Bandit value estimation as an excuse to get some new concentration inequalities

Csaba Szepesvári

November 30, 2021

DeepMind and University of Alberta
BIRS Workshop on Math.Stat & Learning

Collaborators

- Ilja Kuzborskij
- Claire Vernade
- András György

References: [KS19, KVGS21, KS21]

Contextual bandits

- $(X, A, R) \in \mathcal{X} \times [K] \times [0,1]$ X: context, A: action
  R: reward
- Given $x \in \mathcal{X}$, $a \in [K]$,

$$R \sim P_{R|X,A}(\cdot|x,a)$$

  is the reward "generated"
- Value $u(\pi)$ of policy $\pi : \mathcal{X} \to \Delta([K])$ is

$$u(\pi) = \int_{\mathcal{X}} \sum_{a \in [K]} \pi(a|x) r(x,a) \, \mathrm{d}P_X(x)$$

  where

$$r(x,a) = \int y \, P_{R|X,A}(\mathrm{d}y|x,a)$$

Contextual batch bandit value estimation

- **Observed**: $S = ((X_1, A_1, R_1), \ldots, (X_n, A_n, R_n))$ i.i.d.,

$$(X_i, A_i, R_i) \in \mathcal{X} \times [K] \times \mathbb{R}, i \in [n] := \{1, \ldots, n\}$$

- **Given**: randomized *behavior and target policies* $\pi_b, \pi : \mathcal{X} \to \Delta([K])$, with

$$A_i \sim \pi_b(\cdot|X_i), \qquad i \in [n]$$

- **Goal**: Find f s.t
for all $x > 0$, w.p. $1 - e^{-x}$,

$$u(\pi) \geq f(S, \pi, \pi_b, x)$$

and $u(\pi) - f(S, \pi, \pi_b, x)$ is "small"

A 2-step approach

- Step #1: Find f_0 such that $u(\pi)$ is close to

$$U := f_0(S, \pi, \pi_b)$$

- Step #2: Find a high probability lower bound

$$U_{\mathrm{LB}} := f(S, \pi, \pi_b, x)$$

for U.

Many ways to do this...

Mean estimation strategies in bandits

- Importance sampling estimator
- Double-robust estimator
- Weighted importance sampling estimator

Importance sampling estimator

Define the *importance weights*

$$W_i = \frac{\pi(A_i|X_i)}{\pi_b(A_i|X_i)} \qquad i \in [n] \ .$$

The (unbiased) *importance sampling (sicc!)* estimator is

$$U^{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} W_i R_i \ .$$

Importance sampling estimator

Define the *importance weights*

$$W_i = \frac{\pi(A_i|X_i)}{\pi_b(A_i|X_i)} \qquad i \in [n] \ .$$

The (unbiased) *importance sampling (sicc!)* estimator is

$$U^{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} W_i R_i \ .$$

Value lower bounds?

Importance sampling estimator

Define the *importance weights*

$$W_i = \frac{\pi(A_i|X_i)}{\pi_b(A_i|X_i)} \qquad i \in [n] \; .$$

The (unbiased) *importance sampling (sicc!)* estimator is

$$U^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} W_i R_i \; .$$

Value lower bounds?

- Disagreeing policies: W_i could be heavy-tailed

Importance sampling estimator

Define the *importance weights*

$$W_i = \frac{\pi(A_i|X_i)}{\pi_b(A_i|X_i)} \qquad i \in [n] \; .$$

The (unbiased) *importance sampling (sicc!)* estimator is

$$U^{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} W_i R_i \; .$$

Value lower bounds?

- Disagreeing policies: W_i could be heavy-tailed
- Hack? $W_i^{\lambda} = \pi(A_i|X_i)/(\pi_b(A_i|X_i) + \lambda)$, $\lambda > 0$, $\lambda =$??

The doubly-robust (DR) estimator

Choose $\hat{\eta} : \mathcal{X} \times [K] \to [0, 1]$ and let

$$U^{\mathrm{DR}} = \frac{1}{n} \sum_{i,a} \pi(a|X_i)\hat{\eta}(X_i, a) + \frac{1}{n} \sum_{i} W_i(R_i - \hat{\eta}(X_i, A_i))\,.$$

The doubly-robust (DR) estimator

Choose $\hat{\eta} : \mathcal{X} \times [K] \to [0, 1]$ and let

$$U^{\mathrm{DR}} = \frac{1}{n} \sum_{i,a} \pi(a|X_i)\hat{\eta}(X_i, a) + \frac{1}{n} \sum_{i} W_i(R_i - \hat{\eta}(X_i, A_i))\,.$$

- Unbiased for any $W_i \in \sigma(X_i, A_i)$ s.t. one of the following hold:

The doubly-robust (DR) estimator

Choose $\hat{\eta} : \mathcal{X} \times [K] \to [0, 1]$ and let

$$U^{\mathrm{DR}} = \frac{1}{n} \sum_{i,a} \pi(a|X_i)\hat{\eta}(X_i, a) + \frac{1}{n} \sum_{i} W_i(R_i - \hat{\eta}(X_i, A_i))\,.$$

- Unbiased for any $W_i \in \sigma(X_i, A_i)$ s.t. one of the following hold:
    1. $\forall f : [K] \to [0, 1]$: $\mathbb{E}[W_i f(A_i)|X_i] = \sum_a \pi(a|X_i) f(a)$ a.s.

The doubly-robust (DR) estimator

Choose $\hat{\eta} : \mathcal{X} \times [K] \to [0,1]$ and let

$$U^{\mathrm{DR}} = \frac{1}{n} \sum_{i,a} \pi(a|X_i)\hat{\eta}(X_i, a) + \frac{1}{n} \sum_{i} W_i(R_i - \hat{\eta}(X_i, A_i))\,.$$

- Unbiased for any $W_i \in \sigma(X_i, A_i)$ s.t. one of the following hold:
  1. $\forall f : [K] \to [0,1]$: $\mathbb{E}[W_i f(A_i)|X_i] = \sum_a \pi(a|X_i) f(a)$ a.s.
  2. $\mathbb{E}[\hat{\eta}(X_i, A_i)|X_i, A_i] = r(X_i, A_i)$

The doubly-robust (DR) estimator

Choose $\hat{\eta} : \mathcal{X} \times [K] \to [0,1]$ and let

$$U^{\mathrm{DR}} = \frac{1}{n} \sum_{i,a} \pi(a|X_i)\hat{\eta}(X_i, a) + \frac{1}{n} \sum_{i} W_i(R_i - \hat{\eta}(X_i, A_i))\,.$$

- Unbiased for any $W_i \in \sigma(X_i, A_i)$ s.t. one of the following hold:
  1. $\forall f : [K] \to [0,1]$: $\mathbb{E}[W_i f(A_i)|X_i] = \sum_a \pi(a|X_i) f(a)$ a.s.
  2. $\mathbb{E}[\hat{\eta}(X_i, A_i)|X_i, A_i] = r(X_i, A_i)$
- Reduces variance when $\hat{\eta} \approx r$

Weighted importance sampling (WIS)

WIS estimator:

$$U^{\mathrm{WIS}} = \frac{\sum_{i=1}^{n} W_i R_i}{\sum_{i=1}^{n} W_i} \ .$$

Weighted importance sampling (WIS)

WIS estimator:

$$U^{\mathrm{WIS}} = \frac{\sum_{i=1}^{n} W_i R_i}{\sum_{i=1}^{n} W_i} \ .$$

- *Biased* (though bias vanishes as $n \to \infty$)

Weighted importance sampling (WIS)

WIS estimator:

$$U^{\mathrm{WIS}} = \frac{\sum_{i=1}^{n} W_i R_i}{\sum_{i=1}^{n} W_i} \; .$$

- *Biased* (though bias vanishes as $n \to \infty$)
- Empirically much better than IS; "low variance"
  Efron-Stein + calculation:

$$\mathrm{Var}(U^{\mathrm{WIS}}) \leq 4 \, \underbrace{\mathbb{E}\left[\sum_k \left(\frac{W_k}{\sum_i W_i}\right)^2\right]}_{=: \frac{1}{n_{\mathrm{eff}}}}$$

Weighted importance sampling (WIS)

WIS estimator:

$$U^{\mathrm{WIS}} = \frac{\sum_{i=1}^{n} W_i R_i}{\sum_{i=1}^{n} W_i} \; .$$

- *Biased* (though bias vanishes as $n \to \infty$)
- Empirically much better than IS; "low variance"
  Efron-Stein + calculation:

$$\mathrm{Var}(U^{\mathrm{WIS}}) \leq 4\, \underbrace{\mathbb{E}\left[\sum_k \left(\frac{W_k}{\sum_i W_i}\right)^2\right]}_{=: \frac{1}{n_{\mathrm{eff}}}}$$

- How do we get value lower bounds?

Semi-empirical Efron-Stein bound for WIS

WIS value estimate:

$$U^{\mathrm{WIS}} = \frac{1}{Z} \sum_{i=1}^{n} W_i R_i \ , \qquad Z = \sum_{i=1}^{n} W_i \ .$$

Let

$$V = \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{W_k}{Z} + \frac{W'_k}{Z^{(k)}} \right)^2 \,\middle|\, W_1^k, X_1^n \right] \qquad \text{("variance")}$$

$$\beta = \min\left(\mathbb{E}\left[\frac{n}{Z} \,\middle|\, X_1^n \right]^{-1}, 1 \right) \ . \qquad \text{(bias)}$$

Theorem ([KVGS21])

W.h.p.,

$$u(\pi) \geq \left(\beta \cdot \left(U^{\mathrm{WIS}} - \sqrt{c \cdot \left(V + \frac{1}{n} \right)} \right) - \frac{c'}{\sqrt{n}} \right)_+$$

where $Z^{(k)} = Z + (W'_k - W_k)$, *and* W'_k *indep. dist. as* W_k.

Semi-empirical Efron-Stein bound for WIS

$$u(\pi) \geq \left(\beta \cdot \left(U^{\mathrm{WIS}} - \sqrt{c \cdot \left(V + \frac{1}{n} \right)} \right) - \frac{c'}{\sqrt{n}} \right)_+$$

$$V = \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{W_k}{Z} + \frac{W_k'}{Z^{(k)}} \right)^2 \,\middle|\, W_1^k, X_1^n \right]$$

$$\beta = \min\left(\mathbb{E}\left[\frac{n}{Z} \,\middle|\, X_1^n \right]^{-1}, 1 \right)$$

$$Z^{(k)} = Z + (W_k' - W_k), \text{ and } W_k' \text{ indep. dist. as } W_k$$

- No truncation! No hyperparameters.
- Contexts are fixed.
- Needs knowledge of π_b — only partly empirical:

  V and β can be computed exactly. Cost: n^K :-(
  Can approximate using Monte-Carlo simulation! :-)
  ... and is "pretty good"!

Proof sketch

Let $u(\pi|X_1^n) := \frac{1}{n}\sum_{i=1}^n \sum_a \pi(a|X_i) r(X_i, a)$.
Then $u(\pi) - U^{\mathrm{WIS}} =$

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Context concentration}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias (fixed } X_1^n)} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

Proof sketch

Let $u(\pi|X_1^n) := \frac{1}{n}\sum_{i=1}^n \sum_a \pi(a|X_i) r(X_i, a)$.
Then $u(\pi) - U^{\mathrm{WIS}} =$

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Context concentration}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias (fixed } X_1^n\text{)}} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

1. Context concentration: Hoeffding

Proof sketch

Let $u(\pi|X_1^n) := \frac{1}{n}\sum_{i=1}^n \sum_a \pi(a|X_i) r(X_i, a)$.
Then $u(\pi) - U^{\mathrm{WIS}} =$

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Context concentration}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias (fixed } X_1^n)} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

1. Context concentration: Hoeffding
2. Bias:

$$\begin{aligned}
\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] &= \mathbb{E}\left[\frac{\sum_{k=1}^n W_k r(X_k, A_k)}{\sum_{k=1}^n W_k} \,\middle|\, X_1^n\right] \\
&\leq \mathbb{E}\left[\frac{1}{\sum_{k=1}^n W_k} \,\middle|\, X_1^n\right] \mathbb{E}\left[\sum_{k=1}^n W_k r(X_k, A_k) \,\middle|\, X_1^n\right] \\
&= \mathbb{E}\left[\frac{n}{\sum_{k=1}^n W_k} \,\middle|\, X_1^n\right] u(\pi, X_1^n)
\end{aligned}$$

Proof: $\sim$ Harris' inequality.

Proof sketch

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Concentration of contexts}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias}} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

Concentration

(Remember) some challenges

- Even for basic importance sampling $(W_1 R_1 + \cdots + W_n R_n)/n$ it's non-trivial: unbiased, but W_i are **unbounded**
  - Excludes Hoeffding/Bernstein/McDiarmid
  - We can "truncate", e.g. $W_i^\lambda = \pi(A_i|X_i)/(\pi_b(A_i|X_i) + \lambda)$ for some h.p. $\lambda > 0$.
  - Ugly! Needs tuning, doesn't always work...

Proof sketch

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Concentration of contexts}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias}} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

Concentration

(Remember) some challenges

- Even for basic importance sampling $(W_1 R_1 + \cdots + W_n R_n)/n$ it's non-trivial: unbiased, but W_i are **unbounded**
  - Excludes Hoeffding/Bernstein/McDiarmid
  - We can "truncate", e.g. $W_i^\lambda = \pi(A_i|X_i)/(\pi_b(A_i|X_i) + \lambda)$ for some h.p. $\lambda > 0$.
  - Ugly! Needs tuning, doesn't always work...

Proof sketch

$$\underbrace{u(\pi) - u(\pi|X_1^n)}_{\text{Concentration of contexts}} + \underbrace{u(\pi|X_1^n) - \mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right]}_{\text{Bias}} + \underbrace{\mathbb{E}\left[U^{\mathrm{WIS}} \mid X_1^n\right] - U^{\mathrm{WIS}}}_{\text{Concentration}}$$

Concentration

(Remember) some challenges

- Even for basic importance sampling $(W_1 R_1 + \cdots + W_n R_n)/n$ it's non-trivial: unbiased, but W_i are **unbounded**
  - Excludes Hoeffding/Bernstein/McDiarmid
  - We can "truncate", e.g. $W_i^\lambda = \pi(A_i|X_i)/(\pi_b(A_i|X_i) + \lambda)$ for some h.p. $\lambda > 0$.
  - Ugly! Needs tuning, doesn't always work...
- Variance is important: need bounds with empirical variance.
- Sometimes, estimator is not a sum of indep. elements (self-normalization).

(New) Efron-Stein tail bound

Theorem ([KS19, KS21])

Assume elements of $S = (X_1, X_2, \dots, X_n)$ *are independent, and let*

$$\Delta = f(S) - \mathbb{E}[f(S)]\ , \quad V = \sum_{k=1}^{n} \mathbb{E}\left[(f(S) - f(S^{(k)}))^2 \,\middle|\, X_1, \dots, X_k\right]\ .$$

Then, for any $x \geq 0,\ y > 0$*, w.p.* $1 - e^{-x}$*,*

$$|\Delta| < \sqrt{2(V + y)\left(x + \frac{1}{2}\ln(1 + V/y)\right)}$$

Application to WIS tail bounds

Take $f = U^{\mathrm{WIS}}$, condition on X_1^n. Algebra gives that V obeys

$$V \leq \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{W_k}{Z} + \frac{W_k'}{Z^{(k)}} \right)^2 \,\middle|\, W_1^k, X_1^n \right] .$$

Choose $y = 1/n$.

Proof of Efron-Stein tail bound

Step #1: $(\Delta, \sqrt{V})$ is a canonical pair

Step #2: Use self-normalized bounds available for canonical pairs

Canonical pairs – [dlPLS08]

We call (A, B) a canonical pair if $B \geq 0$ and

$$\sup_{\lambda \in \mathbb{R}} \mathbb{E}\left[\exp\left(\lambda A - \frac{\lambda^2}{2} B^2\right)\right] \leq 1 \; .$$

Step #2: Tail bounds for canonical pairs

Let (A, B) be a canonical pair.

Theorem (Thm 2.7 of [dlPLS08])

For all $x > 0$, w.p. $1 - \sqrt{2}e^{-x}$,

$$|A| < 2\sqrt{x(B^2 + (\mathbb{E}[B])^2)}$$

Step #2: Tail bounds for canonical pairs

Let (A, B) be a canonical pair.

Theorem (Thm 2.7 of [dlPLS08])

For all $x > 0$, w.p. $1 - \sqrt{2}e^{-x}$,

$$|A| < 2\sqrt{x(B^2 + (\mathbb{E}[B])^2)}$$

Theorem ([KS21])

For all $x \geq 0$ and $y > 0$, w.p. $1 - e^{-x}$,

$$|A| < \sqrt{2(B^2 + y)\left(x + \frac{1}{2}\ln\left(1 + \frac{B^2}{y}\right)\right)}$$

Proof of 2nd result: Method of mixtures

Proof.

Markov: For $x > 0$, w.p. $1 - e^{-x}$, $X < \ln \mathbb{E}[e^X] + x$.

$\square$

Proof of 2nd result: Method of mixtures

Proof.

Markov: For $x > 0$, w.p. $1 - e^{-x}$, $X < \ln \mathbb{E}[e^X] + x$.
Let $\Lambda \sim \mathcal{N}(0, \sigma^2)$, $\Lambda \perp (A, B)$.
Choose

$$X = \ln \mathbb{E}\left[e^{\Lambda A - \frac{\Lambda^2}{2} B^2} \,\middle|\, A, B \right]$$

Apply previous inequality, calculate (on the RHS use Fubini).
Set $y = 1/\sigma^2$. $\square$

Proof of 2nd result: Method of mixtures

Proof.

Markov: For $x > 0$, w.p. $1 - e^{-x}$, $X < \ln \mathbb{E}[e^X] + x$.
Let $\Lambda \sim \mathcal{N}(0, \sigma^2)$, $\Lambda \perp (A, B)$.
Choose

$$X = \ln \mathbb{E}\left[e^{\Lambda A - \frac{\Lambda^2}{2} B^2} \,\middle|\, A, B \right]$$

Apply previous inequality, calculate (on the RHS use Fubini).
Set $y = 1/\sigma^2$. $\square$

Note: Thm 12.4 of [dlPLS08] is almost the same, the proof here is shorter and the result is slightly improved.

Step #1: $(\Delta, \sqrt{V})$ is a canonical pair. Part I

Let $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot \mid X_1, \ldots, X_k]$. Recall

$$\Delta = f(S) - \mathbb{E}[f(S)] \ , \qquad V = \sum_{k=1}^{n} \underbrace{\mathbb{E}_k\left[(f(S) - f(S^{(k)}))^2\right]}_{=: V_k} \ .$$

Step #1: $(\Delta, \sqrt{V})$ is a canonical pair. Part I

Let $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot \mid X_1, \ldots, X_k]$. Recall

$$\Delta = f(S) - \mathbb{E}[f(S)] \ , \qquad V = \sum_{k=1}^{n} \underbrace{\mathbb{E}_k\left[(f(S) - f(S^{(k)}))^2\right]}_{=: V_k} \ .$$

Proof: We have

$$\Delta = \sum_{k=1}^{n} D_k \qquad \text{and} \qquad V = \sum_{k=1}^{n} V_k$$

where

$$D_k = \mathbb{E}_k[f(S) - f(S^{(k)})]$$

Step #1: $(\Delta, \sqrt{V})$ is a canonical pair. Part I

Let $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot \mid X_1, \ldots, X_k]$. Recall

$$\Delta = f(S) - \mathbb{E}[f(S)] \ , \qquad V = \sum_{k=1}^{n} \underbrace{\mathbb{E}_k \left[(f(S) - f(S^{(k)}))^2 \right]}_{=: V_k} \ .$$

Proof: We have

$$\Delta = \sum_{k=1}^{n} D_k \qquad \text{and} \qquad V = \sum_{k=1}^{n} V_k$$

where

$$D_k = \mathbb{E}_k[f(S) - f(S^{(k)})]$$

Indeed, $\mathbb{E}_{k-1}[f(S)] = \mathbb{E}_k[f(S^{(k)})]$, so
$D_k = \mathbb{E}_k[f(S)] - \mathbb{E}_{k-1}[f(S)]$, use telescoping.

Proof of Step #1: Part II

Assume for now

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1 \quad \text{a.s.} \forall k \in [n] \qquad (1)$$

Proof of Step #1: Part II

Assume for now

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1 \quad \text{a.s.} \forall k \in [n] \tag{1}$$

Then

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \Delta - \frac{\lambda^2}{2} V\right)\right] \\
&\quad = \mathbb{E}\left[\underbrace{\mathbb{E}_{n-1}\left[\exp\left(\lambda D_n - \frac{\lambda^2}{2} V_n\right)\right]}_{\leq 1 \text{ a.s}} \prod_{k=1}^{n-1} \exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \\
&\quad \leq \mathbb{E}\left[\underbrace{\mathbb{E}_{n-2}\left[\exp\left(\lambda D_{n-1} - \frac{\lambda^2}{2} V_{n-1}\right)\right]}_{\leq 1 \text{ a.s}} \prod_{k=1}^{n-2} \exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \\
&\quad \leq \cdots \leq 1\,.
\end{aligned}
$$

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1$ a.s.

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k]$, $V_k = \mathbb{E}_k[\Delta_k]$.

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k],\ V_k = \mathbb{E}_k[\Delta_k]$.

Jensen:

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq \mathbb{E}_{k-1}\cancel{\mathbb{E}_k}\left[\exp\left(\lambda \Delta_k - \frac{\lambda^2}{2} \Delta_k^2\right)\right]$$

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k]$, $V_k = \mathbb{E}_k[\Delta_k]$.

Jensen:

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \leq \mathbb{E}_{k-1} \cancel{\mathbb{E}_k} \left[\exp\left(\lambda \Delta_k - \frac{\lambda^2}{2} \Delta_k^2\right)\right]$$

Let $S_{-k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$,

$$F_k(s) = \exp\left(\lambda(s\Delta_k) - \frac{\lambda^2}{2}(s\Delta_k)^2\right), \qquad s \in \mathbb{R}$$

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2}V_k\right)\right] \leq 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k]$, $V_k = \mathbb{E}_k[\Delta_k]$.

Jensen:

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2}V_k\right)\right] \leq \mathbb{E}_{k-1}\cancel{\mathbb{E}_k}\left[\exp\left(\lambda \Delta_k - \frac{\lambda^2}{2}\Delta_k^2\right)\right]$$

Let $S_{-k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$,

$$F_k(s) = \exp\left(\lambda(s\Delta_k) - \frac{\lambda^2}{2}(s\Delta_k)^2\right), \qquad s \in \mathbb{R}$$

By def. of S', $P_{\Delta_k|S_{-k}} = P_{-\Delta_k|S_{-k}} \Rightarrow$ for $\varepsilon \sim$ Rad, $\varepsilon \perp S, S'$, $P_{F_k(1)|S_{-k}} = P_{F_k(\varepsilon)|S_{-k}}$

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \le 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k],\ V_k = \mathbb{E}_k[\Delta_k]$.
Jensen:

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \le \mathbb{E}_{k-1} \cancel{\mathbb{E}_k}\left[\exp\left(\lambda \Delta_k - \frac{\lambda^2}{2}\Delta_k^2\right)\right]$$

Let $S_{-k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$,

$$F_k(s) = \exp\left(\lambda(s\Delta_k) - \frac{\lambda^2}{2}(s\Delta_k)^2\right), \qquad s \in \mathbb{R}$$

By def. of S', $P_{\Delta_k|S_{-k}} = P_{-\Delta_k|S_{-k}} \Rightarrow$ for $\varepsilon \sim$ Rad, $\varepsilon \perp S, S'$,
$P_{F_k(1)|S_{-k}} = P_{F_k(\varepsilon)|S_{-k}}$ Thus,

$$\mathbb{E}\left[F_k(1)|S_{-k}\right] = \mathbb{E}\left[F_k(\varepsilon)|S_{-k}\right] \qquad \text{(symmetrization)}$$

Proof of Step #1: Part III

Claim: $\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \le 1$ a.s.

Proof: $\Delta_k := f(S) - f(S^{(k)}) \Rightarrow D_k = \mathbb{E}_k[\Delta_k],\ V_k = \mathbb{E}_k[\Delta_k]$.

Jensen:

$$\mathbb{E}_{k-1}\left[\exp\left(\lambda D_k - \frac{\lambda^2}{2} V_k\right)\right] \le \mathbb{E}_{k-1} \cancel{\mathbb{E}_k} \left[\exp\left(\lambda \Delta_k - \frac{\lambda^2}{2} \Delta_k^2\right)\right]$$

Let $S_{-k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$,

$$F_k(s) = \exp\left(\lambda(s\Delta_k) - \frac{\lambda^2}{2}(s\Delta_k)^2\right), \qquad s \in \mathbb{R}$$

By def. of S', $P_{\Delta_k|S_{-k}} = P_{-\Delta_k|S_{-k}} \Rightarrow$ for $\varepsilon \sim$ Rad, $\varepsilon \perp S, S'$, $P_{F_k(1)|S_{-k}} = P_{F_k(\varepsilon)|S_{-k}}$ Thus,

$$\mathbb{E}\left[F_k(1)|S_{-k}\right] = \mathbb{E}\left[F_k(\varepsilon)|S_{-k}\right] \qquad \text{(symmetrization)}$$

and since $x\varepsilon$ is $x^2/2$-subgaussian for $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}_{k-1} F_k(1) &= \mathbb{E}_{k-1}\mathbb{E}\left[F_k(1)|S_{-k}\right] = \mathbb{E}_{k-1}\mathbb{E}\left[F_k(\varepsilon)|S_{-k}\right] \\ &= \mathbb{E}_{k-1}\mathbb{E}\left[F_k(\varepsilon)|S, S'\right] \le 1\,. \quad \square\end{aligned}$$

Conclusions

- Nontrivial tail bounds for the weighted importance sampling (WIS) estimator
  - Bias: Harris inequality
  - Concentration: Novel concentration $\leq$ using an Efron-Stein variance proxy
- PAC-Bayes variants
- Proof: self-normalized inequalities using canonical pairs
- Bandit value estimation: Exploit small $\mathrm{Var}[R]$?
- Other applications?

References

[dlPLS08] V. H. de la Peña, T. L. Lai, and Q.-M. Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science & Business Media, 2008.

[KS19] I. Kuzborskij and C. Szepesvári. Efron-Stein PAC-Bayesian Inequalities. arXiv:1909.01931, 2019.

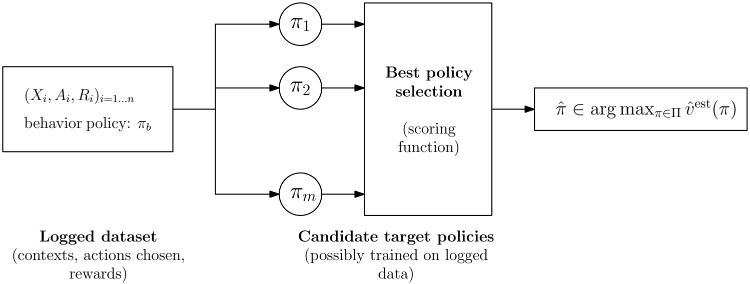
[KS21] I. Kuzborskij and C. Szepesvári. Semi-empirical Efron-Stein concentration inequalities, PAC-Bayes, and applications. under submission, 2021.

[KVGS21] I. Kuzborskij, C. Vernade, A. György, and Cs. Szepesvári. Confident off-policy evaluation and selection through self-normalized importance weighting. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

Is it any good?

The Best Policy Identification problem

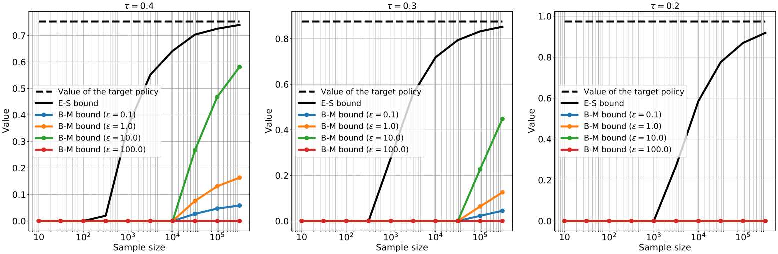
- We have a finite set of target policies Π.
- We do $\hat{\pi} \in \arg\max_{\pi \in \Pi} \hat{v}^{\text{est}}(\pi)$.
- We want to maximize $u(\hat{\pi})$
  — we'll use confidence bounds as $\hat{v}^{\text{est}}$.



Synthetic experiments – setup

- Fix $K > 0$, $\tau > 0$
- $\pi_b(a) \propto e^{\frac{1}{\tau}\mathbb{I}\{a=1\}}$
- $\pi(a) \propto e^{\frac{1}{\tau}\mathbb{I}\{a=2\}}$
- $R_i = \mathbb{I}\{A_i = k\}$, $A_i \sim \pi_b(\cdot)$
- As $\tau \to 0$, π_b and π become increasingly misaligned

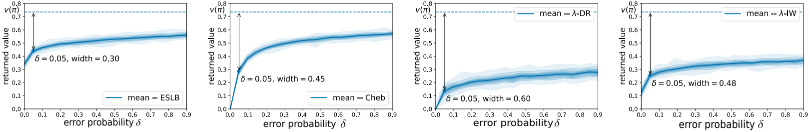
Results



E-S — Our bound

B-M — Empirical Bernstein's bound with ε-truncated weights

Numerical tightness in error probability



Similar setup as before, sample size $= 10^4$, left to right:

- E-S — our bound.
- Chebyshev's ineq.-based CI for WIS.
- Empirical Bernstein's ineq.-based CI for DR estimator with $W_i^\lambda = \frac{\pi(A_i|X_i)}{\pi_b(A_i|X_i)+\lambda}$ for some $\lambda = 1/\sqrt{n}$.
- Empirical Bernstein's ineq.-based CI for IS with W_i^λ.

Nonsynthetic experiments – setup

Target policies are $\left\{\pi^{\mathrm{ideal}}, \pi^{\hat{\mathbf{\Theta}}_{\mathsf{IS}}}, \pi^{\hat{\mathbf{\Theta}}_{\mathsf{WIS}}}\right\}$ where

$$\pi^{\mathbf{\Theta}}(y = k \mid \boldsymbol{x}) \propto e^{\frac{1}{\tau}\boldsymbol{x}^\top \boldsymbol{\theta}_k}$$

with two choices of parameters given by the optimization problems:

$$\hat{\mathbf{\Theta}}_{\mathsf{IS}} \in \operatorname*{arg\,min}_{\mathbf{\Theta} \in \mathbb{R}^{d \times K}} U^{\mathrm{IS}}(\pi^{\mathbf{\Theta}}) \ , \qquad \hat{\mathbf{\Theta}}_{\mathsf{WIS}} \in \operatorname*{arg\,min}_{\mathbf{\Theta} \in \mathbb{R}^{d \times K}} U^{\mathrm{WIS}}(\pi^{\mathbf{\Theta}}) \ .$$

- Trained by GD with $\eta = 0.01, \ T = 10^5$.
- $\tau = 0.1$ — cold! Almost deterministic.

Table: Average test rewards of the target policy when chosen by each method of the benchmark.

name Size	Ecoli 336	Vehicle 846	Yeast 1484	PageBlok 5473	OptDigits 5620	SatImage 6435	PenDigits 10992
ESLB	**0.913 ± 0.263**	**0.716 ± 0.389**	**0.912 ± 0.267**	**0.910 ± 0.270**	**0.843 ± 0.325**	**0.910 ± 0.270**	**0.910 ± 0.270**
DR	0.656 ± 0.410	0.610 ± 0.443	0.563 ± 0.392	0.888 ± 0.291	0.616 ± 0.344	0.423 ± 0.361	0.565 ± 0.382
IS (trunc+Bern)	$-\infty$	$-\infty$	**0.916 ± 0.262**	**0.910 ± 0.270**	0.748 ± 0.404	0.658 ± 0.413	0.810 ± 0.345
Chebyshev-WIS	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
Emp.Lik.	0.511 ± 0.298	0.455 ± 0.405	0.312 ± 0.325	0.669 ± 0.409	0.285 ± 0.359	0.634 ± 0.409	0.549 ± 0.426