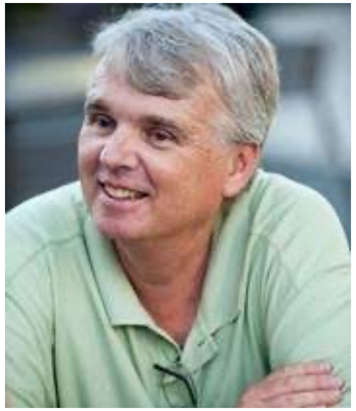
End-to-End Balancing for Causal Continuous Treatment-Effect Estimation

ICML 2022



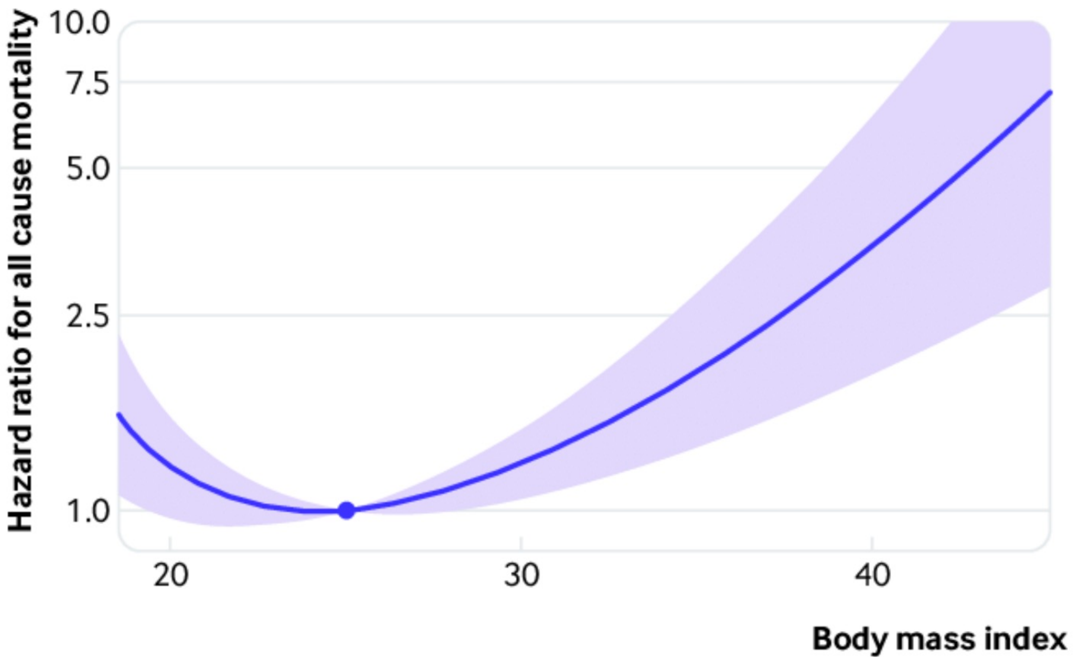


Taha Bahadori Eric Tchetgen Tchetgen David Heckerman

Amazon

Causal Continuous Treatment-Effect Estimation

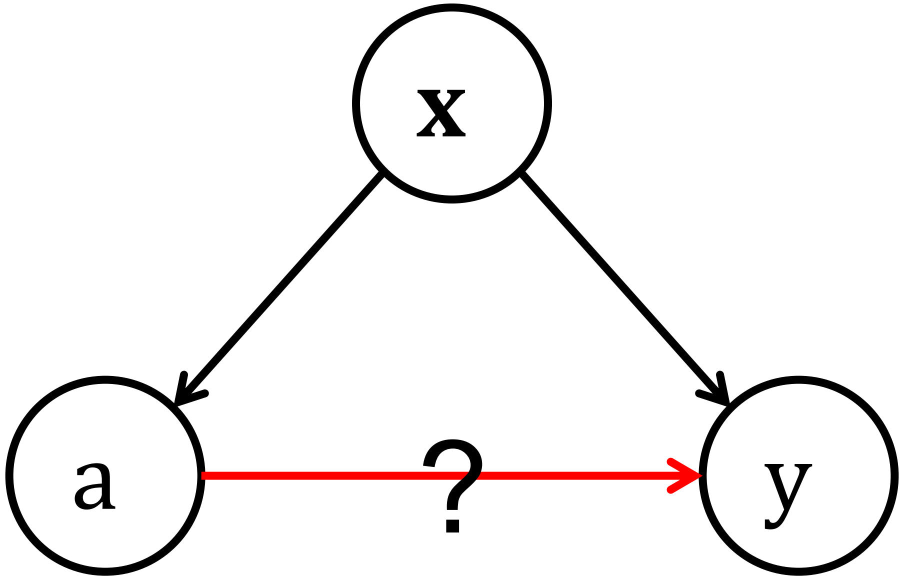
- More challenging than the binary treatment case
  - Uncountably many potential outcomes
  - Only few of them are observed
- Less studied than the binary treatment setting



Sun, Yi-Qian, et al. "Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses." *bmj* 364 (2019).

Problem Setup

- **Data:** Treatments (a), Outcomes (y), and Observed confounders (**x**).
- **Assumptions:** Ignorability, Positivity, and Consistency.
- Common techniques:
  - Outcome regression
  - Weighting approaches
  - Doubly robust combination



Weight Instability and Entropy Balancing

Extreme Weights Problem	Entropy Balancing

- Under severe confounding
- Extreme propensity scores
- Unstable causal inference

- Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical science.

$$\min_{\boldsymbol{w}} -H(\boldsymbol{w})$$
$$s.t.\, Covariate\ Balance$$

- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*.

Learning the Base Weights

$$\min_{w} \sum_{i=1}^{n} w_i \log \frac{w_i}{q_i}$$

Subject to:

$$\sum_{i=1}^{n} w_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(a_i - \bar{a}) = \boldsymbol{0},$$

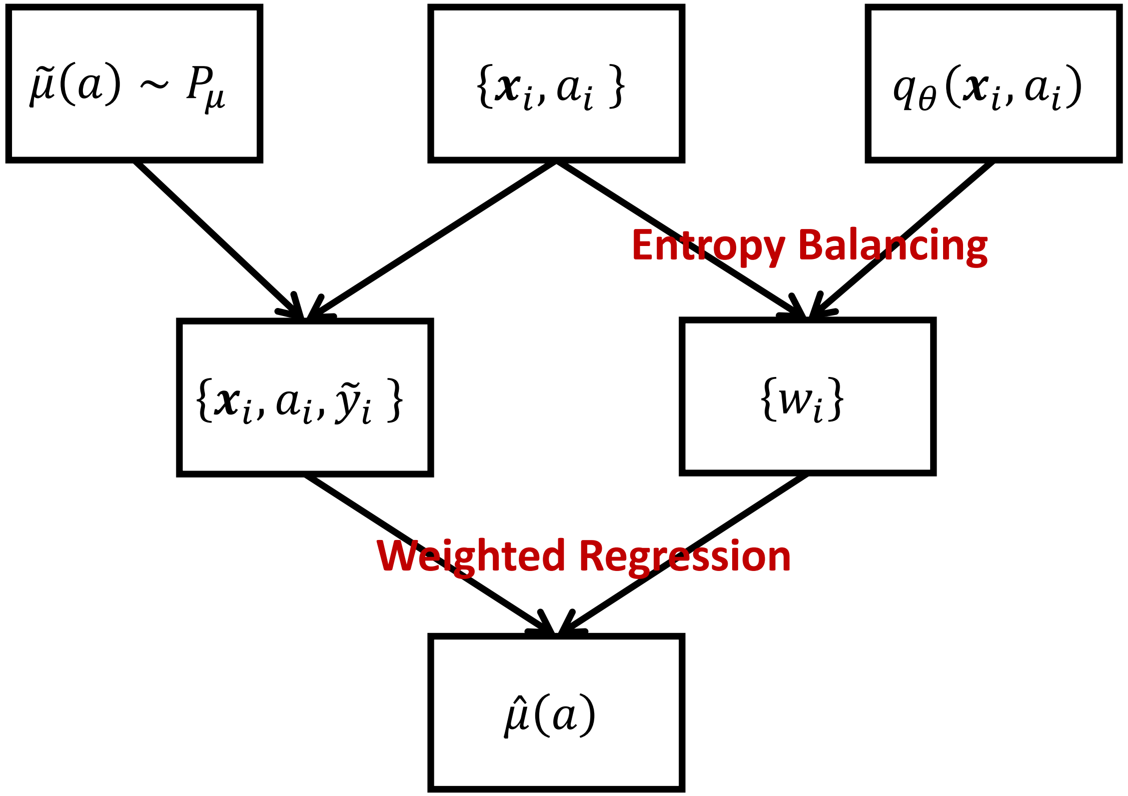
$$\sum_{i=1}^{n} w_i = n,$$

$$w_i \geq 0.$$

- Choice of the base weights (q_i)?
- Proper choice of base weights:
  1. A way to regularize the weights.
  2. Improve the quality of weighted causal regression.
  3. Embed our prior belief about the shape of the response function.

End-to-End Learning for the Base Weights

- Base-weights $q_\theta(x_i, a_i)$ modeled by a neural network
- Generate pseudo-datasets using random response functions $\tilde{\mu}(a)$.
- Train q_θ to minimize estimation error of $\tilde{\mu}(a)$ by $\hat{\mu}(a)$.
- **End-to-end learning**: Learn base weights to directly increase causal inference accuracy.



Theoretical Guarantees

- **Unbiasedness**
  - Regardless of the choice of the base weights, the weights are unbiased estimators of the stable weights: $w(\boldsymbol{x}, a) = {}^{f(a)}\!/_{f(a|\boldsymbol{x})}$.

- **Asymptotic normality**

$$\sqrt{n}\,(\,\widehat{w}_n(a_i, \boldsymbol{x}_i) - w^\star(a_i, \boldsymbol{x}_i)\,) \;\xrightarrow{d}\; \mathcal{N}(0, \sigma^2(a_i, \boldsymbol{x}_i)).$$

Synthetic Experiments

Two datasets:

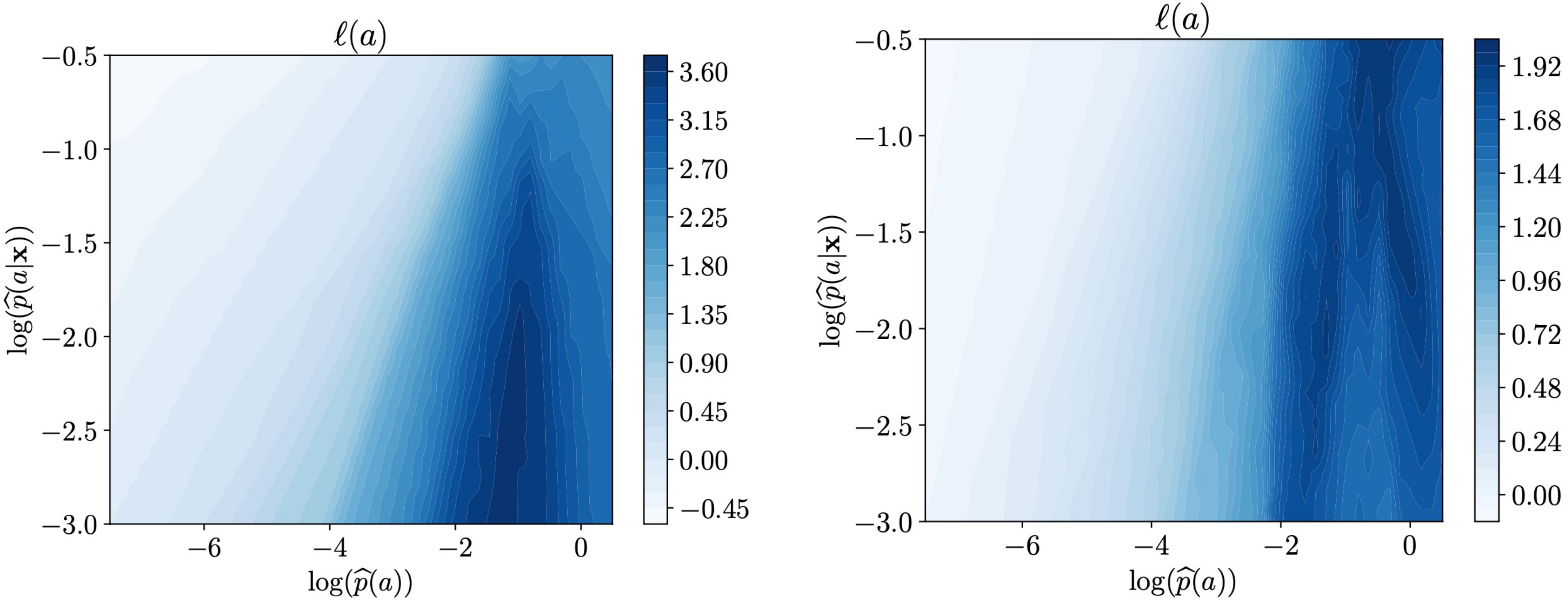
- Linear response curve
- Non-linear response curve modeled with Polynomials of degree 3.

Algorithm	Linear	Non-linear
Inverse Propensity Weighting (SW)	2.057 (0.437)	0.530 (0.025)
Permutation Weighting	1.1543 (6.580)	0.525 (0.250)
Entropy Balancing (Const.)	0.880 (0.072)	0.335 (0.022)
Entropy Balancing (SW)	0.652 (0.059)	0.403 (0.025)
End-to-End Balancing	**0.383** (**0.035**)	**0.276** (**0.014**)

Average RMSE for estimation of the response functions.

Run the analysis 100 times, report the mean and SE of estimating the polynomial coefficients.

Learned Base Weight Functions



(a) Linear Design (b) Non-linear Design

Thank you!