Incremental effects for continuous exposures

Kyle Schindl, Shuying Shen, Edward Kennedy

Department of Statistics & Data Science
Carnegie Mellon University

JSM, 8 Aug 2024

Punchline

Continuous treatments bring substantial challenges

- one solution: incremental effects based on *exponential tilts*, which describe what happens if we shift distribution up/down
- but... unknown how estimation error depends on shift

The main goals of this work are:

1. new minimax lower bound
2. new bias/convergence analysis
3. new methods for dose-response estimation

... all for arbitrary shifts, showing explicit dependence on shift

Continuous treatments

Continuous treatments/exposures are very common in practice

- e.g., measures of dose, duration, frequency
- natural to consider **dose-response curve** $\mathbb{E}(Y^a)$

But this brings substantial challenges vs. binary treatment

- positivity violations more common/extreme
- slower than $\sqrt{n}$ rates, even w/positivity

Lots of work recasting causal effects, w/stochastic interventions

- e.g., what if distribution of treatment changed?
- additive shift $\mathbb{E}(Y^{A+\delta})$ (Diaz/vdL, Haneuse/Rotnitsky)
- other general stochastic effects (Taubman et al, Young et al)

Incremental effects

Incremental effects are another kind of stochastic intervention

- first proposed for binary treatments (Kennedy 2019)
- *what if odds of treatment were multiplied by δ?*
- benefits: (1) no positivity assumption required, (2) gives natural & smooth interpolation between $\mathbb{E}(Y^0) \longleftrightarrow \mathbb{E}(Y^1)$

Q: **How do we extend to continuous treatments?**

- ideally want to keep beneficial properties (1) and (2)....

Diaz & Hejazi (2020): incremental effects are an exponential tilt!

- these tilts have a long history (Esscher 1932, Seigmund 1976)
- their idea: use exponential tilt with treatment *density*

Continuous incremental effects

Diaz & Hejazi (2020): what if trt density $\pi(a \mid x)$ were tilted by δ?

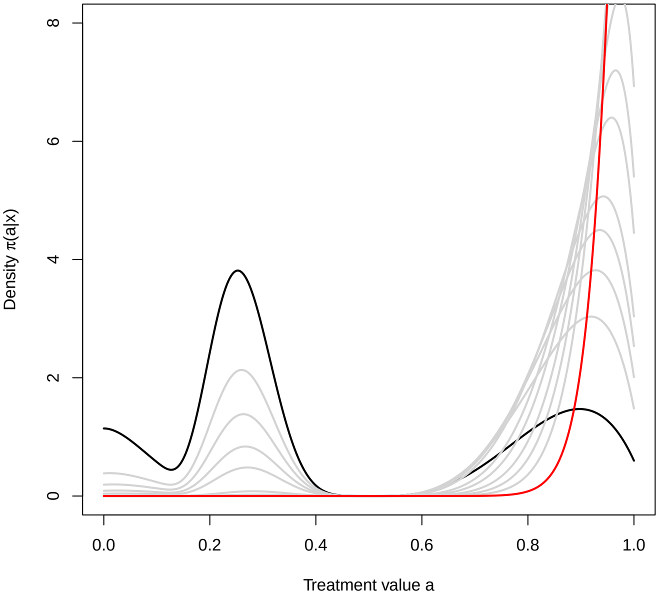
$$\pi(a \mid x) \quad \rightarrow \quad q_\delta(a \mid x) = \frac{\exp(\delta a)\pi(a \mid x)}{\int \exp(\delta t)\pi(t \mid x)\; dt}$$

Note: $q_\delta(a \mid x) = 0$ if $\pi(a \mid x) = 0 \implies$ don't need positivity

Under *only consistency + exchangeability*:

$$\mathbb{E}(Y^{Q(\delta)}) = \int \int \mu(x,a) q_\delta(a \mid x)\; da\; d\mathbb{P}(x) \equiv \psi(\delta)$$

for outcome regression $\mu(x,a) = \mathbb{E}(Y \mid X = x, A = a)$


Density π(a|x)
Treatment value a
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0

More on interpretation

If $A \in \{0, 1\}$ is binary, then $\exp(\delta)$ is an *odds ratio*

- i.e., δ is a change in log-likelihood ratios:

$$\delta = \log \frac{q_\delta(1 \mid x)/q_\delta(0 \mid x)}{\pi(1 \mid x)/\pi(0 \mid x)} = \log \frac{q_\delta(1 \mid x)}{\pi(1 \mid x)} - \log \frac{q_\delta(0 \mid x)}{\pi(0 \mid x)}$$

If A is continuous, δ is the *derivative* of the log-likelihood ratio:

$$\delta = \frac{\partial}{\partial a} \log \left\{ \frac{q_\delta(a \mid x)}{\pi(a \mid x)} \right\}$$

Interpretation is more challenging than in binary setting

- can estimate functionals of $q_\delta(a \mid x)$ to understand

Estimation

Diaz & Hejazi (2020) propose:

$$\widehat{\psi}(\delta) = \mathbb{P}_n\left[\left\{ \frac{\widehat{q}_\delta(A \mid X)}{\widehat{\pi}(A \mid X)} \right\} \left\{ Y - \int \widehat{\mu}(X, a) q_\delta(a \mid X) \; da \right\} + \int \widehat{\mu}(X, a) \widehat{q}_\delta(a \mid X) \; da \right] \equiv \mathbb{P}_n\left\{ \widehat{\varphi}_\delta(Z) \right\}$$

- *one-step / doubly robust / double ML estimator*
- derived from efficient influence function φ_δ

Note the weights have an issue:

$$\frac{q_\delta(A \mid X)}{\pi(A \mid X)} = \frac{\exp(\delta A)}{\int \exp(\delta t) \pi(t \mid X) \; dt} \leq \exp\left\{ \delta(A - \mathbb{E}(A \mid X)) \right\}$$

could be as large as exponential in δ for some A

Dependence on δ

Diaz & Hejazi (2020) apply usual bias analysis

- characterize remainder/bias, apply Cauchy-Schwarz

However: when δ is large, these bounds can explode

- both bias & variance *depend on δ in complicated way*

Why do we care about clarifying dependence on δ ?

- may look at multiple $\psi(\delta)$ curves across varying δ ranges, and pick a range informed by sample size
- taking $\delta \to \infty$ should recover estimate of dose-response, which is interesting in its own right
- fixed δ asymptotics can be misleading, even for modest δ (e.g., say $\delta = \log n \implies \log 1k \approx 6.9$, $\log 10k \approx 9.2$)

Semiparametric efficiency bound – sandwiched by δ

Recall: variance of influence function = efficiency bound

Theorem

Assume:

1. *bounded density:* $\pi_{\min} \leq \pi(a \mid x) \leq \pi_{\max}$ *for all* a, x
2. *some noise:* $\sigma^2_{\min} \leq var(Y \mid X = x, A = a)$ *for all* a, x
3. *bounded outcomes:* $|Y| \leq B$ *with probability one*

Then

$$C_0 \, \delta \;\leq\; var\Big\{\varphi_\delta(Z)\Big\} \;\leq\; C_1 \, \delta$$

for constants C_0, C_1 *and* δ *large enough.*

$\rightarrow$ efficiency bound increases with δ, goes to infinity if δ does

About positivity

Why are we assuming bounded density?

- aren't positivity violations a big motivation?

Yes. But:

- minimax lower bounds for bounded density models are also lower bounds in larger models with unbounded densities
- results still hold if densities are only upper bounded, & have holes in support (see arxiv soon...)

And, *something weird happens* with incremental effects:

- for ATEs, the more positivity/overlap, the better
- for incremental effects, if no overlap then $\psi(\delta) = \mathbb{E}(Y)$
- so we need *some* overlap for this to be interesting

Minimax rates – no faster than $1/\sqrt{n/\delta}$

Q: What is best possible RMSE in a nonparametric model?

Theorem

Let $\mathcal{P}$ denote the model where:

1. *bounded density:* $\pi(a \mid x) \leq \pi_{\max}$ *for all* a, x
2. *some noise:* $\sigma^2_{\min} \leq var(Y \mid X = x, A = a)$ *for all* a, x
3. *bounded outcomes:* $|Y| \leq B$ *with probability one*

Then

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P |\widehat{\psi} - \psi_P(\delta)| \;\geq\; \sqrt{\frac{C}{n/\delta}}$$

So: for $\psi(\delta)$ no estimator can have RMSE better than $1/\sqrt{n/\delta}$

- proof: Le Cam's two-point method

Bias not affected by δ – but need new analysis

Have seen variance is affected by δ - how about bias?

Theorem

Assume boundedness of (π, Y) as before, as well as of $(\widehat{\pi}, \widehat{\mu})$. Then the bias is bounded as

$$\mathbb{E}\left\{\widehat{\psi}(\delta) - \psi(\delta)\right\} \lesssim \sup_a \|\widehat{\pi}_a - \pi_a\| \sup_a \|\widehat{\mu}_a - \mu_a\| + \sup_a \|\widehat{\pi}_a - \pi_a\|^2$$

where

$$\sup_a \|f_a\|^2 \equiv \sup_a \int f(x,a)^2 \, d\mathbb{P}(x)$$

is a mixed $L_2(\mathbb{P})$/sup norm.

→ doubly robust-style second-order error
→ mixed norm is essential for avoiding δ dependence

Asymptotic normality – need right scaling & bias analysis

For inference, do we have to repair estimator with large δ?

Theorem

Assume boundedness from before, and

$$\sup_a \|\widehat{\pi}_a - \pi_a\| \sup_a \|\widehat{\mu}_a - \mu_a\| + \sup_a \|\widehat{\pi}_a - \pi_a\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n/\delta}}\right).$$

Then

$$\frac{\widehat{\psi}(\delta) - \psi(\delta)}{\sqrt{var(\varphi_\delta)/n}} \rightsquigarrow N(0,1).$$

So usual estimator *w/proper scaling* is still asymptotically normal

- but with $\sqrt{n/\delta}$ rate, instead of $\sqrt{n}$

Dose-response estimation

How about $\widehat{\psi}(\delta)$ as an estimator of dose-response $\mathbb{E}(Y^1)$?

Theorem

Assume boundedness of $(\pi, \widehat{\pi}, Y, \widehat{\mu})$ as before, and:

1. $\sup_a \|\widehat{\pi}_a - \pi_a\| \sup_a \|\widehat{\mu}_a - \mu_a\| = o_{\mathbb{P}}(1/\sqrt{n/\delta})$ *and* $\sup_a \|\widehat{\pi}_a - \pi_a\|^2 = o_{\mathbb{P}}(1/\sqrt{n/\delta})$
2. *Lipschitz:* $|\mu(x,a) - \mu(x,a')| \leq L|a - a'|$ *for any* x.

Then for $\psi(\infty) = \mathbb{E}(Y^1)$ *we have*

$$|\widehat{\psi}(\delta) - \psi(\infty)| \; = \; O_{\mathbb{P}}\left(\frac{1}{\delta} + \sqrt{\frac{1}{n/\delta}}\right).$$

Note if $\delta \sim n^{1/3}$, then $\widehat{\psi}(\delta)$ converges to $\mathbb{E}(Y^1)$ at $n^{-1/3}$ rates

- optimal rate for 1-d Lipschitz functions

Summary

Our contributions:

1. new minimax lower bound
2. new bias/convergence analysis
3. new methods for dose-response estimation

... all for arbitrary δ, even $\delta \to \infty$, showing explicit dependence

Ongoing/future work:

- weakening positivity
- time-varying treatments
- lots left to explore

Will be on arxiv soon