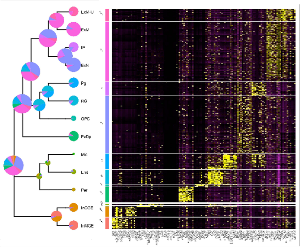
Breaking the sample size barrier in statistical inference and reinforcement learning

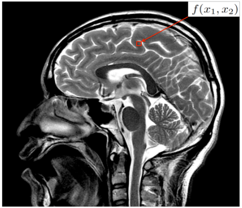
Yuting Wei

Carnegie Mellon University

Princeton, Dec 2020

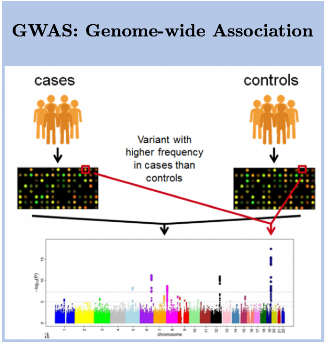
Ubiquity of sample-starved information discovery

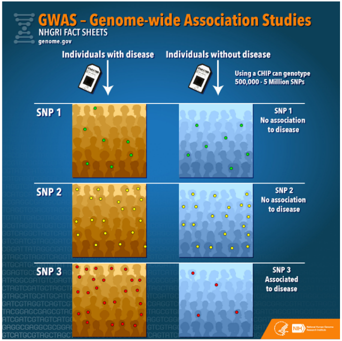




The explosive growth of features outpaces the growth of data samples

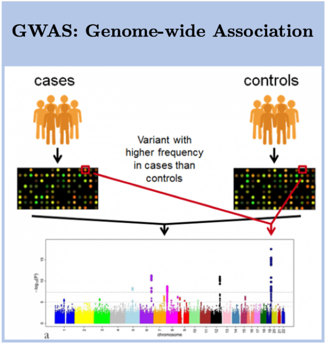
Example: statistical inference in genomics

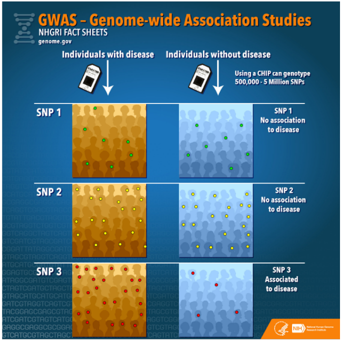




More variables (i.e., genetic variants) than observations (i.e., individuals)

Example: statistical inference in genomics





More variables (i.e., genetic variants) than observations (i.e., individuals)

- lessons from modern statistics: exploit signal sparsity

Example: statistical inference in genomics

Leading Edge
Perspective

Cell

An Expanded View of Complex Traits: From Polygenic to Omnigenic

Evan A. Boyle,[1,*] Yang I. Li,[1,*] and Jonathan K. Pritchard[1,2,3,*]
[1]Department of Genetics
[2]Department of Biology
[3]Howard Hughes Medical Institute
Stanford University, Stanford, CA 94305, USA

matin regions of immune cells (Maurano et al.; 2012; Farh et al., 2015; Kundaje et al., 2015).

These observations are generally interpreted in a paradigm in which complex disease is driven by an accumulation of weak effects on the key genes and regulatory pathways that drive disease risk (Furlong, 2013; Chakravarti and Turner, 2016). This model has motivated many studies that aim to dissect the functional impacts of individual disease-associated variants

True signals might NOT be ultra-sparse

$\longrightarrow$ we have to deal with the sample-limited regime

Example: reinforcement learning (RL)

In RL, an agent learns by interacting with an environment

- decision making in the face of uncertainty (unknown environments)
- enormous state and action spaces

Example: reinforcement learning (RL)

Collecting data samples might be expensive or time-consuming

clinical trials

online ads

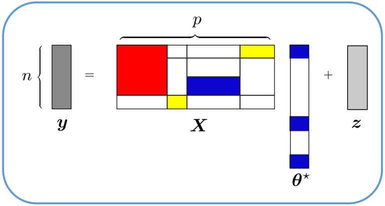
Calls for design of sample-efficient RL algorithms!

A central theme of this talk

Enabling trustworthy inference and learning in sample-starved scenarios

A central theme of this talk

Enabling trustworthy inference and learning in sample-starved scenarios



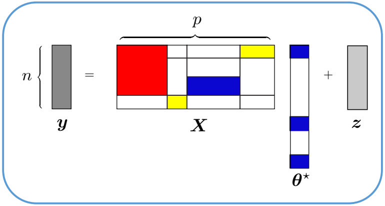
Two vignettes:

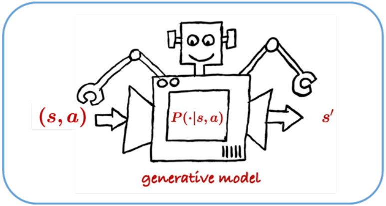
1. Distribution of Lasso with general designs

    — sample-efficient inference via a precise distributional theory

A central theme of this talk

Enabling trustworthy inference and learning in sample-starved scenarios





Two vignettes:

1. Distribution of Lasso with general designs
   — sample-efficient inference via a precise distributional theory
2. Reinforcement learning with a generative model
   — optimal sample efficiency via a model-based approach

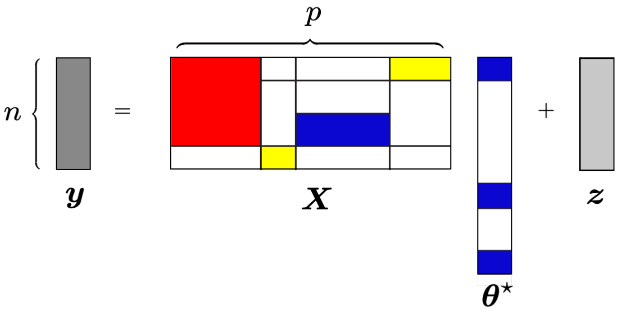
The first vignette: Distribution of Lasso with general designs

Michael Celentano
Stanford Stat

Andrea Montanari
Stanford Stat & EE

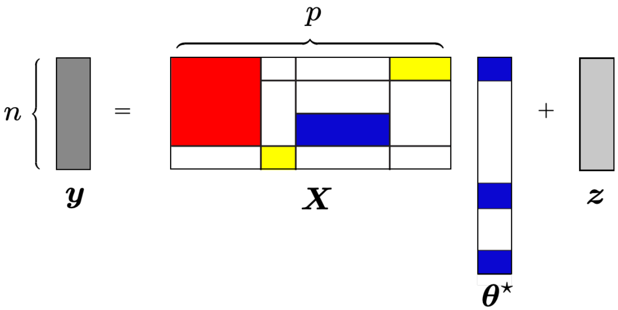
"The Lasso with general Gaussian designs with application to hypothesis testing,"
M. Celentano, A. Montanari, Y. Wei, 2020. https://arxiv.org/abs/2007.13716

Lasso estimator



$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$ [Tibshirani, 1996]

Lasso estimator



$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

[Tibshirani, 1996]

Statistical inference tasks: test $\theta_j^\star = 0$, or construct a confidence interval of $\theta_j^\star$, based on the Lasso estimate $\widehat{\boldsymbol{\theta}}$.

Prior work: Lasso estimation risk

Suppose $\boldsymbol{\theta}^\star$ is s-sparse, $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under certain conditions of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_2 \leq C\sigma\sqrt{\frac{s\log(p)}{n}}$$

Prior work: Lasso estimation risk

Suppose $\boldsymbol{\theta}^\star$ is s-sparse, $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under certain conditions of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_2 \leq C\sigma\sqrt{\frac{s\log(p)}{n}}$$

- unspecified (and possibly enormous) constant

Prior work: Lasso estimation risk

Suppose $\boldsymbol{\theta}^\star$ is s-sparse, $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under certain conditions of design matrix $\boldsymbol{X}$,

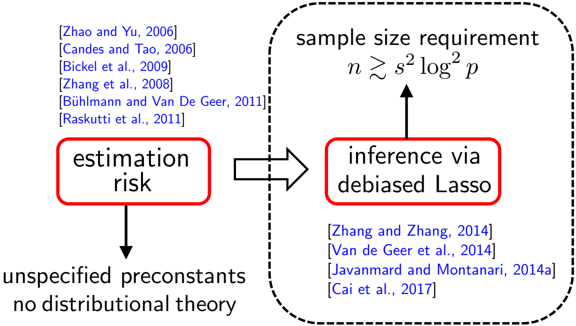
$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_2 \leq C\sigma\sqrt{\frac{s \log(p)}{n}}$$

- unspecified (and possibly enormous) constant
- no distributional characterization of $\widehat{\boldsymbol{\theta}}$

— inadequate for $\underbrace{\text{inference and uncertainty quantification}}_{\text{e.g., confidence intervals, hypothesis testing}}$

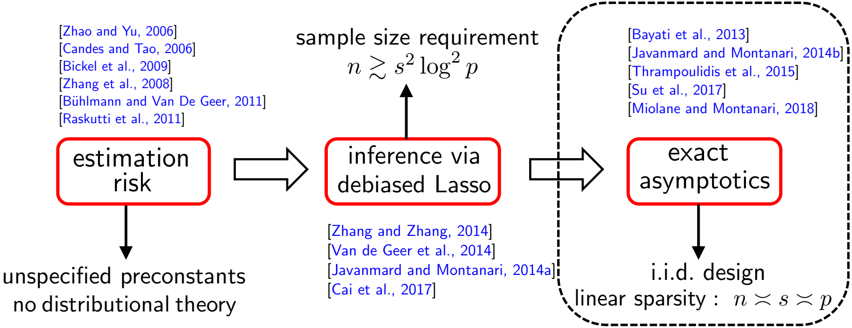
Prior work: inference for Lasso

Construction of confidence intervals via de-biased Lasso



Prior work: inference for Lasso

Tackling the most challenging regime ($n \asymp s$) via exact asymptotics



Prior work: exact asymptotics

Question: can we develop a distributional theory that covers both correlated design & linear sparsity $n/s = \text{const}$?

[Zhao and Yu, 2006]
[Candes and Tao, 2006]
[Bickel et al., 2009]
[Zhang et al., 2008]
[Bühlmann and Van De Geer, 2011]
[Raskutti et al., 2011]

estimation risk → unspecified preconstants, no distributional theory

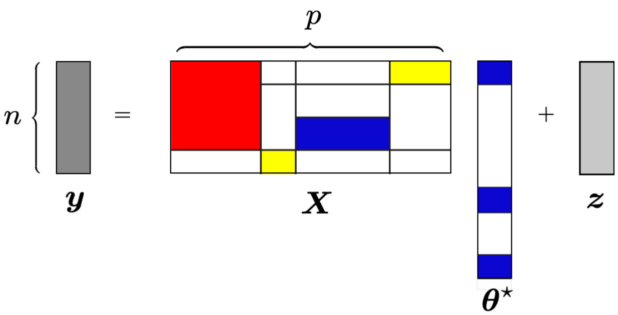
inference via debiased Lasso → sample size requirement $n \gtrsim s^2 \log^2 p$

[Zhang and Zhang, 2014]
[Van de Geer et al., 2014]
[Javanmard and Montanari, 2014a]
[Cai et al., 2017]

exact asymptotics → i.i.d. design, linear sparsity : $n \asymp s \asymp p$

[Bayati et al., 2013]
[Javanmard and Montanari, 2014b]
[Thrampoulidis et al., 2015]
[Su et al., 2017]
[Miolane and Montanari, 2018]

Settings



- $\boldsymbol{\theta}^\star \in \mathbb{R}^p$: s-sparse
- proportional regime: $p/n =$ const, $s/p =$ const
- Gaussian noise: $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$; Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \underbrace{\boldsymbol{\Sigma}/n}_{\text{known}})$

distributional theory
Lasso

$\Downarrow$

distributional theory
debiased Lasso

$\Downarrow$

inference
confidence interval

Key observation

original model
$\widehat{\boldsymbol{\theta}}$

- **original model (random design):** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^{\star} + \boldsymbol{z}$

Key observation

original model
$\widehat{\boldsymbol{\theta}}$

fixed design model
$\widehat{\boldsymbol{\theta}}^f$

- **original model (random design):** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^\star + \boldsymbol{z}$

- (auxiliary) **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

Key observation

original model $\widehat{\boldsymbol{\theta}}$ $\overset{\text{distribution}}{\approx}$ fixed design model $\widehat{\boldsymbol{\theta}}^f$

- **original model (random design):** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^\star + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- (auxiliary) **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{\zeta^\star}\|\boldsymbol{\theta}\|_1 \right\}$$

— $\tau^\star$: effective risk level $\quad$ $\zeta^\star$: effective non-sparsity

Key observation

original model $\widehat{\boldsymbol{\theta}}$ $\overset{\text{distribution}}{\approx}$ fixed design model $\widehat{\boldsymbol{\theta}}^f$

$\Downarrow$

its distribution is much easier to track!

- **original model (random design):** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^\star + \boldsymbol{z}$

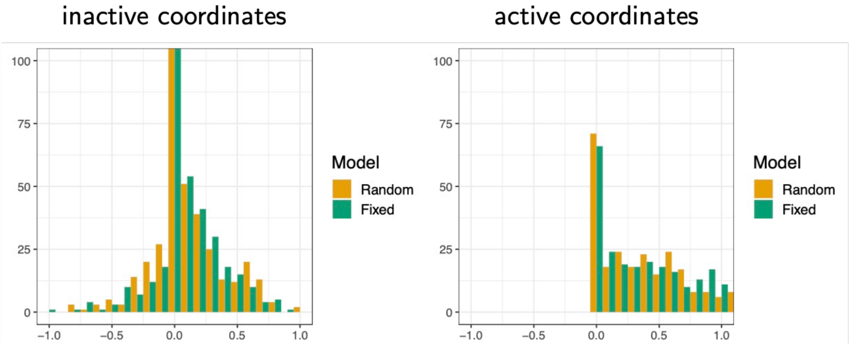
$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right\}$$

- (auxiliary) **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{2} \|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{\zeta^\star} \|\boldsymbol{\theta}\|_1 \right\}$$

— $\tau^\star$: effective risk level $\quad$ $\zeta^\star$: effective non-sparsity

Random designs behave like fixed design



Histogram of $\{\widehat{\theta}_j\}$ vs. histogram of $\{\widehat{\theta}_j^f\}$

Settings: auto-regressive design with $n = 1280, p = 2000, s = 256$, active coordinates $= 1$, λ chosen via cross validation.

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ *is sparse enough, for any* 1*-Lipschitz function* ϕ *and* $\epsilon > 0$

$$\left|\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda^f}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big)\Big]\right| \le \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ is sparse enough, for any 1-Lipschitz function ϕ and $\epsilon > 0$

$$\left|\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}^f_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big)\Big]\right| \leq \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

- informally, empirical-distribution$(\widehat{\boldsymbol{\theta}}_\lambda) \approx$ empirical-distribution$(\widehat{\boldsymbol{\theta}}^f_\lambda)$

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ is sparse enough, for any 1-Lipschitz function ϕ and $\epsilon > 0$

$$\left|\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda^f}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}}\Big)\Big]\right| \le \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

- informally, empirical-distribution($\widehat{\boldsymbol{\theta}}_\lambda$) $\approx$ empirical-distribution($\widehat{\boldsymbol{\theta}}_\lambda^f$)
- **a direct consequence:**

$$\|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^\star\|_2 \approx \mathbb{E}\Big[\|\widehat{\boldsymbol{\theta}}_\lambda^f - \boldsymbol{\theta}^\star\|_2\Big]$$

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ *is sparse enough, for any* 1*-Lipschitz function* ϕ *and* $\epsilon > 0$

$$\forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \qquad \left| \phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}} \Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda^f}{\sqrt{p}}, \frac{\boldsymbol{\theta}^\star}{\sqrt{p}} \Big) \Big] \right| \le \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

- informally, empirical-distribution$(\widehat{\boldsymbol{\theta}}_\lambda) \approx$ empirical-distribution$(\widehat{\boldsymbol{\theta}}_\lambda^f)$
- **a direct consequence:**

$$\forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \qquad \|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^\star\|_2 \approx \mathbb{E}\Big[\|\widehat{\boldsymbol{\theta}}_\lambda^f - \boldsymbol{\theta}^\star\|_2 \Big]$$

- uniform control over regularization parameter λ

  — useful for model selection

Main result: properties for Lasso

- Lasso residual

$$\mathbb{P}\left(\left|\frac{\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\theta}}\|_2}{\sqrt{n}}-\tau^\star\zeta^\star\right|>\epsilon\right)\leq\frac{C}{\epsilon^2}e^{-cn\epsilon^4}\,.$$

- Lasso sparsity

$$\mathbb{P}\left(\left|\frac{\|\widehat{\boldsymbol{\theta}}\|_0}{n}-(1-\zeta^\star)\right|>\epsilon\right)\leq\frac{C}{\epsilon^3}e^{-cn\epsilon^6}\,.$$

distributional theory
Lasso

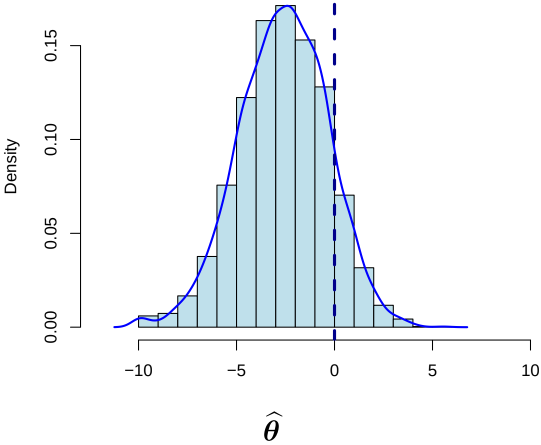
$$\Downarrow$$

distributional theory
debiased Lasso

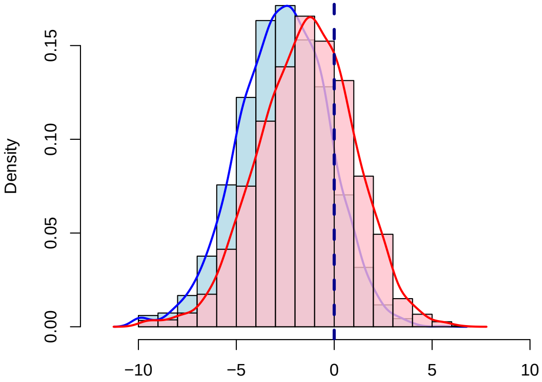
$$\Downarrow$$

inference
confidence interval

Debiased Lasso for statistical inference



Debiased Lasso for statistical inference

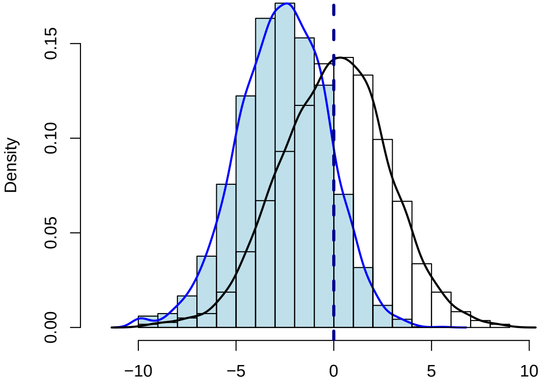


$$\widehat{\boldsymbol{\theta}}^d = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})$$

$\boldsymbol{M}$: surrogate for $\boldsymbol{\Sigma}^{-1} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i^\top]^{-1}$

[Zhang and Zhang, 2014, Van de Geer et al., 2014, Javanmard and Montanari, 2014a]

Debiased Lasso for statistical inference



$$\widehat{\boldsymbol{\theta}}^d = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})$$

$\boldsymbol{M}$: modified version $\boldsymbol{\Sigma}^{-1} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i^\top]^{-1}$

[Javanmard et al., 2018, Bellec and Zhang, 2019a, Bellec and Zhang, 2019b]

Debiased Lasso

- classical debiased Lasso

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{d}} = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$$

Debiased Lasso

- classical debiased Lasso

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{d}} = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$$

- debiased Lasso with degrees-of-freedom (DOF) adjustment

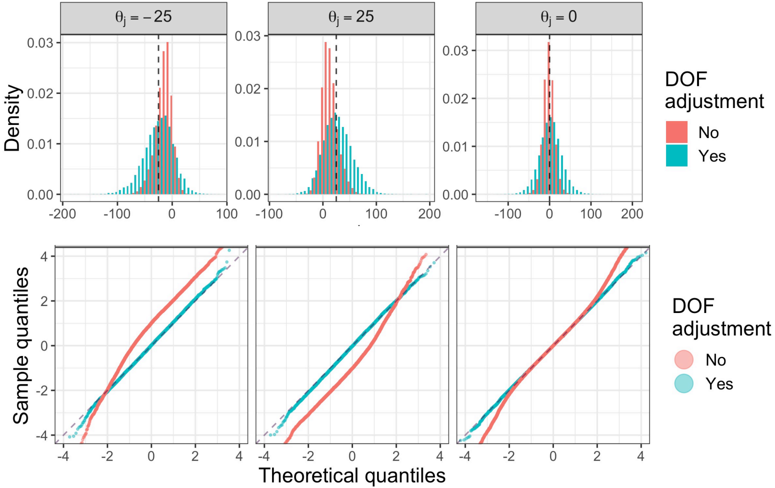
$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \frac{\boldsymbol{\Sigma}^{-1}}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}$$

[Javanmard and Montanari, 2014b, Miolane and Montanari, 2018, Bellec and Zhang, 2019a, Bellec and Zhang, 2019b]

Our result: distribution of $\widehat{\boldsymbol{\theta}}^{\mathrm{d}} \approx$ distribution of $\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{\Sigma}^{-1/2}\boldsymbol{g}$

— generalize prior result to general $\boldsymbol{\Sigma}$

Debiased Lasso with DOF adjustment



Settings: $p = 100,\ n = 25,\ s = 20,\ \Sigma_{ij} = 0.5^{|i-j|},\ \sigma = 1$

distributional theory
Lasso

$$\Downarrow$$

distributional theory
debiased Lasso

$$\Downarrow$$

inference
confidence interval

Degree-of-freedom adjustment is successful

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ is moderately sparse, false coverage proportion (FCP) satisfies

$$\mathbb{P}\left(|\mathsf{FCP} - \alpha| > \epsilon\right) \leq C(\epsilon)e^{-c(\epsilon)n}$$

for the target level $\alpha > 0$.

$$\mathsf{FCP} := \frac{1}{p}\sum_{j=1}^{p} \mathbb{1}\left\{\boldsymbol{\theta}_j^\star \notin \text{confidence-interval}_j\right\}$$

$$\text{confidence-interval}_j := \left[\widehat{\boldsymbol{\theta}}_j^d \ \pm \ \Sigma_{j|-j}^{-1/2}\widehat{\tau} \cdot z_{1-\alpha/2}\right]$$

Degree-of-freedom adjustment is successful

Theorem (Celetano, Montanari, Wei '20)

When $\boldsymbol{\theta}^\star$ is moderately sparse, false coverage proportion (FCP) satisfies

$$\mathbb{P}\left(|\mathsf{FCP} - \alpha| > \epsilon\right) \leq C(\epsilon) e^{-c(\epsilon)n}$$

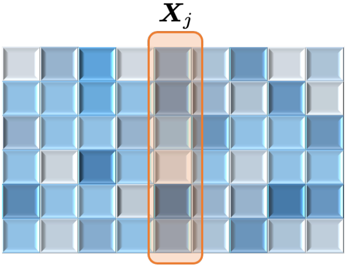
for the target level $\alpha > 0$.

$$\mathsf{FCP} := \frac{1}{p} \sum_{j=1}^{p} \mathbb{1}\left\{\boldsymbol{\theta}_j^\star \notin \text{confidence-interval}_j\right\}$$

— coverage only in the average sense!

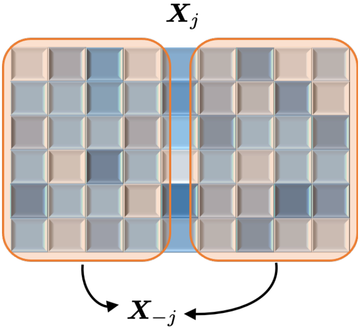
$$\text{confidence-interval}_j := \left[\widehat{\boldsymbol{\theta}}_j^d \ \pm \ \Sigma_{j|-j}^{-1/2} \widehat{\tau} \cdot z_{1-\alpha/2}\right]$$

Confidence interval for a single coordinate



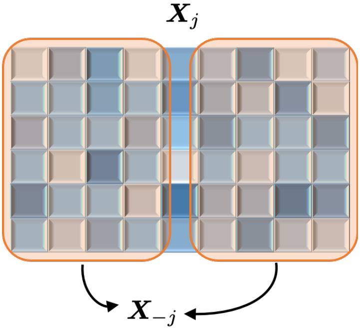
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$

Confidence interval for a single coordinate



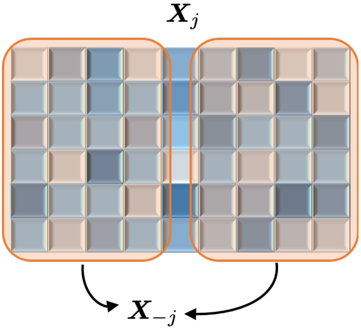
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$

Confidence interval for a single coordinate



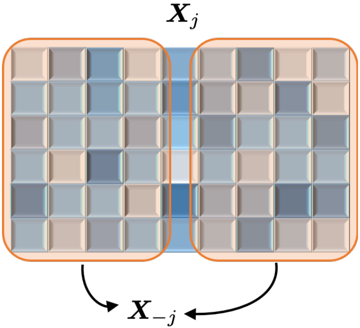
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\quad\longrightarrow\quad$ residual $\boldsymbol{X}_j^{\perp}$

Confidence interval for a single coordinate



- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\quad\longrightarrow\quad$ residual $\boldsymbol{X}_j^{\perp}$
- obtain leave-j^{th}-coordinate-out Lasso $\widehat{\boldsymbol{\theta}}_{\mathrm{loo}}$

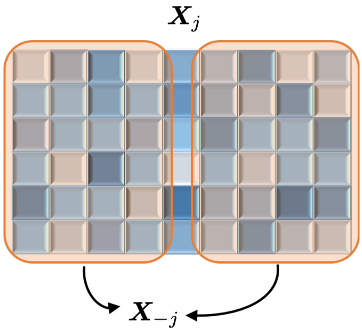
Confidence interval for a single coordinate



- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\quad\longrightarrow\quad$ residual $\boldsymbol{X}_j^{\perp}$
- obtain leave-j^{th}-coordinate-out Lasso $\widehat{\boldsymbol{\theta}}_{\text{loo}}$
- construct confidence interval $\quad \mathsf{CI}_j^{\text{loo}} := \left[\xi_j \;\pm\; \widehat{\mathsf{sd}} \cdot z_{1-\alpha/2}\right]$

$$\xi_j = \text{scaled correlation between } \boldsymbol{X}_j^{\perp} \text{ and } \boldsymbol{y} - \boldsymbol{X}_{-j}\widehat{\boldsymbol{\theta}}_{\text{loo}}$$

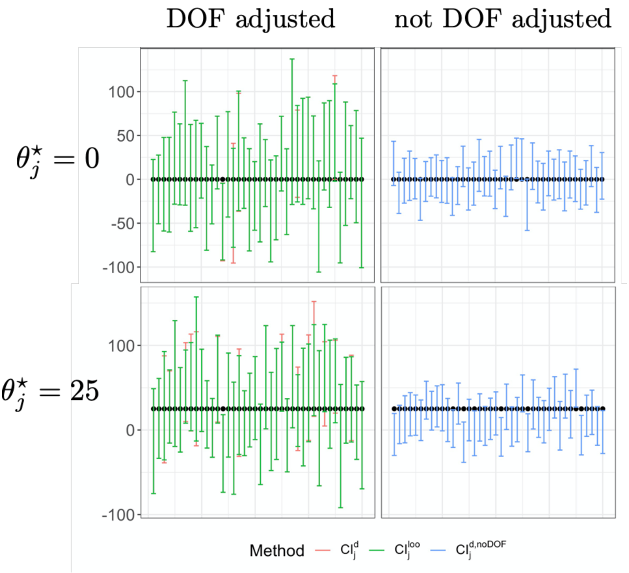
Confidence interval for a single coordinate



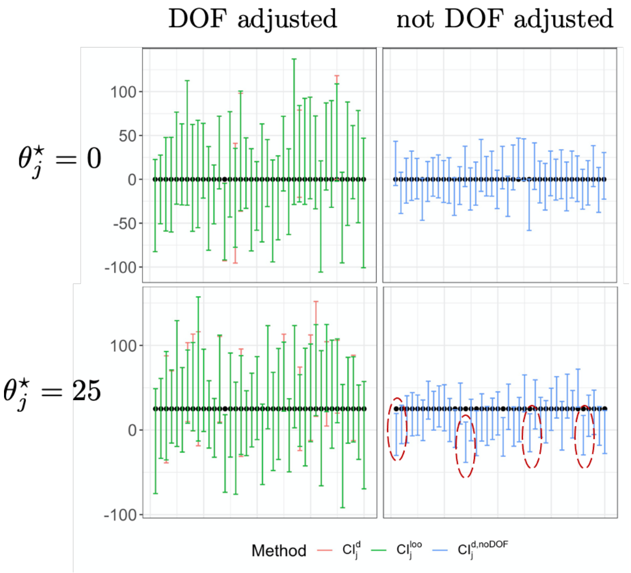
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\longrightarrow$ residual $\boldsymbol{X}_j^{\perp}$
- obtain leave-j^{th}-coordinate-out Lasso $\widehat{\boldsymbol{\theta}}_{\rm loo}$
- construct confidence interval $\mathsf{CI}_j^{\rm loo} := \left[\xi_j \pm \widehat{\mathsf{sd}} \cdot z_{1-\alpha/2}\right]$

Our theory: $\mathbb{P}_{\theta_j^*}\left(\theta_j^* \notin \mathsf{CI}_j^{\rm loo}\right) \approx \alpha$

Confidence interval for a single coordinate



Confidence interval for a single coordinate

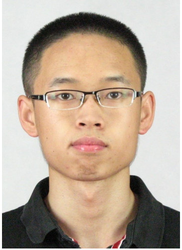


Summary of this part

- distributional theory of Lasso & debiased Lasso
  - general designs
  - sample-limited regime
- fine-grained confidence intervals with mis-coverage rate control

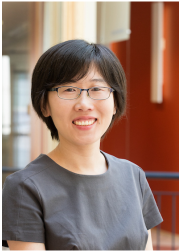
"The Lasso with general Gaussian designs with application to hypothesis testing,"
M. Celentano, A. Montanari, Y. Wei, 2020. https://arxiv.org/abs/2007.13716

The second vignette: RL with a generative model



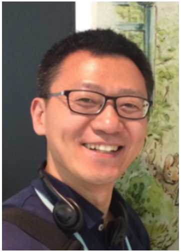
Gen Li

Tsinghua EE



Yuejie Chi

CMU ECE



Yuantao Gu

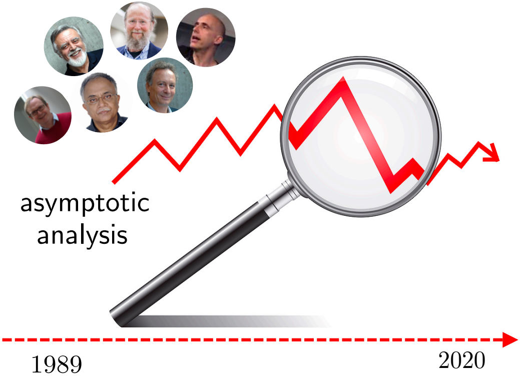
Tsinghua EE

Yuxin Chen

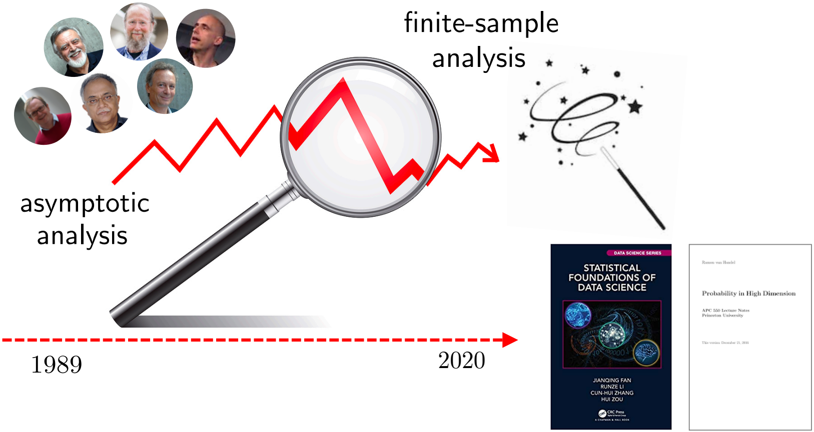
Princeton EE

"Breaking the sample size barrier in model-based reinforcement learning with a generative model," G. Li, Y. Wei, Y. Chi, Y. Gu, Y. Chen, NeurIPS 2020

Statistical foundation of reinforcement learning



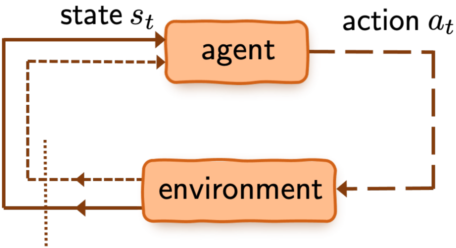
Statistical foundation of reinforcement learning



Understanding sample efficiency of modern RL requires a modern suite of non-asymptotic statistical framework

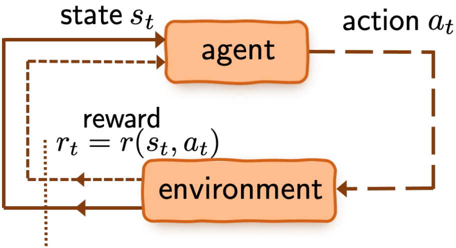
Background: Markov decision processes

Markov decision process (MDP)



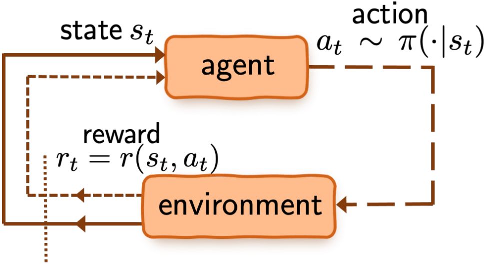
- $\mathcal{S}$: state space
- $\mathcal{A}$: action space

Markov decision process (MDP)



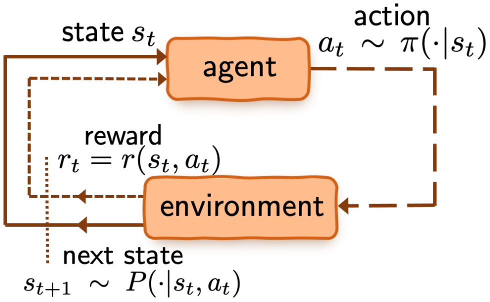
- $\mathcal{S}$: state space
- $\mathcal{A}$: action space
- $r(s,a) \in [0,1]$: immediate reward

Markov decision process (MDP)



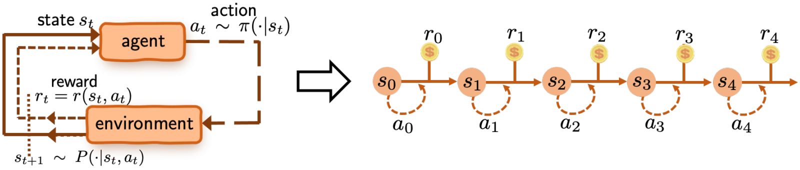
- $\mathcal{S}$: state space
- $\mathcal{A}$: action space
- $r(s, a) \in [0, 1]$: immediate reward
- $\pi(\cdot|s)$: policy (or action selection rule)

Markov decision process (MDP)



- $\mathcal{S}$: state space
- $\mathcal{A}$: action space
- $r(s, a) \in [0, 1]$: immediate reward
- $\pi(\cdot|s)$: policy (or action selection rule)
- $P(\cdot|s, a)$: unknown transition probabilities

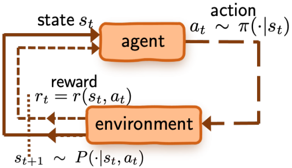
Value function

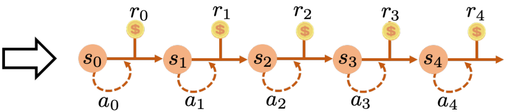


Value of policy π: cumulative discounted reward

$$\forall s \in \mathcal{S}: \quad V^\pi(s) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \,\middle|\, s_0 = s\right]$$

Value function



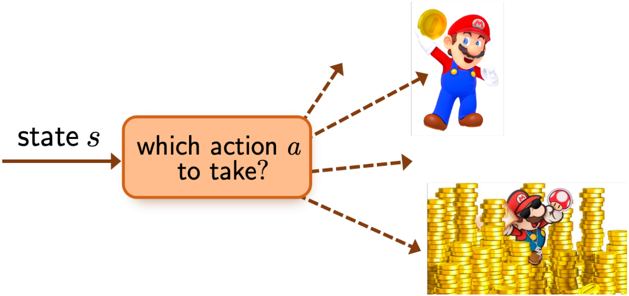


Value of policy π: cumulative discounted reward

$$\forall s \in \mathcal{S}: \quad V^\pi(s) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \,\middle|\, s_0 = s\right]$$

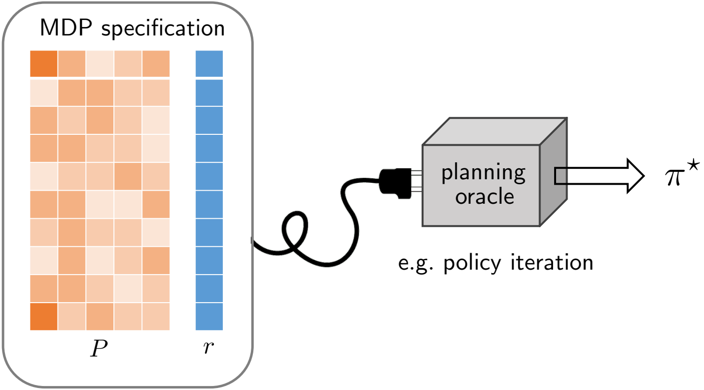
- $\gamma \in [0, 1)$: discount factor
  - take $\gamma \to 1$ to approximate long-horizon MDPs
  - effective horizon: $\frac{1}{1-\gamma}$

Optimal policy



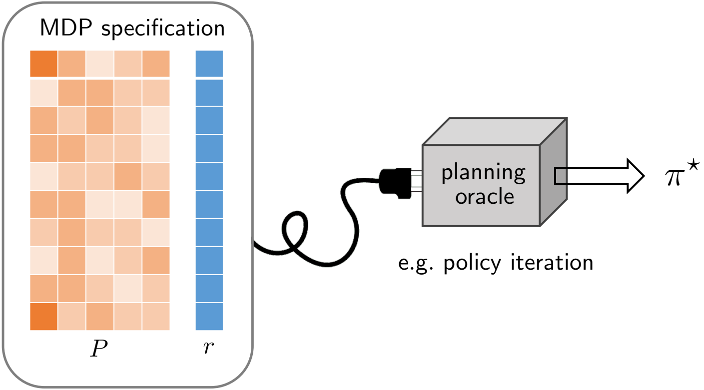
- **optimal policy** $\pi^\star$: maximizing value function $\max_\pi V^\pi(s)$
- How to find this $\pi^\star$?

When the model is known . . .



Planning: computing the optimal policy $\pi^\star$ given the MDP specification

When the model is known . . .

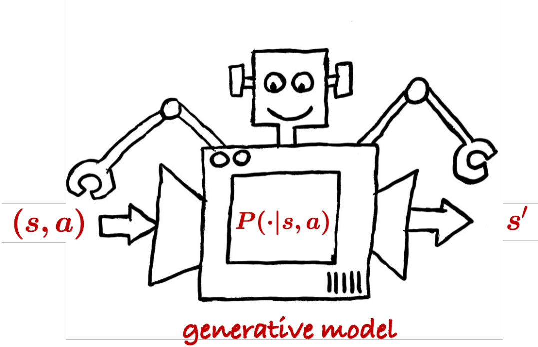


Planning: computing the optimal policy $\pi^\star$ given the MDP specification

In practice, do not know transition matrix P!

This work: sampling from a generative model

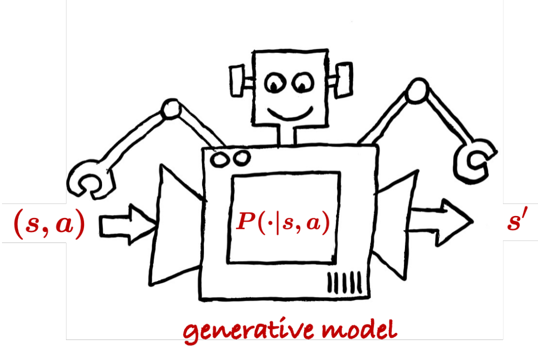
— [Kearns and Singh, 1999]



- **Sampling:** for each (s, a), collect N samples $\{(s, a, s'_{(i)})\}_{1 \le i \le N}$

This work: sampling from a generative model

— [Kearns and Singh, 1999]



- **Sampling:** for each (s, a), collect N samples $\{(s, a, s'_{(i)})\}_{1 \leq i \leq N}$
- construct $\widehat{\pi}$ depending on samples (in total $|\mathcal{S}||\mathcal{A}| \times N$)

Sample complexity: how many samples are required to learn an $\underbrace{\varepsilon\text{-optimal policy}}_{\forall s:\ V^{\hat{\pi}}(s) \geq V^{\star}(s) - \varepsilon}$?

An incomplete list of prior art

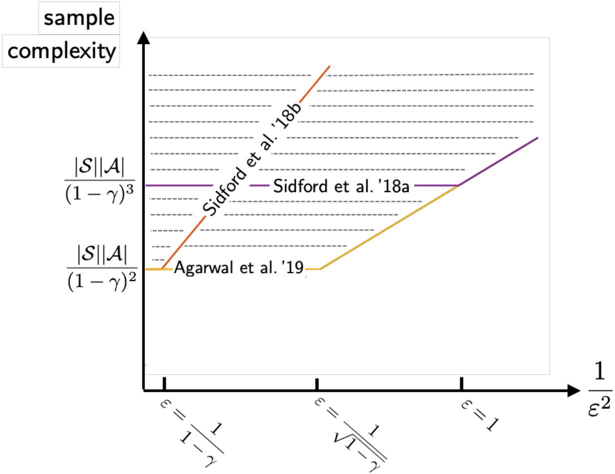
- [Kearns and Singh, 1999]
- [Kakade, 2003]
- [Kearns et al., 2002]
- [Azar et al., 2012]
- [Azar et al., 2013]
- [Sidford et al., 2018a]
- [Sidford et al., 2018b]
- [Wang, 2019]
- [Agarwal et al., 2019]
- [Wainwright, 2019a]
- [Wainwright, 2019b]
- [Pananjady and Wainwright, 2019]
- [Yang and Wang, 2019]
- [Khamaru et al., 2020]
- [Mou et al., 2020]
- ...

An even shorter list of prior art

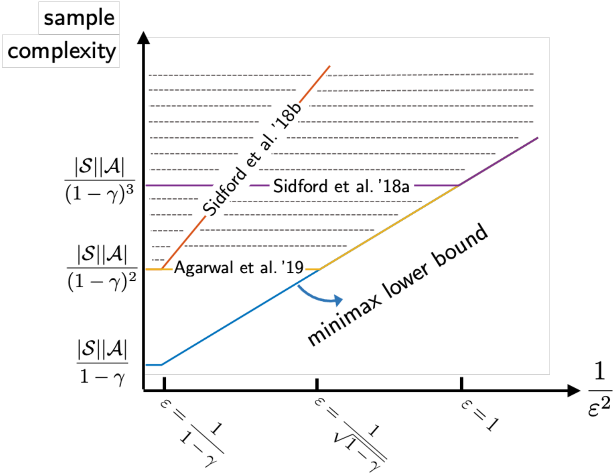
algorithm	sample size range	sample complexity	ε-range
Empirical QVI [Azar et al., 2013]	$\left[\frac{\lvert\mathcal{S}\rvert^2\lvert\mathcal{A}\rvert}{(1-\gamma)^2}, \infty\right)$	$\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^3\varepsilon^2}$	$\left(0, \frac{1}{\sqrt{(1-\gamma)\lvert\mathcal{S}\rvert}}\right]$
Sublinear randomized VI [Sidford et al., 2018b]	$\left[\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^2}, \infty\right)$	$\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^4\varepsilon^2}$	$\left(0, \frac{1}{1-\gamma}\right]$
Variance-reduced QVI [Sidford et al., 2018a]	$\left[\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^3}, \infty\right)$	$\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^3\varepsilon^2}$	$(0, 1]$
Randomized primal-dual [Wang, 2019]	$\left[\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^2}, \infty\right)$	$\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^4\varepsilon^2}$	$\left(0, \frac{1}{1-\gamma}\right]$
Empirical MDP + planning [Agarwal et al., 2019]	$\left[\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^2}, \infty\right)$	$\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^3\varepsilon^2}$	$\left(0, \frac{1}{\sqrt{1-\gamma}}\right]$

important parameters:

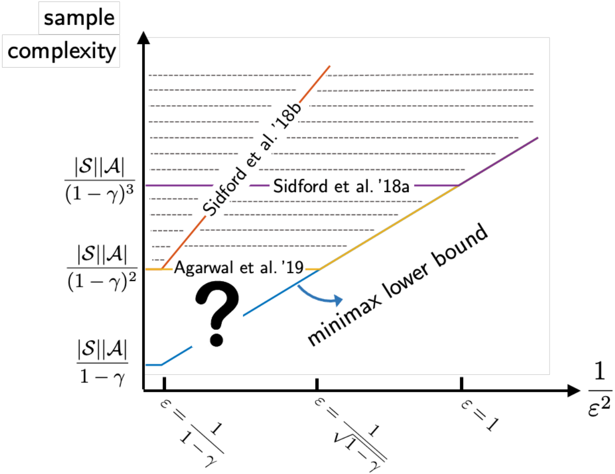
- $\lvert\mathcal{S}\rvert$: # states , $\lvert\mathcal{A}\rvert$: # actions
- $\frac{1}{1-\gamma}$: effective horizon
- $\varepsilon \in \left[0, \frac{1}{1-\gamma}\right]$: approximation error



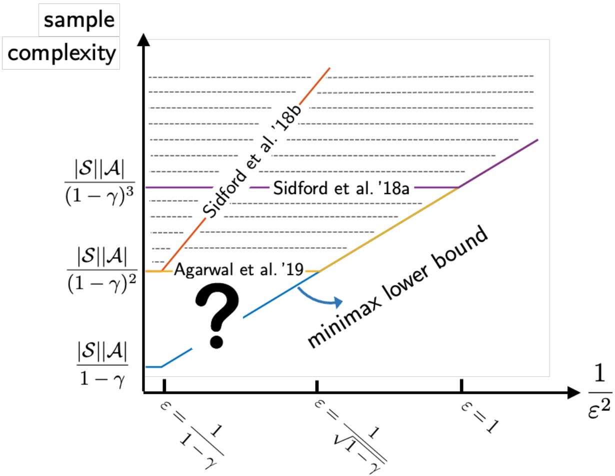
sample complexity
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3}$
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2}$
Sidford et al. '18b
Sidford et al. '18a
Agarwal et al. '19
$\frac{1}{\varepsilon^2}$
$\varepsilon = \frac{1}{1-\gamma}$
$\varepsilon = \frac{1}{\sqrt{1-\gamma}}$
$\varepsilon = 1$



sample
complexity
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3}$
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2}$
$\frac{|\mathcal{S}||\mathcal{A}|}{1-\gamma}$
Sidford et al. '18b
Sidford et al. '18a
Agarwal et al. '19
minimax lower bound
$\varepsilon = \frac{1}{1-\gamma}$
$\varepsilon = \frac{1}{\sqrt{1-\gamma}}$
$\varepsilon = 1$
$\frac{1}{\varepsilon^2}$



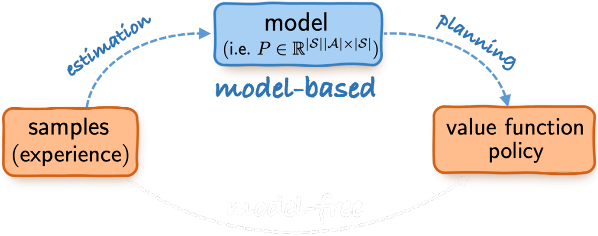
All prior theory requires sample size $\gtrsim \frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2}$



All prior theory requires sample size $\gtrsim \frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2}$

Question: is it possible to break this sample size barrier?

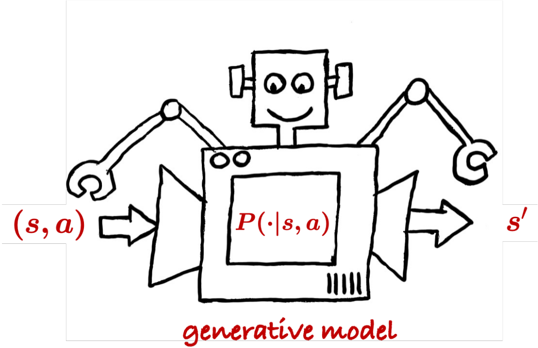
Our algorithm: Model based RL



Model-based approach ("plug-in")

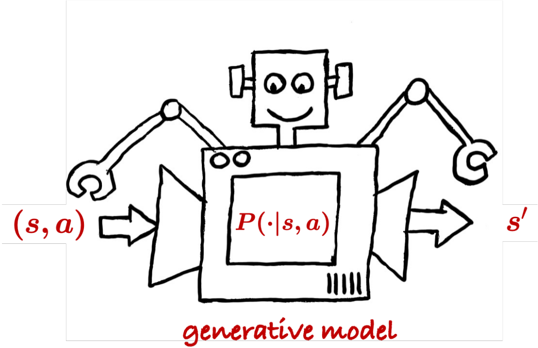
1. build an empirical estimate $\widehat{P}$ for P
2. planning based on empirical $\widehat{P}$

Model estimation



Sampling: for each (s,a), collect N ind. samples $\{(s,a,s'_{(i)})\}_{1\leq i\leq N}$

Model estimation

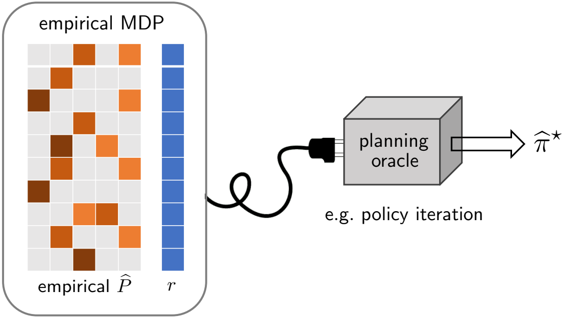


Sampling: for each (s,a), collect N ind. samples $\{(s,a,s'_{(i)})\}_{1\leq i\leq N}$

Empirical estimates: estimate $\widehat{P}(s'|s,a)$ by $\underbrace{\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{s'_{(i)}=s'\}}_{\text{empirical frequency}}$

Model-based (plug-in) estimator

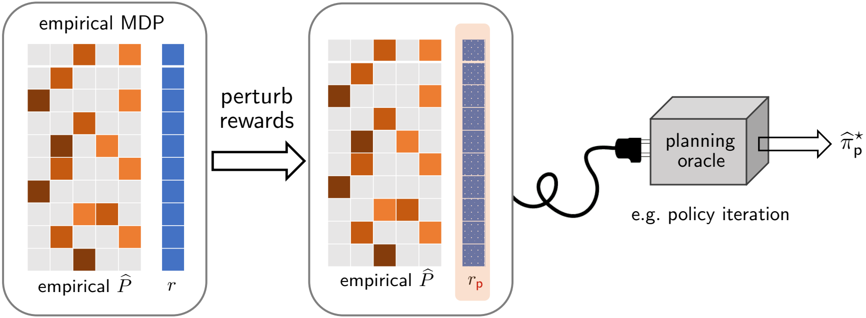
— [Azar et al., 2013, Agarwal et al., 2019, Pananjady and Wainwright, 2019]



Run planning algorithms based on the empirical MDP

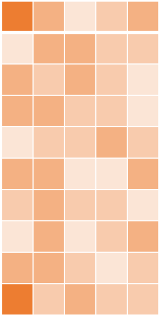
Our method: plug-in estimator + perturbation

— [Li et al., 2020a]

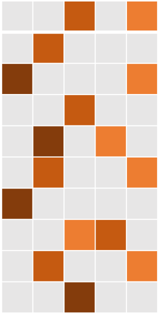


Planning based on the empirical MDP with slightly perturbed rewards

Challenges in the sample-starved regime



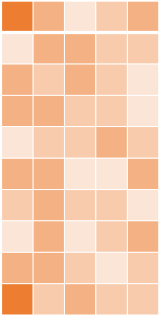
truth: $P \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}| \times |\mathcal{S}|}$



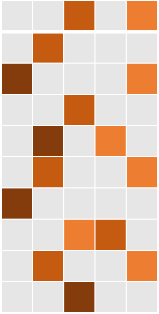
empirical estimate: $\widehat{P}$

- If sample size $\ll |\mathcal{S}|^2|\mathcal{A}|$, then we cannot recover P faithfully.

Challenges in the sample-starved regime



truth: $P \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|\times|\mathcal{S}|}$



empirical estimate: $\widehat{P}$

- If sample size $\ll |\mathcal{S}|^2|\mathcal{A}|$, then we cannot recover P faithfully.
- Can we trust our $\widehat{\pi}$ when $\widehat{P}$ is not accurate?

Main result: ℓ_∞-based sample complexity

Theorem (Li, Wei, Chi, Gu, Chen '20)

For any $0 < \varepsilon \leq \frac{1}{1-\gamma}$, *the optimal policy* $\widehat{\pi}_{\mathrm{p}}^{\star}$ *of perturbed empirical MDP achieves*

$$\|V^{\widehat{\pi}_{\mathrm{p}}^{\star}} - V^{\star}\|_{\infty} \leq \varepsilon$$

with sample complexity at most

$$\widetilde{O}\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3\varepsilon^2}\right)$$

Main result: ℓ_∞-based sample complexity

Theorem (Li, Wei, Chi, Gu, Chen '20)

For any $0 < \varepsilon \leq \frac{1}{1-\gamma}$, *the optimal policy* $\widehat{\pi}_{\mathrm{p}}^{\star}$ *of perturbed empirical MDP achieves*

$$\|V^{\widehat{\pi}_{\mathrm{p}}^{\star}} - V^{\star}\|_{\infty} \leq \varepsilon$$

with sample complexity at most

$$\widetilde{O}\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3\varepsilon^2}\right)$$

- $\varepsilon \in \big(0, \frac{1}{1-\gamma}\big] \quad \rightarrow \quad$ sample size range $\big[\frac{|\mathcal{S}||\mathcal{A}|}{1-\gamma}, \infty\big)$

Main result: ℓ_∞-based sample complexity

Theorem (Li, Wei, Chi, Gu, Chen '20)

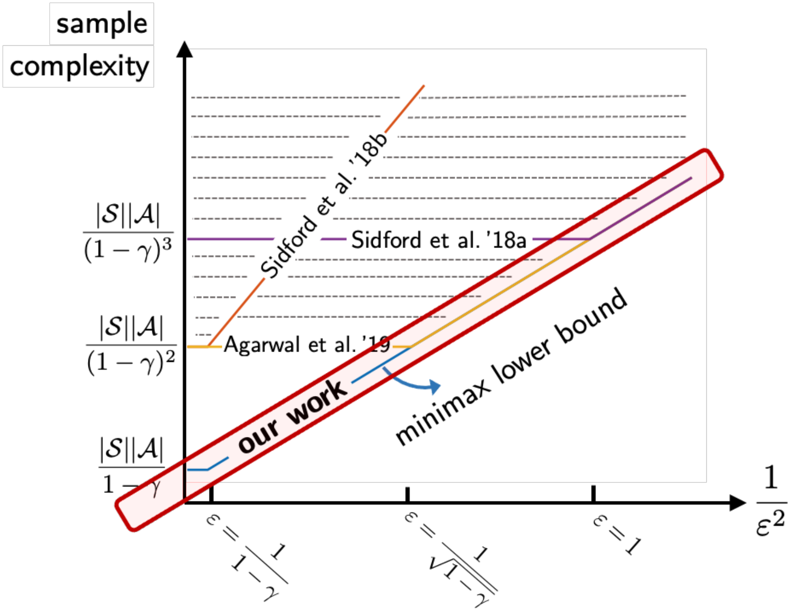
For any $0 < \varepsilon \leq \frac{1}{1-\gamma}$, *the optimal policy* $\widehat{\pi}_{\mathrm{p}}^{\star}$ *of perturbed empirical MDP achieves*

$$\|V^{\widehat{\pi}_{\mathrm{p}}^{\star}} - V^{\star}\|_\infty \leq \varepsilon$$

with sample complexity at most

$$\widetilde{O}\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3\varepsilon^2}\right)$$

- $\varepsilon \in \big(0, \frac{1}{1-\gamma}\big]$ $\quad\rightarrow\quad$ sample size range $\big[\frac{|\mathcal{S}||\mathcal{A}|}{1-\gamma}, \infty\big)$
- minimax lower bound: $\widetilde{\Omega}\big(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3\varepsilon^2}\big)$ [Azar et al., 2013]


sample complexity
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3}$
$\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2}$
$\frac{|\mathcal{S}||\mathcal{A}|}{1-\gamma}$
Sidford et al. '18b
Sidford et al. '18a
Agarwal et al. '19
our work
minimax lower bound
$\frac{1}{\varepsilon^2}$
$\varepsilon = \frac{1}{1-\gamma}$
$\varepsilon = \frac{1}{\sqrt{1-\gamma}}$
$\varepsilon = 1$

A glimpse of the key analysis ideas

Notation and Bellman equation

- V^π: value function under policy π
  - Bellman equation: $V^\pi = (I - P_\pi)^{-1} r$
- $\widehat{V}^\pi$: empirical version value function under policy π
  - Bellman equation: $\widehat{V}^\pi = (I - \widehat{P}_\pi)^{-1} r$

Notation and Bellman equation

- V^π: value function under policy π
  - Bellman equation: $V^\pi = (I - P_\pi)^{-1} r$
- $\widehat{V}^\pi$: empirical version value function under policy π
  - Bellman equation: $\widehat{V}^\pi = (I - \widehat{P}_\pi)^{-1} r$
- $\pi^\star$: optimal policy for V^π
- $\widehat{\pi}^\star$: optimal policy for $\widehat{V}^\pi$

Main steps

Elementary decomposition:

$$
\begin{aligned}
V^{\star} - V^{\widehat{\pi}^{\star}} &= \left(V^{\star} - \widehat{V}^{\pi^{\star}}\right) + \left(\widehat{V}^{\pi^{\star}} - \widehat{V}^{\widehat{\pi}^{\star}}\right) + \left(\widehat{V}^{\widehat{\pi}^{\star}} - V^{\widehat{\pi}^{\star}}\right) \\
&\leq \left(V^{\pi^{\star}} - \widehat{V}^{\pi^{\star}}\right) + 0 + \left(\widehat{V}^{\widehat{\pi}^{\star}} - V^{\widehat{\pi}^{\star}}\right)
\end{aligned}
$$

Main steps

Elementary decomposition:

$$V^\star - V^{\widehat{\pi}^\star} = \left(V^\star - \widehat{V}^{\pi^\star}\right) + \left(\widehat{V}^{\pi^\star} - \widehat{V}^{\widehat{\pi}^\star}\right) + \left(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star}\right)$$
$$\leq \left(V^{\pi^\star} - \widehat{V}^{\pi^\star}\right) + 0 + \left(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star}\right)$$

- **Step 1:** control $V^\pi - \widehat{V}^\pi$ for a fixed π (called "policy evaluation") (high-order decomposition + Bernstein inequality)

Main steps

Elementary decomposition:

$$\begin{aligned} V^\star - V^{\widehat{\pi}^\star} &= \big(V^\star - \widehat{V}^{\pi^\star}\big) + \big(\widehat{V}^{\pi^\star} - \widehat{V}^{\widehat{\pi}^\star}\big) + \big(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star}\big) \\ &\le \big(V^{\pi^\star} - \widehat{V}^{\pi^\star}\big) + 0 + \big(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star}\big) \end{aligned}$$

- **Step 1:** control $V^\pi - \widehat{V}^\pi$ for a fixed π (called "policy evaluation") (high-order decomposition + Bernstein inequality)
- **Step 2:** extend it to control $\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star}$ ($\widehat{\pi}^\star$ depends on samples) (decouple statistical dependency)

Key idea 1: a peeling argument (for fixed policy)

[Agarwal et al., 2019] first-order expansion

$$\widehat{V}^{\pi} - V^{\pi} = \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\widehat{V}^{\pi}$$

Key idea 1: a peeling argument (for fixed policy)

[Agarwal et al., 2019] first-order expansion

$$\widehat{V}^{\pi} - V^{\pi} = \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\widehat{V}^{\pi}$$

Ours: higher-order expansion + Bernstein $\longrightarrow$ tighter control

$$\begin{aligned}\widehat{V}^{\pi} - V^{\pi} = {}& \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)V^{\pi} + \\ & + \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\big(\widehat{V}^{\pi} - V^{\pi}\big)\end{aligned}$$

Key idea 1: a peeling argument (for fixed policy)

[Agarwal et al., 2019] first-order expansion

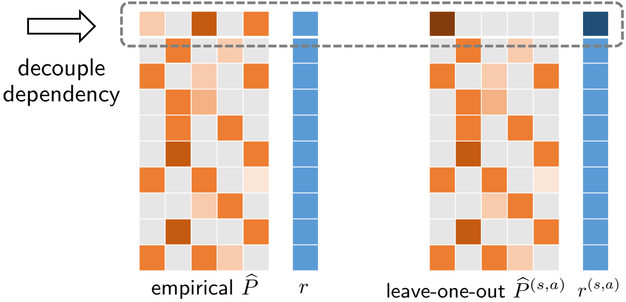
$$\widehat{V}^{\pi} - V^{\pi} = \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\widehat{V}^{\pi}$$

Ours: higher-order expansion + Bernstein $\longrightarrow$ tighter control

$$\begin{aligned}\widehat{V}^{\pi} - V^{\pi} = \gamma\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big){\color{red}V^{\pi}} + \\ + \gamma^{2}\Big(\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\Big)^{2}{\color{red}V^{\pi}} \\ + \gamma^{3}\Big(\big(I - \gamma P_{\pi}\big)^{-1}\big(\widehat{P}_{\pi} - P_{\pi}\big)\Big)^{3}{\color{red}V^{\pi}} \\ + \ldots\end{aligned}$$

Key idea 2: leave-one-out analysis for $(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star})_{s,a}$

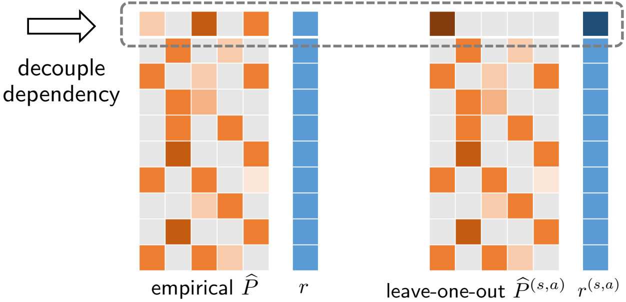
— inspired by [Agarwal et al., 2019] but quite different . . .



- define $\widehat{\pi}^\star_{(s,a)} \longrightarrow (\widehat{P}^{(s,a)}, r^{(s,a)})$
  — decouple dependency by dropping randomness for each (s, a)

Key idea 2: leave-one-out analysis for $(\widehat{V}^{\widehat{\pi}^\star} - V^{\widehat{\pi}^\star})_{s,a}$

— inspired by [Agarwal et al., 2019] but quite different . . .



- define $\widehat{\pi}^\star_{(s,a)} \longrightarrow (\widehat{P}^{(s,a)}, r^{(s,a)})$
  — decouple dependency by dropping randomness for each (s,a)
- works under the separation condition

$$\forall s \in \mathcal{S}, \quad \widehat{Q}^\star(s, \widehat{\pi}^\star(s)) - \max_{a:\, a \neq \widehat{\pi}^\star(s)} \widehat{Q}^\star(s,a) > 0$$

Key idea 3: tie-breaking via reward perturbation

- How to ensure separation between the optimal policy and others?

$$\forall s \in \mathcal{S}, \quad \widehat{Q}^{\star}(s, \widehat{\pi}^{\star}(s)) - \max_{a:\, a \neq \widehat{\pi}^{\star}(s)} \widehat{Q}^{\star}(s, a) > 0$$

Key idea 3: tie-breaking via reward perturbation

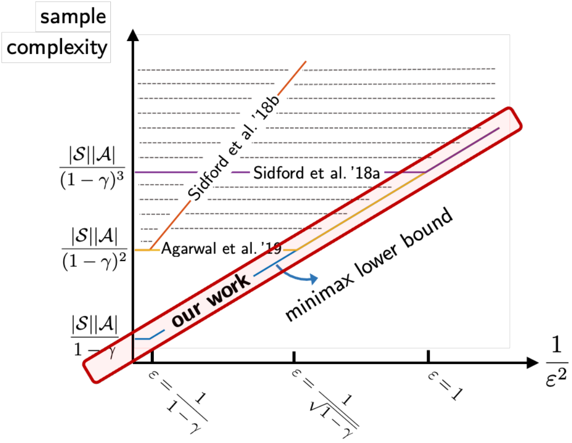
- How to ensure separation between the optimal policy and others?

$$\forall s \in \mathcal{S}, \quad \widehat{Q}^{\star}(s, \widehat{\pi}^{\star}(s)) - \max_{a:\, a \neq \widehat{\pi}^{\star}(s)} \widehat{Q}^{\star}(s, a) > 0$$

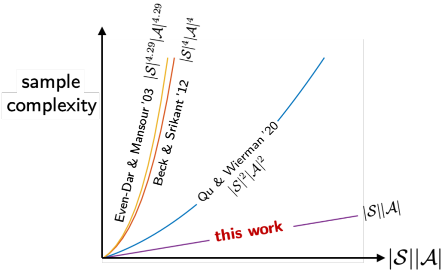
- **Solution:** slightly perturb rewards $r \implies \widehat{\pi}_{\mathrm{p}}^{\star}$
  - ensures $\widehat{\pi}_{\mathrm{p}}^{\star}$ can be differentiated from others
  - $V^{\widehat{\pi}_{\mathrm{p}}^{\star}} \approx V^{\widehat{\pi}^{\star}}$

Summary of this part

Model-based RL is minimax optimal and does not suffer from a sample size barrier!



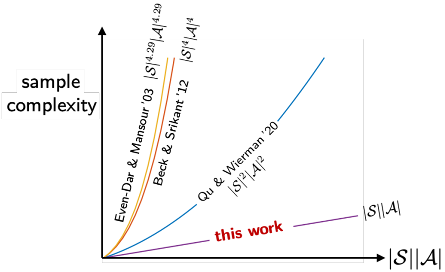
Summary of this part

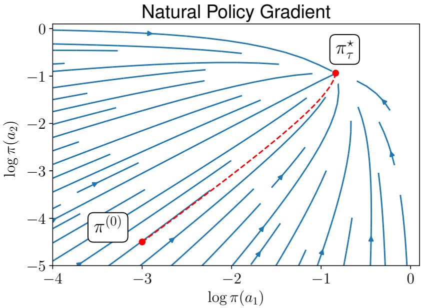


Other directions we have explored:

- *Model-free approach:* [Li et al., 2020b]
  — sharpened sample complexity of Q-learning on Markovian data

Summary of this part



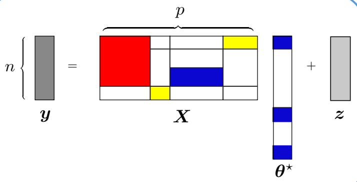


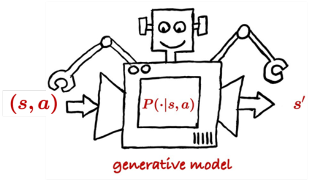
Other directions we have explored:

- *Model-free approach:* [Li et al., 2020b]
  — sharpened sample complexity of Q-learning on Markovian data
- *Policy-based approach:* [Cen et al., 2020]
  — linear convergence of entropy-regularized NPG methods

Concluding remarks

Modern statistical thinking plays a major role in breaking the sample complexity barrier in big-data applications





Thanks for your attention!

Other technical details

Key parameters via fixed point equations

$$(\tau^\star, \zeta^\star) \quad \xrightarrow{\text{solution}} \quad \begin{aligned} \tau^2 &= \sigma^2 + \mathsf{R}(\tau^2, \zeta) \\ \zeta &= 1 - \mathsf{df}(\tau^2, \zeta) \end{aligned}$$

$$\mathsf{R}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\,\mathbb{E}\left[\left\|\boldsymbol{\Sigma}^{1/2}(\widehat{\boldsymbol{\theta}}^f(\tau,\zeta) - \boldsymbol{\theta}^\star)\right\|_2^2\right]}_{\text{in-sample prediction risk}}$$

$$\mathsf{df}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\,\mathbb{E}\left[\left\|\widehat{\boldsymbol{\theta}}^f(\tau,\zeta)\right\|_0\right]}_{\text{degrees of freedom}}$$

Property: solution is unique and bounded for moderately sparse $\boldsymbol{\theta}^\star$

(Gaussian width $< \sqrt{n/p}$)

Coverage and power

Theorem (Celetano, Montanari, Wei '20)

There exist constants $C, c, c' > 0$ such that for all $\epsilon < c'$,

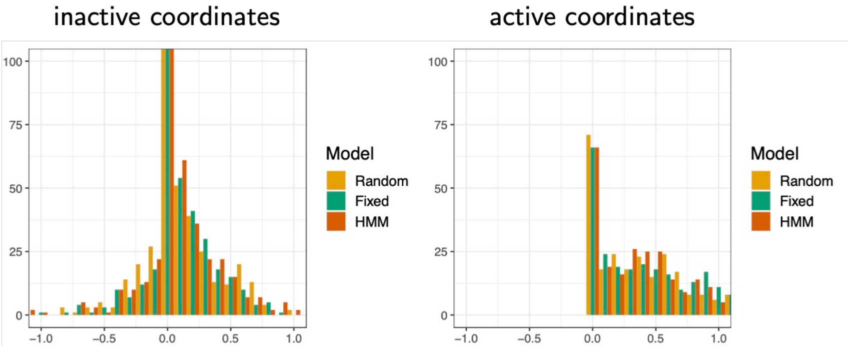
$$\left| \mathbb{P}_{\theta_j^*}\left(\theta \notin \mathsf{CI}_j^{\mathrm{loo}}\right) - \mathbb{P}_{\theta_j^*}\left(|\theta_j^* + \tau_{\mathrm{loo}}^* G - \theta| > \tau_{\mathrm{loo}}^* z_{1-\alpha/2}\right) \right| \leq C\left((1+|\theta_j^*|)\epsilon + \frac{1}{\epsilon^3} e^{-cn\epsilon^6} + \frac{1}{n\epsilon^2}\right),$$

where $G \sim \mathsf{N}(0,1)$.

$$\mathsf{CI}_j^{\mathrm{loo}} := \left[\xi_j \pm \widehat{\mathsf{sd}} \cdot z_{1-\alpha/2}\right]$$

ξ_j = scaled correlation between $\boldsymbol{X}_j^{\perp}$ and $\boldsymbol{y} - \boldsymbol{X}_{-j}\widehat{\boldsymbol{\theta}}_{\mathrm{loo}}$

Universality



Settings: auto-regressive design with $n = 1280, p = 2000, s = .128p$, active coordinates $= 1$, fixed λ_{cv}, plot histogram of $\widehat{\theta}_j$ vs. $\widehat{\theta}_j^f$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^\star}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^\star}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \underbrace{\widehat{\boldsymbol{\theta}}}_{\widehat{\boldsymbol{\theta}}^f} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^\star}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \underbrace{\widehat{\boldsymbol{\theta}}}_{\widehat{\boldsymbol{\theta}}^f} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{\underbrace{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}_{\zeta^*}}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^\star}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\boldsymbol{\Sigma}^{-1}\cdot\zeta^*\boldsymbol{\Sigma}^{1/2}(\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{\theta}}^f) = \zeta^*(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}^f - \widehat{\boldsymbol{\theta}}^f)$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \underbrace{\widehat{\boldsymbol{\theta}}}_{\widehat{\boldsymbol{\theta}}^f} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{\underbrace{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}_{\zeta^*}}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^\star + \tau^\star \boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^\star}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\boldsymbol{\Sigma}^{-1}\cdot\zeta^*\boldsymbol{\Sigma}^{1/2}(\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{\theta}}^f) = \zeta^*(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}^f - \widehat{\boldsymbol{\theta}}^f)$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \underbrace{\widehat{\boldsymbol{\theta}}}_{\widehat{\boldsymbol{\theta}}^f} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{\underbrace{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}_{\zeta^*}} \approx \boldsymbol{\theta}^* + \tau^*\boldsymbol{\Sigma}^{-1/2}\boldsymbol{g}$$

Analysis for model-based RL

Step 1: improved theory for policy evaluation

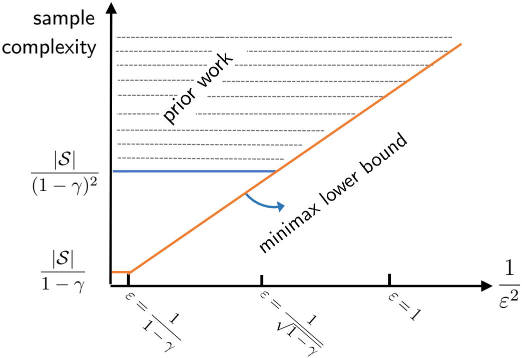
Model-based policy evaluation:

— given a fixed policy π, estimate V^{π} via the plug-in estimate $\widehat{V}^{\pi}$

Step 1: improved theory for policy evaluation

Model-based policy evaluation:

— given a fixed policy π, estimate V^{π} via the plug-in estimate $\widehat{V}^{\pi}$



- A sample size barrier $\frac{|\mathcal{S}|}{(1-\gamma)^2}$ already appeared in prior work (Agarwal et al. '19, Pananjady & Wainwright '19, Khamaru et al. '20)

Step 1: improved theory for policy evaluation

Model-based policy evaluation:

— given a fixed policy π, estimate V^π via the plug-in estimate $\widehat{V}^\pi$

Theorem (Li, Wei, Chi, Gu, Chen'20)

Fix any policy π. For $0 < \varepsilon \leq \frac{1}{1-\gamma}$, the plug-in estimator $\widehat{V}^\pi$ obeys

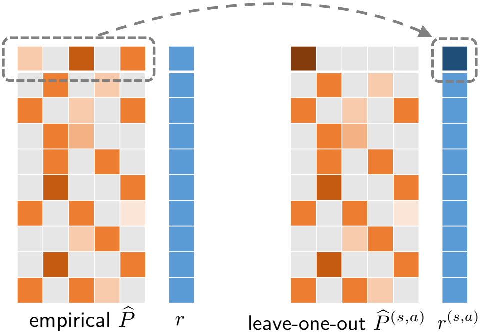
$$\|\widehat{V}^\pi - V^\pi\|_\infty \leq \varepsilon$$

with sample complexity at most

$$\widetilde{O}\Big(\frac{|\mathcal{S}|}{(1-\gamma)^3\varepsilon^2}\Big)$$

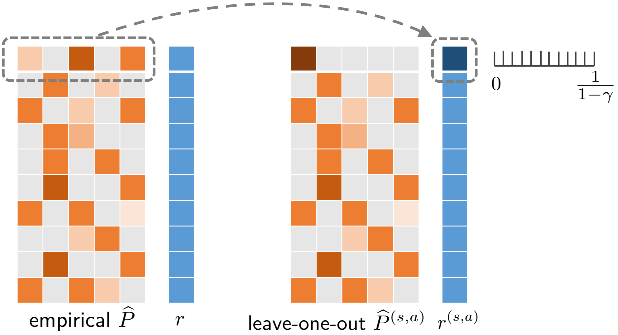
- Minimax optimal for all ε (Azar et al. '13, Pananjady & Wainwright '19)

Key idea 2: leave-one-out analysis



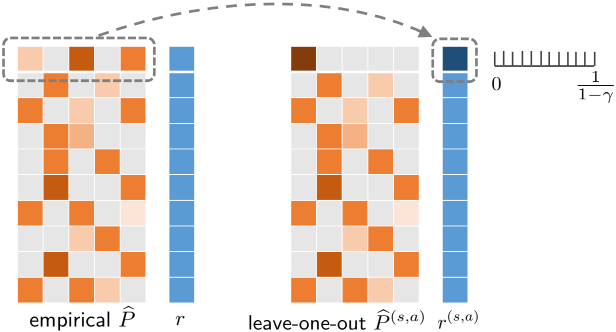
1. embed all randomness from $\widehat{P}_{s,a}$ into a single scalar (i.e. $r_{s,a}^{(s,a)}$)

Key idea 2: leave-one-out analysis



1. embed all randomness from $\widehat{P}_{s,a}$ into a single scalar (i.e. $r_{s,a}^{(s,a)}$)
2. build an ϵ-net for this scalar

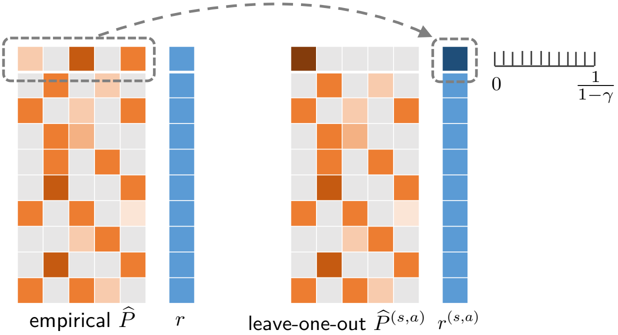
Key idea 2: leave-one-out analysis



1. embed all randomness from $\widehat{P}_{s,a}$ into a single scalar (i.e. $r_{s,a}^{(s,a)}$)
2. build an ϵ-net for this scalar
3. $\widehat{\pi}^\star$ can be determined by this ϵ-net under separation condition

$$\forall s \in \mathcal{S}, \quad \widehat{Q}^\star(s, \widehat{\pi}^\star(s)) - \max_{a:\, a \neq \widehat{\pi}^\star(s)} \widehat{Q}^\star(s,a) > 0$$

Key idea 2: leave-one-out analysis



Our decoupling argument vs. [Agarwal et al., 2019]

- [Agarwal et al., 2019]: dependency btw value $\widehat{V}$ & samples
- **Ours:** dependency btw policy $\widehat{\pi}$ & samples