernethy and Shivani Agarwal. Vol. 125. Proceedings of Machine Learning Research. PMLR, Sept. 2020, pp. 2079–2084. URL: https://proceedings.mlr.press/v125/jeong20a.html.

[11] Kevin P Josey et al. "Transporting experimental results with entropy balancing". en. In: *Statistics in medicine* 40.19 (Aug. 2021), pp. 4310–4326. URL: http://dx.doi.org/10.1002/sim.9031.

[12] Nathan Kallus and Xiaojie Mao. "On the role of surrogates in the efficient estimation of treatment effects with limited outcome data". In: (Dec. 2022). arXiv: 2003.12408 [stat.ML]. URL: http://arxiv.org/abs/2003.12408.

[13] Xinkun Nie, Guido Imbens, and Stefan Wager. "Covariate Balancing Sensitivity Analysis for Extrapolating Randomized Trials across Locations". In: (Dec. 2021). arXiv: 2112.04723 [econ.EM]. URL: http://arxiv.org/abs/2112.04723.

[14] My Phan et al. "Designing Transportable Experiments Under S-admissability". In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. 2021, pp. 2539–2547.

[15] Roshni Sahoo, Lihua Lei, and Stefan Wager. "Learning from a Biased Sample". en. In: *WP* (2022). URL: https://github.com/roshni714/ru_regression.

[16] Yixin Wang and Jose R Zubizarreta. "Minimal dispersion approximately balancing weights: asymptotic properties and practical considerations". en. In: *Biometrika* (Oct. 2019). URL: http://jrzubizarreta.com/minimal.pdf.