Statistical Learning of **Distributionally Robust Stochastic Control** in Continuous State Spaces

Shengbo Wang
Joint work with *Nian Si, Jose Blanchet, and Zhengyuan Zhou*
Stanford MS&E

This work is supported generously by NSF 1915967, 2118199, 2229012, 2312204, 2312205, and Air Force Office of Scientific Research FA9550-20-1-0397.

Stanford University

Motivation

Controlled Stochastic Dynamics

$$X_{t+1} = f(X_t, A_t, W_t)$$

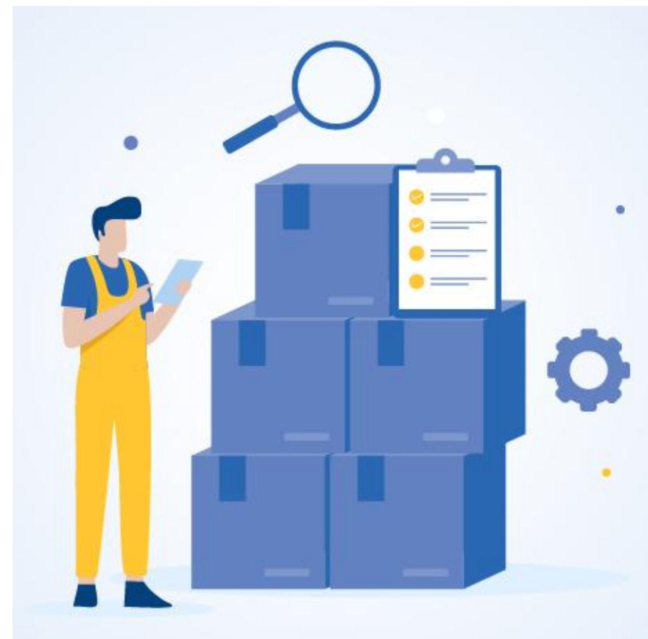
$$E_x^\pi \sum_{t=0}^{\infty} \alpha^t r(X_t, A_t, W_t)$$

State: $x, X_t \in \mathbb{X} \subset \mathbb{R}^{d_{\mathbb{X}}}$
Action: $A_t \in \mathbb{A}$
Noisy input: $W_t \in \mathbb{W} \subset \mathbb{R}^{d_{\mathbb{W}}}$
Control policy: $\pi \in \Pi$
State transition function: f
Reward function: r

$\{W_t : t \geq 0\}$ are i.i.d.

$$X_{t+1} = (X_t + A_t - W_t)_+$$

A simple inventory model



Systems in Operations Research

$$X_{t+1} = f(X_t, A_t, W_t)$$

f is **known**

Data for W_t is avalible

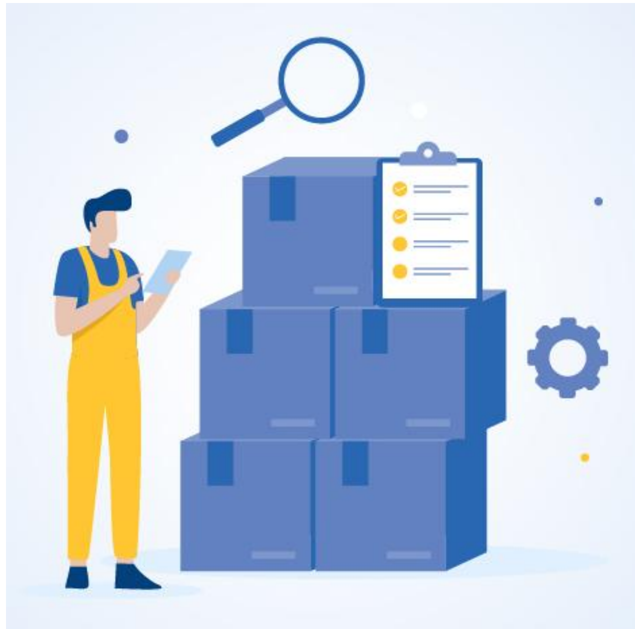
Systems in Operations Research

$$W_t \stackrel{d}{=} D \text{ i.i.d.?}$$

In dynamic decision-making context, model misspecification due to:

- Distribution shifts.
- Temporal correlation within $\{W_t : t \geq 0\}$. E.g. AR(1) $W_t = D_t + \delta W_{t-1}$
- Input might depend on history. E.g. $W_t = D_t + \delta g(X_t, A_t, X_{t-1} \dots)$

where $\{D_t : t \geq 0\}$ an i.i.d. sequence.



$$X_{t+1} = (X_t + A_t - W_t)_+$$

Systems in Operations Research

$$W_t \overset{d}{=} D \text{ i.i}$$

In dynamic decisi

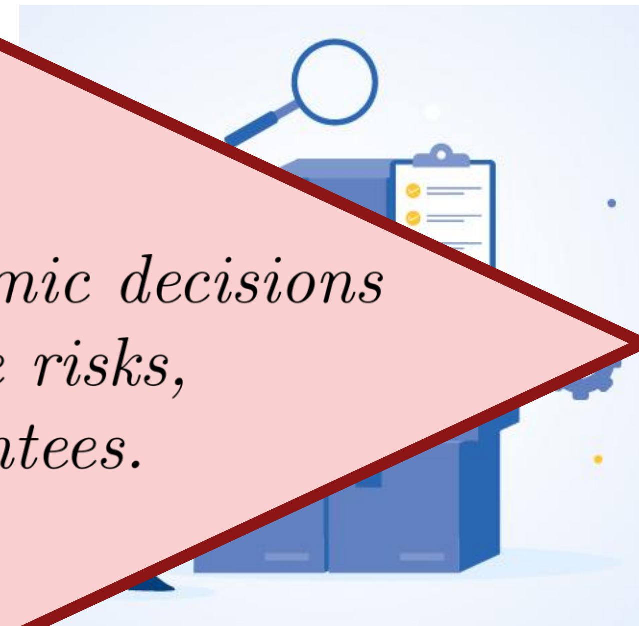
model missp

E.g

- Input migh
  E.g. $W_t = D_t +$

where $\{D_t : t \geq 0\}$ an i.i.d. seq

Learn good and reliable dynamic decisions
that are robust to these risks,
with statistical guarantees.



$$X_{t+1} = (X_t + A_t - W_t)_+$$

Literature on DR Policy Learning

Sample complexity of DRRL in finite state and action spaces: [Zhou et al. 21], [Panaganti and Kalathil 21], [Yang et al. 22], [W et al. 23a], [Shi et al. 23], [W et al. 23b], [Shi and Chi 24]......

DR Contextual Bandit: [Mu et al. 22], [Si et al. 23], [Shen et al. 23]

Linear and or kernel based DRRL in continuous state spaces: [Blanchet et al. 23], [Ma et al. 22]

Statistical analysis of DR single stage optimal decisions: [Duchi and Namkoong 21], [Lee and Raginsky 18]

Statistical analysis of DR stochastic control in continuous state spaces: [W et al. 24b] (this paper)

Formulation

Adversarial Robustness Approach

$$\sup_{\pi\in\Pi} E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t r(X_t,A_t,W_t)\right] \quad \text{vs.} \quad \boxed{\sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma} E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty}\alpha^t r(X_t,A_t,W_t)\right]}$$

Subject to $X_{t+1} = f(X_t, A_t, W_t)$

Under $E_x^{\pi,\gamma}$:
Given $\pi = (\pi_0, \pi_1 \dots), \gamma = (\gamma_0, \gamma_1, \dots)$ and start from $t = 0$, $X_t = x$

1. Simulate A_t from $\pi_t(\cdot|X_t, X_{t-1}, A_{t-1} \dots)$
2. Simulate W_t from $\gamma_t(\cdot|X_t, A_t, X_{t-1}, A_{t-1} \dots)$
3. Compute $X_{t+1} = f(X_t, A_t, W_t)$, $t \leftarrow t + 1$, and go back to 1.

Distributional Robustness Constraints

$$\sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma} E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty}\alpha^t r(X_t, A_t, W_t)\right], \quad \text{subject to } X_{t+1} = f(X_t, A_t, W_t)$$

$$W_t \approx D_t + \delta g_t(U_t, X_t, A_t, (W_{t-1})...)?$$

Given $\pi = (\pi_0, \pi_1 \ldots), \gamma = (\gamma_0, \gamma_1, \ldots)$ and start from $t = 0$, $X_t = x$

1. Simulate A_t from $\pi_t(\cdot|X_t, X_{t-1}, A_{t-1} \ldots)$
2. Simulate W_t from $\gamma_t(\cdot|X_t, A_t, X_{t-1}, A_{t-1} \ldots)$
3. Compute $X_{t+1} = f(X_t, A_t, W_t)$, $t \leftarrow t + 1$, and go back to 1.

Distributional Robustness Constraints

$$\sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty} \alpha^t r(X_t, A_t, W_t)\right], \quad \text{subject to } X_{t+1} = f(X_t, A_t, W_t)$$

$$W_t \approx D_t + \delta g_t(U_t, X_t, A_t, (W_{t-1})\ldots)?$$

Given $\pi = (\pi_0,$ … m $t = 0$, $X_t = x$

$$\{\mu \in \Delta(\mathbb{W}) : d(\mu \| \mathcal{L}(D)) \leq \delta\}$$

1. Simulate A_t …
2. Simulate W_t from $\gamma_t(\cdot | X_t, A_t, X_{t-1}, A_{t-1} \ldots) \in \mathcal{P}_\delta(D)$
3. Compute $X_{t+1} = f(X_t, A_t, W_t)$, $t \leftarrow t + 1$, and go back to 1.

Hence, $\Gamma = \{\gamma = (\gamma_t : t \geq 0) : \gamma_t(\cdot|\text{history}) \in \mathcal{P}_\delta(D)\}$.

Current Action Aware vs. Unaware Adversary

Current Action Awareness: $\gamma = (\gamma_t : t \geq 0) \in \Gamma$

$$W_t \sim \gamma_t(\cdot | X_t, A_t, X_{t-1}, A_{t-1} \dots) \in \mathcal{P}_\delta(D)$$

Current Action Unawareness: $\gamma = (\gamma_t : t \geq 0) \in \Gamma$

$$W_t \sim \gamma_t(\cdot | X_t, a, X_{t-1}, A_{t-1} \dots) = \mu \in \mathcal{P}_\delta(D) \text{ for every } a \in \mathbb{A}$$

Characterizing the Optimal Robust Value

$$v(x) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty} \alpha^t r(X_t, A_t, W_t)\right], \quad \text{s.t. } X_{t+1} = f(X_t, A_t, W_t)$$

Theorem: Assume appropriate regularity conditions.
If the adversary is current-action-aware, $v = u^*$ where u^* uniquely solve:

$$u^*(x) = \sup_{\phi\in\Delta(\mathbb{A})} \int_{\mathbb{A}} \inf_{\psi\in\mathcal{P}_\delta(D)} \int_{\mathbb{W}} r(x,a,w) + \alpha u^*(f(x,a,w))\psi(dw)\phi(da)$$

If the adversary is current-action-unaware, $v = \bar{u}$ where $\bar{u}$ uniquely solve:

$$\bar{u}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \inf_{\psi\in\mathcal{P}_\delta(D)} \int_{\mathbb{A}\times\mathbb{W}} r(x,a,w) + \alpha\bar{u}(f(x,a,w))\phi\times\psi(da,dw)$$

This result is an adaptation of one of the main theorems in our previous work [W et al. 23c]

Statistical complexity

Stanford University

Minimax Complexity for Uniform Learning

$W_t \approx D_t + \delta g_t(U_t, X_t, A_t, (W_{t-1})\dots);$
$\{D_t : t \geq 0\}$ are i.i.d. with unknown distribution.

$\implies \mathcal{P}_\delta(D) := \{\mu \in \Delta(\mathbb{W}) : d(\mu \| \mathcal{L}(D)) \leq \delta\}$

Data set: $\mathcal{D} := \left\{\hat{D}_k : k = 1 \dots n\right\}$ i.i.d. $\hat{D}_1 \sim \mathcal{L}(D)$

$$v(x) = \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty} \alpha^t r(X_t, A_t, W_t)\right]$$

s.t. $X_{t+1} = f(X_t, A_t, W_t)$, where f is **known**.

Learn the value function uniformly: $\sup_{x \in \mathbb{X}} |\hat{v}_{\mathcal{D}}(x) - v(x)|$

Empirical Robust Bellman Equations

Learn the value function uniformly: $\sup_{x\in\mathbb{X}} |\hat{v}_{\mathcal{D}}(x) - v(x)| < \epsilon(n)$

For the current-action-aware (CAA) case: $v = u^*$

$$u^*(x) = \sup_{\phi\in\Delta(\mathbb{A})} \int_{\mathbb{A}} \inf_{\psi\in\mathcal{P}_\delta(D)} \int_{\mathbb{W}} r(x,a,w) + \alpha u^*(f(x,a,w))\psi(dw)\phi(da)$$

For the current-action-unaware (CAU) case: $v = \bar{u}$

$$\bar{u}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \inf_{\psi\in\mathcal{P}_\delta(D)} \int_{\mathbb{A}\times\mathbb{W}} r(x,a,w) + \alpha\bar{u}(f(x,a,w))\phi\times\psi(da,dw)$$

Empirical Robust Bellman Equations

Learn the value function uniformly: $\sup_{x\in\mathbb{X}} |\hat{v}_{\mathcal{D}}(x) - v(x)| < \epsilon(n)$

For the CAA: use estimator $\hat{v}_{\mathcal{D}} := u^*_{\mathcal{D}}$

$$u^*_{\mathcal{D}}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \int_{\mathbb{A}} \inf_{\psi\in\mathcal{P}_\delta(\mathcal{D})} \int_{\mathbb{W}} r(x,a,w) + \alpha u^*_{\mathcal{D}}(f(x,a,w))\psi(dw)\phi(da)$$

For the CAA: use estimator $\hat{v}_{\mathcal{D}} := \bar{u}_{\mathcal{D}}$

$$\bar{u}_{\mathcal{D}}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \inf_{\psi\in\mathcal{P}_\delta(\mathcal{D})} \int_{\mathbb{A}\times\mathbb{W}} r(x,a,w) + \alpha\bar{u}_{\mathcal{D}}(f(x,a,w))\phi\times\psi(da,dw)$$

Minimax Complexity for Uniform Learning

Theorem: If the underlying spaces $\mathbb{X}, \mathbb{A}, \mathbb{W}$
f, r, v are Lipschitz, and $k' = k/(k-1)$, then:

Ambiguity Set $\mathcal{P}_\delta(D)$	Type	Action	Rate $\epsilon(n)$
Wasserstein	CAA $(v = u^*)$	Continuum	$\Theta\left(n^{-1/2}\right)$
	CAU $(v = \bar{u})$	Finite	
f_k-divergence	CAA $(v = u^*)$	Continuum	$\widetilde{\Theta}\left(n^{-\frac{1}{k' \vee 2}}\right)$
	CAU $(v = \bar{u})$	Finite	

Θ implies that we prove a matching lower bound

Doesn't depend on $d_{\mathbb{X}}$, $d_{\mathbb{W}}$!

χ^2 is a special case where $k = k' = 2$

Slides and the paper can be found here

Thank You

Your questions and thoughts are most welcome!

Slides and the paper can be found here

Summary

$v(x) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} E_x^{\pi,\gamma} \sum_{t=0}^{\infty} \alpha^t r(X_t, A_t, W_t)$
s.t. $X_{t+1} = f(X_t, A_t, W_t)$, where f is **known**.

Learn the value function uniformly: $\sup_{x\in\mathbb{X}} |\hat{v}_{\mathcal{D}}(x) - v(x)| < \epsilon(n)$

For the CAA: use estimator $\hat{v}_{\mathcal{D}} := u^*_{\mathcal{D}}$

$$u^*_{\mathcal{D}}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \int_{\mathbb{A}} \inf_{\psi\in\mathcal{P}_\delta(\mathcal{D})} \int_{\mathbb{W}} r(x,a,w) + \alpha u^*_{\mathcal{D}}(f(x,a,w))\psi(dw)\phi(da)$$

For the CAA: use estimator $\hat{v}_{\mathcal{D}} := \bar{u}_{\mathcal{D}}$

$$\bar{u}_{\mathcal{D}}(x) = \sup_{\phi\in\Delta(\mathbb{A})} \inf_{\psi\in\mathcal{P}_\delta(\mathcal{D})} \int_{\mathbb{A}\times\mathbb{W}} r(x,a,w) + \alpha\bar{u}_{\mathcal{D}}(f(x,a,w))\phi\times\psi(da,dw)$$

Ambiguity Set $\mathcal{P}_\delta(D)$	Type	Action	Rate $\epsilon(n)$
Wasserstein	CAA $(v = u^*)$ CAU $(v = \bar{u})$	Continuum Finite	$\Theta\left(n^{-1/2}\right)$
f_k-divergence	CAA $(v = u^*)$ CAU $(v = \bar{u})$	Continuum Finite	$\tilde{\Theta}\left(n^{-\frac{1}{k'\vee 2}}\right)$

Minimax Complexity for Uniform Learning

Learn the value function uniformly *efficiently*:

$$\sup_{x\in\mathbb{X}} |\hat{v}_{\mathcal{D}}(x) - v(x)| \leq \widetilde{O}_P(n^{-1/p})$$

where p doesn't depend on the dimension of $\mathbb{X}, \mathbb{W}$.

Hardest problem for that estimator

Lower bound:

$$\inf_{\tilde{v}} \sup_{D} E^D \sup_{x\in\mathbb{X}} |\tilde{v}_{\mathcal{D}}(x) - v(x)| \geq \widetilde{\Omega}(n^{-1/p}),$$

where $\mathcal{D} = \{D_1, \ldots, D_n\}$ i.i.d. and $D_1 \stackrel{d}{=} D$ under E^D.

"Best" possible estimator/algorithm