ABC3: Active Bayesian Causal Inference with Cohn Criteria in Randomized Experiments

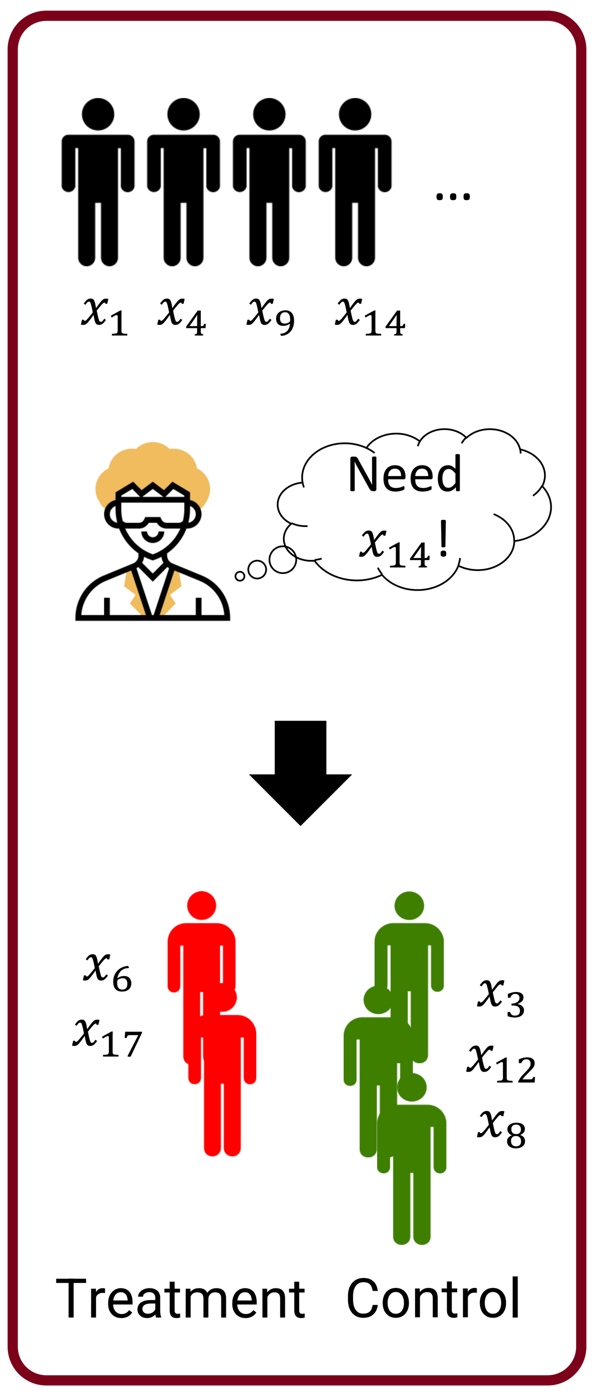
Taehun Cha and Donghun Lee
Department of Mathematics, Korea University

A I M L @ K

AAAI

Motivation

- Randomized experiment is usually expensive, so an efficient experiment design is desirable
- What if we already know the covariate information in prior, can we utilize it?
- Active learning: a practitioner choose unlabeled data points and ask an oracle to label them



Problem Setting

- X: Covariate, Y^a: Potential outcome for treatment ($a = 1$) and control ($a = 0$) groups
- X_Ω, X_t^a: covariate sets of the whole subjects (Ω) and treated/controlled subjects at t
- $\hat{y}_\Omega^a$, $\hat{y}_t^a$: a regressor trained on oracle (Ω) or collected data set at t
- Goal: Estimate $CATE(x) = E[Y^1 - Y^0 | X = x]$ with an estimator $\widehat{CATE}_t = \hat{y}_t^1 - \hat{y}_t^0$
- Original expected precision in estimation of heterogeneous effect (PEHE) includes the population $CATE(x)$, which makes analysis tricky. Define $\widehat{CATE}_\Omega = \hat{y}_\Omega^1 - \hat{y}_\Omega^0$.
- Suggest Bayesian PEHE:

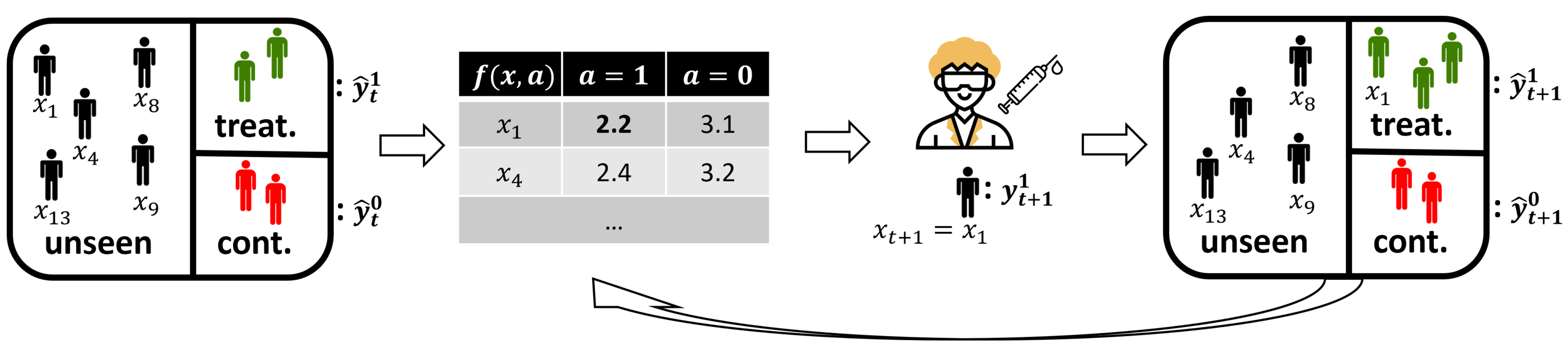
$$\epsilon_{PEHE}^{\Omega}\left(\widehat{CATE}_t\right) = \int_X \left(\widehat{CATE}_t(x) - \widehat{CATE}_\Omega(x)\right)^2 dP(x)$$

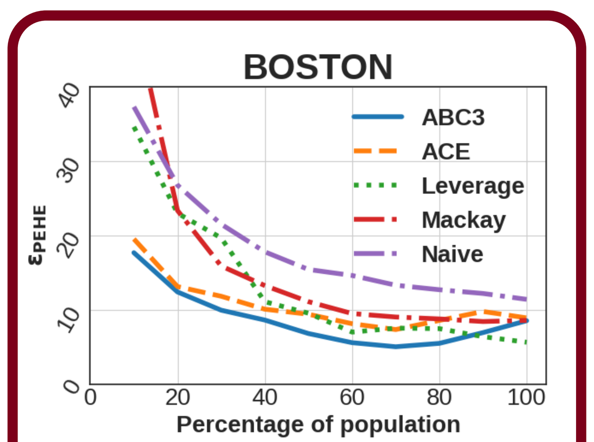
Algorithm: ABC3

- Goal: Find $x_{t+1} \in X_\Omega \setminus (X_t^1 \cup X_t^0), a_{t+1} \in \{0,1\}$, s.t. $argmin_{x_{t+1},a_{t+1}} E_{t+1}[\epsilon_{PEHE}^{\Omega}(\widehat{CATE}_t)]$
- **(Theorem)** Assume $Y^a \sim GP(0, k(x, x')), \hat{y}_t^a(x) = E_t[Y^a(x)]$. Then under mild assumptions,

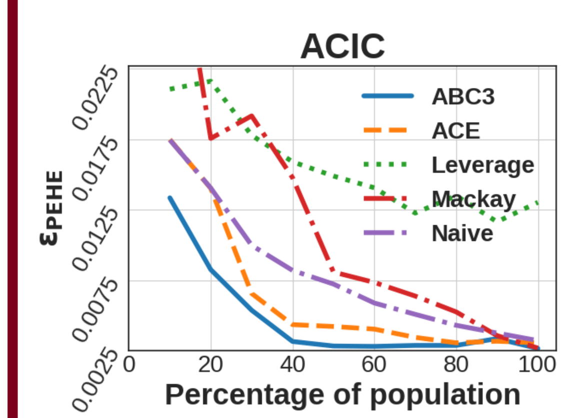
$$argmin_{x_{t+1},a_{t+1}} E_{t+1}[\epsilon_{PEHE}^{\Omega}(\widehat{CATE}_t)] = argmin_{x_{t+1},a_{t+1}} \int_X V_{t+1}[Y^1(x)] + V_{t+1}[Y^0(x)] dP(x)$$

  which is similar to Cohn criteria
- Thanks to GP, $V_{t+1}[Y^a(x)]$ does not depend on future outcome y_{t+1} (computable in prior)
- We also propose an efficient way to compute the target quantity (Proposition 4.2.)





- ABC3 at 20% > Naïve at 100%

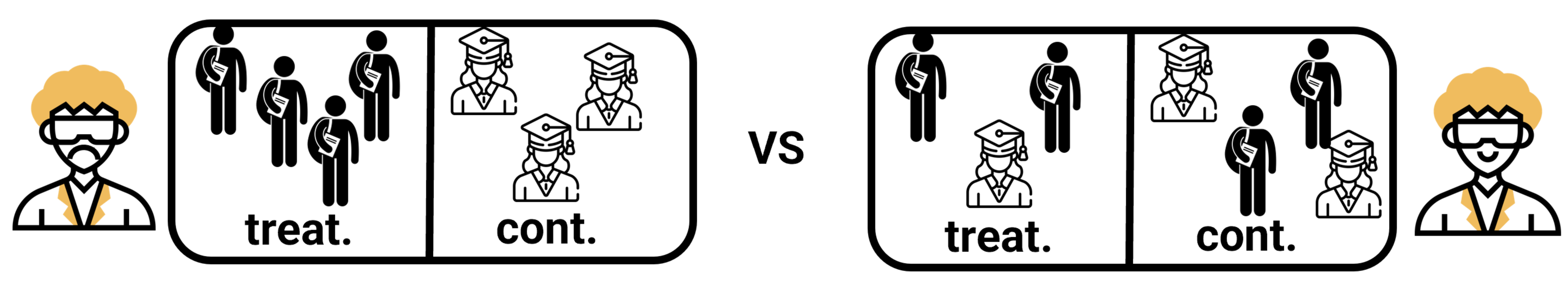


- ABC3 at 40% > Naïve at 100%

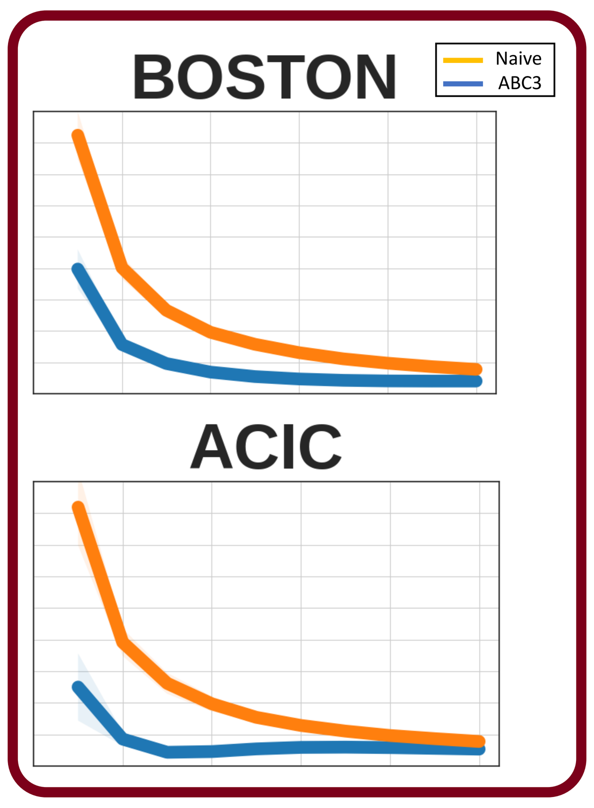
Theoretical Property

Balancing

- Balance between the treatment/control groups is crucial



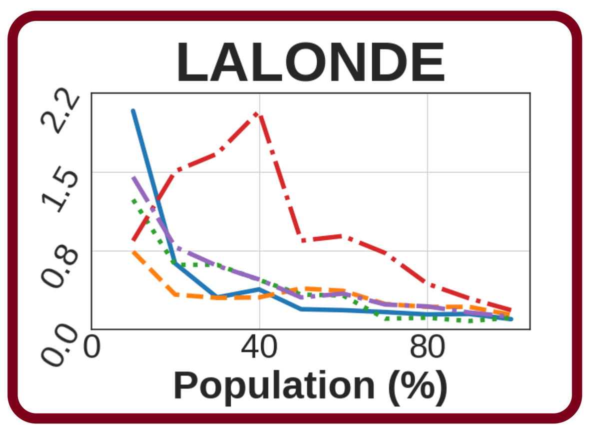
- Maximum Mean Discrepancy(P, Q): difference of mean embeddings of P and Q in RKHS
- **(Theorem)** MMD(Treatment, Control) $\leq 4\left(\frac{\lambda^*}{|X_t^1|} + \frac{\lambda^*}{|X_t^0|}\right)$: minimized as time proceed
  $+2\int_X V_{t+1}[Y^1(x)] + V_{t+1}[Y^0(x)] dP(x)$: our optimization target



Type 1 Error

- Type 1 Error rate: $P_t[Type\ 1\ Error(x)] = P_t[|Y^1(x) - Y^0(x)| > \alpha]$, α: decision threshold
- **(Theorem)** Under Fisher's Sharp Null (i.e. $CATE(x) = 0$), ABC3 minimizes the upper bound of

$$\int_X P_{t+1}[Type\ 1\ Error(x)]\, dP(x)$$



Conclusion

- We introduce ABC3, active learning algorithm for causal inference from Bayesian perspective
- ABC3 outperforms other algorithms by balancing treatment/control groups and minimizes type 1 error rate

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea Institute for Advancement of Technology (KIAT) through P0022334 and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1G1A1102828).
We have our oral presentation at ???

Paper Code Personal