Distribution-Free Assessment of Population Overlap in Observational Studies

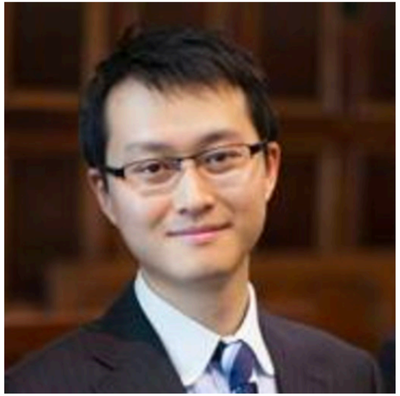
Lihua Lei

ASSA 2022 Virtual Annual Meeting

Collaborators

Alexander D'Amour

(Google Brain)



Peng Ding

(UC Berkeley)

Avi Feller

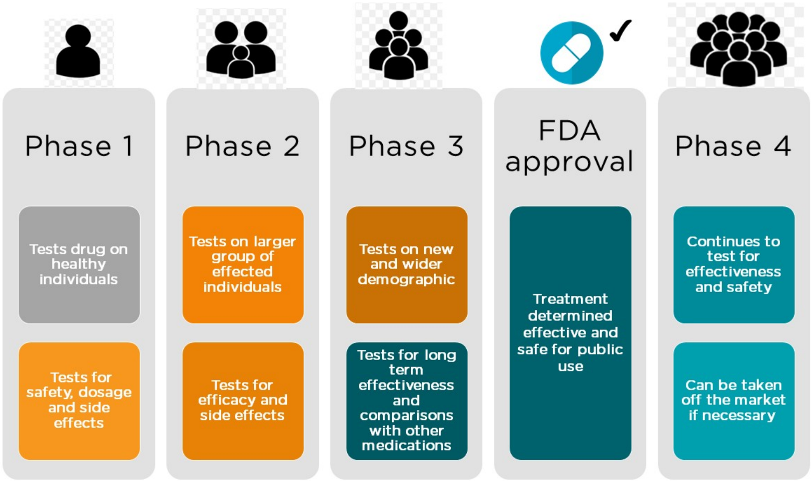
(UC Berkeley)

Jasjeet Sekhon

(Yale)

Causal inference in observational studies

Randomized experiments

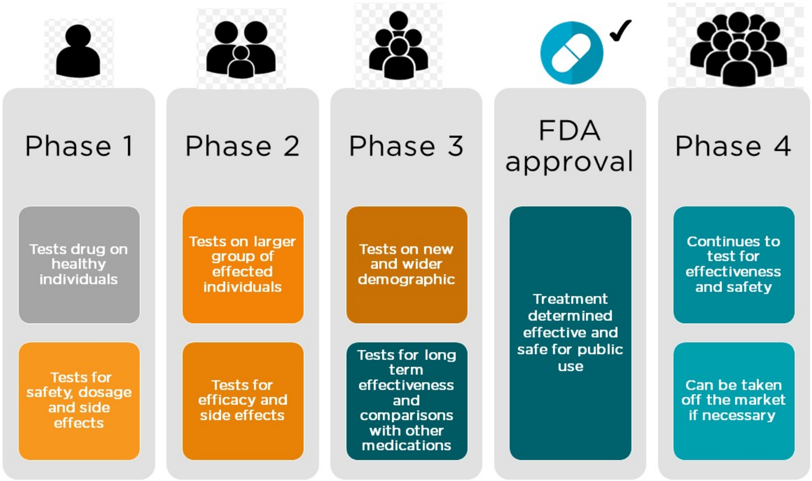


Everyone has a chance to be exposed to treatment(s)

Reliable inference of causal effect w/o modeling outcomes

Causal inference in observational studies

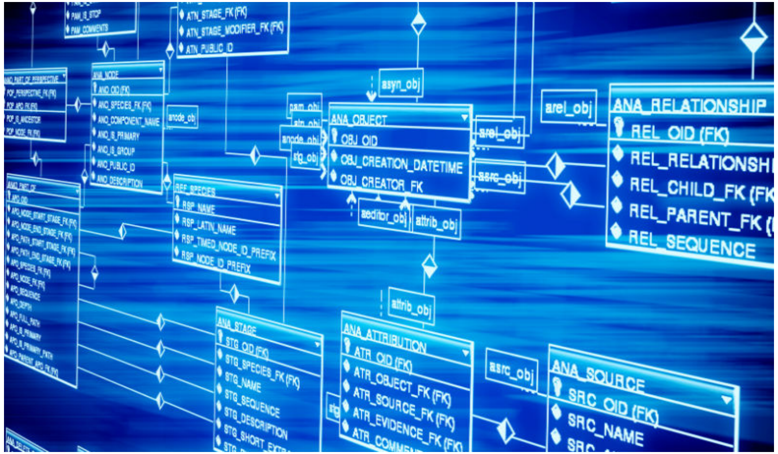
Randomized experiments



Everyone has a chance to be exposed to treatment(s)

Reliable inference of causal effect w/o modeling outcomes

Observational studies

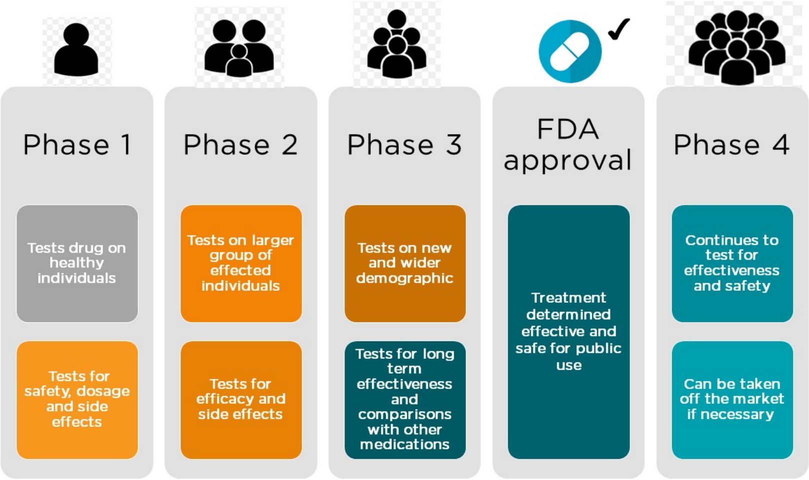


Everyone has a chance to be exposed **conditional on covariates**

Reliable inference of causal effect w/o modeling outcomes

Causal inference in observational studies

Randomized experiments

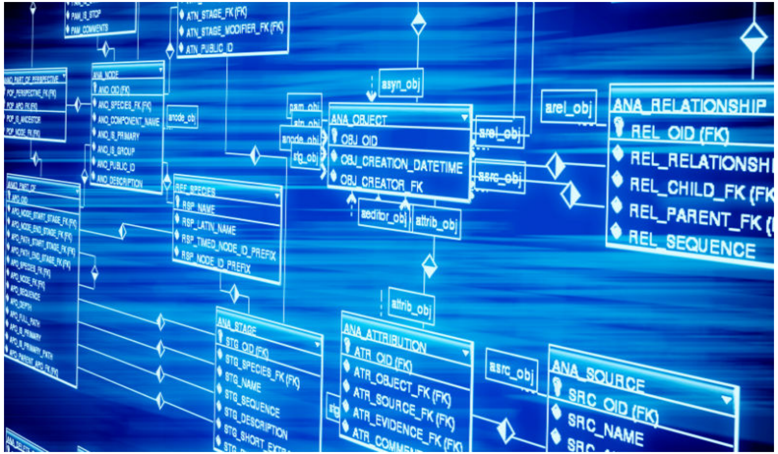


Everyone has a chance to be exposed to treatment(s)

Reliable inference of causal effect w/o modeling outcomes

★ Need sufficiently many treated/control subjects

Observational studies

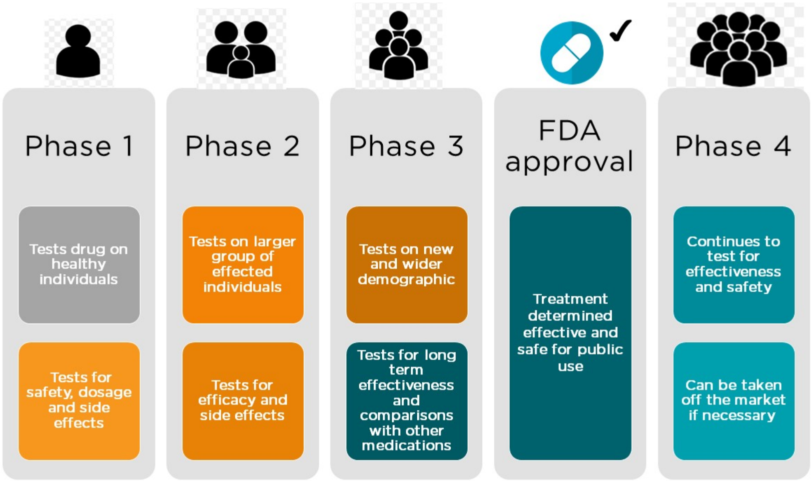


Everyone has a chance to be exposed **conditional on covariates**

Reliable inference of causal effect w/o modeling outcomes

Causal inference in observational studies

Randomized experiments

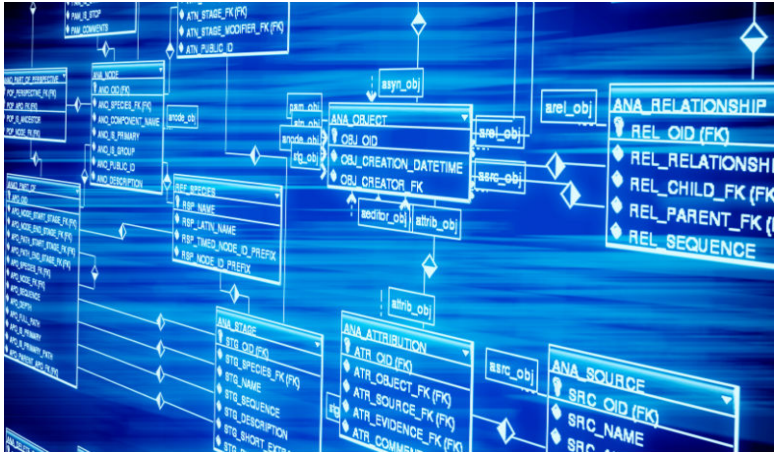


Everyone has a chance to be exposed to treatment(s)

Reliable inference of causal effect w/o modeling outcomes

★ Need sufficiently many treated/control subjects

Observational studies



Everyone has a chance to be exposed **conditional on covariates**

Reliable inference of causal effect w/o modeling outcomes

★ Need sufficiently many treated/control subjects **conditionally**

Strict overlap condition and population overlap slack

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^{n} \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition and population overlap slack

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition (example):

$$0.1 \leq e(X) \leq 0.9, \text{ a.s.}$$

One of the most fundamental condition!

Strict overlap condition and population overlap slack

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition (example):

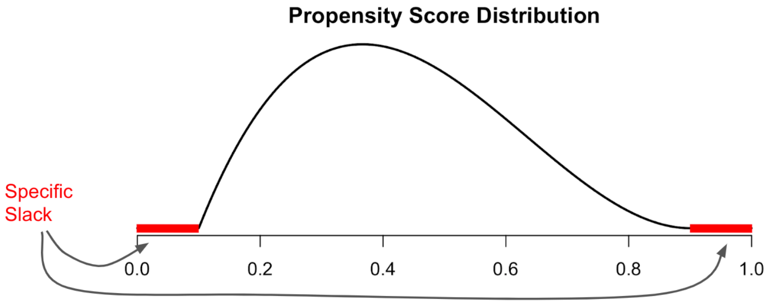
$$0.1 \le e(X) \le 0.9, \text{ a.s.}$$

One of the most fundamental condition!

Definition (population overlap slack)

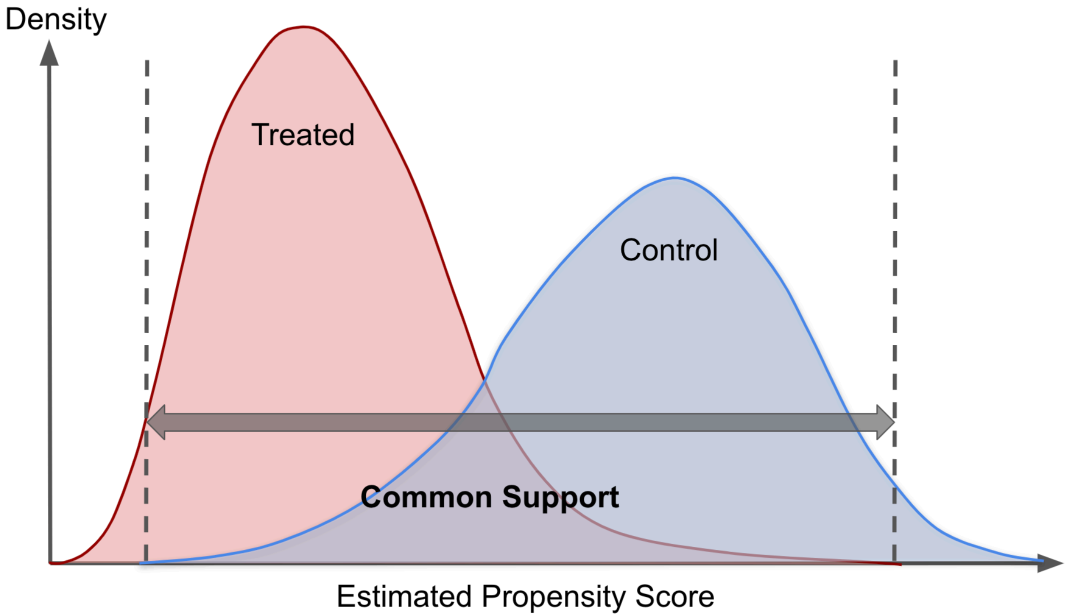
$$\mathcal{O}^* \triangleq \min_x \min\{e(x), 1 - e(x)\}$$

Strict overlap condition $\iff \mathcal{O}^* \ge 0.1$

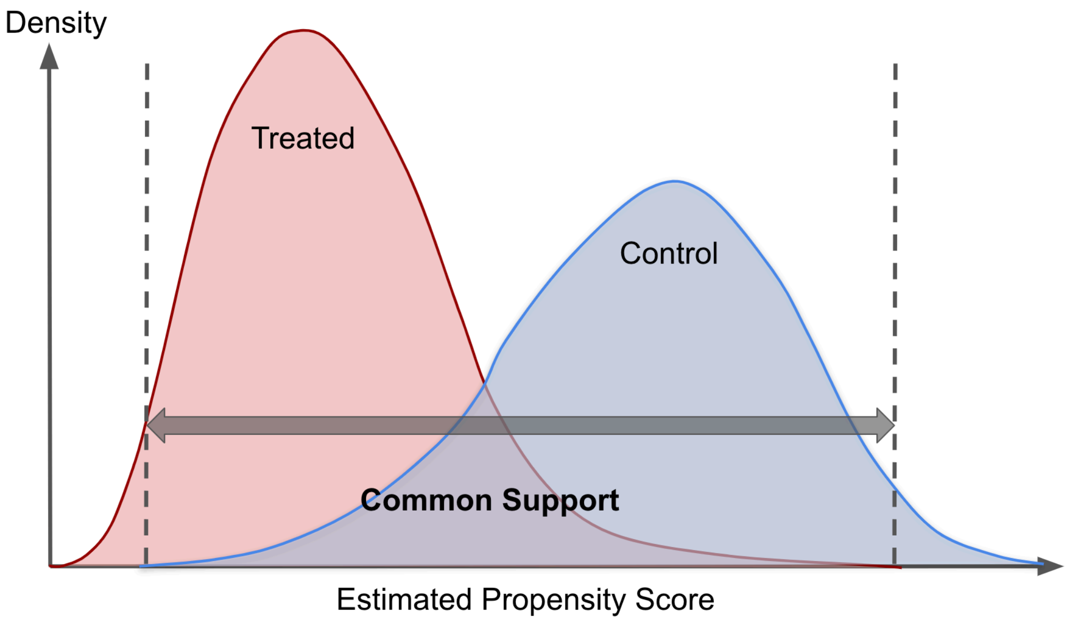


$n\mathcal{O}^*$ is the **effective samples size** without outcome restrictions (Hong et al. '20)

Current assessment of overlap



Current assessment of overlap



Misspecification error:

- wrong model for $e(x)$
- bad hyper-parameter tuning
- optimization issues
- …

Finite-sample error:

- $\mathcal{O}^*$ is an irregular parameter
- uncertainty quantification for function estimation is hard (Barber '20)

O-value: distribution-free assessment of population overlap

We propose **O-values** as **upper confidence bounds** of $\mathcal{O}^*$, denoted by $\hat{\mathcal{O}}$, that

+ **guarantees coverage**, i.e., $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

+ **in finite samples** (no asymptotics!)

+ **without any assumption** on $e(x)$ or X

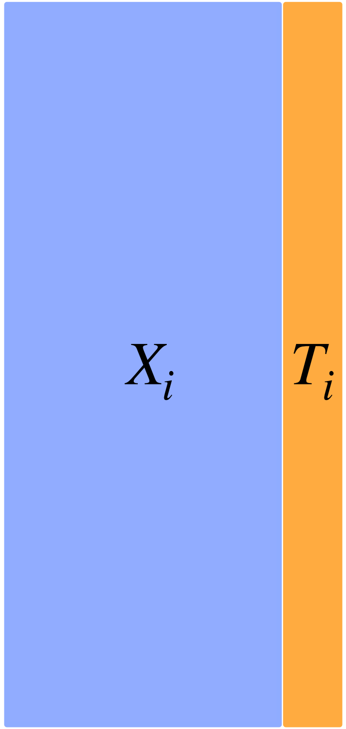
+ is able to wrap around **any black-box** algorithm to estimate $e(X)$

Analogous to p-value:

- small $\hat{\mathcal{O}} \Rightarrow$ strong evidence against overlap
- large $\hat{\mathcal{O}} \not\Rightarrow$ sufficient overlap

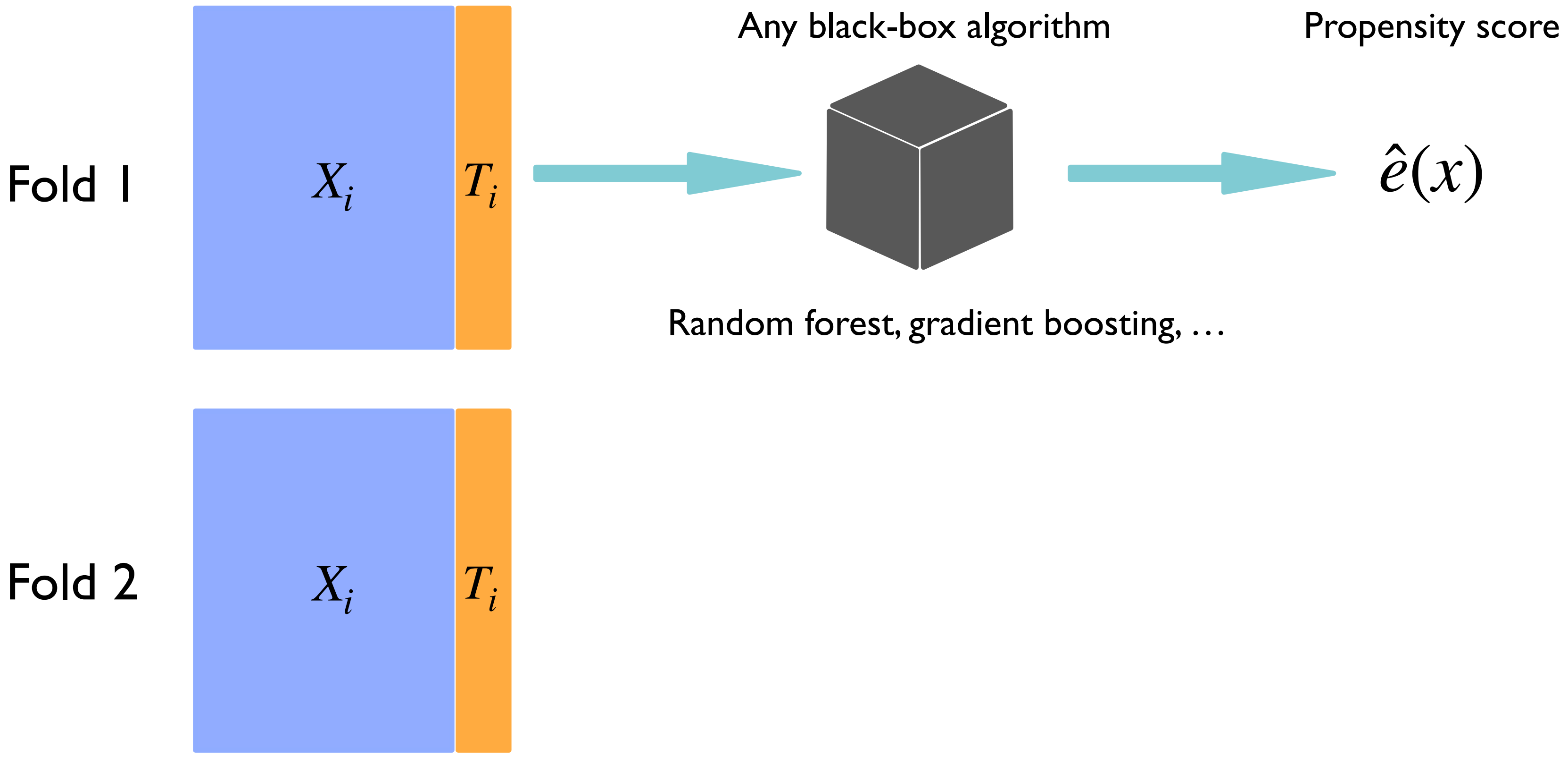
Step 1: covariate standardization

Data

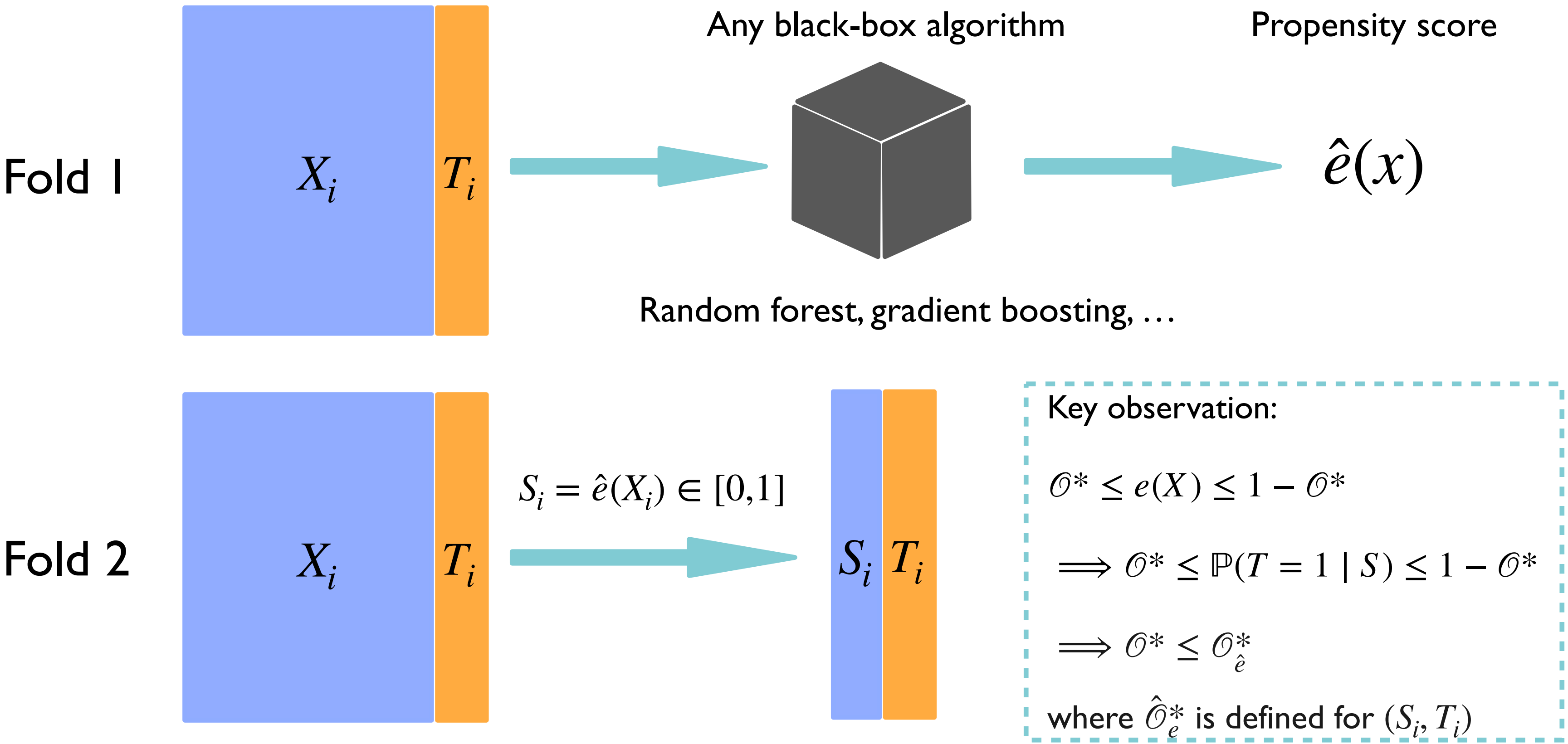


First challenge: X may be mixed-typed, high-dimensional, non-numeric, …

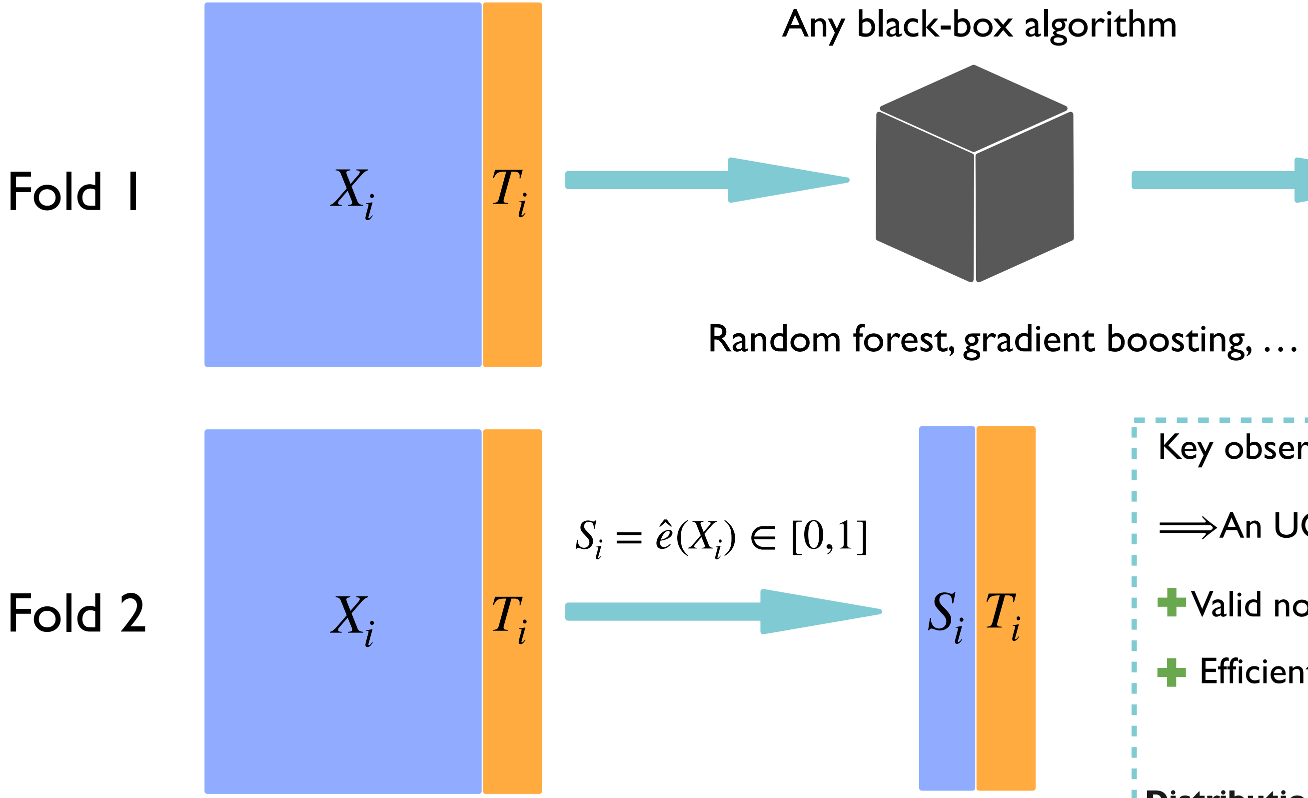
Step 1: covariate standardization



Step I: covariate standardization



Step I: covariate standardization



Key observation: $\mathcal{O}^* \leq \mathcal{O}^*_{\hat{e}}$

$\Longrightarrow$An UCB of $\mathcal{O}^*_{\hat{e}}$ is an UCB of $\mathcal{O}^*$

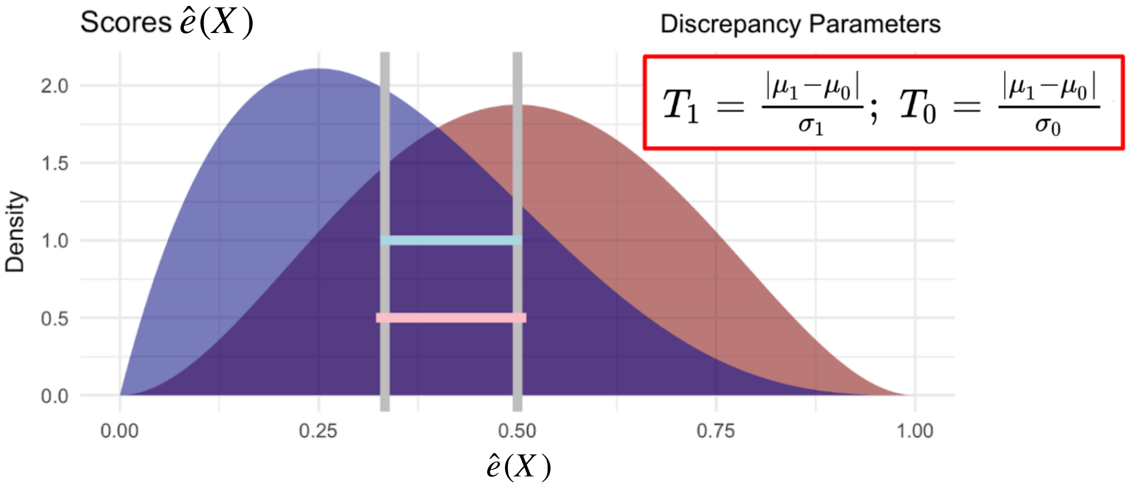
➕ Valid no matter how poor $\hat{e}(x)$ is

➕ Efficient when $\hat{e}(x)$ is good

$$\mathcal{O}^* = \mathcal{O}^*_e$$

Distribution-free validity + adaptivity!

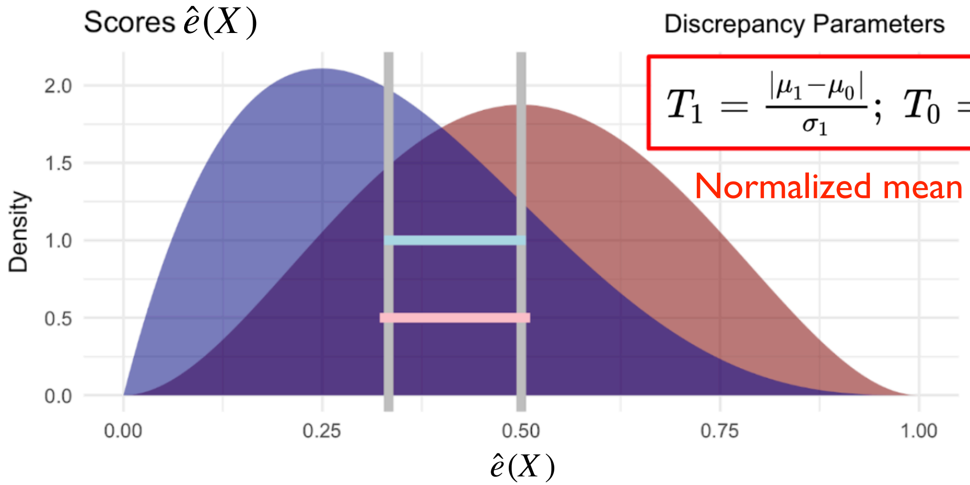
Step 2: careful balance check (difference-in-means O-value)



Intuition:

large $\mathcal{O}^*$ $\Longrightarrow$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (difference-in-means O-value)



Discrepancy Parameters

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}; \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Normalized mean difference

Theorem (using information theory)

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

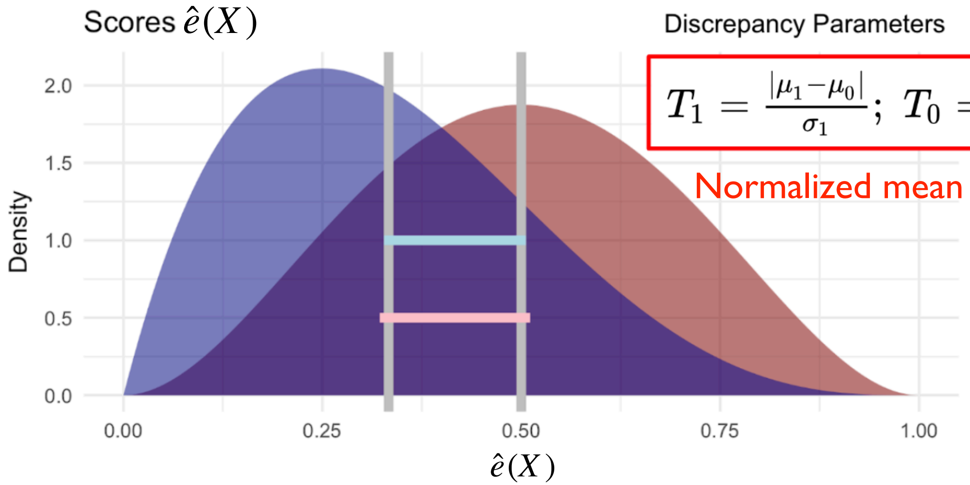
$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}, \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

Intuition:

large $\mathcal{O}^* \implies$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (difference-in-means O-value)



Discrepancy Parameters

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}; \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Normalized mean difference

Theorem (using information theory)

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}, \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

Empirical Bernstein's inequality

$\Longrightarrow$ Joint confidence region of $(\mu_1, \sigma_1, \mu_0, \sigma_0)$

$\Longrightarrow$ Upper confidence bound on $f(T_1, T_0)$

$\Longrightarrow$ DiM O-value

Intuition:

large $\mathcal{O}^* \Longrightarrow$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

EBI is loose; we use other tools instead

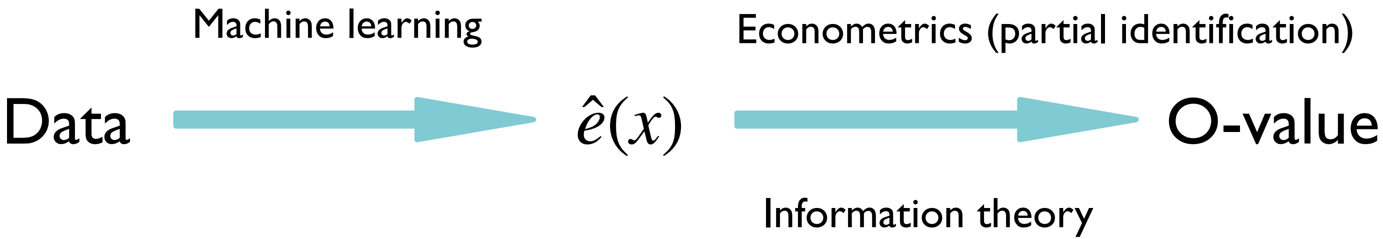
Application: O-values for Lalonde data

- National Supported Work Demonstration program (Lalonde '86)
- Treatment group has 185 units
- 7 control groups: 6 observational and 1 experimental

	CPS			PSID			RCT		
	n_0	$\hat{\mathcal{O}}$	$\hat{L}$	n_0	$\hat{\mathcal{O}}$	$\hat{L}$	n_0	$\hat{\mathcal{O}}$	$\hat{L}$
Raw	15992	0.003	77%	2490	0.018	75%	260	0.483	0%
V2	2369	0.021	71%	253	0.234	45%			
V3	429	0.143	53%	128	0.313	23%			

Efficiency loss due to imbalance : $\hat{L} = 1 - \underbrace{n\hat{\mathcal{O}}}_{\text{effective sample size}} \;/\; \underbrace{\min\{n_1, n_0\}}_{\text{effective sample size in an RCT}}$

Summary



Thank you!

I am on the 2021-22 job market. Check out my CV and other works at

https://lihualei71.github.io/

lihua_lei_stat