Value Function Inference

A Mildly Technical Introduction

John Sperger

March 8th 2024

Outline

Learning Objectives

Learning Objectives

At the end of today's talk you should be able to:

- Identify the fundamental challenge facing value function inference that applies to all temporal settings.

Learning Objectives

At the end of today's talk you should be able to:

- Identify the fundamental challenge facing value function inference that applies to all temporal settings.

- Categorize the major approaches to inference in this setting

Learning Objectives

At the end of today's talk you should be able to:

- Identify the fundamental challenge facing value function inference that applies to all temporal settings.
- Categorize the major approaches to inference in this setting
- Provide reasons when projection intervals may be preferable to bootstrap intervals and vice versa.

Learning Objectives

At the end of today's talk you should be able to:

- Identify the fundamental challenge facing value function inference that applies to all temporal settings.
- Categorize the major approaches to inference in this setting
- Provide reasons when projection intervals may be preferable to bootstrap intervals and vice versa.
- Explain the connection between the finite-sample approach and the asymptotic approaches

Notation

Let $[K]$ denote the set $\{1, \ldots, K\}$ for a positive integer K. Assume the data is comprised of iid replicates:

$$\{X_t, A_t, Y_t\}_{t=1}^{T}$$

$X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ denotes the covariates (contexts)
$A_t \in \mathcal{A}$ denotes the treatment arm (arm, intervention, action), and
$Y_t \in \mathbb{R}$ denotes the response (reward)

Notation

Let $[K]$ denote the set $\{1, \ldots, K\}$ for a positive integer K. Assume the data is comprised of iid replicates:

$$\{X_t, A_t, Y_t\}_{t=1}^{T}$$

$X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ denotes the covariates (contexts)
$A_t \in \mathcal{A}$ denotes the treatment arm (arm, intervention, action), and
$Y_t \in \mathbb{R}$ denotes the response (reward)

We'll use the potential outcome framework and denote the potential outcome $Y^\star(a)$

Notation

Let $[K]$ denote the set $\{1, \ldots, K\}$ for a positive integer K. Assume the data is comprised of iid replicates:

$$\{X_t, A_t, Y_t\}_{t=1}^T$$

$X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ denotes the covariates (contexts)
$A_t \in \mathcal{A}$ denotes the treatment arm (arm, intervention, action), and
$Y_t \in \mathbb{R}$ denotes the response (reward)

We'll use the potential outcome framework and denote the potential outcome $Y^\star(a)$

I'll use the index n when discussing asymptotics and t when discussing finite samples. I'll try to maintain consistency but might slip up.

Operator notation

- $X_1, \dots, X_T$ is an iid random sample from a fixed but unknown distribution P
- g is a generic parametric function indexed by $\theta \in \Theta$
- $\hat{\theta} \in \Theta$ is a random variable constructed from the sample $X_1, \dots, X_n$

Operator notation

- $X_1, \ldots, X_T$ is an iid random sample from a fixed but unknown distribution P
- g is a generic parametric function indexed by $\theta \in \Theta$
- $\hat{\theta} \in \Theta$ is a random variable constructed from the sample $X_1, \ldots, X_n$

P denotes the probability measure:

$$\mathrm{P}\, g(X; \hat{\theta}) = \int g(x; \widehat{\theta}) \,\mathrm{d}\,\mathrm{P}(x))$$

Operator notation

- $X_1, \dots, X_T$ is an iid random sample from a fixed but unknown distribution P
- g is a generic parametric function indexed by $\theta \in \Theta$
- $\hat{\theta} \in \Theta$ is a random variable constructed from the sample $X_1, \dots, X_n$

P denotes the probability measure:

$$\mathrm{P}\, g(X; \hat{\theta}) = \int g(x; \widehat{\theta})\, \mathrm{d}\, \mathrm{P}(x))$$

$\mathbb{P}_\mathrm{n}$ denotes the corresponding empirical measure:

$$\mathbb{P}_\mathrm{n}\, g(X; \hat{\theta}) = n^{-1} \sum_{i=1}^{n} g(x_i; \hat{\theta})$$

Operator notation

- $X_1, \dots, X_T$ is an iid random sample from a fixed but unknown distribution P
- g is a generic parametric function indexed by $\theta \in \Theta$
- $\hat{\theta} \in \Theta$ is a random variable constructed from the sample $X_1, \dots, X_n$

P denotes the probability measure:

$$\mathrm{P}\, g(X; \hat{\theta}) = \int g(x; \hat{\theta})\, \mathrm{d}\, \mathrm{P}(x))$$

$\mathbb{P}_\mathrm{n}$ denotes the corresponding empirical measure:

$$\mathbb{P}_\mathrm{n}\, g(X; \hat{\theta}) = n^{-1} \sum_{i=1}^{n} g(x_i; \hat{\theta})$$

$\rightsquigarrow$ denotes convergence in distribution.

Table of Contents

Nonregularity

Nonregular is the catch-all for when standard regularity conditions do not hold.

Nonregularity

Nonregular is the catch-all for when standard regularity conditions do not hold.

The general problem we'll be addressing is when the limiting distribution depends sharply on the parameter values.

A common cause is a lack of

Smoothness: *A function* f *is a smooth if it has continuous derivatives up to some desired order over some domain. The number of derivatives is problem-specific.*

Examples of Nonregularity

Suppose $X_1, \ldots, X_n$ are iid copies of a random vector $X \in \mathbb{R}^p$ drawn from an unknown distribution P. Let $\mu_0 = \mathrm{P}\, X = (\mu_{01}, \ldots, \mu_{0p})$

Examples of Nonregularity

Suppose $X_1, \ldots, X_n$ are iid copies of a random vector $X \in \mathbb{R}^p$ drawn from an unknown distribution P. Let $\mu_0 = \mathrm{P}\, X = (\mu_{01}, \ldots, \mu_{0p})$

Superefficient estimators Define the estimator $\widetilde{\mu}_n$

$$\widetilde{\mu}_n = \begin{cases} \mathbb{P}_\mathrm{n}\, X & \text{if } \mathbb{P}_\mathrm{n}\, X \geq 1/4 \\ 0 & \text{if } \mathbb{P}_\mathrm{n}\, X < 1/4 \end{cases}$$

Examples of Nonregularity

Suppose $X_1, \ldots, X_n$ are iid copies of a random vector $X \in \mathbb{R}^p$ drawn from an unknown distribution P. Let $\mu_0 = \mathrm{P}\, X = (\mu_{01}, \ldots, \mu_{0p})$

Superefficient estimators Define the estimator $\tilde{\mu}_n$

$$\tilde{\mu}_n = \begin{cases} \mathbb{P}_\mathrm{n}\, X & \text{if } \mathbb{P}_\mathrm{n}\, X \geq 1/4 \\ 0 & \text{if } \mathbb{P}_\mathrm{n}\, X < 1/4 \end{cases}$$

Max of Means Define θ_0 as the component-wise maximum mean

$$\theta_0 = \bigvee_{j=1}^{p} \mu_{0j} = \max_{j \in [p]} \mu_{0j}$$

Is Nonregularity Avoidable?

"All ~~happy families~~ regular estimators are alike; each ~~unhappy family~~ nonregular estimator is unhappy in its own way." — Markov, quoting Tolstoy

Is Nonregularity Avoidable?

“All ~~happy families~~ regular estimators are alike; each ~~unhappy family~~ nonregular estimator is unhappy in its own way.” — Markov, quoting Tolstoy... maybe

Is Nonregularity Avoidable?

"All ~~happy families~~ regular estimators are alike; each ~~unhappy family~~ nonregular estimator is unhappy in its own way." — Markov, quoting Tolstoy... maybe

For the superefficient estimator $\tilde{\mu}$ the form of the estimator, not the underlying estimand was the source of the nonregularity.

Is Nonregularity Avoidable?

"All ~~happy families~~ regular estimators are alike; each ~~unhappy family~~ nonregular estimator is unhappy in its own way." — Markov, quoting Tolstoy... maybe

For the superefficient estimator $\tilde{\mu}$ the form of the estimator, not the underlying estimand was the source of the nonregularity.

The nonregularity in value function inference is due to the nondifferentiability of the treatment policy/DTR, and this kind of nonregularity is unavoidable [1 2].

[1] Keisuke Hirano and Jack R. Porter. "Impossibility Results for Nondifferentiable Functionals". In: *Econometrica* 80.4 (2012), pp. 1769–1790

[2] Or additional assumptions that change the problem.

Max of Means Limiting Distribution

Consider the estimators $\widehat{\mu}_n = \mathbb{P}_\mathrm{n} X$ and $\widehat{\theta}_n$

$$\widehat{\theta}_n = \bigvee_{j=1}^{p} \widehat{\mu}_{nj}$$

Lemma (Limiting Distribution of $\widehat{\theta}_n$)

Define the set $\mathfrak{U}(\mu_0) = \operatorname{argmax}_j \mu_{0j}$ *and assume* $\widehat{\mu}_n$ *is regular* $\sqrt{n}(\mathbb{P}_\mathrm{n} - \mathrm{P})X \rightsquigarrow \mathrm{MVN}(\mathbf{0}, \Sigma)$*. Then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \rightsquigarrow \bigvee_{j \in \mathfrak{U}(\mu_0)} Z_j$$

where $Z \sim \mathrm{MVN}(\mathbf{0}, \Sigma)$

Details of the Limiting Distribution

Define the event $E_n = \mathbf{1}\left\{\max_{k \notin \mathfrak{U}(\mu_0)} \widehat{\mu}_{nk} \geq \max_{j \in \mathfrak{U}(\mu_0)} \widehat{\mu}_{nj}\right\}$. Note that

1. $E_n = o_P(1)$
2. When E_n holds the maximizer(s) is not in $\mathfrak{U}(\mu_0)$ and vice versa

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_n - \theta_0) &= \bigvee_{j=1}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0) \\
&= (1 - E_n + E_n) \bigvee_{j \in \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0) \\
&= \bigvee_{j \in \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0) + E_n \left(\bigvee_{k \notin \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nk} - \theta_0) - \bigvee_{j \in \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0) \right)
\end{aligned}
$$

Proof Concept

$$\overbrace{\bigvee_{j \in \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0)}^{\text{The desired result}}$$

$$+ \quad \underbrace{E_n\left(\bigvee_{k \notin \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nk} - \theta_0) - \bigvee_{j \in \mathfrak{U}(\mu_0)}^{p} \sqrt{n}(\widehat{\mu}_{nj} - \theta_0)\right)}_{\text{A nuisance that's } o_P(1)}$$

Max of Means Simulations

Let $\mu_0 = (1, 1-\delta, \ldots, 1-\delta)$ comprised of a unique maximizer of 1 and p equally sub-optimal components with difference δ.

Input: Sample size n, Mean vector μ_0

foreach *simulation replicate* **do**

  Sample $X_i \sim N(\mu_0, I)$ for $i = 1, \ldots, n$

  Estimate $\widehat{\mu}_n = \mathbb{P}_\mathrm{n} X$

  Estimate $\widehat{\theta}_n = \max_{j \in [p]} \widehat{\mu}_{nj}$

  Calculate standardized estimate: $\tilde{\theta}_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$

end foreach

Plot the esimated density of $\tilde{\theta}_n$ against the density of $N(0, 1)$

In the following simulations we'll use $p = 6$ and by construction $\tilde{\theta}_n = \sqrt{n}(\widehat{\theta}_n - 1)$.

Max of Means with Fixed δ & Varying n

Let $\delta = .01$ so $\mu_0 = (1, .99, \ldots, .99)$, and we'll investigate $n = 100,\ 800,\ 1200,\ 2000.$ Before we look at the densities for $\widehat{\theta}_n$, as a refresher let's look at the same plot for $\sqrt{n}(\widehat{\mu}_{n1} - \mu_{01})$

PLACEHOLDER

Max of Means with Fixed δ & Varying n

Let $\delta = .01$ so $\mu_0 = (1, .99, \dots, .99)$, and we'll investigate $n = 100,\ 800,\ 1200,\ 2000.$ Before we look at the densities for $\widehat{\theta}_n$, as a refresher let's look at the same plot for $\sqrt{n}(\widehat{\mu}_{n1} - \mu_{01})$

PLACEHOLDER

As expected, the distribution of $\sqrt{n}(\widehat{\mu}_{n1} - \mu_{01})$ is approximately normal at all simulated sample sizes, with slightly heavier tails for smaller n.

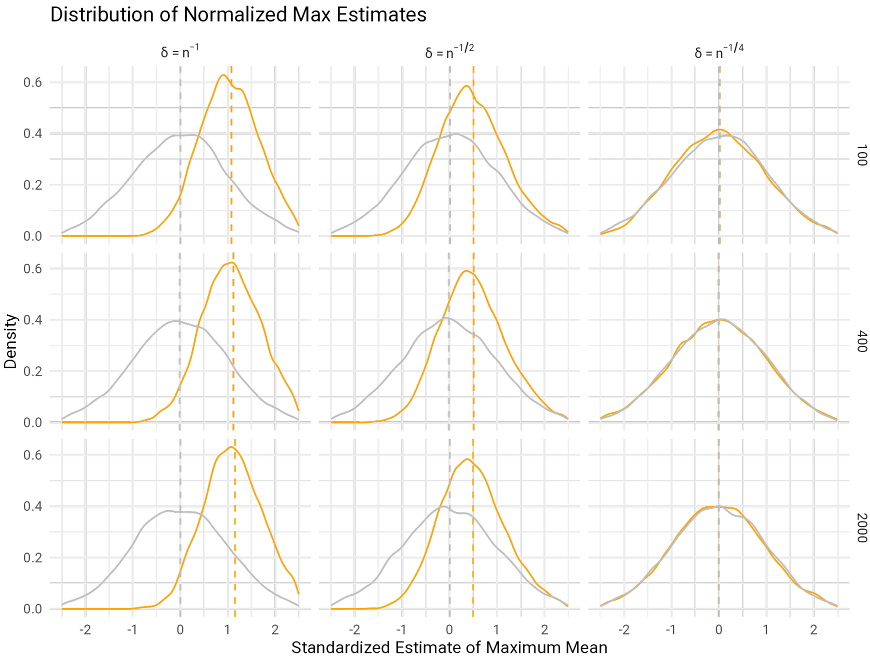
Max of Means with Fixed δ & Varying n

PLACEHOLDER
How does this relate to the assertion that "no effects, however small, are ever really zero?"

Max of Means with Sample-size-dependent δ

Now consider different values of δ that scale with n: $\delta = n^{-1}$, $\delta = n^{-1/2}$, and $\delta = n^{-1/4}$.

TODO: The plots on the next slide use estimated densities for $N(0, 1)$ as well as for the estimated max. Replace the estimated normal densities with the exact density



Distribution of Normalized Max Estimates
δ = n^-1
δ = n^-1/2
δ = n^-1/4
100
400
2000
Density
0.0
0.2
0.4
0.6
-2
-1
0
1
2
Standardized Estimate of Maximum Mean

What makes a good approximation?

Asymptotically valid does not guarantee a good approximation in finite samples.

Recall that the limiting distribution of $\widehat{\theta}_n$ didn't depend on the differences in μ_0 just the relevant sub-elements of Σ that are in $\mathfrak{U}(\mu_0)$

In finite samples, the sample size limits how large the differences need to be for us to distinguish them (remember the simulation plots).

Our approximation ought to reflect our uncertainty about the set of maximizers.

The Local Alternative

Idea: allow the data generating model to change with n. The new models are comprised of a static part and a part that changes with n which will go to zero in the limit.

$$\mu_{0n} = \mu_0 + s \times h(n)$$

- Where μ_0, $s \in \mathbb{R}^p$ are both fixed.
- $h(n)$ controls the perturbations. It is often, but not exclusively, $n^{-1/2}$ depending on the problem.

When we're trying to recover regular estimators $n^{-1/2}$ will be the target

Triangular Array

For each n, $X_{1,n}, \ldots, X_{n,n} \sim_{i.i.d.} P_n$

$$
\begin{array}{lllll l}
\underline{\text{Observations}} & & & & & \underline{\text{Distribution}} \\
X_{1,1} & & & & & P_1 \\
X_{1,2} & X_{2,2} & & & & P_2 \\
X_{1,3} & X_{2,3} & X_{3,3} & & & P_3 \\
X_{1,4} & X_{2,4} & X_{3,4} & X_{4,4} & & P_4 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots
\end{array}
$$

Define $\mu_{0,n} = P_n X$ and $\theta_{0,n} = \bigvee_{j=1}^{p} \mu_{0,j}$

- Assume $\mu_{0,n} = \mu_0 + s/\sqrt{n}$ where $s \in \mathbb{R}^p$ called local parameter
- Assume $\sqrt{n}(\mathbb{P}_n - P_n)X \rightsquigarrow \mathrm{Normal}(0, \Sigma)^4$

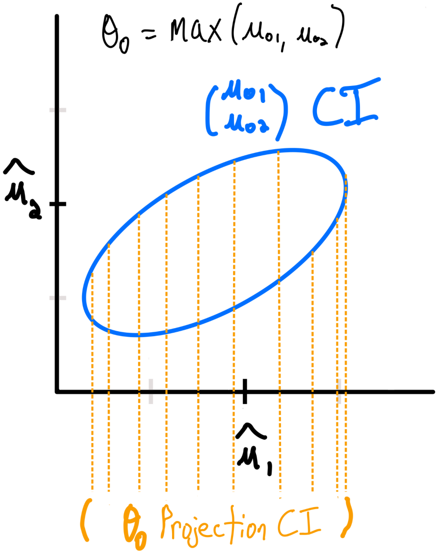
Table of Contents

Asymptotic Approach Overview

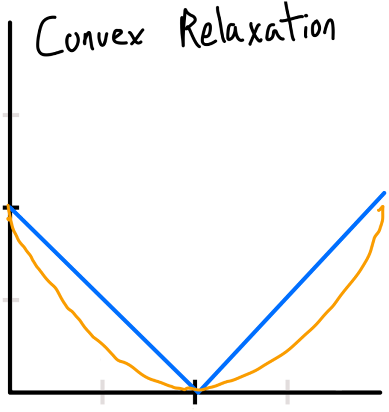
	Theoretical Guarantees	Easy to Implement	Conservative
Projection sets	$\checkmark^+$	☐	$\checkmark^+$
Bounding	$\checkmark^+$	☐	$\checkmark$
m-out-of-n Bootstrap	$\checkmark$	$\checkmark^+$	☐
Regularization	!!	$\checkmark$	☐
The Jackknife	☐	$\checkmark^+$	

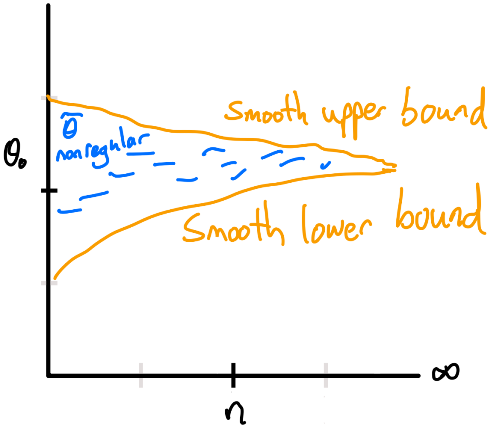
!! Regularization may induce infinite bias in certain scenarios. Has had empirical success in some applications but this approach is no longer recommended because of the potential for bias.

Visual Example — Projection

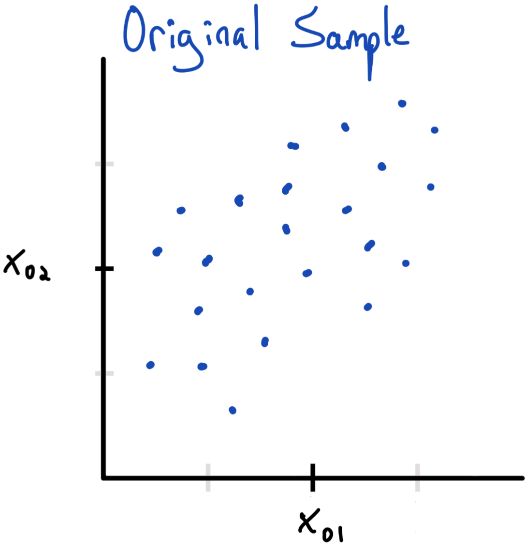


Visual Example — Bounding

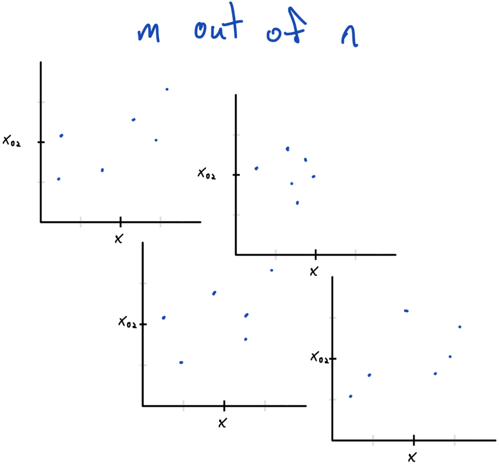


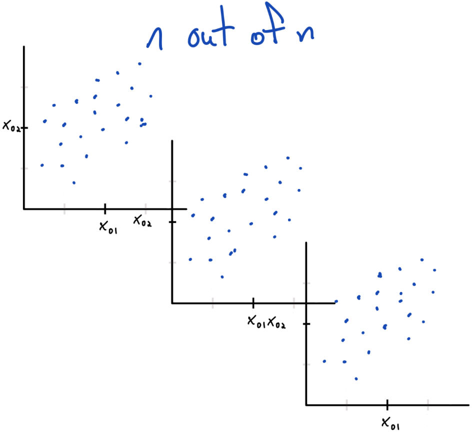


Visual Example — m-out-of-n Bootstrap

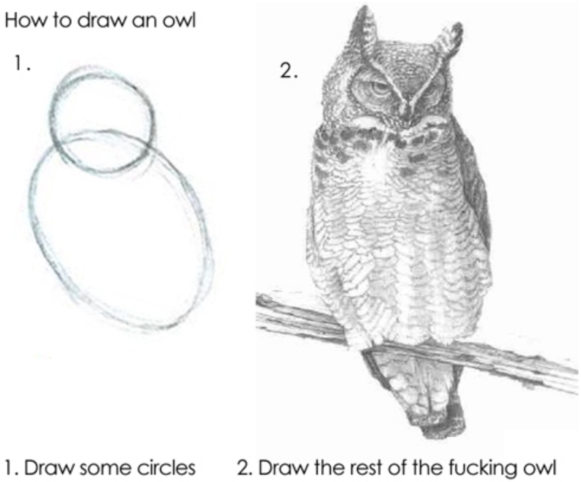


Visual Example — m-out-of-n Bootstrap





OWL


How to draw an owl
1.
2.
1. Draw some circles
2. Draw the rest of the fucking owl

Projection Region Big Picture

Recall θ_0 and μ_0 from the max of means problem. What do we know?

- The empirical estimator $\widehat{\mu}_n = \mathbb{P}_\mathrm{n}\, X$ is well-behaved (regular)

Projection Region Big Picture

Recall θ_0 and μ_0 from the max of means problem. What do we know?

- The empirical estimator $\widehat{\mu}_n = \mathbb{P}_\text{n}\, X$ is well-behaved (regular)
- If we knew μ_0 forming confidence intervals for θ_0 would be simple

Projection Region Big Picture

Recall θ_0 and μ_0 from the max of means problem. What do we know?

- The empirical estimator $\widehat{\mu}_n = \mathbb{P}_\mathrm{n} X$ is well-behaved (regular)

- If we knew μ_0 forming confidence intervals for θ_0 would be simple

Solution:

1. Determine all plausible values of μ_0 using the $(1-\alpha)$ CI
2. For each plausible value of μ_0, construct a CI for θ_0 treating the plausible μ_0 as fixed
3. Take a union over over all of these CIs

Projection Region Big Picture

Recall θ_0 and μ_0 from the max of means problem. What do we know?

- The empirical estimator $\widehat{\mu}_n = \mathbb{P}_n X$ is well-behaved (regular)

- If we knew μ_0 forming confidence intervals for θ_0 would be simple

Solution:

1. Determine all plausible values of μ_0 using the $(1-\alpha)$ CI
2. For each plausible value of μ_0, construct a CI for θ_0 treating the plausible μ_0 as fixed
3. Take a union over over all of these CIs

Can be very conservative

Projection Confidence Set

Let $\zeta_{n,1-\alpha}$ denote a $(1-\alpha)$ confidence set μ_0. For concreteness, take the Wald CI

$$\zeta_{n,1-\alpha} = \left\{\mu \in \mathbb{R}^p : n(\widehat{\mu}_n - \mu)^\top \widehat{\Sigma}_n^{-1}(\widehat{\mu}_n - \mu) \leq \chi^2_{p,1-\alpha}\right\}$$

Projection Confidence Set

Let $\zeta_{n,1-\alpha}$ denote a $(1-\alpha)$ confidence set μ_0. For concreteness, take the Wald CI

$$\zeta_{n,1-\alpha} = \left\{\mu \in \mathbb{R}^p : n(\widehat{\mu}_n - \mu)^\top \widehat{\Sigma}_n^{-1}(\widehat{\mu}_n - \mu) \leq \chi^2_{p,1-\alpha}\right\}$$

Projection Confidence Set:

$$\Gamma_{n,1-\alpha} = \left\{\theta \in \mathbb{R} : \theta = \bigvee_{j=1}^{p} \mu_j \textit{ for some } \mu \in \zeta_{n,1-\alpha}\right\}$$

is a valid confidence interval for θ_0

Bound-based CIs

Try to sandwich the nonsmooth functional between smooth upper and lower bounds.

Fertile ground seems to looking at the inf and sup of the nonsmooth functional

The m-out-of-n Bootstrap

It's the bootstrap, but with subsamples of size m_n instead of samples of size n

m_n is $o(n)$ i.e. as $m_n \to \infty$ $\frac{m}{n} \to 0$

The m-out-of-n Bootstrap

It's the bootstrap, but with subsamples of size m_n instead of samples of size n

m_n is $o(n)$ i.e. as $m_n \to \infty$ $\frac{m}{n} \to 0$

This allows the m-out-of-nBootstrap to directly approximates nonsmooth functionals by dark magic (even more magical than the standard bootstrap)

The m-out-of-n Bootstrap

It's the bootstrap, but with subsamples of size m_n instead of samples of size n

m_n is $o(n)$ i.e. as $m_n \to \infty$ $\frac{m}{n} \to 0$

This allows the m-out-of-nBootstrap to directly approximates nonsmooth functionals by dark magic (even more magical than the standard bootstrap)

Intuition: bootstrap samples will tend to have the same characteristics as the original sample (very similar means, variances etc.). Subsamples get weird.

The m-out-of-n Bootstrap

It's the bootstrap, but with subsamples of size m_n instead of samples of size n

m_n is $o(n)$ i.e. as $m_n \to \infty$ $\frac{m}{n} \to 0$

This allows the m-out-of-nBootstrap to directly approximates nonsmooth functionals by dark magic (even more magical than the standard bootstrap)

Intuition: bootstrap samples will tend to have the same characteristics as the original sample (very similar means, variances etc.). Subsamples get weird.

Caveat: Not as straightforwardly valid as Projection and Bounding approaches (modulo a certain definition of "straightforward"), and can fail [3].

[3] Donald W. K. Andrews and Patrik Guggenberger. "Asymptotic Size and a Problem With Subsampling and with the m out of n Bootstrap". In: *Econometric Theory* 26.2 (Apr. 2010), pp. 426–468

Table of Contents

Treatment Policy Notation

A policy (aka DTR, treatment rule) π[4] is a function which maps contexts to actions
$\pi : \mathcal{X} \mapsto \mathcal{A}$

[4] Authors may write $\pi(x, a)$, $\pi(a \mid x)$, $\pi(x)$, or simply π depending on what the author wishes to emphasize.

Treatment Policy Notation

A policy (aka DTR, treatment rule) π[4] is a function which maps contexts to actions $\pi : \mathcal{X} \mapsto \mathcal{A}$

Define the value of a policy π as

$$\mathrm{V}(\pi) \doteq \mathbf{E}[Y^\star(a = \pi(X))]$$

[4] Authors may write $\pi(x, a)$, $\pi(a \mid x)$, $\pi(x)$, or simply π depending on what the author wishes to emphasize.

Treatment Policy Notation

A policy (aka DTR, treatment rule) π[4] is a function which maps contexts to actions $\pi : \mathcal{X} \mapsto \mathcal{A}$

Define the value of a policy π as

$$V(\pi) \doteq \mathbf{E}[Y^\star(a = \pi(X))]$$

An optimal policy $\pi^\star$ is any policy that satisfies

$$V(\pi^\star) \geq V(\pi) \qquad \text{for all } \pi \in \Pi$$

[4] Authors may write $\pi(x, a)$, $\pi(a \mid x)$, $\pi(x)$, or simply π depending on what the author wishes to emphasize.

Treatment Policy Notation

A policy (aka DTR, treatment rule) π[4] is a function which maps contexts to actions $\pi : \mathcal{X} \mapsto \mathcal{A}$

Define the value of a policy π as

$$V(\pi) \doteq \mathbf{E}[Y^\star(a = \pi(X))]$$

An optimal policy $\pi^\star$ is any policy that satisfies

$$V(\pi^\star) \geq V(\pi) \qquad \text{for all } \pi \in \Pi$$

Denote the estimated optimal policy $\widehat{\pi}_n$, and for conciseness, assume that there is a unique maximizer for every covariate value.

[4] Authors may write $\pi(x, a)$, $\pi(a \mid x)$, $\pi(x)$, or simply π depending on what the author wishes to emphasize.

Value Functions

1. (Conditional) Value of the estimated optimal policy

$$\mathrm{V}(\widehat{\pi}_n) = \mathbf{E}[Y^\star(\widehat{\pi}_n(x)) \mid \widehat{\pi}_n(x)] = \mathrm{P}\, Y^\star(\widehat{\pi}_n(x))$$

[5] Not truly conditional because the policy is not a random variable.

Value Functions

1. (Conditional) Value of the estimated optimal policy

$$V(\widehat{\pi}_n) = \mathbf{E}[Y^\star(\widehat{\pi}_n(x)) \mid \widehat{\pi}_n(x)] = \mathrm{P}\, Y^\star(\widehat{\pi}_n(x))$$

2. Expected Value of an estimated optimal policy

$$\mathbf{E}[Y^\star(\widehat{\pi}_n(x))]$$

[5] Not truly conditional because the policy is not a random variable.

Value Functions

1. (Conditional) Value of the estimated optimal policy

$$V(\hat{\pi}_n) = \mathbf{E}[Y^\star(\hat{\pi}_n(x)) \mid \hat{\pi}_n(x)] = \mathrm{P}\, Y^\star(\hat{\pi}_n(x))$$

2. Expected Value of an estimated optimal policy

$$\mathbf{E}[Y^\star(\hat{\pi}_n(x))]$$

3. (Conditional[5])Value of the optimal policy (or, more broadly, any fixed policy)

$$V(\pi^\star) = \mathbf{E}[Y^\star(\pi^\star); \pi^\star]$$

[5] Not truly conditional because the policy is not a random variable.

Value Functions

1. (Conditional) Value of the estimated optimal policy

$$\mathrm{V}(\widehat{\pi}_n) = \mathbf{E}[Y^\star(\widehat{\pi}_n(x)) \mid \widehat{\pi}_n(x)] = \mathrm{P}\, Y^\star(\widehat{\pi}_n(x))$$

2. Expected Value of an estimated optimal policy

$$\mathbf{E}[Y^\star(\widehat{\pi}_n(x))]$$

3. (Conditional[5])Value of the optimal policy (or, more broadly, any fixed policy)

$$\mathrm{V}(\pi^\star) = \mathbf{E}[Y^\star(\pi^\star); \pi^\star]$$

These three will not generally be equivalent even asymptotically.
The conditional value function is often simply called the value function in many papers, you have to look at the formula to know.

[5] Not truly conditional because the policy is not a random variable.

Other Distinctions in Evaluating Policies

Let $h(a \mid x)$ denote the policy that was used to assign treatment during the experiment and $\pi(a \mid x)$ the policy we are intersested in evaluating.

On/Off-Policy:

- *On-policy evaluation:* $\pi = h$ *for all* x
- *Off-policy evaluation:* $\pi(a \mid x) \neq h(a \mid x)$ *for some* x *and* a.
  Requires additional assumptions. Ex:

$$w_{\max} \doteq \operatorname*{ess\,sup}_{t \in \mathbb{N}, a \in \mathcal{A}, x \in \mathcal{X}} \frac{\pi(a \mid x)}{h_t(a \mid x)} < \infty$$

Other Distinctions in Evaluating Policies

Let $h(a \mid x)$ denote the policy that was used to assign treatment during the experiment and $\pi(a \mid x)$ the policy we are intersested in evaluating.

On/Off-Policy:

- *On-policy evaluation:* $\pi = h$ *for all* x
- *Off-policy evaluation:* $\pi(a \mid x) \neq h(a \mid x)$ *for some* x *and* a.
  Requires additional assumptions. Ex:

$$w_{\max} \doteq \operatorname*{ess\,sup}_{t \in \mathbb{N}, a \in \mathcal{A}, x \in \mathcal{X}} \frac{\pi(a \mid x)}{h_t(a \mid x)} < \infty$$

Fixed or Estimated:

- *Data-derived or estimated policy*
- *Fixed policy that's pre-specified before looking at the data.*

Setup Specifics

- Two arms $A \in \{-1, 1\}$
- $\widehat{\pi}_n(x) = \mathrm{sign}(x^\top \widehat{\beta}_n)$
- Assume $\widehat{\Sigma}_n$ is a consistent estimator of the asymptotic variance of $\widehat{\beta}_n$

Setup Specifics

- Two arms $A \in \{-1, 1\}$
- $\widehat{\pi}_n(x) = \text{sign}(x^\top \widehat{\beta}_n)$
- Assume $\widehat{\Sigma}_n$ is a consistent estimator of the asymptotic variance of $\widehat{\beta}_n$

Define $\mathcal{G} = \{g(X, A, Y; \delta) = Y \mathbf{1}\{AX^\top \delta > 0\} \mathbf{1}\{X^\top \beta_0 = 0\} : \delta \in \mathbb{R}^p\}$.

We'll think of $\sqrt{n}(\mathbb{P}_\text{n} - \text{P}_\text{n})$ as a random element of $l^\infty(\mathbb{R}^p)$

See Anastasios A. Tsiatis et al. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine.* New York: Chapman and Hall/CRC, Dec. 19, 2019. 618 pp. for the assumptions. They're long and Donsker makes an appearance

Joint Distribution Before Maximizing

$$\sqrt{n}\begin{bmatrix} \mathbb{P}_{\mathrm{n}} - \mathrm{P}_n \\ \widehat{\beta}_n - \beta_0 \\ (\mathbb{P}_{\mathrm{n}} - \mathrm{P}_n) Y \mathbf{1}\left\{AX^\top \beta_0 > 0\right\} \end{bmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{T} \\ \mathbb{Z} \\ \mathbb{W} \end{pmatrix}$$

where

- $\mathbb{T}$ is a Brownian Bridge indexed by $\mathbb{R}^p$
- $\mathbb{Z}$ and $\mathbb{W}$ are normal

Distribution after Maximizing

$$\sqrt{n}(\widehat{\mathrm{V}}_n(\widehat{\beta}_n) - \mathrm{V}(\widehat{\beta}_n)) \rightsquigarrow \mathbb{T}(\mathbb{Z} + s) + \mathbb{W}$$

where s is a local parameter. Again $\mathbb{T}$ is a Brownian Bridge, and $\mathbb{Z}$, $\mathbb{W}$ normal.

[6] Note it wasn't in the joint distribution

Distribution after Maximizing

$$\sqrt{n}(\widehat{\mathrm{V}}_n(\widehat{\beta}_n) - \mathrm{V}(\widehat{\beta}_n)) \rightsquigarrow \mathbb{T}(\mathbb{Z} + s) + \mathbb{W}$$

where s is a local parameter. Again $\mathbb{T}$ is a Brownian Bridge, and $\mathbb{Z}$, $\mathbb{W}$ normal.

Having s in the limit shows this is nonregular[6].

[6]Note it wasn't in the joint distribution

Distribution after Maximizing

$$\sqrt{n}(\widehat{\mathbb{V}}_n(\widehat{\beta}_n) - \mathbb{V}(\widehat{\beta}_n)) \rightsquigarrow \mathbb{T}(\mathbb{Z} + s) + \mathbb{W}$$

where s is a local parameter. Again $\mathbb{T}$ is a Brownian Bridge, and $\mathbb{Z}$, $\mathbb{W}$ normal.

Having s in the limit shows this is nonregular[6].

We can construct a bound-based CI by

1. Partition $\mathcal{X}$ into xs near the decision boundary ($x^\top \widehat{\beta}_n \approx 0$) and those far away
2. Take sup/inf over local perturbations in the group close to the boundary

[6] Note it wasn't in the joint distribution

Upper Bound

Let τ_n be a sequence of tuning parameters that satisfies $\tau_n \to \infty$ and $\tau_n = o(n)$ as $n \to \infty$

$$U_n = \sup_{\omega \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - \mathrm{P}_n) Y \mathbf{1}\left\{AX^\top \omega > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} \leq \tau_n\right\}$$
$$+ \sqrt{n}(\mathbb{P}_n - \mathrm{P}_n) Y \mathbf{1}\left\{AX^\top \widehat{\beta}_n > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} > \tau_n\right\}$$

Upper Bound

Let τ_n be a sequence of tuning parameters that satisfies $\tau_n \to \infty$ and $\tau_n = o(n)$ as $n \to \infty$

$$
\begin{aligned}
U_n = \sup_{\omega \in \mathbb{R}^p} & \sqrt{n}(\mathbb{P}_n - \mathrm{P}_n) Y \mathbf{1}\left\{A X^\top \omega > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} \le \tau_n\right\} \\
& + \sqrt{n}(\mathbb{P}_n - \mathrm{P}_n) Y \mathbf{1}\left\{A X^\top \widehat{\beta}_n > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} > \tau_n\right\}
\end{aligned}
$$

Note that ω has been replaced with $\widehat{\beta}_n$ in the second line. Why?

Upper Bound

Let τ_n be a sequence of tuning parameters that satisfies $\tau_n \to \infty$ and $\tau_n = o(n)$ as $n \to \infty$

$$
\begin{aligned}
U_n = \sup_{\omega \in \mathbb{R}^p} & \sqrt{n}(\mathbb{P}_n - P_n) Y \mathbf{1}\left\{A X^\top \omega > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} \le \tau_n\right\} \\
& + \sqrt{n}(\mathbb{P}_n - P_n) Y \mathbf{1}\left\{A X^\top \widehat{\beta}_n > 0\right\} \mathbf{1}\left\{\frac{n(X^\top \widehat{\beta}_n)^2}{X^\top \widehat{\Sigma}_n X} > \tau_n\right\}
\end{aligned}
$$

Note that ω has been replaced with $\widehat{\beta}_n$ in the second line. Why?

Lower bound is analagous, replace sup with inf

Table of Contents

Alternatives to Asymptotics

If the problems happen in the limit, what if we just don't take it to the limit?

Finite-sample bounds

Similar in spirit to the asymptotic bound-based approach, but with E-processes and test supermartingales playing the role of the nicely behaved functions.

$$\text{Test supermartingale} \quad \mathbf{E}[Z_{t+1}|Z_t] \leq Z_t \; \forall t \in \mathbb{N}^+$$

Finite-sample bounds

Similar in spirit to the asymptotic bound-based approach, but with E-processes and test supermartingales playing the role of the nicely behaved functions.

$$\text{Test supermartingale} \qquad \mathbf{E}[Z_{t+1}|Z_t] \leq Z_t \; \forall t \in \mathbb{N}^+$$

A test supermartingale M_t is a non-negative supermartingale that under the null P_0 satisfies $\mathbf{E}_{P_0}[M_t] \leq 1$ at any time t

Assumptions in this Setting

- Y is bounded, and for convenience we'll assume $Y \in [0,1]$
- $\nu_t \doteq \mathbf{E}_\pi[Y_t \mid \mathcal{H}_{t-1}]$ and is adapted to the filtration $\mathcal{H}_{t-1}$
- Exogenous treatment

$$\Pr(A_t \mid \mathcal{H}_{t-1}) = \Pr(A_t \mid X_t)$$

- Finite importance weights

$$w_{\max} \doteq \operatorname*{ess\,sup}_{t \in \mathbb{N}, a \in \mathcal{A}, x \in \mathcal{X}} \frac{\pi(a \mid x)}{h_t(a \mid x)} < \infty$$

Confidence Sequences

Definition (Confidence Sequence)

We say that a sequence of intervals $[L_t, U_t]_{t=1}^{\infty}$ is a confidence sequence for the parameter $\theta \in \mathbb{R}$ if

$$\Pr\left(\forall t \in \mathbb{N},\ \theta \in [L_t, U_t]\right) \geq 1 - \alpha$$

or equivalently,

$$\Pr\left(\exists t \in \mathbb{N} : \theta \notin [L_t, U_t]\right) \leq \alpha$$

Confidence Sequences

Definition (Confidence Sequence)

We say that a sequence of intervals $[L_t, U_t]_{t=1}^{\infty}$ is a confidence sequence for the parameter $\theta \in \mathbb{R}$ if

$$\Pr\left(\forall t \in \mathbb{N},\ \theta \in [L_t, U_t]\right) \geq 1 - \alpha$$

or equivalently,

$$\Pr\left(\exists t \in \mathbb{N} : \theta \notin [L_t, U_t]\right) \leq \alpha$$

For comparison, recall that a $(1 - \alpha)$ confidence interval (CI) satisfies

$$\forall t \in \mathbb{N},\ \Pr(\theta \in [L_t, U_t]) \geq 1 - \alpha$$

Key Result

Define the weighted rewards $\phi_t^{(\text{IW-}\ell)} \doteq w_t Y_t$ and $\phi_t^{(\text{IW-}u)} \doteq w_t(1 - Y_t)$

Let $(\lambda_t^L(\nu'))_{t=1}^{\infty}$ be any $[0, 1/\nu')$-valued predictable sequence

$$L_t^{\text{IW}} \doteq \inf\left\{\nu' \in [0,1] : \prod_{i=1}^{t}\Big(1 + \lambda_i^L(\nu') \cdot (\phi_t^{(\text{IW-}\ell)} - \nu')\Big) < \frac{1}{\alpha}\right\} \quad (1)$$

forms a lower $(1-\alpha)$ confidence sequence for ν, $\Pr(\forall t \in \mathbb{N},\ \nu \geq L_t^{\text{IW}}) \geq 1 - \alpha$.

Key Result

Define the weighted rewards $\phi_t^{(\text{IW-}\ell)} \doteq w_t Y_t$ and $\phi_t^{(\text{IW-}u)} \doteq w_t(1 - Y_t)$

Let $(\lambda_t^L(\nu'))_{t=1}^{\infty}$ be any $[0, 1/\nu')$-valued predictable sequence

$$L_t^{\text{IW}} \doteq \inf\left\{\nu' \in [0,1] : \prod_{i=1}^{t}\left(1 + \lambda_i^L(\nu') \cdot (\phi_t^{(\text{IW-}\ell)} - \nu')\right) < \frac{1}{\alpha}\right\} \tag{1}$$

forms a lower $(1-\alpha)$ confidence sequence for ν, $\Pr(\forall t \in \mathbb{N},\ \nu \geq L_t^{\text{IW}}) \geq 1 - \alpha$.

ν' represents candidate policy value estimates

Key Tools: Ville's Theoerem & Inequality

For any discrete-time stochastic process P an event A has measure zero under P if and only if there is a test martingale for P that grows to infinity on all of A.

Key Tools: Ville's Theoerem & Inequality

For any discrete-time stochastic process P an event A has measure zero under P if and only if there is a test martingale for P that grows to infinity on all of A.

Moreover, $P(A) < \epsilon$ if and only if there is a test martingale for P that exceeds $1/\epsilon$ on all of A.

Key Tools: Ville's Theoerem & Inequality

For any discrete-time stochastic process P an event A has measure zero under P if and only if there is a test martingale for P that grows to infinity on all of A.

Moreover, $P(A) < \epsilon$ if and only if there is a test martingale for P that exceeds $1/\epsilon$ on all of A.

The Gambler's Ruin (Ville's Inequality): *If M is a test martingale for P then*

$$\Pr\left(\sup_t M_t \geq \alpha\right) \leq \frac{1}{\alpha} \qquad \textit{for any } \alpha \geq 1$$

From Sequence to Interval

Suppose that all we care about is a $(1-\alpha)$ CI after T observations.

A CS is also trivially a CI at a fixed time, but the width of the interval will be wider than if we only needed to guarantee coverage at one point in time.

From Sequence to Interval

Suppose that all we care about is a $(1-\alpha)$ CI after T observations.

A CS is also trivially a CI at a fixed time, but the width of the interval will be wider than if we only needed to guarantee coverage at one point in time.

Lemma (The minimum and maximum bounds of a $1-\alpha$ confidence sequence form a $1-\alpha$ confidence interval.)

Define lower and upper bounds, L^{CI} and L^{CI}, as

$$L^{CI} = \max_{t \leq T} L_t^{CS} \qquad \text{and } U^{CI} = \max_{t \leq T} U_t^{CS}$$

Then (L^{CI}, U^{CI}) is a $(1-\alpha)$ confidence interval for $\mathrm{V}(\pi)$

Table of Contents

Practical Improvements

Pre-testing/Screening for Projection Intervals

Double bootstrap, other ways of selecting m iteratively

E-process Confidence Sequences

- Double robustness
- Trimming the observed rewards
- Empirical weights

Where to Start

Start with Chapter 10 of Tsiatis et al., *Dynamic Treatment Regimes*

General Theory – Asymptotic

- Chapter 3 of Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. New York, NY: Springer, 2006
- Chapters 6 and 7 of Aad W Van der Vaart. *Asymptotic Statistics*. Vol. 3. Cambridge university press, 2000
- Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. New York, NY: Springer, 2008
- Important Paper: Keisuke Hirano and Jack R. Porter. "Impossibility Results for Nondifferentiable Functionals". In: *Econometrica* 80.4 (2012), pp. 1769–1790
- Aurelien Bibaut et al. "Post-Contextual-Bandit Inference". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 28548–28559

Precision Medicine

- Eric B. Laber et al. "Dynamic Treatment Regimes: Technical Challenges and Applications". In: *Electronic Journal of Statistics* 8.1 (Jan. 2014), pp. 1225–1272
- Alexander R. Luedtke and Mark J. van der Laan. "Statistical Inference for the Mean Outcome under a Possibly Non-Unique Optimal Treatment Strategy". In: *The Annals of Statistics* 44.2 (Apr. 2016), pp. 713–742
- Chengchun Shi et al. "Statistical Inference of the Value Function for Reinforcement Learning in Infinite-Horizon Settings". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.3 (July 1, 2022), pp. 765–793
- Vitor Hadad et al. *Confidence Intervals for Policy Evaluation in Adaptive Experiments.* 2021

General Theory – Non-asymptotic

- Ian Waudby-Smith et al. *Anytime-Valid off-Policy Inference for Contextual Bandits.* Nov. 17, 2022. URL: http://arxiv.org/abs/2210.10768 (visited on 03/04/2024). preprint

- Steven R Howard et al. “Time-Uniform, Nonparametric, Nonasymptotic Confidence Sequences”. In: *The Annals of Statistics* 49.2 (2021), pp. 1055–1080

Conformal Inference?

Stuff that might be relevant but I haven't looked into it yet. Prediction intervals not CIs. More cites because this list isn't as filtered.

- Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. "On-Line Predictive Linear Regression". In: *The Annals of Statistics* 37.3 (June 2009), pp. 1566–1590
- Victor Chernozhukov, Kaspar Wüthrich, and Zhu Yinchu. "Exact and Robust Conformal Inference Methods for Predictive Machine Learning with Dependent Data". In: *Proceedings of the 31st Conference On Learning Theory*. Conference On Learning Theory. PMLR, July 3, 2018, pp. 732–749
- Rina Foygel Barber et al. "Predictive Inference with the Jackknife+". In: *The Annals of Statistics* 49.1 (Feb. 2021), pp. 486–507
- Lihua Lei and Emmanuel J. Candès. "Conformal Inference of Counterfactuals and Individual Treatment Effects". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.5 (Nov. 1, 2021), pp. 911–938
- Ying Jin, Zhimei Ren, and Emmanuel J. Candès. "Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach". In: *Proceedings of the National Academy of Sciences* 120.6 (Feb. 7, 2023),

Andrews, Donald W. K. and Patrik Guggenberger. “Asymptotic Size and a Problem With Subsampling and with the m out of n Bootstrap”. In: *Econometric Theory* 26.2 (Apr. 2010), pp. 426–468.

Barber, Rina Foygel et al. “Predictive Inference with the Jackknife+”. In: *The Annals of Statistics* 49.1 (Feb. 2021), pp. 486–507.

Bibaut, Aurelien et al. “Post-Contextual-Bandit Inference”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 28548–28559.

Chernozhukov, Victor, Kaspar Wüthrich, and Zhu Yinchu. “Exact and Robust Conformal Inference Methods for Predictive Machine Learning with Dependent Data”. In: *Proceedings of the 31st Conference On Learning Theory*. Conference On Learning Theory. PMLR, July 3, 2018, pp. 732–749.

Hadad, Vitor et al. *Confidence Intervals for Policy Evaluation in Adaptive Experiments*. 2021.

Hirano, Keisuke and Jack R. Porter. “Impossibility Results for Nondifferentiable Functionals”. In: *Econometrica* 80.4 (2012), pp. 1769–1790.

Howard, Steven R et al. "Time-Uniform, Nonparametric, Nonasymptotic Confidence Sequences". In: *The Annals of Statistics* 49.2 (2021), pp. 1055–1080.

Jin, Ying, Zhimei Ren, and Emmanuel J. Candès. "Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach". In: *Proceedings of the National Academy of Sciences* 120.6 (Feb. 7, 2023), e2214889120.

Kosorok, Michael R. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. New York, NY: Springer, 2008.

Laber, Eric B. et al. "Dynamic Treatment Regimes: Technical Challenges and Applications". In: *Electronic Journal of Statistics* 8.1 (Jan. 2014), pp. 1225–1272.

Lei, Lihua and Emmanuel J. Candès. "Conformal Inference of Counterfactuals and Individual Treatment Effects". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.5 (Nov. 1, 2021), pp. 911–938.

Luedtke, Alexander R. and Mark J. van der Laan. "Statistical Inference for the Mean Outcome under a Possibly Non-Unique Optimal Treatment Strategy". In: *The Annals of Statistics* 44.2 (Apr. 2016), pp. 713–742.

- Shi, Chengchun et al. "Statistical Inference of the Value Function for Reinforcement Learning in Infinite-Horizon Settings". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.3 (July 1, 2022), pp. 765–793.
- Tsiatis, Anastasios A. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. New York, NY: Springer, 2006.
- Tsiatis, Anastasios A. et al. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. New York: Chapman and Hall/CRC, Dec. 19, 2019. 618 pp.
- Van der Vaart, Aad W. *Asymptotic Statistics*. Vol. 3. Cambridge university press, 2000.
- Vovk, Vladimir, Ilia Nouretdinov, and Alex Gammerman. "On-Line Predictive Linear Regression". In: *The Annals of Statistics* 37.3 (June 2009), pp. 1566–1590.
- Waudby-Smith, Ian et al. *Anytime-Valid off-Policy Inference for Contextual Bandits*. Nov. 17, 2022. URL: `http://arxiv.org/abs/2210.10768` (visited on 03/04/2024). preprint.