Sensitivity analysis under the f-sensitivity models: A distributional robustness perspective

Ying Jin

Department of Statistics, Stanford University

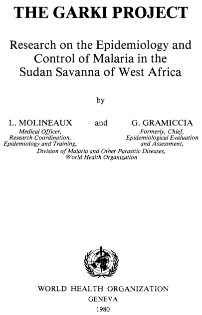
Data Driven Decision Making Seminar, November 16, 2022

Joint work with

Zhimei Ren
UChicago Statistics

Zhengyuan Zhou
NYU Stern

Estimating treatment effects


THE GARKI PROJECT

Research on the Epidemiology and
Control of Malaria in the
Sudan Savanna of West Africa

by

L. MOLINEAUX
Medical Officer,
Research Coordination,
Epidemiology and Training,

and

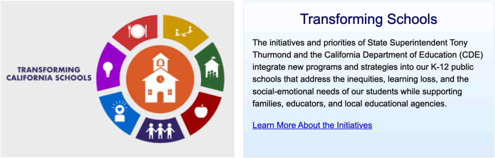
G. GRAMICCIA
Formerly, Chief,
Epidemiological Evaluation
and Assessment,

Division of Malaria and Other Parasitic Diseases,
World Health Organization

WORLD HEALTH ORGANIZATION
GENEVA
1980

Source: the World Health Organization

Public health policies



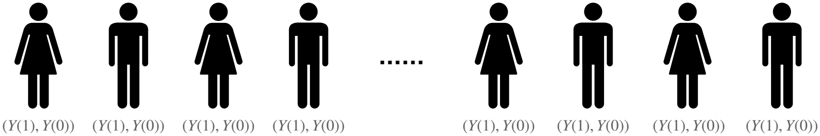
Source: `cde.ca.gov`

Education programs

Treatment effects in observational studies

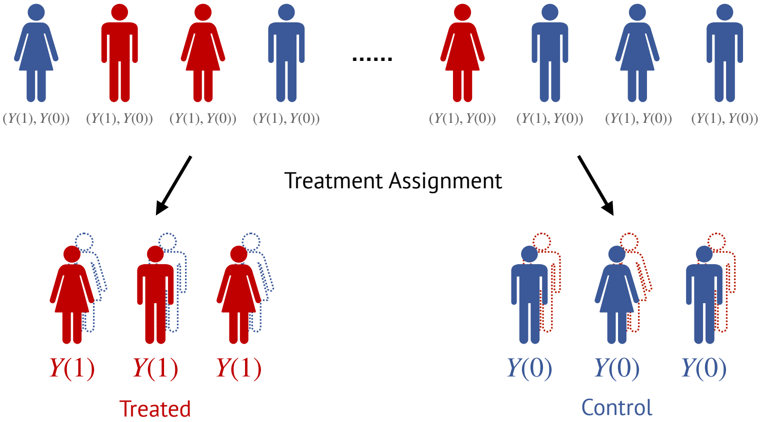
- Randomized experiment is the golden rule, but not always feasible
- Opportunities for observational data

Potential outcome framework



Sample from population

Potential outcome framework



Y(1)	✓	✓	✓
Y(0)	?	?	?

Y(1)	?	?	?
Y(0)	✓	✓	✓

Potential outcome framework

- Population $(X_i, Y_i(1), Y_i(0), T_i) \overset{\text{i.i.d.}}{\sim} \mathbb{P}$
- Subjects $\big(X_i, Y_i(1), Y_i(0)\big)$, Treatment $T_i \in \{0, 1\}$ (unknown mechanism)
- Partial observations: (X_i, T_i, Y_i), where $Y_i = Y(T_i)$ (SUTVA)

Potential outcome framework

- Population $(X_i, Y_i(1), Y_i(0), T_i) \overset{\text{i.i.d.}}{\sim} \mathbb{P}$
- Subjects $\big(X_i, Y_i(1), Y_i(0)\big)$, Treatment $T_i \in \{0, 1\}$ (unknown mechanism)
- Partial observations: (X_i, T_i, Y_i), where $Y_i = Y(T_i)$ (SUTVA)
- Target estimands
  - Average treatment effect (ATE): $\mathbb{E}[Y(1) - Y(0)]$
  - Average treatment effect on the treated (ATT): $\mathbb{E}[Y(1) - Y(0) \mid T = 1]$
  - Average treatment effect on the control (ATC): $\mathbb{E}[Y(1) - Y(0) \mid T = 0]$

Standard assumption: strong ignorability (unconfoundedness)

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

- Not testable but violation is consequential

Standard assumption: strong ignorability (unconfoundedness)

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

- T: admission to ICU
- X: demographics + examination results upon admission
- $Y(1), Y(0)$: mortality if admitted / not admitted to ICU

Standard assumption: strong ignorability (unconfoundedness)

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

- T: admission to ICU
- X: demographics + examination results upon admission
- $Y(1), Y(0)$: mortality if admitted / not admitted to ICU
- Any two patients with the same features are equally likely to be admitted to ICU
- Undocumented symptoms? Doctor's judgement? ... Unmeasured confounding

Unmeasured confounding

- Unmeasured confounder U that affects both outcomes and treatment

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid (X, U)$$

- Impact of confounding: selection bias

$$\mathrm{OR}(X, U) = \underbrace{\frac{\mathbb{P}(T=1 \mid X)}{\mathbb{P}(T=0 \mid X)}}_{\text{Observed odds}} \Big/ \underbrace{\frac{\mathbb{P}(T=1 \mid X, U)}{\mathbb{P}(T=0 \mid X, U)}}_{\text{Actual odds}}$$

$\mathrm{OR}(X, U) = 1 \quad \Leftrightarrow \quad$ strong ignorability

$\mathrm{OR}(X, U) \neq 1 \quad \Leftrightarrow \quad$ confounding at (X, U)

Sensitivity analysis

$$\begin{aligned} \text{assume some degree of confounding} \quad &\Rightarrow \quad \text{bounds on treatment effects} \\ &\Rightarrow \quad \text{robustness of conclusions} \end{aligned}$$

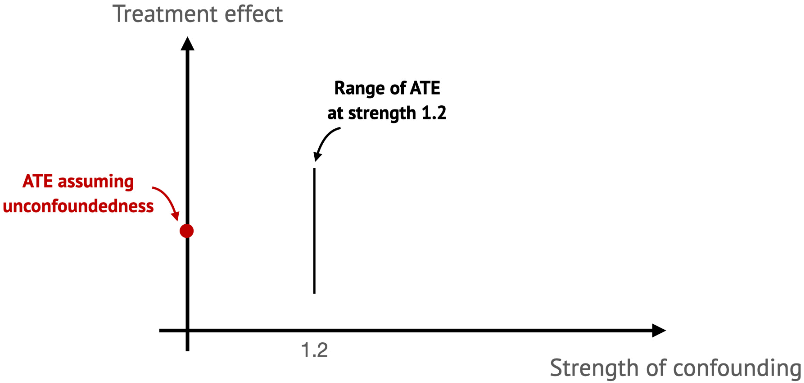
Sensitivity analysis

$$\text{assume some degree of confounding} \quad \Rightarrow \quad \text{bounds on treatment effects}$$
$$\Rightarrow \quad \text{robustness of conclusions}$$

Sensitivity analysis

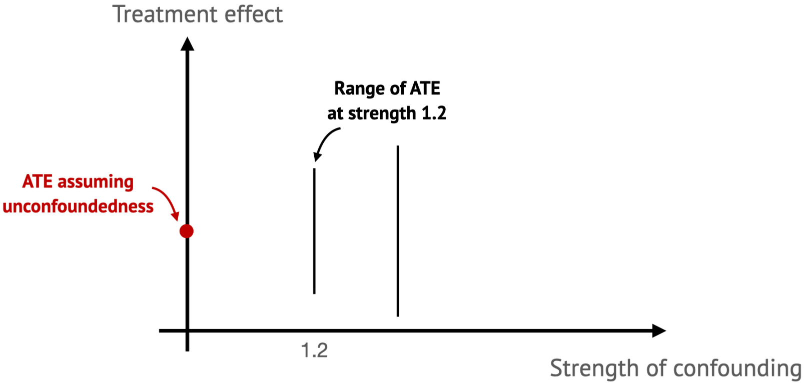
assume some degree of confounding $\Rightarrow$ bounds on treatment effects

$\Rightarrow$ robustness of conclusions



Sensitivity analysis

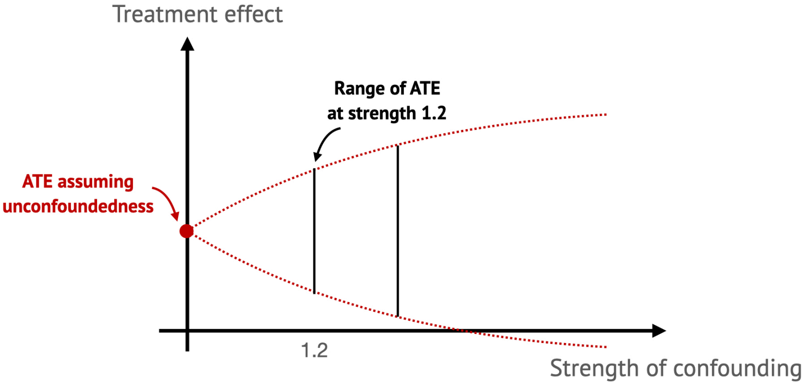
assume some degree of confounding $\Rightarrow$ bounds on treatment effects
$\Rightarrow$ robustness of conclusions



Sensitivity analysis

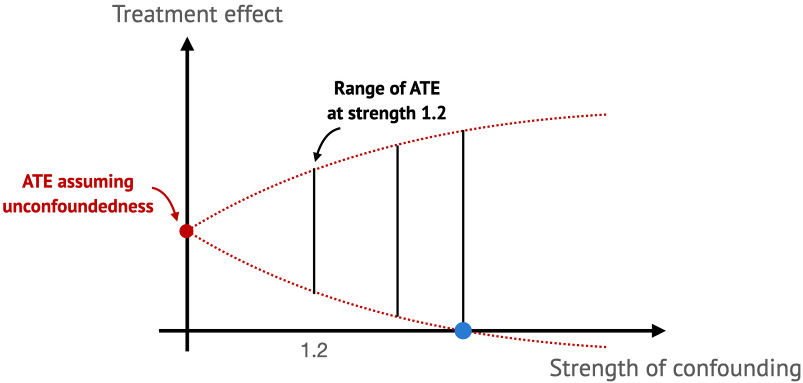
assume some degree of confounding $\Rightarrow$ bounds on treatment effects

$\Rightarrow$ robustness of conclusions



Sensitivity analysis

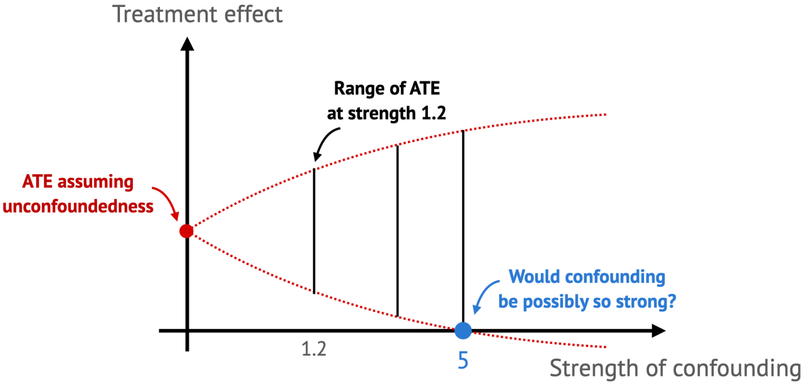
assume some degree of confounding $\Rightarrow$ bounds on treatment effects
$\Rightarrow$ robustness of conclusions



Sensitivity analysis

assume some degree of confounding $\Rightarrow$ bounds on treatment effects

$\Rightarrow$ robustness of conclusions



Sensitivity models on selection bias

- Uniform bounds on the selection bias (odds ratio) Rosenbaum and Rubin (1983); Tan (2006); Zhao et al. (2017); Dorn and Guo (2021); Dorn et al. (2021); Jin et al. (2021)

$$1/\Gamma \leq \mathrm{OR}(x, u) \leq \Gamma, \quad \forall x, u$$

- For any two patients with the same features and arbitrarily different confounders, their likelihood of admission to ICU can be off up to a constant

Sensitivity models on selection bias

- A practitioner imagining a parametric model... (Imbens 2003; Franks et al. 2019)

$$\mathrm{OR}(x, u) = \frac{e^{\theta_1^\top x + \theta_2 u}}{1 + e^{\theta_1^\top x + \theta_2 u}}, \quad U \sim N(0, 1)$$

- If U Gaussian? NO uniform bound on $\mathrm{OR}(x, u)$

  $\leftrightsquigarrow$ Small region of severe confounding?

This work:

- Sensitivity model that characterizes the **overall** strength of confounding
- Estimation and statistical inference on the bounds of treatment effects

This talk

- **A new sensitivity model**
- Sensitivity analysis: estimand under our model

  $\leftrightsquigarrow$ new class of distributionally robust optimization (DRO) problems
- Estimation and inference

  blessings from DRO $\rightsquigarrow$ more than doubly robust

The new f-sensitivity model

Use integral to measure average scale of selection bias

- Let f be any strongly convex function with $f(1) = 0$, and $\rho \geq 0$ any constant
- The (f, ρ) sensitivity model assumes for a.s. x,

$$\int f\big(\mathrm{OR}(x, U)\big)\, \mathrm{d}\mathbb{P}_{U \mid X=x, T=1} \leq \rho, \quad \int f\big(1/\mathrm{OR}(x, U)\big)\, \mathrm{d}\mathbb{P}_{U \mid X=x, T=0} \leq \rho$$

- ρ measures the **overall** deviation of $\mathrm{OR}(x, u)$ from 1

The new f-sensitivity model

Use integral to measure average scale of selection bias

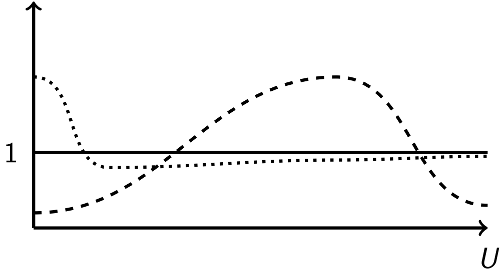
- Let f be any strongly convex function with $f(1) = 0$, and $\rho \geq 0$ any constant
- The (f, ρ) sensitivity model assumes for a.s. x,

$$\int f\big(\mathrm{OR}(x, U)\big) \, \mathrm{d}\mathbb{P}_{U \mid X=x, T=1} \leq \rho, \quad \int f\big(1/\mathrm{OR}(x, U)\big) \, \mathrm{d}\mathbb{P}_{U \mid X=x, T=0} \leq \rho$$

- Examples: KL-divergence $f(x) = -x \log x$, second moment bound $f(x) = (x-1)^2$...

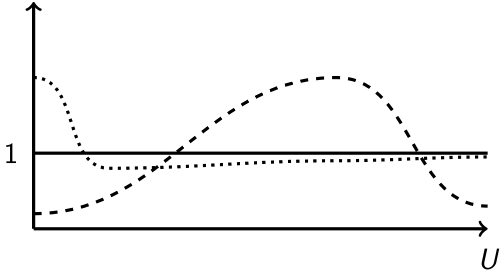
The new f-sensitivity model: interpretation

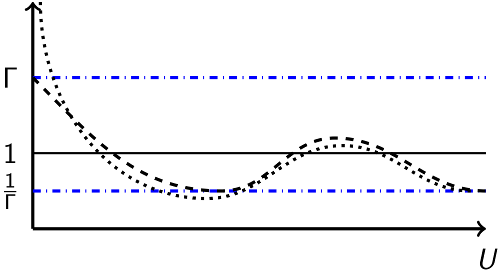
- First moment: $\mathbb{E}[\mathrm{OR}(x, U) \mid T = 1, X = x] \equiv 1$
- $\int f\big(\mathrm{OR}(x, U)\big)\, \mathrm{d}\mathbb{P}_{U \mid X=x, T=1}$ measures a "distance" between $\mathrm{OR}(x, U)$ and constant 1



The new *f*-sensitivity model: interpretation

- First moment: $\mathbb{E}[\mathrm{OR}(x, U) \mid T = 1, X = x] \equiv 1$
- $\int f\big(\mathrm{OR}(x, U)\big)\, \mathrm{d}\mathbb{P}_{U \mid X = x, T = 1}$ measures a "distance" between $\mathrm{OR}(x, U)$ and constant 1
  - Integral/expectation: scale + probability of such a scale
  - Uniform bound: only the scale





This talk

- A new sensitivity model
- **Sensitivity analysis: estimand under our model**

  $\leftrightsquigarrow$ new class of distributionally robust optimization (DRO) problems
- Estimation and inference

  blessings from DRO $\rightsquigarrow$ more than doubly robust

Causal inference as a counterfactual inference problem

- Observations are from either $\mathbb{P}_{X,Y(1)\,|\,T=1}$ or $\mathbb{P}_{X,Y(0)\,|\,T=0}$
- The essense is counterfactuals

$$(\text{ATC}) = \underbrace{\mathbb{E}[Y(1) \mid T = 0]}_{\text{counterfactual}} - \underbrace{\mathbb{E}[Y(0) \mid T = 0]}_{\text{observable}}$$

Range of counterfactual distribution

To infer $Y(1)$:

- the counterfactual distribution is $\mathbb{Q}_{X,Y} := \mathbb{P}_{X,Y(1)\,|\,T=0}$
- the observable distribution is $\mathbb{P}_{X,Y} := \mathbb{P}_{X,Y(1)\,|\,T=1}$

Under strong ignorability

It is a pure covariate shift:

$$\frac{\mathrm{d}\mathbb{Q}_{X,Y}}{\mathrm{d}\mathbb{P}_{X,Y}}(x,y) = \frac{1-e(x)}{e(x)}\frac{p}{1-p}$$

where $p = \mathbb{P}(T=1)$, and $e(x) = \mathbb{P}(T=1\,|\,X=x)$

Range of counterfactual distribution

To infer $Y(1)$:

- the counterfactual distribution is $\mathbb{Q}_{X,Y} := \mathbb{P}_{X,Y(1)\,|\,T=0}$
- the observable distribution is $\mathbb{P}_{X,Y} := \mathbb{P}_{X,Y(1)\,|\,T=1}$

With unmeasured confounding (not identifiable) [J. Ren, Zhou' 22]

Under (f,ρ)-selection condition,

$$\begin{aligned} \text{Covariate shift:} \quad & \frac{\mathrm{d}\mathbb{Q}_X}{\mathrm{d}\mathbb{P}_X}(x) = \frac{1-e(x)}{e(x)}\frac{p}{1-p}, \\ + \text{ bounded } Y\,|\,X \text{ shift:} \quad & D_f(\mathbb{Q}_{Y\,|\,X=x}\,\|\,\mathbb{P}_{Y\,|\,X=x}) \leq \rho, \quad \forall x \end{aligned}$$

where $D_f(Q\|P) = \mathbb{E}_P[f(\mathrm{d}Q/\mathrm{d}P)]$ is the f-divergence.

A distributionally robust optimization (DRO) perspective

- The range of the unknown target (counterfactual) distribution

$$\mathcal{Q} = \left\{ \mathbb{Q} \colon \frac{\mathrm{d}\mathbb{Q}_X}{\mathrm{d}\mathbb{P}_X}(x) = w(x),\ D_f\big(\mathbb{Q}_{Y \mid X} \,\|\, \mathbb{P}_{Y \mid X}\big) \ \leq \rho \right\},$$

- Partial identification bound is the optimal objective of a DRO problem:

$$\min_{\mathbb{P} \text{ satisfies } (f,\rho)} \mathbb{E}[Y(1) \mid T = 0] = \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[Y]$$

A distributionally robust optimization (DRO) perspective

- Sensitivity analysis defines a new class of DRO problems

$$\mathcal{Q} = \left\{ \mathbb{Q} \colon \frac{\mathrm{d}\mathbb{Q}_X}{\mathrm{d}\mathbb{P}_X}(x) = w(x),\ D_f\left(\mathbb{Q}_{Y|X} \,\|\, \mathbb{P}_{Y|X}\right) \le \rho \right\},$$

- Robust inference in the literature

$$\left\{ \mathbb{Q} \colon D_f\left(\mathbb{Q}_{X,Y} \,\|\, \mathbb{P}_{X,Y}\right) \le \rho \right\}$$

A distributionally robust optimization (DRO) perspective

- Sensitivity analysis defines a new class of DRO problems

$$\mathcal{Q} = \left\{ \mathbb{Q} \colon \frac{\mathrm{d}\mathbb{Q}_X}{\mathrm{d}\mathbb{P}_X}(x) = w(x),\ D_f\left(\mathbb{Q}_{Y|X} \,\|\, \mathbb{P}_{Y|X}\right) \leq \rho \right\},$$

- Robust inference in the literature

$$\left\{ \mathbb{Q} \colon D_f\left(\mathbb{Q}_{X,Y} \,\|\, \mathbb{P}_{X,Y}\right) \leq \rho \right\}$$

- New robust inference: good knowledge of $\mathbb{Q}_X$, expect $\mathbb{Q}_{Y|X}$ to be close to $\mathbb{P}_{Y|X}$
  - Counterfactual (causal) inference
  - Transfer learning, demographic information in census...
  - Other statistical inference / learning tasks under the new DRO model?

Dual of DRO is an ERM problem

- A known loss function $\ell(\cdot)$ and a estimable weight function $w(\cdot)$

ERM problem as dual of DRO (J. Ren, Zhou' 22)

The lower bound on $\mathbb{E}[Y(1) \mid T=0]$ under (f, ρ) sensitivity model equals

$$-\mathbb{E}\left[w(X) \cdot \ell\left(\alpha^*(X), \eta^*(X), X, Y(1)\right) \,\middle|\, T=1\right],$$

where $(\alpha^*(x), \eta^*(x)) = \operatorname{argmin}_{\alpha \geq 0, \eta \in \mathbb{R}} \mathbb{E}\left[\ell(\alpha, \eta, X, Y(1)) \,\middle|\, X=x, T=1\right]$ for all x.

Dual of DRO is an ERM problem

- A known loss function $\ell(\cdot)$ and a estimable weight function $w(\cdot)$

ERM problem as dual of DRO (J. Ren, Zhou' 22)

The lower bound on $\mathbb{E}[Y(1) \mid T=0]$ under (f,ρ) sensitivity model equals

$$-\mathbb{E}\left[w(X) \cdot \ell\left(\alpha^*(X), \eta^*(X), X, Y(1)\right) \,\middle|\, T=1\right],$$

where $(\alpha^*(x), \eta^*(x)) = \operatorname{argmin}_{\alpha \geq 0, \eta \in \mathbb{R}} \mathbb{E}\left[\ell(\alpha, \eta, X, Y(1)) \,\middle|\, X=x, T=1\right]$ for all x.

- Basic idea: plug in estimated quantities + bias correction

This talk

- A new sensitivity model
- Sensitivity analysis: estimand under our model

  $\leftrightsquigarrow$ new class of distributionally robust optimization (DRO) problems
- **Estimation and inference**

  blessings from DRO $\rightsquigarrow$ more than doubly robust

Estimating lower bound on $\mathbb{E}[Y(1) \mid T = 0]$

- Naive plug-in estimator:

$$-\widehat{\mathbb{E}}\left[\widehat{w}(X) \cdot \ell\left(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\right) \,\middle|\, T = 1\right]$$

Estimating lower bound on $\mathbb{E}[Y(1) \mid T = 0]$

- Naive plug-in estimator:

$$-\widehat{\mathbb{E}}\left[\widehat{w}(X) \cdot \ell\left(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\right) \,\middle|\, T = 1\right]$$
$$\approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\| + \|\widehat{\eta} - \eta^*\|$$

- Slow convergence in $\widehat{w}$, $\widehat{\alpha}$ and $\widehat{\eta}$ hinders root-n statistical inference

Estimating lower bound on $\mathbb{E}[Y(1) \mid T = 0]$

- Naive plug-in estimator:

$$
\begin{aligned}
&-\widehat{\mathbb{E}}\big[\widehat{w}(X) \cdot \ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big) \bigm| T = 1\big] \\
&\qquad \approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\| + \|\widehat{\eta} - \eta^*\|
\end{aligned}
$$

- Slow convergence in $\widehat{w}$, $\widehat{\alpha}$ and $\widehat{\eta}$ hinders root-n statistical inference

- Techniques from convex optimization + semiparametric stats

Estimating lower bound on $\mathbb{E}[Y(1) \mid T = 0]$

- Naive plug-in estimator:

$$-\widehat{\mathbb{E}}\left[\widehat{w}(X) \cdot \ell\left(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\right) \,\middle|\, T = 1\right]$$
$$\approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\| + \|\widehat{\eta} - \eta^*\|$$

- Slow convergence in $\widehat{w}$, $\widehat{\alpha}$ and $\widehat{\eta}$ hinders root-n statistical inference

- Techniques from convex optimization + semiparametric stats
  - Impact of $\widehat{\alpha}, \widehat{\eta}$ is second-order due to convexity & smoothness

$$\approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\|^2 + \|\widehat{\eta} - \eta^*\|^2$$

Estimating lower bound on $\mathbb{E}[Y(1) \mid T = 0]$

- Naive plug-in estimator:

$$
\begin{aligned}
-\widehat{\mathbb{E}}\big[\widehat{w}(X) \cdot \ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big) \mid T = 1\big] \\
\approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\| + \|\widehat{\eta} - \eta^*\|
\end{aligned}
$$

- Slow convergence in $\widehat{w}$, $\widehat{\alpha}$ and $\widehat{\eta}$ hinders root-n statistical inference

- Techniques from convex optimization + semiparametric stats
  - Impact of $\widehat{\alpha}, \widehat{\eta}$ is second-order due to convexity & smoothness

    $$\approx \text{target} + \|\widehat{w} - w\| + \|\widehat{\alpha} - \alpha^*\|^2 + \|\widehat{\eta} - \eta^*\|^2$$

  - Impact of $\widehat{w}$ can be made to be second-order using regression adjustment

    $$\approx \text{target} + \|\widehat{w} - w\| \cdot \|\text{regression error}\| + \|\widehat{\alpha} - \alpha^*\|^2 + \|\widehat{\eta} - \eta^*\|^2$$

The procedure for estimating $\mu^-_{1,0}$

1. Split the data into three disjoint folds: $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_3$
2. Estimate the covariate shift $\widehat{w}(\cdot)$ using $\mathcal{I}_1$
3. ERM to estimate $\widehat{\alpha}(\cdot), \widehat{\eta}(\cdot)$ for $\alpha^*(\cdot), \eta^*(\cdot)$ using $\mathcal{I}_1$
4. Debias: cond. regression of $\widehat{H}(X, Y(1)) := \ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big)$ on X using $\mathcal{I}_2$
5. Plug in estimation, and cross-fit (switching roles of three folds)

The procedure for estimating $\mu^-_{1,0}$

1. Split the data into three disjoint folds: $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_3$
2. Estimate the covariate shift $\widehat{w}(\cdot)$ using $\mathcal{I}_1$
3. ERM to estimate $\widehat{\alpha}(\cdot), \widehat{\eta}(\cdot)$ for $\alpha^*(\cdot), \eta^*(\cdot)$ using $\mathcal{I}_1$
4. Debias: cond. regression of $\widehat{H}(X, Y(1)) := \ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big)$ on X using $\mathcal{I}_2$
5. Plug in estimation, and cross-fit (switching roles of three folds)

$$\widehat{\mu}^{(j)}_{1,0} = \frac{1}{|\mathcal{I}^{(j)}_1|} \sum_{i \in \mathcal{I}^{(j)}_1} \underbrace{\widehat{w}^{(j)}(X_i)}_{\text{reweight}} \underbrace{\left(\widehat{H}^{(j)}(X_i, Y_i) - \widehat{h}^{(j)}(X_i)\right)}_{\text{debias}} + \frac{1}{|\mathcal{I}^{(j)}_0|} \sum_{i \in \mathcal{I}^{(j)}_0} \underbrace{\widehat{h}^{(j)}(X_i)}_{\text{debias}}.$$

Subroutine: Sieve estimation for ERM

- Obtaining $(\widehat{\alpha}(\cdot), \widehat{\eta}(\cdot))$: optimize over a function class
- Example: sieve estimator (polynomials, splines...)
  - J-th order polynomials on $[0, 1]$:
$$\mathrm{Pol}(J, \epsilon) = \left\{x \mapsto \sum_{k=0}^{J} a_k x^k : a_k \in \mathbb{R}\right\},$$
  - r-th order splines with J knots
$$\mathrm{Spl}(r, J) = \left\{x \mapsto \sum_{k=0}^{r-1} a_k x^k + \sum_{j=1}^{J} b_j (x - t_j)_+^{r-1} : a_k, b_k \in \mathbb{R}\right\}$$
- Faster than $o(n^{-1/4})$ under proper smoothness conditions

Estimation consistency

Three parts of estimation:

(1) Estimate covariate shift $\widehat{w}$

(2) ERM: fit $\widehat{\alpha}(\cdot), \widehat{\eta}(\cdot)$ for $\alpha^*(\cdot), \eta^*(\cdot)$

(3) Conditional regression of $\ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big)$

- **Double robustness**: If (2) is consistent, our estimator is consistent if either (1) or (3) is consistent
- **One-side validity**: If (2) is inconsistent, our estimator is valid but conservative if either (1) or (3) is consistent

Statistical inference

Three parts of estimation:

(1) Estimate covariate shift $\widehat{w}$

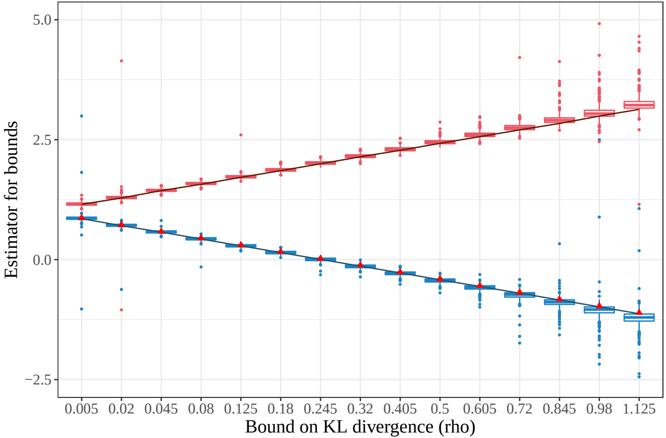
(2) ERM: fit $\widehat{\alpha}(\cdot), \widehat{\eta}(\cdot)$ for $\alpha^*(\cdot), \eta^*(\cdot)$

(3) Conditional regression of $\ell\big(\widehat{\alpha}(X), \widehat{\eta}(X), X, Y(1)\big)$

- **Double robustness**: If (2) is $n^{-1/4}$-consistent, then our estimator is asymptotically normal $\sqrt{n}(\widehat{\theta} - \theta^*) \xrightarrow{d} N(0, \sigma^2)$ if the product of errors in (1) and (3) is $o(n^{-1/2})$
- **One-side validity**: If (2) is inconsistent, then our estimator has valid but conservative inference if the product of errors in (1) and (3) is $o(n^{-1/2})$

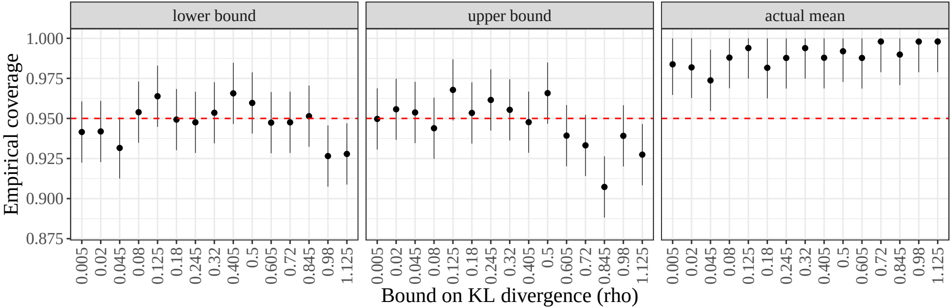
Simulations: validity and sharpness

- Simulate a confounded dataset for $\rho \in \{0.1, 0.2, \ldots, 1.5\}$ and f for KL-divergence
- Apply our method at the true ρ, repeat $N = 500$ runs
- Sieve (cubic spline) for ERM, random forest for regression



Simulations: validity and sharpness

- Coverage of confidence intervals



Summary

- **New sensitivity model** on average selection bias
- **New pespective** to sensitivity analysis from DRO
- **New class of DRO problems**: known X-shift, bounded $Y \mid X$-shift (Jin et al. 2021)
- **New DRO techniques and guarantees**
  - Doubly robust inference by 'adjusting with another group' (Jin and Rothenhäusler 2021)
  - 'Wrong but valid' guarantee for partial identification (Dorn et al. 2021)

Thanks!

More details in the manuscript: `https://arxiv.org/abs/2203.04373`

References

J. Dorn and K. Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *arXiv preprint arXiv:2102.04543*, 2021.

J. Dorn, K. Guo, and N. Kallus. Doubly-valid/doubly-sharp sensitivity analysis for causal inference with unmeasured confounding. *arXiv preprint arXiv:2112.11449*, 2021.

A. Franks, A. D'Amour, and A. Feller. Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 2019.

G. W. Imbens. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2): 126–132, 2003.

Y. Jin and D. Rothenhäusler. One estimator, many estimands: fine-grained quantification of uncertainty using conditional inference. *arXiv preprint arXiv:2104.04565*, 2021.

Y. Jin, Z. Ren, and E. J. Candès. Sensitivity analysis of individual treatment effects: A robust conformal inference approach. *arXiv preprint arXiv:2111.12161*, 2021.

P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218, 1983.

Z. Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

Q. Zhao, D. S. Small, and B. B. Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *arXiv preprint arXiv:1711.11286*, 2017.

Distributionally robust optimization problem

Denote $w(x) = \frac{1-e(x)}{e(x)} \frac{p}{1-p}$,

Proposition (J. Ren, Zhou' 22)

Let $\mu^-_{1,0}$ (resp. $\mu^+_{1,0}$) be the optimal objective function of the convex optimization problem

$$\min_{L(x)\text{ measurable}}(\text{resp. } \max)\ \mathbb{E}\big[Y(1)L(X)\,\big|\,T=1\big]$$

$$\text{s.t. } \mathbb{E}[L(x)\,|\,X=x, T=1] = w(x)$$

$$\mathbb{E}\big[f\big(L(x)/w(x)\big)\,\big|\,X=x, T=1\big] \le \rho, \quad \text{for almost all } x,$$

where all the expectations are induced by the observed distribution. Then $\mu^-_{1,0} \le \mathbb{E}[Y(1)\,|\,T=0] \le \mu^+_{1,0}$ under the (f,ρ)-selection condition.

Dual problem of DRO

Proposition (J. Ren, Zhou' 22)

The optimal objective of the previous DRO problem is given by

$$\mu_{1,0}^{-} = -\inf_{\alpha(X)\geq 0, \eta(X)\in\mathbb{R}} \mathbb{E}\left[w(X)\left\{ \alpha(X) f^{*}\left(\frac{Y(1)+\eta(X)}{-\alpha(X)} \right) + \eta(X) + \alpha(X)\rho \right\} \,\middle|\, T=1 \right],$$

where $f^{*}(s) = \sup_{t\geq 0}\{st - f(t)\}$ is the conjugate function of f. In particular, denoting $\ell(\alpha,\eta,x,y) = \alpha f^{*}(\frac{y+\eta}{-\alpha}) + \eta + \alpha\rho$ for $(\alpha,\eta)\in\mathbb{R}^{+}\times\mathbb{R}$, we have $\mu_{1,0}^{-} = -\mathbb{E}\left[\ell(\alpha^{*}(X), \eta^{*}(X), X, Y(1)) \,\middle|\, T=1\right]$, where for $\mathbb{P}_{X|\,T=1}$-almost all x,

$$(\alpha^{*}(x), \eta^{*}(x)) \in \operatorname{argmin} \alpha \geq 0, \eta\in\mathbb{R}\ \mathbb{E}\left[\alpha f^{*}\left(\frac{Y(1)+\eta}{-\alpha} \right) + \eta + \alpha\rho \,\middle|\, X=x, T=1 \right].$$

Estimation of the lower bound

Define quantities:

- true conditional mean $\bar{h}$ for step 3 (Debiasing), $\theta^\diamond$ the limit of $(\widehat{\alpha}, \widehat{\eta})$ in step 2 (ERM)
- μ^- for the true lower bound, $\widehat{\mu}^-$ for our estimator

Theorem (Informal, J. Ren, Zhou' 22)

Under regularity conditions, suppose either (i) $\|\widehat{w} - w\|_{L_2(\mathbb{P}_{X \mid T=1})} = o_P(1)$ *or (ii)* $\|\widehat{h} - \bar{h}\|_{L_2(\mathbb{P}_{X \mid T=1})} = o_P(1)$*. Then*

- *if* $\theta^\diamond = \theta^*$*, i.e., the ERM step is consistent, then* $\widehat{\mu}^- = \mu^- + o_P(1)$*;*
- *otherwise,* $\widehat{\mu}^- = \mu^\diamond + o_P(1)$ *for some constant* $\mu^\diamond \leq \mu^-$*.*

CLT-type inference for the lower bound

Define quantities:

- true conditional mean $\bar{h}$ for step 3 (Debiasing), $\theta^\diamond$ the limit of $(\widehat{\alpha}, \widehat{\eta})$ in step 2 (ERM)
- μ^- for the true lower bound, $\widehat{\mu}^-$ for our estimator

Theorem (Informal, J. Ren, Zhou' 22)

Under regularity cnditions, suppose $\|\widehat{w} - w\|_{L_2(\mathbb{P}_{X \mid T=1})} \cdot \|\widehat{h} - \bar{h}\|_{L_2(\mathbb{P}_{X \mid T=1})} = o_P(n^{-1/2})$, *and* $\|(\widehat{\alpha} - \alpha^*, \widehat{\eta} - \eta^*)\|_{L_2(\mathbb{P}_{X \mid T=1})} = o_P(n^{-1/4})$ *for some optimizer* $(\alpha^*(x), \eta^*(x))$. *Then* $\sqrt{n}(\widehat{\mu}^-_{1,0} - \mu^-_{1,0}) \rightsquigarrow N(0, \mathrm{Var}(\phi_{1,-}(X, Y, T)))$, *where*

$$\phi_{1,-}(X_i, Y_i, T_i) = \frac{T_i}{p_1} w(X_i)\big[H(X_i, Y_i(1)) - h(X_i)\big] + \frac{1 - T_i}{p_0} h(X_i).$$

Here $p_1 = \mathbb{P}(T = 1) = 1 - p_0$, $H(x, y) = \ell(\theta^*, x, y)$, $h(x) = \mathbb{E}\big[H(X, Y(1)) \,\big|\, X = x, T = 1\big]$. *All the expectations (variances) are induced by the observed distribution.*