Inference on Winners

Isaiah Andrews[1] Toru Kitagawa[2] Adam McCloskey[3]

[1]Harvard University

[2]Brown University

[3]University of Colorado

October 26, 2022

Inference on Winners

- We study inference on the best-performing treatment in an experiment
  - Data-driven choice of target parameter leads to bias and undercoverage for conventional estimators and confidence sets, respectively
  - Previously noted by many others, e.g. Lee and Shen (2018)
- To illustrate, consider a stylized example

Setup

- Consider a researcher who runs a randomized trial and estimates average outcomes under two treatments, $\theta \in \Theta = \{\theta_1, \theta_2\}$

$$\begin{pmatrix} X(\theta_1) \\ X(\theta_2) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\theta_1) \\ \mu(\theta_2) \end{pmatrix}, I_2 \right),$$

  where $\mu(\theta)$ is population average outcome under θ

- Which treatment to recommend? Natural to maximize observed outcomes:

$$\hat{\theta} = \arg\max_{\theta} X(\theta)$$

- Along with recommendation, want assessment of effectiveness: estimates and confidence sets for $\mu\left(\hat{\theta}\right)$

Winner's Curse

- $\hat{\theta} = \theta_1 \Rightarrow X(\theta_1) \geq X(\theta_2)$
- Distribution of $X(\theta_1)$ conditional on $\hat{\theta} = \theta_1$ is truncated below, and

$$Pr\left\{X(\theta_1) > \mu(\theta_1) | \hat{\theta} = \theta_1\right\} > \frac{1}{2}$$

- Since same true for θ_2, holds unconditionally as well

$$P\left\{X(\hat{\theta}) > \mu(\hat{\theta})\right\} > \frac{1}{2}$$

- Hence, $X(\hat{\theta})$ is upwards median biased as an estimator of $\mu(\hat{\theta})$. Confidence set

$$\left[X(\hat{\theta}) - 1.96, X(\hat{\theta}) + 1.96\right]$$

  may undercover

Simulation Designs

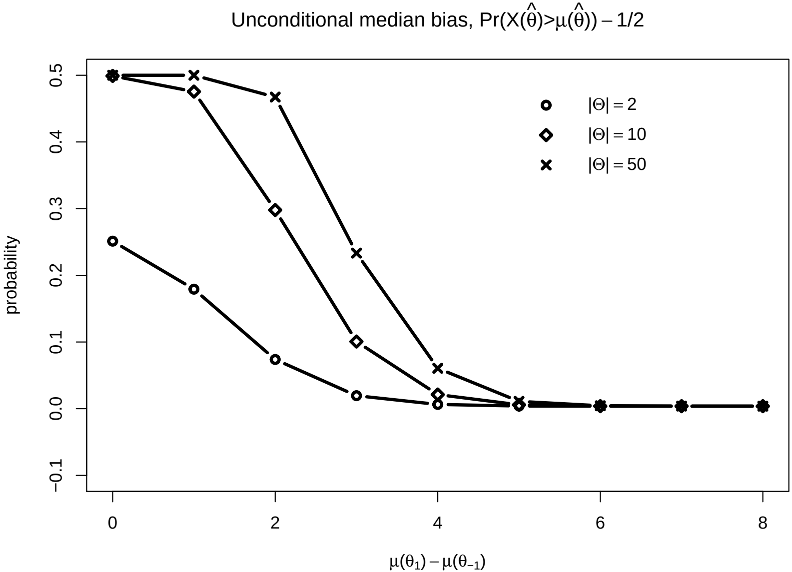
- Consider cases with $|\Theta| = 2, 10, 50$ treatments
  - Still assume identity covariance matrix
- Consider $\mu(\theta_1) \geq \mu(\theta_2) = ... = \mu(\theta_{-1})$
  - First treatment (weakly) most effective
- Examine performance of conventional estimator

$$\hat{\mu} = X(\hat{\theta})$$

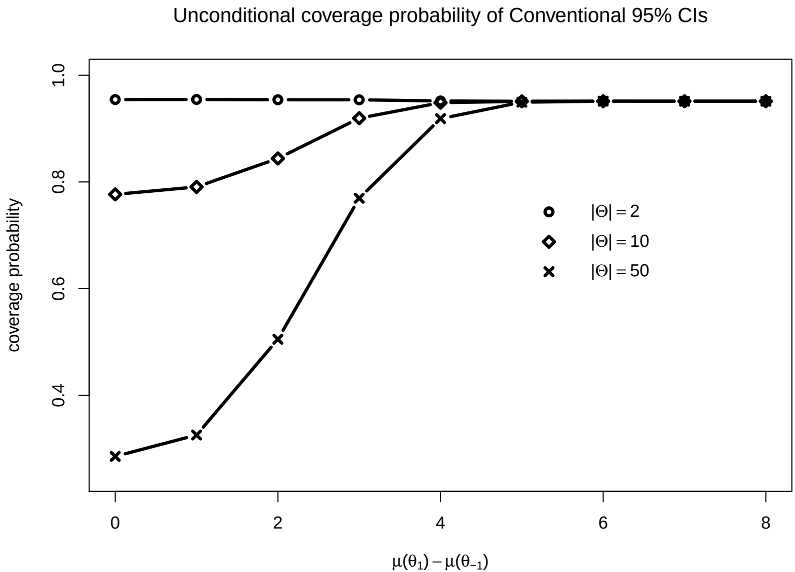
and conventional ("naive") confidence set

$$\left[X(\hat{\theta}) - 1.96, X(\hat{\theta}) + 1.96\right]$$

Winner's Curse



Winner's Curse



Stylized Example: Conditional Inference

Two possible goals for corrected inference

- First: Conditional Inference
  - Want procedures to be valid conditional on $\hat{\theta}$
  - Requires validity conditional on the recommendation made
- For confidence sets, conditional coverage

$$Pr_{\mu}\left\{\mu(\hat{\theta}) \in CS | \hat{\theta} = \tilde{\theta}\right\} \geq 1 - \alpha \text{ for all } \tilde{\theta} \in \Theta, \mu$$

- For estimators, conditional median unbiasedness

$$Pr_{\mu}\left\{\hat{\mu} > \mu(\hat{\theta}) | \hat{\theta} = \tilde{\theta}\right\} = \frac{1}{2} \text{ for all } \tilde{\theta} \in \Theta, \mu$$

Unconditional Inference

- Second: Unconditional Inference
  - Require validity only on average across values of $\hat{\theta}$
  - Valid on average, but not conditional on recommendation
- For confidence sets, unconditional coverage

$$Pr_{\mu}\left\{\mu(\hat{\theta}) \in CS\right\} \geq 1-\alpha \text{ for all } \mu$$

- For estimators, unconditional median unbiasedness

$$Pr_{\mu}\left\{\hat{\mu} > \mu(\hat{\theta})\right\} = \frac{1}{2} \text{ for all } \mu$$

- Less demanding than conditional inference
  - Any valid conditional procedure is also valid unconditionally
  - Class of unconditional procedures is larger. May allow (unconditional) performance improvements

Conditional vs. Unconditional

Why would we want to impose conditional validity?

- Suppose θ_1 is a new treatment, and θ_2 is control
  - Baseline is the control
- If impose only unconditional validity, and $Pr\left\{\hat{\theta} = \theta_1\right\}$ is small, may have

$$Pr\left\{\mu(\hat{\theta}) \in CS | \hat{\theta} = \theta_1\right\} \ll 1 - \alpha$$

  so confidence set is too optimistic when recommend new treatment
  - If implement, results may lie below CS with high probability
- Conditional vs. unconditional validity: does this bother us?
  - Yes: want coverage conditional on recommendation
    $\Rightarrow$ impose conditional validity
  - No: only care about performance on average across cases where do and don't recommend the new treatment
    $\Rightarrow$ impose unconditional validity

Conditional Inference Results

- Consider inference on $\mu(\hat{\theta})$ conditional on $\hat{\theta} = \theta_1$
- Conditional distribution of X multivariate truncated normal
  - Exponential family $\Rightarrow$ optimal median-unbiased estimator $\hat{\mu}_{\frac{1}{2}}$, equal tailed confidence set CS_{ET}
  - "Equal tailed" in sense that equally likely to over- and under-estimate $\mu(\hat{\theta})$

Unconditional Inference Results

Two options:

1. Conditional confidence set CS_{ET}
2. Projection confidence set (existing approach)

$$CS_P = [X(\hat{\theta}) - c_{1-\alpha}, X(\hat{\theta}) + c_{1-\alpha}],$$

for $c_{1-\alpha}$ the $(1-\alpha)$ quantile of $\max\{|Z_1|, |Z_2|\}$, $Z \sim N(0, I_2)$

- Forms a rectangular joint confidence for $(\mu(\theta_1), \mu(\theta_2))$ and projects to obtain confidence set for $\mu(\hat{\theta})$

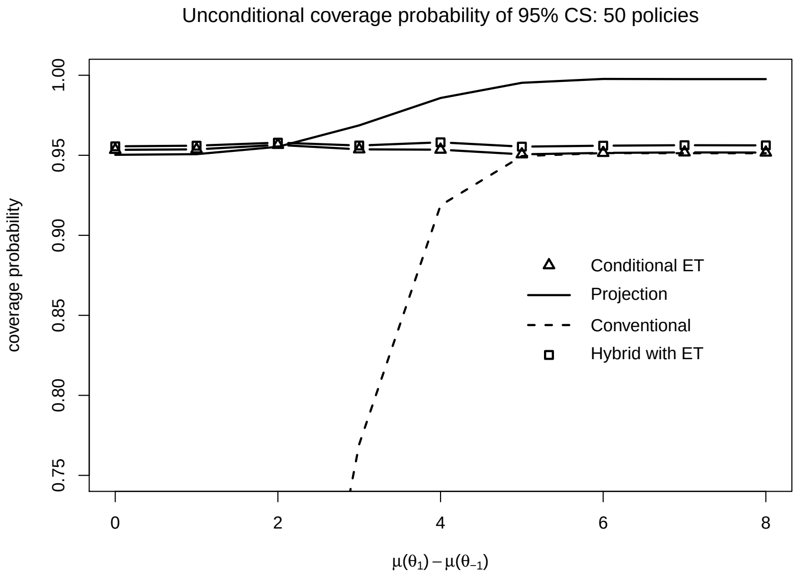
Neither confidence set fully satisfactory:

- Conditional confidence sets performs well when $\mu(\theta_1) \gg \mu(\theta_2)$ or the reverse, poorly when $\mu(\theta_1) \approx \mu(\theta_2)$
- CS_P performs reasonably well when $\mu(\theta_1) \approx \mu(\theta_2)$, but unnecessarily wide when $\mu(\theta_1) \gg \mu(\theta_2)$

Unconditional Inference Results

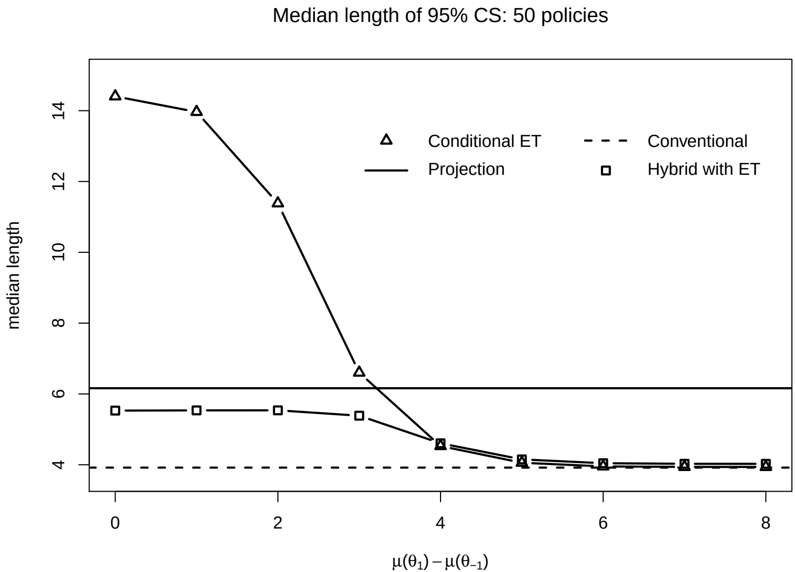
- If we require conditional coverage, conditional procedures optimal so little scope for improvement
- If we only care about unconditional coverage, propose Hybrid confidence set CS^H_{ET}, and corresponding estimator
  - Combine conditioning and projection (details in paper)
  - Allows substantial (unconditional) performance improvements

Coverage



Unconditional coverage probability of 95% CS: 50 policies
coverage probability
1.00
0.95
0.90
0.85
0.80
0.75
0
2
4
6
8
μ(θ1) − μ(θ−1)
Conditional ET
Projection
Conventional
Hybrid with ET

Median Length



Wrapping Up

Takeaways:

- Inference on the best-performing treatment invalidates conventional inference
- We develop optimal inference procedures that are valid conditional on treatment selected
- If satisfied with unconditional validity, we propose Hybrid inference procedures with better performance

In the paper we:

- Extend our results to more general settings (general correlation structures, asymptotic results)
- Show how similar ideas dominate sample-splitting
- Illustrate with applications

The End

Thanks very much!