Distributionally Robust Policy Evaluation and Learning

Assured Autonomy in Contested Environments (AACE)
Spring 2025 Review

Yi Shen, Michael M. Zavlanos

Duke University

Outline

1. Problem setting and motivations
2. DRO formulation and its computational challenges
3. Examples: contextual bandit
4. Outlier-robust DRO

Stochastic Optimization Problems

Given a data distribution of interest $\xi \sim \mathbb{P}^*(\Xi)$ (testing distribution), a learning parameter $\theta \in \Theta$, and a loss function $l(\xi, \theta)$, the goal of many stochastic optimization problems is to minimize the expected loss:

$$\inf_{\theta \in \Theta} \ \mathbb{E}_{\xi \sim \mathbb{P}^*(\Xi)} \left[l(\xi, \theta) \right]$$

Stochastic Optimization Problems

Given a data distribution of interest $\xi \sim \mathbb{P}^*(\Xi)$ (testing distribution), a learning parameter $\theta \in \Theta$, and a loss function $l(\xi, \theta)$, the goal of many stochastic optimization problems is to minimize the expected loss:

$$\inf_{\theta \in \Theta} \mathbb{E}_{\xi \sim \mathbb{P}^*(\Xi)} [l(\xi, \theta)]$$

Empirical Risk Minimization

Given finite i.i.d. samples $\xi_1, \xi_2, ..., \xi_n \sim \mathbb{P}_0(\Xi)$ (training distribution), denoted the empirical distribution as $\hat{\mathbb{P}}_0(\Xi)$, the empirical risk minimization (ERM) minimizes the empirical loss:

$$\inf_{\theta \in \Theta} \mathbb{E}_{\xi \sim \hat{\mathbb{P}}_0(\Xi)} [l(\xi, \theta)] = \inf_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} [l(\xi_i, \theta)]$$

Ideal goal: ERM $\rightarrow \hat{\theta} \rightarrow \mathbb{E}_{\xi \sim \mathbb{P}^*(\Xi)} \left[l(\xi, \hat{\theta}) \right] \approx \mathbb{E}_{\xi \sim \hat{\mathbb{P}}_0(\Xi)} \left[l(\xi, \hat{\theta}) \right]$

Failures: ERM Goes Wrong

ERM can go wrong even when the size of the training set goes to infinity due to distribution shifts.

Distribution shifts: $\mathbb{P}^*(\Xi) \neq \mathbb{P}_0(\Xi)$

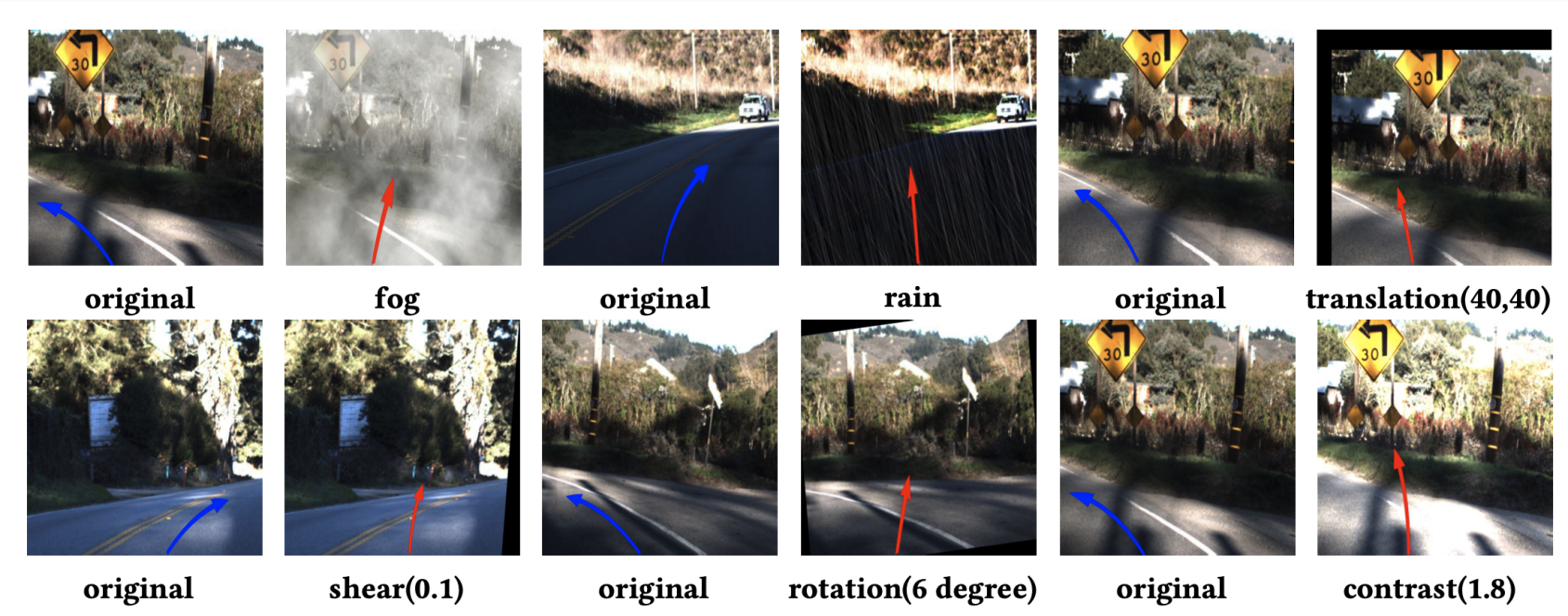


Figure 1: Sample images showing erroneous behaviors detected by DeepTest[1] using synthetic images. For original images the arrows are marked in blue, while for the synthetic images they are marked in red.

[1] [1] Tian, Yuchi, et al. "Deeptest: Automated testing of deep-neural-network-driven autonomous cars."

Failures: ERM Goes Wrong

ERM can go wrong even when the size of the training set goes to infinity due to distribution shifts.

Distribution shifts: $\mathbb{P}^*(\Xi) \neq \mathbb{P}_0(\Xi)$

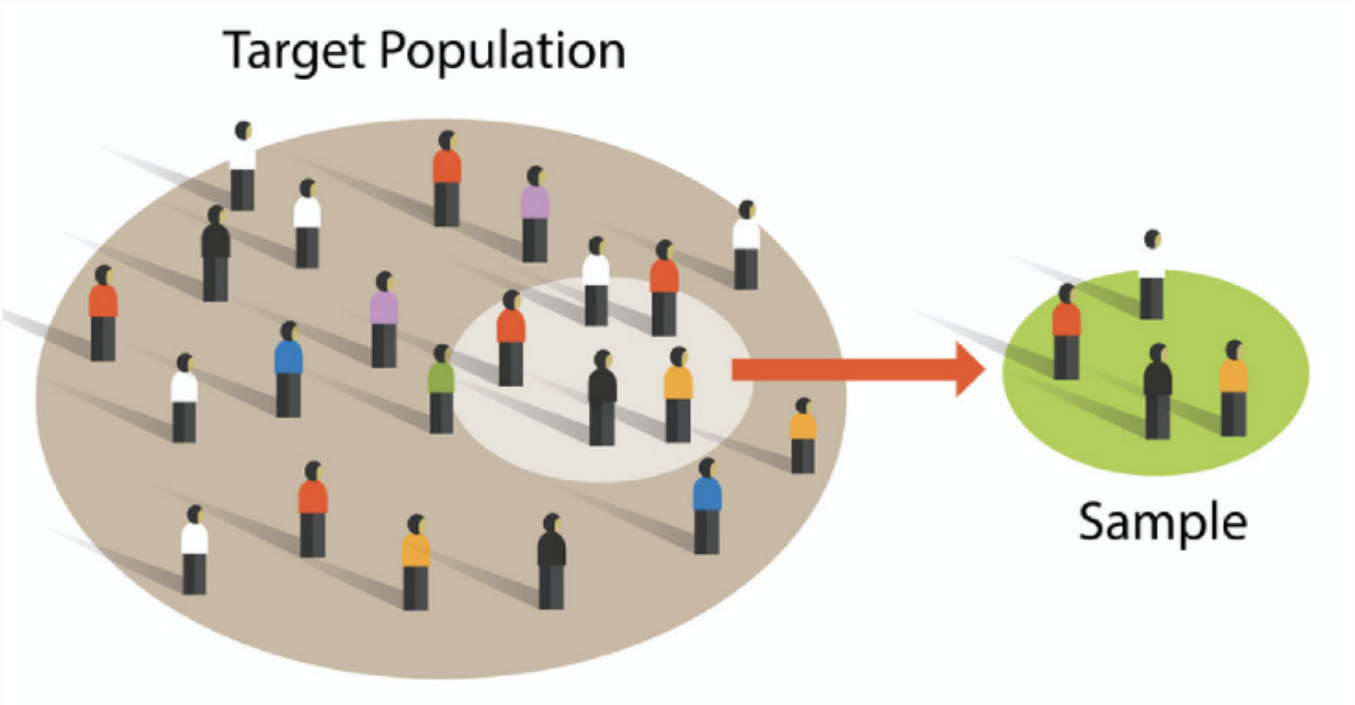


Figure 2: Human AI collaboration: volunteers (training population) who are willing to participate and provide the data (sample) do not represent the whole population (target population). [2]

[2] image credit: www.simplypsychology.org/sampling.html

Distributionally Robust Optimization (DRO)

Recall, empirical Risk Minimization minimizes the following loss:

$$\text{ERM:} \quad \inf_{\theta \in \Theta} \mathbb{E}_{\xi \sim \hat{\mathbb{P}}_0} [l(\xi, \theta)]$$

Distributionally robust optimization aims to find a robust solution by solving the following problem:

$$\text{DRO:} \quad \inf_{\theta \in \Theta} \sup_{\mathbb{Q} \in \mathcal{B}(\hat{\mathbb{P}}_0)} \mathbb{E}_{\xi \sim \mathbb{Q}} [l(\xi, \theta)] ,$$

where $\mathcal{B}(\hat{\mathbb{P}}_0)$ is the **ambiguity set** (uncertainty ball) around the empirical distribution $\hat{\mathbb{P}}_0$.

We design the ambiguity set $\mathcal{B}(\hat{\mathbb{P}}_0)$ such that it includes the testing distribution, i.e., $\mathbb{P}^* \in \mathcal{B}(\hat{\mathbb{P}}_0)$.

Ambiguity Set Selection

The ambiguity set is usually defined by a radius ϵ ball around the emperical distribution $\hat{\mathbb{P}}_0$, i.e.,

$$\mathcal{B}_\epsilon(\hat{\mathbb{P}}_0) : \{\mathbb{Q} : dist(\mathbb{Q}, \hat{\mathbb{P}}_0) \leq \epsilon\}.$$

Two popular distribution distance (divergence) metrics

- The f-divergence between distribution P and Q is

$$D_f(P||Q) := \int f\left(dP/dQ\right) dQ,$$

  where f is a convex function with $f(1) = 0$, e.g., Kullback-Leibler (KL) divergence by taking $f(x) = x \log x$ [3].

- The 1-wasserstein distance between distribution P and Q is

$$W(P, Q) := \inf_{\pi \in \mathcal{M}(\Xi, \Xi)} \mathbb{E}_{(\xi, \xi') \sim \pi} \left[\|\xi - \xi'\|\right], \text{ s.t. } \pi_1 = P, \pi_2 = Q.$$

[3] The popular conditional value at risk (CVaR) can also be derived from the dual of a DRO problem under a special uncertainty set.

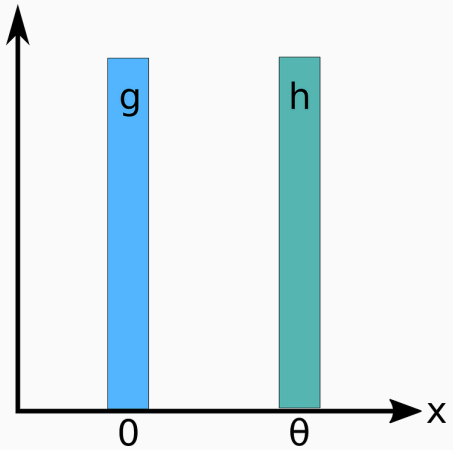
Ambiguity Set Selection

- The KL-divergence between distribution P and Q is

$$D_{KL}(P||Q) := \int f(dP/dQ)dQ,$$

  where $f(x) = x \log x$.
- The 1-wasserstein distance between distribution P and Q is

$$W(P, Q) := \inf_{\pi \in \mathcal{M}(\Xi, \Xi)} \mathbb{E}_{(\xi, \xi') \sim \pi} \left[\|\xi - \xi'\| \right], \text{ s.t. } \pi_1 = P, \pi_2 = Q.$$



$$D_{KL}(g||h) = \infty$$
$$D_{KL}(h||g) = \infty$$
$$D_{JS}(g||h) = log\, 2$$
$$W(g, h) = |\theta|$$

Figure 3: Distances of two discrete distributions g and h under different metrics.[4]

[4] JS is the Jensen–Shannon divergence. $D_{JS}(P||Q) = 1/2 D_{KL}(P||M) + 1/2 D_{KL}(Q||M)$, where $M = 1/2P + 1/2Q$ is a mixture distribution of P and Q. Image credit: medium.com/@sunil7545

Wasserstein DRO

When both P and Q are discrete, the wasserstein distance $W(Q, P)$ can be solved by a linear program.

W-DRO:

$$\inf_{\theta \in \Theta} \sup_{\mathbb{Q} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_0)} \mathbb{E}_{\xi \sim \mathbb{Q}} \left[l(\xi, \theta) \right], \text{ where } \mathcal{B}_\epsilon(\hat{\mathbb{P}}_0) : \{\mathbb{Q} : W(\mathbb{Q}, \hat{\mathbb{P}}_0) \leq \epsilon\}.$$

The wasserstein distance is hard to solve when the target distribution Q is **continuous**.

The wasserstein DRO problem is **harder (in general, intractable)** as it is optimizing over all distributions (infinite-dimensional) inside the ambiguity set!

Wasserstein Distributionally Robust Offline Contextual Bandit

W-DRO for Offline Contextual Bandit

Problem Setup (robotic navigation in unknown and varying terrains)

1. the robot observes a type of terrain i with contextual information $X_i :=$ {percentages of the terrain of sand, gravel, mud, rocky surfaces}, $X_i \sim P_x^0$;

W-DRO for Offline Contextual Bandit

Problem Setup (robotic navigation in unknown and varying terrains)

1. the robot observes a type of terrain i with contextual information $X_i :=$ {percentages of the terrain of sand, gravel, mud, rocky surfaces}, $X_i \sim P_x^0$;
2. the robot selects one policy from a set of pre-learned locomotion policies advisor recommends according to a policy $\pi_0(\cdot|X_i)$, e.g., $A_i \sim \pi_0(\cdot|X_i)$ and $A_i \in \{$ High-frequency stepping (good for mud), Low-frequency, energy-efficient gait (good for flat surfaces), etc., $\}$;

W-DRO for Offline Contextual Bandit

Problem Setup (robotic navigation in unknown and varying terrains)

1. the robot observes a type of terrain i with contextual information $X_i :=$ {percentages of the terrain of sand, gravel, mud, rocky surfaces}, $X_i \sim P_x^0$;
2. the robot selects one policy from a set of pre-learned locomotion policies advisor recommends according to a policy $\pi_0(\cdot|X_i)$, e.g., $A_i \sim \pi_0(\cdot|X_i)$ and $A_i \in \{$ High-frequency stepping (good for mud), Low-frequency, energy-efficient gait (good for flat surfaces), etc., $\}$;
3. a cost is revealed after executing the policy for a period of time $Y_i \sim P^0_{X_i,A_i}$, e.g., energy consumption, penalty for falls or unsafe states ;

W-DRO for Offline Contextual Bandit

Problem Setup (robotic navigation in unknown and varying terrains)

1. the robot observes a type of terrain i with contextual information $X_i :=$ {percentages of the terrain of sand, gravel, mud, rocky surfaces}, $X_i \sim P_x^0$;
2. the robot selects one policy from a set of pre-learned locomotion policies advisor recommends according to a policy $\pi_0(\cdot|X_i)$, e.g., $A_i \sim \pi_0(\cdot|X_i)$ and $A_i \in \{$ High-frequency stepping (good for mud), Low-frequency, energy-efficient gait (good for flat surfaces), etc., $\}$;
3. a cost is revealed after executing the policy for a period of time $Y_i \sim P^0_{X_i,A_i}$, e.g., energy consumption, penalty for falls or unsafe states ;
4. an offline dataset $\mathcal{D}_n := \{(X_i, A_i, Y_i)\}_{i=1}^n$ is collected by running steps 1-3 n times.

Policy Evaluation and Learning

Given the offline dataset $\mathcal{D}_n := \{(X_i, A_i, Y_i)\}_{i=1}^n$, we study offline policy evaluation (OPE) and offline policy learning (OPL)

- OPE returns the expected return of a given policy,
- OPL finds the optimal policy that maximized the expected return.

OPE and OPL problems have been studied in causal inference [2][5] and offline reinforcement learning [3][6]. Yet, most of the works do not consider distribution shifts.

Why distributional robustness?

1. avoid overfitting due to the finite dataset $\mathcal{D}_n$;
2. selection biases: offline datasets are collected in a specfic funding office that do not represent the population of interests;
3. cost (reward) shifts.

[5] Athey, Susan, and Guido W. Imbens. "The state of applied econometrics: Causality and policy evaluation." Journal of Economic perspectives.

[6] Jiang, Nan, and Lihong Li. "Doubly robust off-policy value evaluation for reinforcement learning." ICML 2016.

Distributionally Robust Policy Evaluation and Learning [4][7]

Structures in the offline dataset $\mathcal{D}_n := \{(X_i, A_i, Y_i)\}_{i=1}^n$

The first trick: $\mathbb{P}(X, A, Y) = \mathbb{P}(X) \times \mathbb{P}(A|X) \times \mathbb{P}(Y|X, A)$

Q: Where should we add the ambiguity sets?

DR-OPE: For given $\epsilon_x, \epsilon_c > 0$ and policy π, we define the distributionally robust policy value $V(\pi)$ as:

$$V(\pi) = \sup_{P_x \in \mathcal{U}(\epsilon_x; P_x^0)} \mathbb{E}_{x \sim P_x} \left[\mathbb{E}_{a \sim \pi(\cdot|x)} \left[\sup_{P_{x,a} \in \mathcal{U}(\epsilon_c; P_{x,a}^0)} \mathbb{E}_{P_{x,a}} [Y] \right] \right] \quad (1)$$

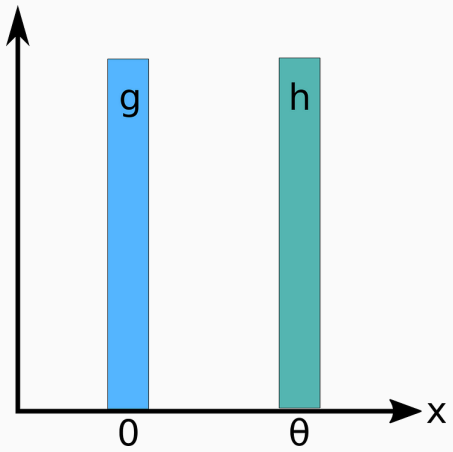
DR-OPL: Given a learning policy space Π, we define the distributionally robust policy learning problem:

$$\begin{aligned} &inf_{\pi \in \Pi} V(\pi) = inf_{\theta \in \Theta} V(\pi_\theta) \\ &= inf_{\theta \in \Theta} \sup_{P_x \in \mathcal{U}(\epsilon_x; P_x^0)} \mathbb{E}_{x \sim P_x} \left[\mathbb{E}_{a \sim \pi_\theta(\cdot|x)} \left[\sup_{P_{x,a} \in \mathcal{U}(\epsilon_c; P_{x,a}^0)} \mathbb{E}_{P_{x,a}} [Y] \right] \right] \quad (2) \end{aligned}$$

[7] Shen, Xu, Zavlanos, Wasserstein distributionally robust policy evaluation and learning for contextual bandits, TMLR 2024

Distributionally Robust Policy Evaluation and Learning

Q: Can you define the DR-OPE and DR-OPL by using KL ambiguity set?
A: Yes, [5, 6] [8] study DR-OPE and DR-OPL with KL ambiguity set. However, the KL ambiguity set will only include supports that have been observed in the offline set and does not consider the geometry of the support set.



$$D_{KL}(g||h) = \infty$$
$$D_{KL}(h||g) = \infty$$
$$D_{JS}(g||h) = log\, 2$$
$$W(g,h) = |\theta|$$

Figure 4: Distances of two discrete distributions g and h under different metrics.

[8] Si, Nian, et al. "Distributionally robust batch contextual bandits." Management Science; Mu, Tong, et al. "Factored DRO: Factored distributionally robust policies for contextual bandits." NeurIPS (2022)

Regularized Wasserstein DRO

$$\text{DR-OPE: } V(\pi) = \sup_{P_x \in \mathcal{U}(\epsilon_x; P_x^0)} \mathbb{E}_{x \sim P_x} \left[\mathbb{E}_{a \sim \pi(\cdot|x)} \left[\sup_{P_{x,a} \in \mathcal{U}(\epsilon_c; P_{x,a}^0)} \mathbb{E}_{P_{x,a}} [Y] \right] \right]$$

By duality, we have that:

$$V(\pi) = \inf_{\lambda \geq 0} \left\{ \epsilon_x \lambda + \mathbb{E}_{x \sim P_x^0} \left[\sup_{\zeta \in \mathcal{X}} \left(\mathbb{E}_{a \sim \pi(\cdot|\zeta)} [m(\zeta, a)] - \lambda (x - \zeta)^2 \right) \right] \right\},$$

$$m(x, a) = \inf_{\lambda \geq 0} \left\{ \epsilon_c \lambda + \mathbb{E}_{\xi \sim P^0_{\Xi_{x,a}}} \sup_{\zeta \in \Xi_{x,a}} \left(y_{x,a}(\zeta) - \lambda (\xi - \zeta)^2 \right) \right\}, \forall x \in \mathcal{X}, a \in \mathcal{A}.$$

Computational challenges

1. two inner maximization problems: discretized space is large
2. when considering policy learning θ, we might use different numerical optimization packages, e.g., pytorch for θ on GPUs while GUROBI for maximization on CPUs.

Regularized Wasserstein DRO

The second trick: smoothing

$$(\text{WDRO}) = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

We define the smoothed dual problem as:

$$(\text{WDRO}_\eta) := \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\frac{1}{\eta} \log \left(\sum_{\zeta \in \Xi} \frac{1}{|\Xi|} e^{\eta(l(\zeta) - \lambda(\xi - \zeta)^2)} \right) \right] \right\},$$

where $\eta > 0$ is a hyper-parameter that controls the distance between the "softmax" and the maximum.

Now, the (WDRO_η) problem is a smoothed problem and we can use stochastic gradient descent (SGD) to solve it!

Outlier-robust Wasserstein Distributionally Robust Optimization

DRO with Outliers

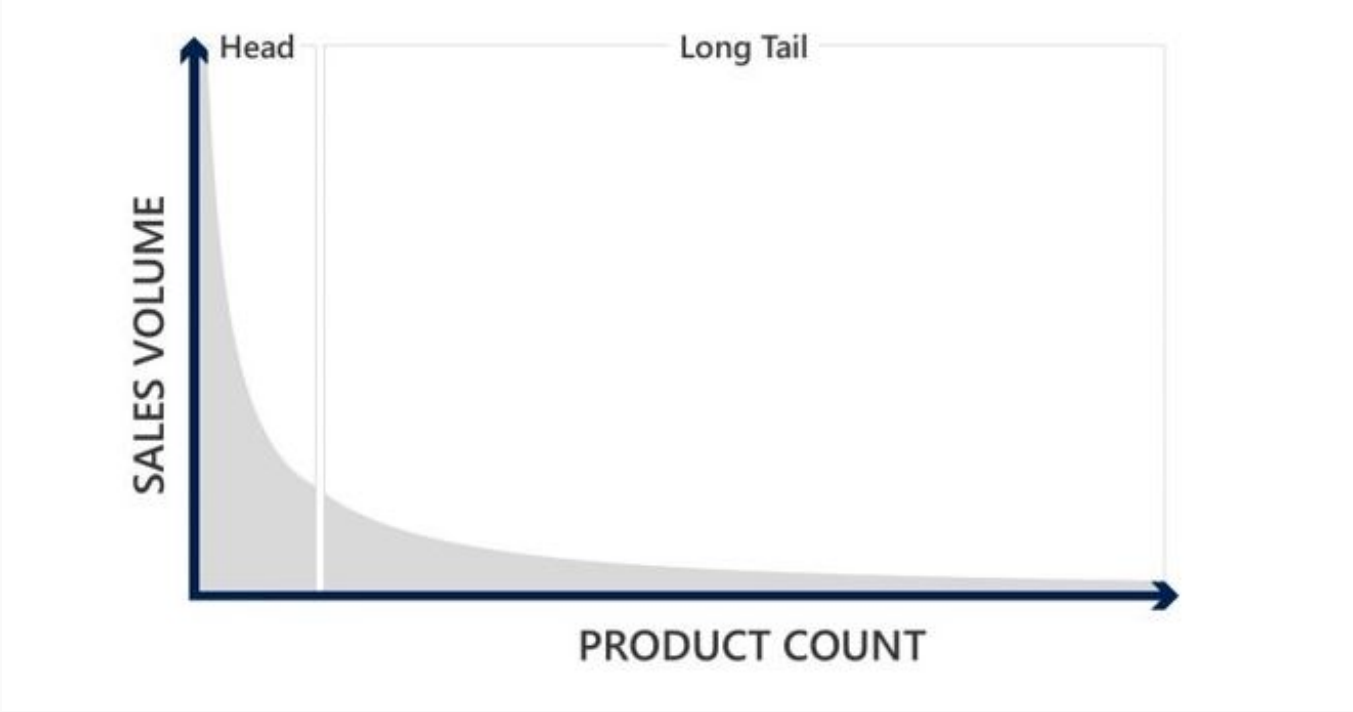


Figure 5: The long tail: Intermittent demand and "unforecastable" SKUs (hard machine learning training samples in general). [9]

[9] Image credit: www.microsoft.com/en-us/industry/blog/manufacturing-and-mobility/2018/08/15/the-three-primary-pains-of-modern-inventory-optimization/

DRO with Outliers

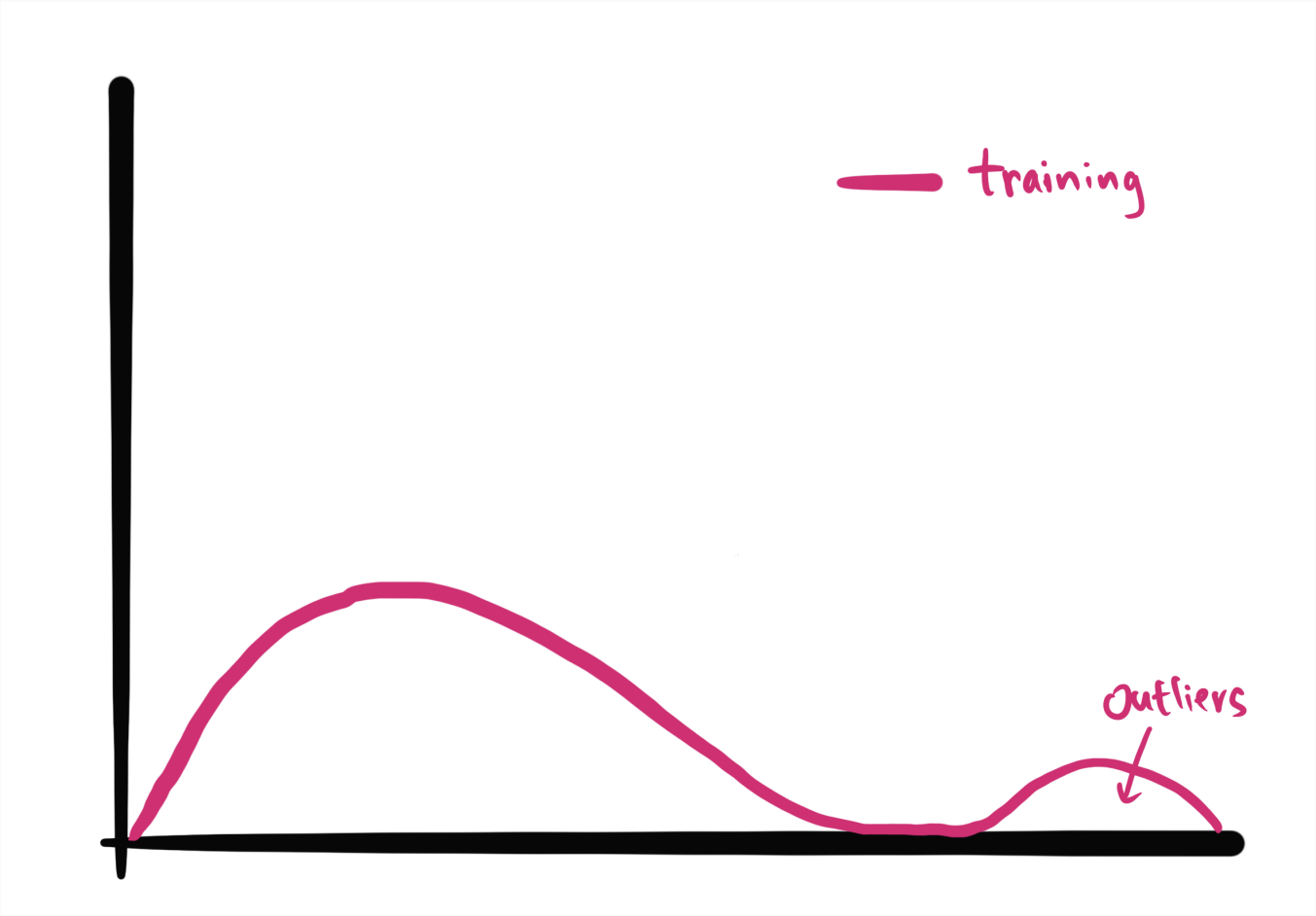


Figure 6: Wasserstein ball fails to include distirbutions of interests when outliers exist.

DRO with Outliers

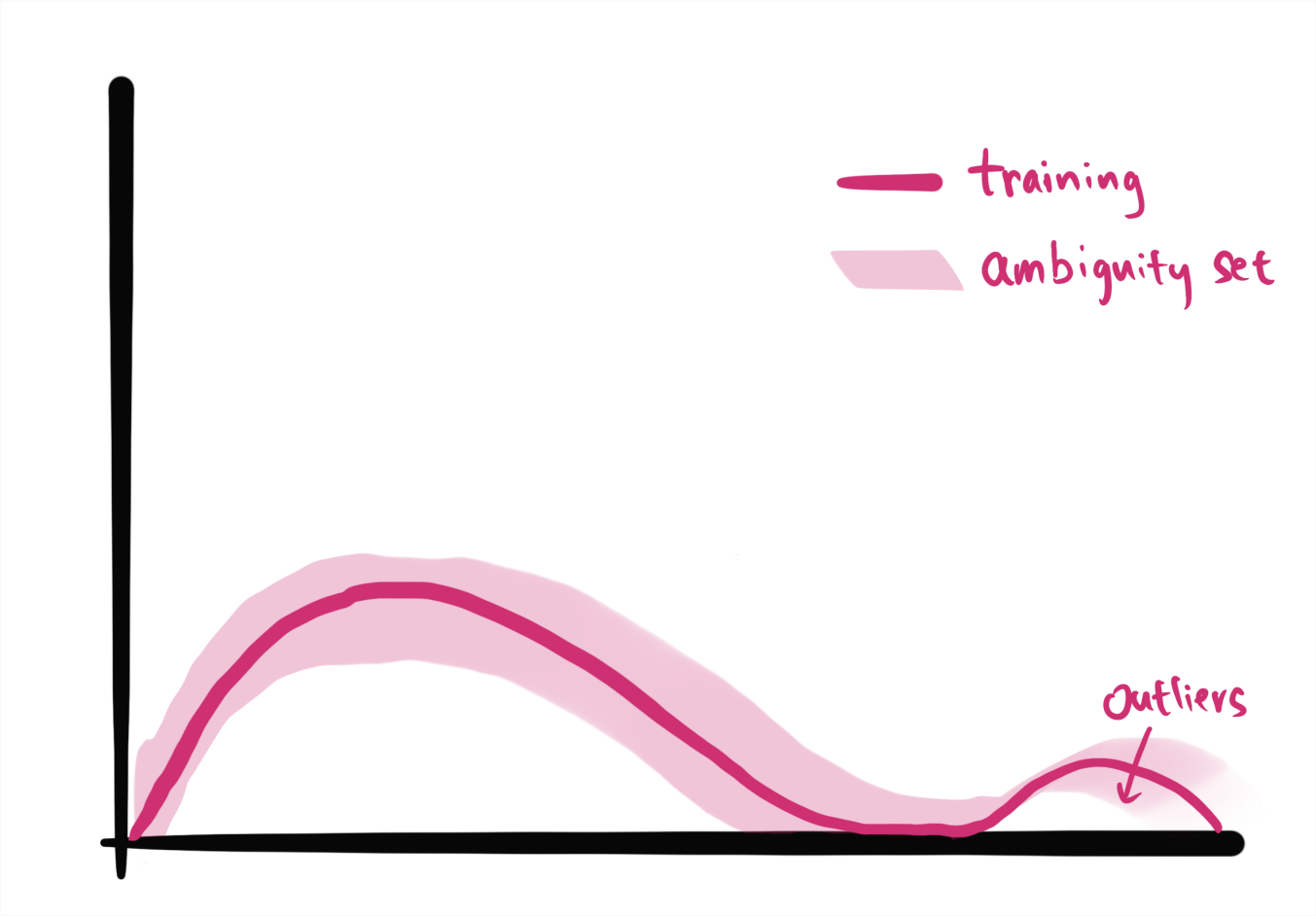


Figure 7: Wasserstein ball fails to include distirbutions of interests when outliers exist.

DRO with Outliers

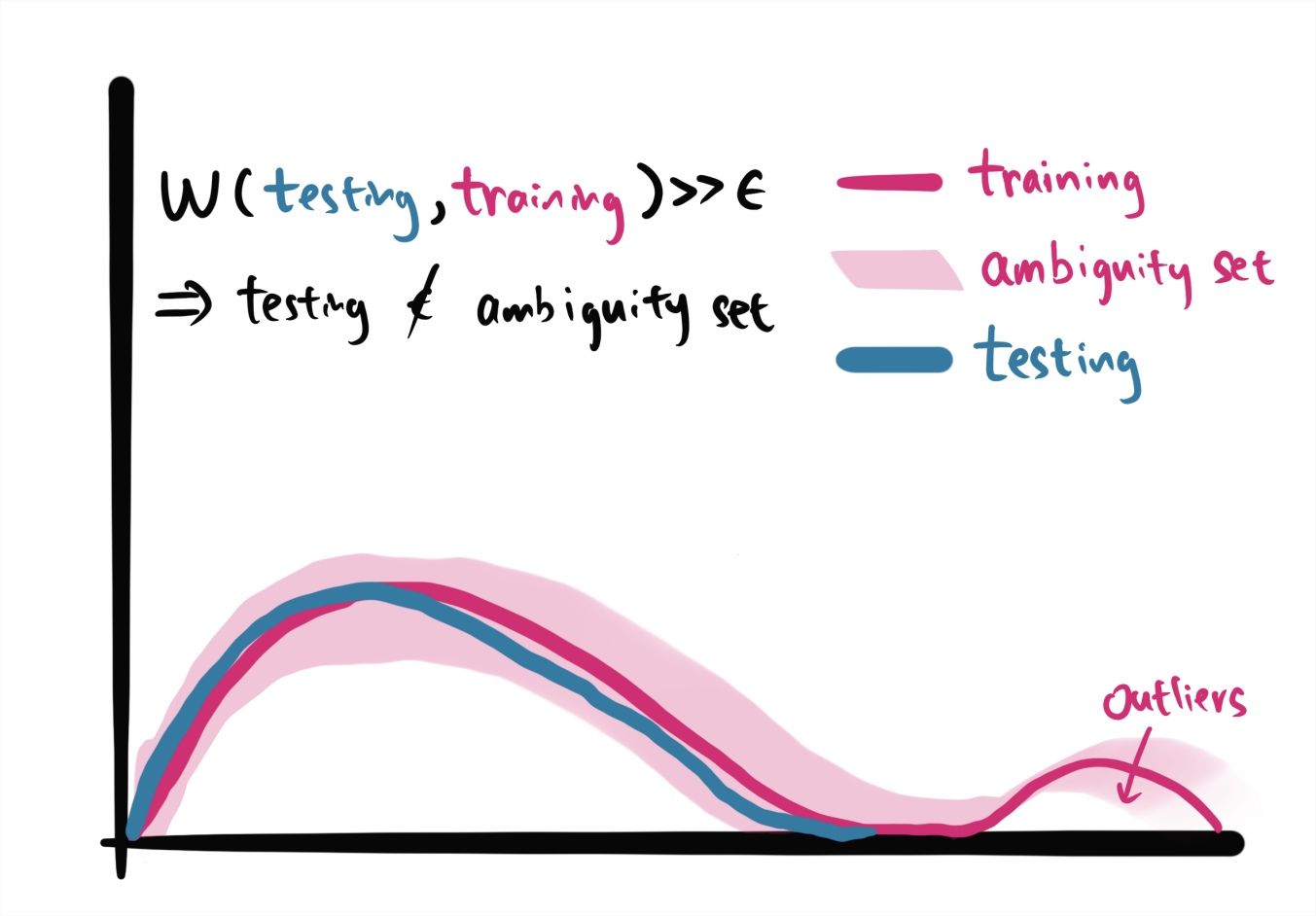


Figure 8: Wasserstein ball fails to include distirbutions of interests when outliers exist.

DRO with Outliers

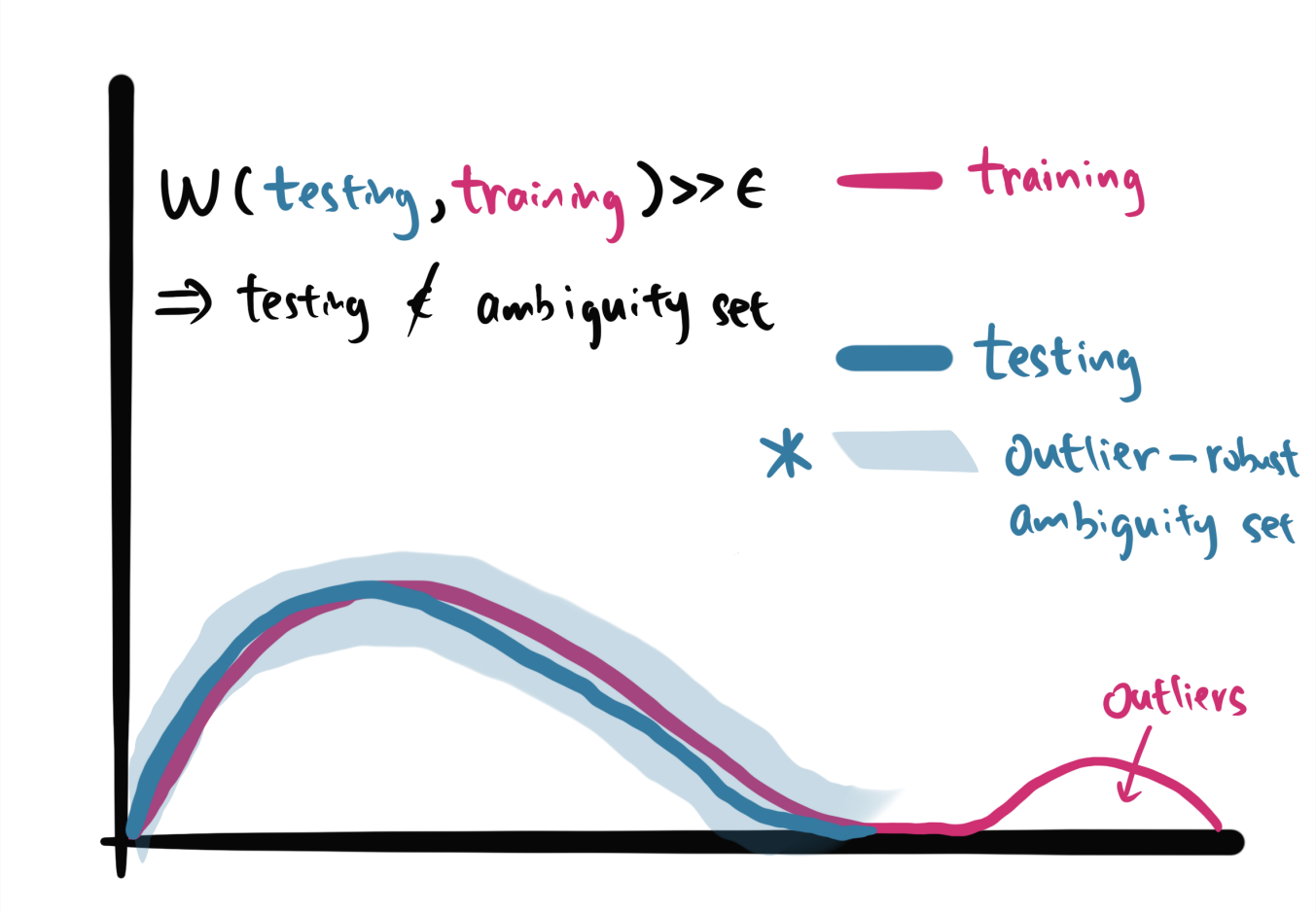
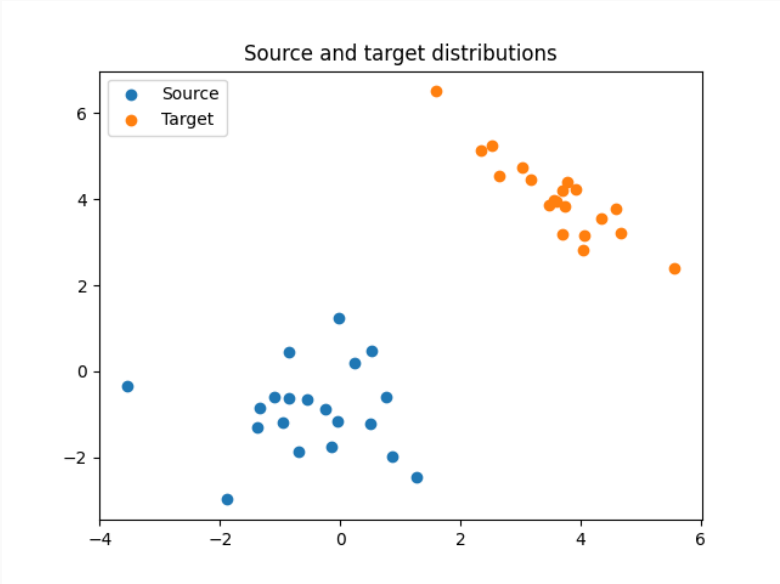


Figure 9: Wasserstein ball fails to include distirbutions of interests when outliers exist.

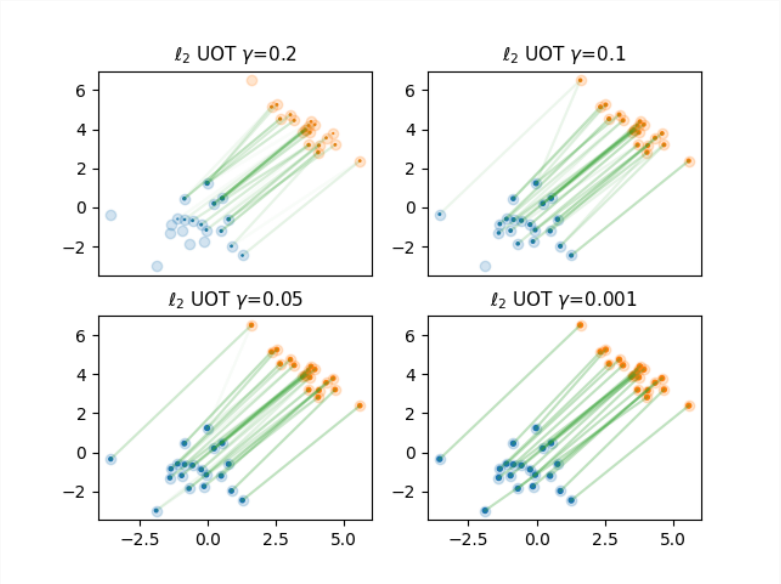
Task: design a distribution distance metric such that the distributions of interests are close to the training distribution (with outliers).

DRO with Outliers

If the offline dataset contains outliers, then any ambiguity set around the empirical distribution will include the outliers.



(a) Two discrete distributions



(b) Unbalanced optimal transport

Figure 10: An visualization of optimal transport plans using unbalanced optimal transport. [10]

[10] Image credit: pythonot.github.io

DRO with Outliers

Recall, the Wasserstein distance between P and Q is given by

$$W(P,Q) := \inf_{\pi \in \mathcal{M}(\Xi,\Xi)} \mathbb{E}_{(\xi,\xi')\sim\pi} \left[\|\xi - \xi'\|\right], \text{ s.t. } \pi_1 = P, \pi_2 = Q.$$

The unbalanced wasserstein distance is defined as [7]

$$\mathrm{UW}(P,Q) = \inf_{\pi \in \mathcal{M}(\Xi,\Xi)} \mathbb{E}_{(\xi,\xi')\sim\pi} \left[\|\xi - \xi'\|\right] + D_{\varphi_1}(\pi_1|P) + D_{\varphi_2}(\pi_2|Q),$$

where D_{φ_1} and D_{φ_2} are given f-divergence metrics.

For ease of analysis and computation, we select $\varphi_1 = \iota_{\{1\}}$ (i.e., $\varphi_1(1) = 0$ and ∞ otherwise) and $\varphi_2(x) = x \log x - x + 1$ (KL divergence).

DRO with Outliers

Recall, we have the Wasserstein DRO as

$$\text{(W-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

By taking the unbalanced wasserstein divergence as the ambiguity set metric, in [8] [11] we show that

$$\text{(UW-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \lambda \log \mathbb{E}_{\xi \sim P^0} \left[\exp \left(\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) / \lambda \right) \right] \right\}.$$

[11] Wang, Shen, Zavalnos, Johansson, Outlier-Robust Distributionally Robust Optimization via Unbalanced Optimal Transport, NeurIPS 2024

DRO with Outliers

Recall, we have the Wasserstein DRO as

$$\text{(W-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

By taking the unbalanced wasserstein divergence as the ambiguity set metric, in [8] [11] we show that

$$\text{(UW-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \lambda \log \mathbb{E}_{\xi \sim P^0} \left[\exp \left(\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) / \lambda \right) \right] \right\}.$$

Problem:

The UW distance allows distributions without outliers to be close to the training distribution. However, distributions that contains outliers are still close to the training distribution. As a result, both W-DRO and UW-DRO may fail as the inner-max term $\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right)$ could be very large due to the existence of outliers, e.g., $l(\text{outlier}) = \infty$.

[11] Wang, Shen, Zavalnos, Johansson, Outlier-Robust Distributionally Robust Optimization via Unbalanced Optimal Transport, NeurIPS 2024

DRO with Outliers

Assumption

We are given a function $h(\zeta)$ s.t. it penalizes outliers.

The (UW-DRO) becomes

$$\inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \lambda \log \mathbb{E}_{\xi \sim P^0} \left[\exp \left(\sup_{\zeta \in \Xi} \left(l(\zeta) - h(\zeta) - \lambda(\xi - \zeta)^2 \right) / \lambda \right) \right] \right\}.$$

DRO with Outliers

Assumption

We are given a function $h(\zeta)$ s.t. it penalizes outliers.

The (UW-DRO) becomes

$$\inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \lambda \log \mathbb{E}_{\xi \sim P^0} \left[\exp \left(\sup_{\zeta \in \Xi} \left(l(\zeta) - h(\zeta) - \lambda(\xi - \zeta)^2 \right) / \lambda \right) \right] \right\}.$$

Wait! Can't you do the same for WDRO? For example:

$$\text{(W-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - h(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

DRO with Outliers

Assumption

We are given a function $h(\zeta)$ s.t. it penalizes outliers.

The (UW-DRO) becomes

$$\inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \lambda \log \mathbb{E}_{\xi \sim P^0} \left[\exp \left(\sup_{\zeta \in \Xi} \left(l(\zeta) - h(\zeta) - \lambda(\xi - \zeta)^2 \right) / \lambda \right) \right] \right\}.$$

Wait! Can't you do the same for WDRO? For example:

$$\text{(W-DRO)} = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - h(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

Question: Does this work?

Answer: No! Recall, ϵ is the uncertainty ball radius, to include the testing distribution $\mathbb{P}^*$, the radius for WDRO needs to be set very large! As a result, WDRO will provide overly conservative results.

In contrast, a small radius uncertainty ball can include $\mathbb{P}^*$ due to UOT.

Conclusion

Summary

1. We can apply Wasserstein DRO to address distribution shifts in decision-making problems.

Summary

1. We can apply Wasserstein DRO to address distribution shifts in decision-making problems.
2. The Wasserstein DRO problem is in general computationally intractable. 1. Dataset structures should be considered when placing ambiguity sets (conditional probability). 2. Approximation methods, e.g, smoothing, can be used to enable an end-to-end training.

Summary

1. We can apply Wasserstein DRO to address distribution shifts in decision-making problems.
2. The Wasserstein DRO problem is in general computationally intractable. 1. Dataset structures should be considered when placing ambiguity sets (conditional probability). 2. Approximation methods, e.g, smoothing, can be used to enable an end-to-end training.
3. Wasserstein DRO may provide overly conservative results when outliers exist. Unbalanced optimal transport can help.

Summary

1. We can apply Wasserstein DRO to address distribution shifts in decision-making problems.
2. The Wasserstein DRO problem is in general computationally intractable. 1. Dataset structures should be considered when placing ambiguity sets (conditional probability). 2. Approximation methods, e.g, smoothing, can be used to enable an end-to-end training.
3. Wasserstein DRO may provide overly conservative results when outliers exist. Unbalanced optimal transport can help.
4. Robustness is the core of assured autonomy. Distributional robustness provides one solution to data-driven assured autonomy problems.

Acknowledgments

- [4][12] is supported in part by AFOSR under award #FA9550-19-1-0169 and by NSF under award CNS-1932011. Xu is supported by the Whitehead Scholars Program and the Department of Biostatistics and Bioinformatics at Duke University.
- [8][13] is supported in part by Swedish Research Council Distinguished Professor Grant 2017-01078, Knut and Alice Wallenberg Foundation, Wallenberg Scholar Grant, the Swedish Strategic Research Foundation SUCCESS Grant, and AFOSR under award #FA9550-19-1-0169.

[12] Shen, Xu, Zavlanos, Wasserstein distributionally robust policy evaluation and learning for contextual bandits, TMLR 2024

[13] Wang, Shen, Zavalnos, Johansson, Outlier-Robust Distributionally Robust Optimization via Unbalanced Optimal Transport, NeurIPS 2024

Questions?

References i

Yuchi Tian, Kexin Pei, Suman Jana, and Baishakhi Ray.
Deeptest: Automated testing of deep-neural-network-driven autonomous cars.
In *Proceedings of the 40th international conference on software engineering*, pages 303–314, 2018.

Susan Athey and Guido W Imbens.
The state of applied econometrics: Causality and policy evaluation.
Journal of Economic perspectives, 31(2):3–32, 2017.

Nan Jiang and Lihong Li.
Doubly robust off-policy value evaluation for reinforcement learning.
In *International conference on machine learning*, pages 652–661. PMLR, 2016.

References ii


Yi Shen, Pan Xu, and Michael Zavlanos.
Wasserstein distributionally robust policy evaluation and learning for contextual bandits.
Transactions on Machine Learning Research, 2024.


Nian Si, Fan Zhang, Zhengyuan Zhou, and Jose Blanchet.
Distributionally robust batch contextual bandits.
Management Science, 69(10):5772–5793, 2023.


Tong Mu, Yash Chandak, Tatsunori B Hashimoto, and Emma Brunskill.
Factored dro: Factored distributionally robust policies for contextual bandits.
Advances in Neural Information Processing Systems, 35:8318–8331, 2022.

References iii

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard.
Unbalanced optimal transport: Dynamic and Kantorovich formulations.
Journal of Functional Analysis, 274(11):3090–3123, 2018.

Zifan Wang, Yi Shen, Michael M Zavlanos, and Karl Henrik Johansson.
Outlier-robust distributionally robust optimization via unbalanced optimal transport.
In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

References iv


International Stroke Trial Collaborative Group et al.
The international stroke trial (ist): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke.
The Lancet, 349(9065):1569–1581, 1997.


Peter AG Sandercock, Maciej Niewada, and Anna Członkowska.
The international stroke trial database.
Trials, 12(1):1–7, 2011.


Sloan Nietert, Ziv Goldfeld, and Soroosh Shafiee.
Outlier-robust wasserstein dro.
Advances in Neural Information Processing Systems, 36, 2024.

References v

Ruidi Chen and Ioannis Ch Paschalidis.
A robust learning approach for regression models based on distributionally robust optimization.
Journal of Machine Learning Research, 19(13):1–48, 2018.

Soroosh Shafieezadeh-Abadeh, Daniel Kuhn, and Peyman Mohajerin Esfahani.
Regularization via mass transportation.
Journal of Machine Learning Research, 20(103):1–68, 2019.

OPE/L for healthcare

The International Stroke Trial (IST) [9] is a randomized controlled trial that aims to study the effects of early administration of aspirin and/or heparin on the clinical course of acute ischaemic stroke.

The IST dataset [10] includes 19,435 patients with:

- contextual information: age, gender, and level of consciousness before treatment admissions, as well as follow-up results on day 14, including the occurrence of recurrent stroke, pulmonary embolism, and death.
- actions: prescribing both aspirin and heparin (high or medium doses) (a_1) and the control action of of not administering any treatment, neither aspirin or heparin (a_2). The behavioral policy $\pi_0(a = a_1) = 0.5$.
- cost: the cost function is calculated based on the recorded follow-up events on day 14.

OPE/L for healthcare

The International Stroke Trial (IST) [9] is a randomized controlled trial that aims to study the effects of early administration of aspirin and/or heparin on the clinical course of acute ischemic stroke.

Distribution shift: we split the offline dataset into a training set and a testing set, and we introduce a **selection bias** into the training set. Specifically, we randomly remove 50% of the patients in the training set who are not fully conscious. This creates a difference in the context distribution between the training set and the testing set, with the patients in the testing set being more likely to be unconscious before treatment than those in the training set.

OPE/L for healthcare

Q: How is the uncertainty set radius ϵ selected?

A: We can split the training set into two parts and calculate the Wasserstein distance between them, which yields the uncertainty set $\epsilon_{\text{w}} = 0.03$.

Table 1: Policy evaluation and learning results

Method / Uncertainty set radius (ϵ_{W})	0.03	0.05	0.1
OPE: KL DRO (ϵ_{KL})	0.30 (0.01)	0.374 (0.1)	0.74 (1.0)
OPE: Wasserstein DRO (LP on sub-support)	0.53	0.61	0.79
OPE: Regulated Wasserstein DRO (BSGD)	0.59	0.76	1.05
OPL: KL DRO (ϵ_{KL})	0.28 (0.01)	0.368 (0.1)	0.69 (1.0)
OPL: Regulated Wasserstein DRO (BSGD)	0.54	0.62	0.92
Expectation under $\hat{P}$ (training set, random policy)		0.28	
Expectation under Q (testing set, random policy)		0.38	

Optimal transport (exact matching)

```python
import gurobipy as gp

def wasserstein_distance(p, q, C):
    """
    Computes the Wasserstein distance between two discrete distributions
    given their support set and a distance matrix C.

    Args:
    - p: numpy array of shape (n,) representing the first distribution
    - q: numpy array of shape (m,) representing the second distribution
    - C: numpy array of shape (n, m) representing the distance matrix

    Equations:
    min sum_{ij} T[i,j]*C[i,j]
        s.t. sum_j T[i,j] == p[i]   for all i
             sum_i T[i,j] == q[j]   for all j
             T[i,j] >= 0            for all i,j

    Returns:
    - The Wasserstein distance between the two distributions
    """

    # Create a new model
    model = gp.Model()

    # Define the decision variables
    n = p.size
    m = q.size
    T = model.addVars(n, m, lb=0.0, ub=GRB.INFINITY)

    # Define the objective function
    obj = gp.quicksum(C[i, j] * T[i, j] for i in range(n) for j in range(m))
    model.setObjective(obj, GRB.MINIMIZE)

    # Add the constraints
    model.addConstrs((gp.quicksum(T[i, j] for j in range(m)) == p[i]) for i in range(n))
    model.addConstrs((gp.quicksum(T[i, j] for i in range(n)) == q[j]) for j in range(m))

    # Optimize the model
    model.optimize()

    return model.objVal
```

Figure 11: Python code example.

Outlier-robust W-DRO for Linear Regression

Consider a linear regression problem with the loss $l(\xi, \theta) =: |\theta^\intercal x - y|$, where a data point $\xi = (x, y)$ includes features x and a label y.

Clean data $\{X_i, \theta_*^\intercal X_i\}_{i=1}^n$, where $X_1, \ldots, X_n$ are i.i.d. from $\mathcal{N}(0, I_d)$.

Drawing a uniform random subset $S \subset [n]$ of size $\lfloor 0.1n \rfloor$, the corrupted data distribution $\hat{\mathbb{P}}_0$ is defined to be uniform over

$$\left\{ \left(C^{1\{i \in S\}} X_i, \left(-C^2\right)^{1\{i \in S\}} \theta_*^\intercal X_i + \rho \right) \right\}_{i=1}^n,$$

where $C = 8$ is a corruption scaling coefficient and $\rho = 0.1$ is a shift coefficient.

We consider prior knowledge on the mean of the clean distribution, denoted as $\bar{\xi}$, and design the function $h(\xi) = \lambda_2 \left\| \xi - \bar{\xi} \right\|$.

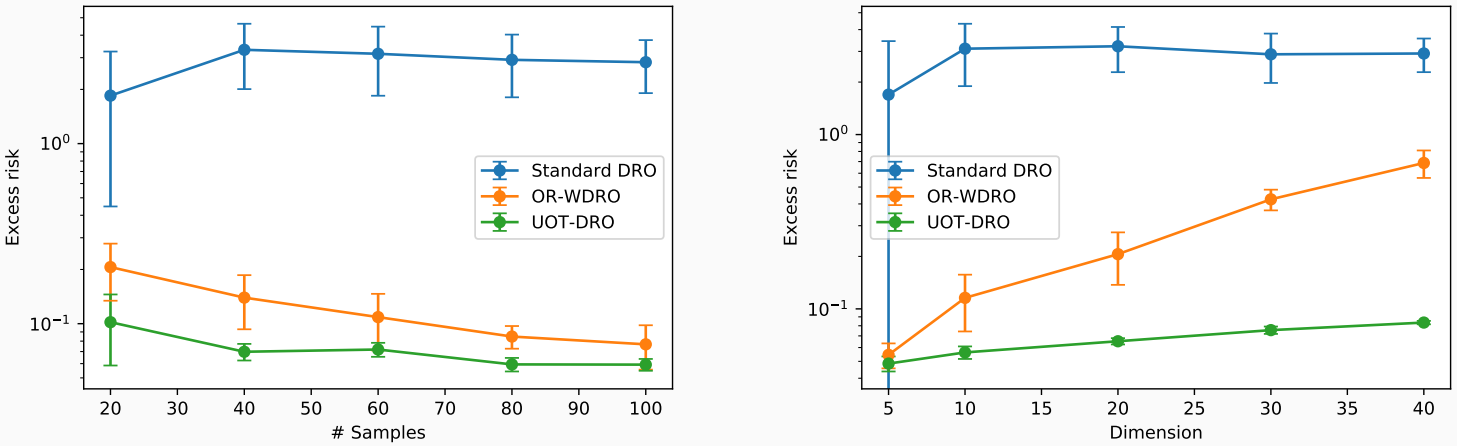
Wasserstein DRO in practice v.s. KL DRO

$$(\text{WDRO}) = \inf_{\lambda \geq 0} \left\{ \epsilon\lambda + \mathbb{E}_{\xi \sim P^0} \left[\sup_{\zeta \in \Xi} \left(l(\zeta) - \lambda(\xi - \zeta)^2 \right) \right] \right\}.$$

Benefits of using Wasserstein DRO

1. extra tuning parameters via the choice of distance metric in OT, e.g, 2-norm as above
2. the ambiguity set radius ϵ can be estimated if testing set is given; it can be lower bounded if testing set is not given

Outlier-robust W-DRO for Linear Regression



(a) Excess risk with various samples. (b) Excess risk with varied dimensions.

Figure 12: Excess risk (performance gap compared to θ^*) of standard DRO, OR-WDRO [11], and UOT-DRO [8] [14] with varied sample size and dimension for linear regression. The error bar denotes $\pm$ standard deviation.

[14] Wang, Shen, Zavalnos, Johansson, Outlier-Robust Distributionally Robust Optimization via Unbalanced Optimal Transport, NeurIPS 2024

Outlier-robust W-DRO for Linear Regression

Table 2: Comparison of running time and excess risk of different methods for linear regression. The symbol '*' indicates that running time is over 12000 seconds.

Sample Size n	Standard DRO		OR-WDRO [11]		UOT-DRO [8] [15]	
	Time	Excess risk	Time	Excess risk	Time	Excess risk
80	0.1	3.230	0.4	0.103	2.7	0.060
200	0.2	2.298	1.3	0.064	3.4	0.040
2000	3.3	0.441	29.8	0.050	4.7	0.038
5000	9.2	0.371	259.5	0.040	7.7	0.034
10000	28.9	0.352	1438.7	0.033	11.9	0.033
20000	110.8	0.380	*	*	22.2	0.031

[15] Wang, Shen, Zavalnos, Johansson, Outlier-Robust Distributionally Robust Optimization via Unbalanced Optimal Transport, NeurIPS 2024

Wasserstein Distributionally Robust Linear Regression

W-DRO for Linear Regression

Consider a linear regression problem with the loss $l(\xi, \theta) =: |\theta^\intercal x - y|$, where a data point $\xi = (x, y)$ includes features x and a label y.

classic robust optimization via regularization

$$\inf_{\theta \in \Theta} \mathbb{E}_{\xi \sim \hat{\mathbb{P}}_0} [l(\xi, \theta)] = \inf_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} |\theta^\intercal x_i - y_i| + \lambda \|\theta\|$$

W-DRO as regularization [12, 13]

$$(P) = \inf_{\theta \in \Theta} \sup_{\mathbb{Q} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_0)} \mathbb{E}_{\xi \sim \mathbb{Q}} [l(\xi, \theta)], \text{ where } \mathcal{B}_\epsilon(\hat{\mathbb{P}}_0) : \{\mathbb{Q} : W(\mathbb{Q}, \hat{\mathbb{P}}_0) \leq \epsilon\}$$

Q: How can you express the expectation in Q using offline data points?

W-DRO for Linear Regression

classic robust optimization via regularization

$$\inf_{\theta\in\Theta} \mathbb{E}_{\xi\sim\hat{\mathbb{P}}_0}\left[l(\xi,\theta)\right] = \inf_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^{N} |\theta^\intercal x_i - y_i| + \lambda \|\theta\|$$

W-DRO as regularization [12, 13]

$$(P) = \inf_{\theta\in\Theta} \sup_{\mathbb{Q}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_0)} \mathbb{E}_{\xi\sim\mathbb{Q}}\left[l(\xi,\theta)\right], \text{ where } \mathcal{B}_\epsilon(\hat{\mathbb{P}}_0) : \{\mathbb{Q} : W(\mathbb{Q},\hat{\mathbb{P}}_0) \le \epsilon\}$$

$$(D) = \inf_{\theta\in\Theta} \inf_{\lambda\ge 0, s_i} \lambda\epsilon + \frac{1}{N}\sum_{i=1}^{N} s_i, \text{ s.t. } \sup_{\zeta\in X\times Y} l(\zeta,\theta) - \lambda\|\zeta - \xi_i\| \le s_i, \ \forall i$$

zero duality gap (P)=(D),

$$(D) = \inf_{\theta\in\Theta} \epsilon\|(-\theta,1)\|_* + \frac{1}{N}\sum_{i=1}^{N} |\theta^\intercal x_i - y_i| \quad \text{this is a special case!}$$

What We Learned from W-DRO for Linear Regression

- W-DRO for linear regression is similar to classic regression methods with a regularization term.

What We Learned from W-DRO for Linear Regression

- W-DRO for linear regression is similar to classic regression methods with a regularization term.
- W-DRO in general involves an inner-maximization problem and is computationally challenging (linear regression has a closed-form solution to its inner-maximization problem).

$$\sup_{\zeta \in X \times Y} l(\zeta, \theta) - \lambda \|\zeta - \xi_i\| \leq s_i.$$

NO CLOSED-FORM SOLUTION!