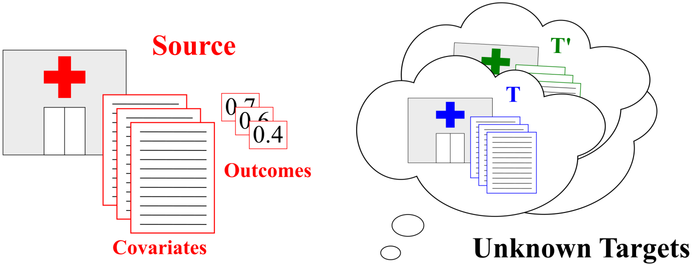
Universal Adaptability: A New Method to Draw Inference from Non-Probability Surveys and Other Data Sources

Christoph Kern

School of Social Sciences, University of Mannheim

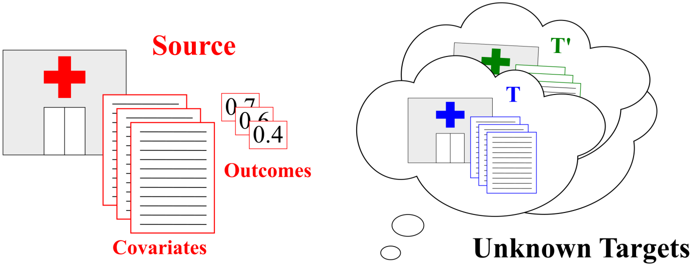
AAPOR 2022

Kim, M. P., Kern, C., Goldwasser, S., Kreuter, F. and Reingold, O. (2022). Universal Adaptability: Target-Independent Inference that Competes with Propensity Scoring. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 119(4). `https://doi.org/10.1073/pnas.2108097119`.

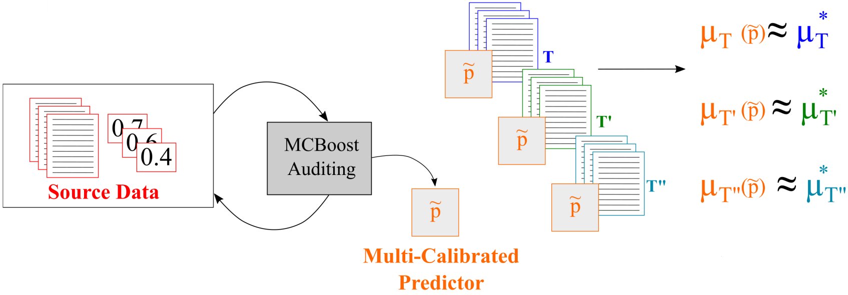


(a) Problem setup

Figure: Inference task and universal adaptability via multi-calibration



(a) Problem setup



(b) Universal adaptability

Figure: Inference task and universal adaptability via multi-calibration

Setting and Notation

$$\underbrace{\begin{array}{l|l} \underline{\text{Source distribution } (\mathcal{D}_s)} & \underline{\text{Target distribution } (\mathcal{D}_t)} \\ \text{Covariates } X\text{, outcome } Y & \text{Covariates } X \end{array}}$$

Sampling in source vs. target: $Z \in \{s, t\}$
Inference task: $\mu_{\mathrm{t}}^* = E_{(X,Y)\sim\mathcal{D}_t}[\ Y\ ]$
Estimation error: $\mathrm{er}_{\mathrm{t}}(\tilde{\mu}) = |\ \tilde{\mu} - \mu_{\mathrm{t}}^*\ |$

Propensity score: $e_{\mathrm{st}}(x) = P[\ Z = s, X = x\ ]$
Class of propensity scoring functions: Σ
Best-fit propensity score: $\sigma_{\mathrm{st}}^* \in \Sigma$
Propensity odds ratio: $c_\sigma(x) = \frac{1-\sigma(x)}{\sigma(x)}$
Class of propensity odds ratios: $\mathcal{C}(\Sigma)$

Key Challenge

Single source $\rightarrow$ many different targets!

- Challenge: Reweighting for every target is costly
- Goal: Provide insights in a "universal" format

Target-Specific Inference	Target-Independent Inference
e.g., propensity scoring	
Training Time: unlabeled samples from s, t	Training Time: labeled samples from s
Evaluation Time: labeled samples from s	Evaluation Time: unlabeled samples from t

Target-Independent Inference?

Imputation (e.g., Chen et al. 2020)

Given a predictor $p : \mathcal{X} \to [\ 0, 1\ ]$, estimate $E[\ Y|Z = t\ ]$ as

$$\hat{\mu}_t(p) = E[\ p(X)|Z = t\ ]$$

1. Learn an outcome predictor $p : \mathcal{X} \to \mathcal{Y}$ from source data
2. Average the "imputed" value in target distribution

Predictor trained on source may give bad predictions on target!

Multi-Calibration

Definition (Multi-Calibration)

For a given distribution $\mathcal{D}$ and class of functions $\mathcal{C}$, a predictor $\tilde{p} : \mathcal{X} \to [0,1]$ is $(\mathcal{C}, \alpha)$-multi-calibrated if

$$\left| E_{(X,Y)\sim\mathcal{D}} \left[\, c(X) \cdot (Y - \tilde{p}(X)) \, \right] \right| \leq \alpha.$$

- Multi-calibration (Hebert-Johnson et al., 2018; Kim et al., 2019) ensures that predictions are unbiased across every (weighted) subpopulation defined by $c \in \mathcal{C}$
- We derive a direct correspondence between protecting many subpopulations from **miscalibration** and ensuring **unbiased estimates** over a vast collection of target populations

Multi-Calibration guarantees Universal Adaptability

Definition (Universal Adaptability)

For a source distribution $\mathcal{D}_s$ and a class of propensity scores Σ, a predictor $\tilde{p} : \mathcal{X} \to [0,1]$ is (Σ, β)-universally adaptable if for any target distribution $\mathcal{D}_t$

$$\mathrm{er}_{\mathrm{t}}\left(\mu_{\mathrm{t}}(\tilde{p})\right) \leq \mathrm{er}_{\mathrm{t}}\left(\mu_{\mathrm{t}}^{\mathrm{ps}}(\sigma_{\mathrm{st}}^{*})\right) + \beta$$

Theorem

Suppose $\tilde{p} : \mathcal{X} \to [0,1]$ is a $(\mathcal{C}(\Sigma), \alpha)$-multi-calibrated prediction function over source distribution $\mathcal{D}_s$. Then, for any target distribution $\mathcal{D}_t$, and for any $\sigma \in \Sigma$, the estimator $\mu_{\mathrm{t}}(\tilde{p})$ is $(\Sigma, \alpha + \Delta_{\mathrm{st}}(\sigma))$-universally-adaptable.

MCBoost

Given:

- Initial predictor $\tilde{p}$
- Validation data D
- An auditor to search for subpopulations c
  - find largest residuals
  - e.g. ridge regression, decision tree

Repeat:

- Search over $c \in \mathcal{C}$
- If $|E_{x \sim D}[c(x) \cdot (y - \tilde{p}(x))]| > \alpha$
  - update as $\tilde{p}(x) \leftarrow \tilde{p}(x) - \eta \cdot c(x)$

R package on CRAN (Pfisterer et al., 2021) – `https://github.com/mlr-org/mcboost`

Application Setup

- **Data**
  - Source: unweighted NHANES III
  - Target: weighted NHIS
  - Linked to death certificates records (NDI)
- **Analytical Statistic**

  Estimate of 15-year all-cause mortality rate, by subpopulation
- **Inference Methods**
  - IPSW-Overall: Reweighting with global propensity scores
  - IPSW-Subgroup: Reweighting with subgroup-specific propensity scores
  - RF-Naive: Mortality prediction with random forest
  - RF-MCBoost: Mortality prediction with multi-calibrated RF

Application Results

Table: Estimation error in inferred mortality rate (% error in parentheses)

	IPSW		RF	
	Overall	Subgroup	Naive	MC-Boost
Overall	2.37 (13.5%)	—	1.11 (6.3%)	**0.52 (3.0%)**
Male	2.51 (13.4)	0.91 (4.9)	-0.34 (1.8)	**0.11 (0.6)**
Female	2.40 (14.6)	3.99 (24.2)	2.43 (14.8)	**0.90 (5.4)**
Age 18-24	**0.00 (0.1)**	-0.39 (17.5)	6.03 (270.2)	1.76 (79.0)
Age 25-44	**-0.20 (5.2)**	-0.41 (10.6)	0.82 (21.2)	0.66 (17.2)
Age 45-64	-0.75 (4.2)	-0.41 (2.3)	0.86 (4.8)	-0.29 (1.6)
Age 65-69	-4.23 (9.3)	-5.23 (11.5)	**-3.52 (7.7)**	**-1.99 (4.4)**
Age 70-74	-1.36 (2.3)	**0.47 (0.8)**	-3.02 (5.0)	**0.61 (1.0)**
Age 75+	3.53 (4.1)	2.85 (3.3)	0.51 (0.6)	2.19 (2.5)
White	3.53 (18.9)	0.75 (4.0)	1.03 (5.5)	0.69 (3.7)
Black	-4.00 (21.1)	**-0.48 (2.5)**	**-0.66 (3.5)**	**-0.52 (2.7)**
Hispanic	1.73 (17.0)	**0.48 (4.7)**	2.91 (28.6)	1.55 (15.2)
Other	**-0.02 (0.2)**	-3.54 (39.5)	3.52 (39.3)	-2.06 (23.0)

Semi-synthetic Simulation

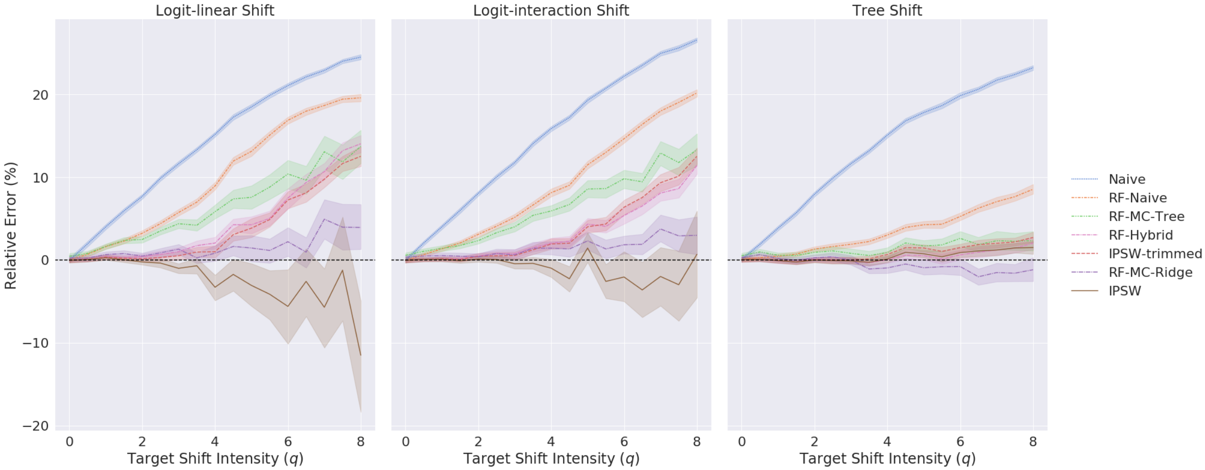


Figure: Relative error in inferred voting rates under synthetic shift

Discussion

- **Universal Adaptability**
  Valid inferences across a rich class of targets
- **General Result**
  Multicalibration persists under covariate shift
- **Meta-Takeaway**
  Algorithmic fairness useful beyond “fairness”

Thanks!

c.kern@uni-mannheim.de

References

Chen, S., Yang, S., and Kim, J. K. (2020). Nonparametric Mass Imputation for Data Integration. *Journal of Survey Statistics and Methodology*, 10(1):1–24.

Hebert-Johnson, U., Kim, M. P., Reingold, O., and Rothblum, G. (2018). Multicalibration: Calibration for the (computationally-identifiable) masses. In *Proceedings of the 35th International Conference on Machine Learning, PMLR 80*, pages 1939–1948.

Kim, M. P., Ghorbani, A., and Zou, J. (2019). Multiaccuracy: Black-box post-processing for fairness in classification. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES 19)*, pages 247–254. Association for Computing Machinery.

Pfisterer, F., Kern, C., Dandl, S., Sun, M., Kim, M. P., and Bischl, B. (2021). mcboost: Multi-Calibration Boosting for R. *Journal of Open Source Software*, 6(64).