Sinkhorn Distributionally Robust Optimization

Jie Wang[1], Rui Gao[2], and Yao Xie[1]

[1]Georgia Institute of Technology

[2]The University of Texas at Austin

International Conference on Continuous Optimization (ICCOPT)

Decision-Making Under Uncertainty

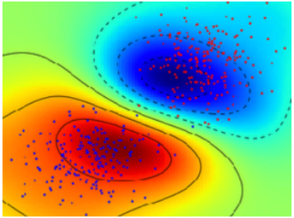
$$\text{Risk}: \qquad \mathscr{R}(\theta;\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[f_\theta(z)]$$

$$\text{Optimal Risk}: \qquad \mathscr{R}(\Theta;\mathbb{P}) = \inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}[f_\theta(z)]$$

Applications

Supply Chain Mgmt.

Portfolio Mgmt.



Machine Learning

Data-driven Decision-Making

- Available Information:

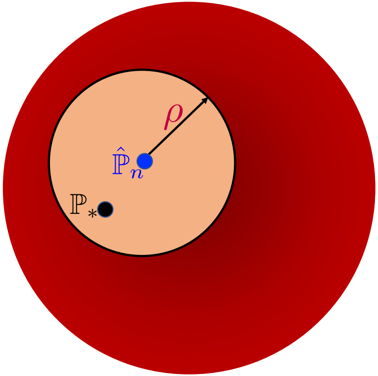
$$\begin{array}{ll} \text{Structual :} & \mathbb{P} \text{ is supported on } \Omega \subseteq \mathbb{R}^d \\ \text{Statistical :} & \hat{x}_1, \ldots, \hat{x}_n \sim \mathbb{P} \end{array}$$

- Nominal Problem:

  - Non-parametric estimators: $\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\hat{x}_i}$.
  - Kernel density estimators: $\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^n K(\hat{x}_i)$.

Wasserstein DRO

Definition: $\mathscr{P} = \{\mathbb{P} : \; W(\mathbb{P}, \hat{\mathbb{P}}_n) \leq \rho\}$.



Contain each $\mathbb{P}$ such that $W(\mathbb{P}, \hat{\mathbb{P}}_n) \leq \rho$

Worst-case risk : $\sup_{\mathbb{P} \in \mathscr{P}} \mathbb{E}_{\mathbb{P}}[f_{\theta}(z)]$

Robust Optimal Risk : $\inf_{\theta \in \Theta} \sup_{\mathbb{P} \in \mathscr{P}} \mathbb{E}_{\mathbb{P}}[f_{\theta}(z)]$

Limitations of Wasserstein DRO

- Worst-case distribution is discrete:

  For WDRO with n-point nominal distribution, the worst-case distribution is supported on $n+1$ points[1]*.*

- Tractability for limited scenarios:

  Finite-dimensional convex reformulation is available if the objective is a pointwise maximum of finitely many concave functions[2]*.*

- Some cases the same performance as SAA[2].

[1] Rui Gao and Anton J. Kleywegt. "Distributionally Robust Stochastic Optimization with Wasserstein Distance". In: *arXiv preprint arXiv:1604.02199* (Apr. 2016).

[2] Peyman Mohajerin Esfahani and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations". In: *Mathematical Programming* 171.1 (July 2017), pp. 115–166.

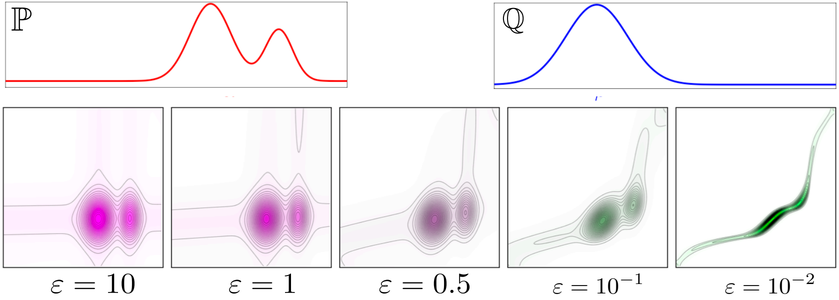
Sinkhorn Distance

- Sinkhorn Distance [Cuturi 2013]:

$$W_{\varepsilon}(\mathbb{P},\mathbb{Q}) = \inf_{\gamma \in \Gamma(\mathbb{P},\mathbb{Q})} \left\{ \mathbb{E}_{(X,Y)\sim\gamma}[c(X,Y)] + \varepsilon H(\gamma \,|\, \mathbb{P}\otimes\nu) \right\}.$$

- Relative Entropy between γ and $\mathbb{P}\otimes\nu$:

$$H(\gamma \,|\, \mathbb{P}\otimes\nu) = \int \log\left(\frac{\mathrm{d}\gamma(x,y)}{\mathrm{d}\mathbb{P}(x)\,\mathrm{d}\nu(y)}\right) \mathrm{d}\gamma(x,y).$$



Highlights of Sinkhorn Distance

Probability distance between distributions in $\mathbb{R}^d$ using n samples:

	MMD	Wasserstein	Sinkhorn
Computation	$O(n)$	$\tilde{O}(n^3)$	$\tilde{O}(n^2)$ [Altschuler, Niles-Weed, and Rigollet 2017]
Sample Complexity	$O(n^{-1/2})$	$O(n^{-1/d})$	$O(e^{\kappa/\varepsilon} n^{-1/2} \varepsilon^{-\lfloor d/2 \rfloor})$[Genevay et al. 2019]

- Fast algorithms for implementation;
- Sharp sample complexity rate;
- Encourage stochastic optimal transport (helpful in some applications, e.g., domain adaptation [Courty, Flamary, and Tuia 2014]).

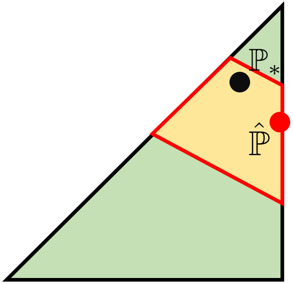
Sinkhorn DRO

- Sinkhorn DRO:

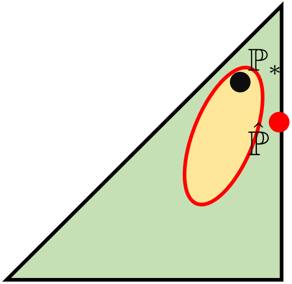
$$\inf_{\theta} \sup_{\mathbb{P}\in\mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}})} \mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)],$$

$$\text{where } \mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}}) = \left\{\mathbb{P} : \; W_\varepsilon(\widehat{\mathbb{P}},\mathbb{P}) \leq \rho\right\}.$$

where $\mu = \hat{\mathbb{P}}$ and ν is a measure independent of $\mathbb{P}$.



Ambiguity set for Wasserstein DRO



Ambiguity set for Sinkhorn DRO

Sinkhorn DRO

- Sinkhorn DRO:
$$\inf_{\theta} \sup_{\mathbb{P}\in\mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}})} \mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)],$$
$$\text{where } \mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}}) = \left\{\mathbb{P}:\ W_\varepsilon(\widehat{\mathbb{P}},\mathbb{P}) \le \rho\right\}.$$
where $\mu = \hat{\mathbb{P}}$ and ν is a measure independent of $\mathbb{P}$.
- Outline:
  - Duality Formulation for Sinkhorn DRO
  - Optimization Algorithm
  - Numerical Results

Tractable Formulation

Assume that

(I) $\nu\{z:\ 0 \le c(x,z) < \infty\} = 1$ for $\widehat{\mathbb{P}}$-almost every x;

(II) The integral $\int e^{-c(x,z)/\varepsilon}\, \mathrm{d}\nu(z) < \infty$ for $\widehat{\mathbb{P}}$-almost every x;

(III) Ω is a measurable space, and the function $f:\ \Omega \to \mathbb{R} \cup \{\infty\}$ is measurable.

Consider the primal of the worst-risk evaluation problem:

$$V_{\mathrm{P}} = \sup_{\mathbb{P} \in \mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}})} \mathbb{E}_{z\sim\mathbb{P}}[f(z)], \quad \text{where } \mathbb{B}_{\rho,\varepsilon}(\widehat{\mathbb{P}}) = \left\{\mathbb{P}:\ W_{\varepsilon}(\widehat{\mathbb{P}},\mathbb{P}) \le \rho\right\}. \qquad \texttt{(Sinkhorn DRO)}$$

It admits the **strong dual reformulation**:

$$V_{\mathrm{D}} = \inf_{\lambda>0}\ \lambda\overline{\rho} + \lambda\varepsilon \int_{\Omega} \log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda\varepsilon)}\right]\right) \mathrm{d}\widehat{\mathbb{P}}(x),$$

where

$$\overline{\rho} = \rho + \varepsilon \int_{\Omega} \log\left(\int_{\Omega} e^{-c(x,z)/\varepsilon}\, \mathrm{d}\nu(z)\right) \mathrm{d}\widehat{\mathbb{P}}(x),$$

$$\mathrm{d}\mathbb{Q}_x(z) = \frac{e^{-c(x,z)/\varepsilon}}{\int_{\Omega} e^{-c(x,u)/\varepsilon}\, \mathrm{d}\nu(u)}\, \mathrm{d}\nu(z).$$

Duality for General Nominal Distributions

For light-tailed distribution $\widehat{\mathbb{P}}$, it holds that $V_{\mathrm{P}} = V_{\mathrm{D}} < \infty$:

$$V_{\mathrm{P}} = \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] : \; W_{\varepsilon}(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}$$

$$V_{\mathrm{D}} = \inf_{\lambda > 0} \; \lambda\overline{\rho} + \lambda\varepsilon \int_{\Omega} \log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda\varepsilon)} \right] \right) \mathrm{d}\widehat{\mathbb{P}}(x)$$

General procedure for showing the strong duality:

- First show the weak duality result $V_{\mathrm{P}} \le V_{\mathrm{D}}$.
- Show the existence of dual minimizer (take the limit of Lebesgue integration)
- Show optimality conditions for the dual problem.
- Construct a primal feasible solution $\tilde{\mathbb{P}}$ that is optimal, e.g., for $\lambda^* > 0$,

$$\mathrm{d}\tilde{\mathbb{P}} = \int \frac{e^{f(z)/(\lambda^*\varepsilon)}\,\mathrm{d}\mathbb{Q}_x(z)}{\mathbb{E}_{\mathbb{Q}_x}[e^{f(z)/(\lambda^*\varepsilon)}]} \mathrm{d}\widehat{\mathbb{P}}(x) \implies V_{\mathrm{P}} \ge \mathbb{E}_{z\sim\tilde{\mathbb{P}}}[f(z)] = V_{\mathrm{D}}.$$

Duality for General Nominal Distributions

For light-tailed distribution $\widehat{\mathbb{P}}$, it holds that $V_{\text{P}} = V_{\text{D}} < \infty$:

$$V_{\text{P}} = \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] : \; W_{\varepsilon}(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}$$

$$V_{\text{D}} = \inf_{\lambda > 0} \; \lambda\overline{\rho} + \lambda\varepsilon \int_{\Omega} \log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda\varepsilon)} \right] \right) \mathrm{d}\widehat{\mathbb{P}}(x)$$

General procedure for showing the strong duality:

- First show the weak duality result $V_{\text{P}} \le V_{\text{D}}$.
- Show the existence of dual minimizer (take the limit of Lebesgue integration)
- Show optimality conditions for the dual problem.
- Construct a primal feasible solution $\tilde{\mathbb{P}}$ that is optimal, e.g., for $\lambda^* > 0$,

$$\mathrm{d}\tilde{\mathbb{P}} = \int \frac{e^{f(z)/(\lambda^*\varepsilon)}\,\mathrm{d}\mathbb{Q}_x(z)}{\mathbb{E}_{\mathbb{Q}_x}[e^{f(z)/(\lambda^*\varepsilon)}]}\,\mathrm{d}\widehat{\mathbb{P}}(x) \implies V_{\text{P}} \ge \mathbb{E}_{z\sim\tilde{\mathbb{P}}}[f(z)] = V_{\text{D}}.$$

Duality for General Nominal Distributions

For light-tailed distribution $\widehat{\mathbb{P}}$, it holds that $V_{\mathrm{P}} = V_{\mathrm{D}} < \infty$:

$$V_{\mathrm{P}} = \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] : \ W_\varepsilon(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}$$

$$V_{\mathrm{D}} = \inf_{\lambda > 0} \ \lambda\overline{\rho} + \lambda\varepsilon \int_{\Omega} \log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda\varepsilon)} \right] \right) \mathrm{d}\widehat{\mathbb{P}}(x)$$

General procedure for showing the strong duality:

- First show the weak duality result $V_{\mathrm{P}} \le V_{\mathrm{D}}$.
- Show the existence of dual minimizer (take the limit of Lebesgue integration)
- Show optimality conditions for the dual problem.
- Construct a primal feasible solution $\tilde{\mathbb{P}}$ that is optimal, e.g., for $\lambda^* > 0$,

$$\mathrm{d}\tilde{\mathbb{P}} = \int \frac{e^{f(z)/(\lambda^*\varepsilon)}\,\mathrm{d}\mathbb{Q}_x(z)}{\mathbb{E}_{\mathbb{Q}_x}[e^{f(z)/(\lambda^*\varepsilon)}]} \mathrm{d}\widehat{\mathbb{P}}(x) \implies V_{\mathrm{P}} \ge \mathbb{E}_{z\sim\tilde{\mathbb{P}}}[f(z)] = V_{\mathrm{D}}.$$

Duality for General Nominal Distributions

For light-tailed distribution $\widehat{\mathbb{P}}$, it holds that $V_{\mathrm{P}} = V_{\mathrm{D}} < \infty$:

$$V_{\mathrm{P}} = \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] : \; W_{\varepsilon}(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}$$

$$V_{\mathrm{D}} = \inf_{\lambda>0} \; \lambda\overline{\rho} + \lambda\varepsilon \int_{\Omega} \log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda\varepsilon)} \right] \right) \mathrm{d}\widehat{\mathbb{P}}(x)$$

General procedure for showing the strong duality:

- First show the weak duality result $V_{\mathrm{P}} \le V_{\mathrm{D}}$.
- Show the existence of dual minimizer (take the limit of Lebesgue integration)
- Show optimality conditions for the dual problem.
- Construct a primal feasible solution $\tilde{\mathbb{P}}$ that is optimal, e.g., for $\lambda^* > 0$,

$$\mathrm{d}\tilde{\mathbb{P}} = \int \frac{e^{f(z)/(\lambda^*\varepsilon)}\,\mathrm{d}\mathbb{Q}_x(z)}{\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(z)/(\lambda^*\varepsilon)}\right]} \mathrm{d}\widehat{\mathbb{P}}(x) \implies V_{\mathrm{P}} \ge \mathbb{E}_{z\sim\tilde{\mathbb{P}}}[f(z)] = V_{\mathrm{D}}.$$

Worst-case distribution $\tilde{\mathbb{P}}$ support on whole space, while W-DRO is discrete.

Connection of Sinkhorn DRO with Wasserstein DRO

When $\varepsilon \to 0$, the dual objective of Sinkhorn DRO converges into

$$\lambda\rho + \int \text{ess-sup}_{\nu}\, \big(f(\cdot) - \lambda c(x,\cdot)\big)\, \mathrm{d}\widehat{\mathbb{P}}(x).$$

When $\mathbf{supp}(\nu) = \Omega$,

Sinkhorn DRO

$$\begin{aligned} \sup_{\mathbb{P}} \quad & \mathbb{E}_{z\sim\mathbb{P}}[f(z)] \\ \text{s.t.} \quad & W_\epsilon(\hat{\mathbb{P}}, \mathbb{P}) \le \rho \end{aligned}$$

Take $\epsilon \to 0$ $\longrightarrow$

Wasserstein DRO

$$\begin{aligned} \sup_{\mathbb{P}} \quad & \mathbb{E}_{z\sim\mathbb{P}}[f(z)] \\ \text{s.t.} \quad & W(\hat{\mathbb{P}}, \mathbb{P}) \le \rho \end{aligned}$$

Connection of Sinkhorn DRO with KL DRO

Upper bound of Sinkhorn DRO:

$$
\begin{aligned}
V_{\mathrm{D}} &\triangleq \inf_{\lambda>0}\ \lambda\overline{\rho}+\lambda\varepsilon\int_{\Omega}\log\left(\mathbb{E}_{\mathbb{Q}_x}\left[e^{f(y)/(\lambda\varepsilon)}\right]\right)\mathrm{d}\widehat{\mathbb{P}}(x)\\
&\le \inf_{\lambda>0}\ \lambda\overline{\rho}+\lambda\varepsilon\log\left(\mathbb{E}_{\mathbb{P}^0}\left[e^{f(y)/(\lambda\varepsilon)}\right]\right)
\end{aligned}
$$

$\mathbb{P}^0$: kernel density estimate based on $\widehat{\mathbb{P}}$:

$$
\mathrm{d}\mathbb{P}^0(z)=\int_x \mathrm{d}\mathbb{Q}_x(z)\,\mathrm{d}\widehat{\mathbb{P}}(x).
$$

Sinkhorn DRO

$$
\sup_{\mathbb{P}}\ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] \quad \text{s.t.}\quad W_\epsilon(\hat{\mathbb{P}},\mathbb{P})\le\rho
$$

$\xrightarrow{\text{Take } \mathbb{P}^0 \text{ as the KDE estimate of } \hat{\mathbb{P}}}$

KL DRO

$$
\sup_{\mathbb{P}}\ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] \quad \text{s.t.}\quad D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^0)\le\bar{\rho}/\epsilon
$$

Connection of Sinkhorn DRO with SAA

When $\overline{\rho} = 0$, Sinkhorn becomes SAA:

$$V_{\mathrm{P}} = \mathbb{E}_{z\sim\mathbb{P}^0}[f(z)]$$

$\mathbb{P}^0$: kernel density estimate based on $\widehat{\mathbb{P}}$:

$$\mathrm{d}\mathbb{P}^0(z) = \int_x \mathrm{d}\mathbb{Q}_x(z)\,\mathrm{d}\widehat{\mathbb{P}}(x).$$

Sinkhorn DRO

$$\sup_{\mathbb{P}} \quad \mathbb{E}_{z\sim\mathbb{P}}[f(z)]$$
$$\text{s.t.} \quad W_\epsilon(\hat{\mathbb{P}}, \mathbb{P}) \le \rho$$

$\xrightarrow{\text{Take } \bar{\rho} = 0}$

SAA

$$\mathbb{E}_{z\sim\mathbb{P}^0}[f(z)]$$

Choice of Hyper-parameters $(\varepsilon, \overline{\rho})$

Sinkhorn DRO

$$\min_{\theta} \sup_{\mathbb{P}} \quad \mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)] \quad \text{s.t.} \quad \mathcal{W}_\epsilon(\hat{\mathbb{P}}, \mathbb{P}) \le \rho$$

$$\xrightarrow{\text{Take } \bar{\rho} = 0}$$

SAA

$$\min_{\theta} \quad \int \mathbb{E}_{z\sim\mathbb{Q}_x}[f_\theta(z)] d\hat{\mathbb{P}}(x)$$

- First choose ε to optimize the hold-out performance for

$$\operatorname{argmin}_\theta \int \mathbb{E}_{z\sim\mathbb{Q}_x}[f(z)] \, \mathrm{d}\widehat{\mathbb{P}}(x), \quad \mathrm{d}\mathbb{Q}_x(z) \propto e^{-c(\hat{x}_i,z)/\varepsilon} \mathrm{d}\nu(z).$$

- For fixed ε, choose $\overline{\rho}$ to optimize the hold-out performance for

$$\operatorname{argmin}_\theta \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)] : \quad W_\varepsilon(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}.$$

Optimization Algorithm for Sinkhorn DRO

- Based on strong duality,

$$\min_{\theta\in\Theta}\sup_{\mathbb{P}} \left\{\mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)]: \quad W_\varepsilon(\widehat{\mathbb{P}},\mathbb{P})\le\rho\right\}$$
$$=\min_{\theta\in\Theta,\lambda\ge 0}\lambda\overline{\rho}+\frac{1}{n}\sum_{i=1}^{n}\lambda\varepsilon\log\left(\mathbb{E}_{\mathbb{Q}_{\hat{x}_i}}\left[e^{f_\theta(z)/(\lambda\varepsilon)}\right]\right)$$

- Solve the Monte-Carlo approximated formulation[3]:

$$\min_{\theta\in\Theta,\lambda\ge 0}\ \lambda\overline{\rho}+\frac{1}{n}\sum_{i=1}^{n}\lambda\varepsilon\log\left(\frac{1}{m}\sum_{j=1}^{m}e^{f_\theta(z_{i,j})/(\lambda\varepsilon)}\right),$$

where $\{z_{i,j}\}_j$ are i.i.d. samples generated from $\mathbb{Q}_{\hat{x}_i}$.

[3] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

Optimization Algorithm for Sinkhorn DRO: Biased Gradient Update

- Based on strong duality:

$$\min_{\theta\in\Theta}\left\{\sup_{\mathbb{P}}\ \mathbb{E}_{z\sim\mathbb{P}}[f_\theta(z)] - \lambda W_\varepsilon(\widehat{\mathbb{P}},\mathbb{P})\right\} = \min_{\theta\in\Theta}\left\{F(\theta) := \frac{1}{n}\sum_{i=1}^{n}\lambda\varepsilon\log\left(\mathbb{E}_{\mathbb{Q}_{\hat{x}_i}}[e^{f_\theta(z)/\lambda\varepsilon}]\right)\right\}.$$

- Biased gradient update: for each iteration t,
  - Construct a subgradient estimate[4] of $F(\theta_t)$, denoted as $v(\theta_t)$;
  - Update $\theta_{t+1} = \text{Proximal}_{\theta_t}\left(\gamma_t v(\theta_t)\right)$.

Estimators	Convex (Possibly Nonsmooth)			Nonconvex Smooth		
	Iteration	Per-iteration cost	Total Cost	Iteration	Per-iteration cost	Total Cost
L-SGD	$O(\delta^{-2})$	$O(\delta^{-1})$	$O(\delta^{-3})$	$O(\delta^{-4})$	$O(\delta^{-2})$	$O(\delta^{-6})$
V-MLMC	$O(\delta^{-1})$	$\widetilde{O}(\delta^{-1})$	$\widetilde{O}(\delta^{-2})$	$O(\delta^{-2})$	$\widetilde{O}(\delta^{-2})$	$\widetilde{O}(\delta^{-4})$
RT-MLMC	$\widetilde{O}(\delta^{-2})$	$O(1)$	$\widetilde{O}(\delta^{-2})$	$\widetilde{O}(\delta^{-4})$	$O(1)$	$\widetilde{O}(\delta^{-4})$

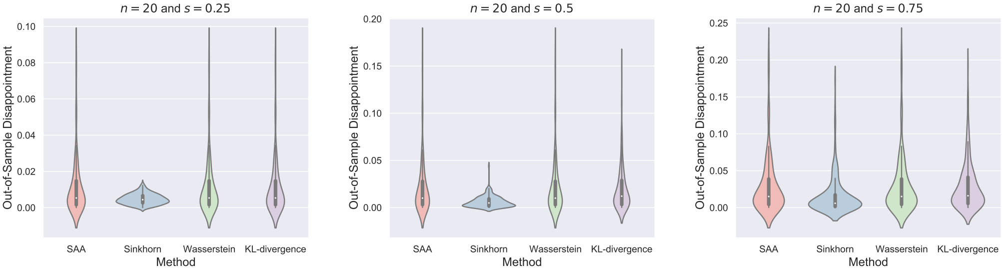
[4] Yifan Hu, Xin Chen, and Niao He. "On the Bias-Variance-Cost Tradeoff of Stochastic Optimization". In: *Advances in Neural Information Processing Systems*. Dec. 2021.

Numerical Results

Newsvendor problem:

$$\min_{\beta} \ \mathbb{E}_{\mathbb{P}_*}\left[k\beta - u\min\{\beta,\zeta\}\right], \quad k=5, u=7.$$

$\mathbb{P}_* \sim \exp(1/s)$ with $s \in \{0.25, 0.5, 0.75\}$. Access to $n = 20$ samples.

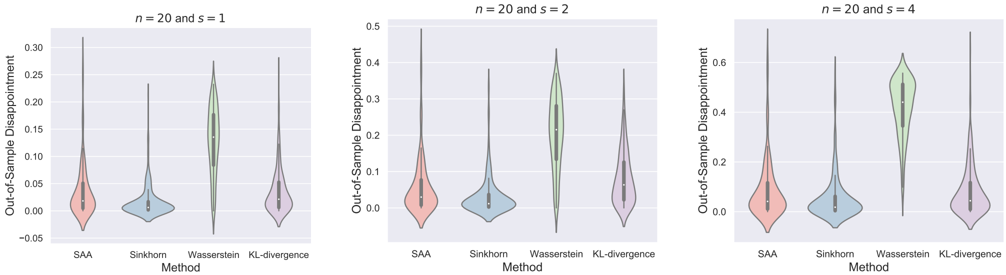


Numerical Results

Newsvendor problem:

$$\min_{\beta} \mathbb{E}_{\mathbb{P}_*}\left[k\beta - u\min\{\beta,\zeta\}\right], \quad k=5, u=7.$$

$\mathbb{P}_* \sim \exp(1/s)$ with $s \in \{1,2,4\}$. Access to $n=20$ samples.

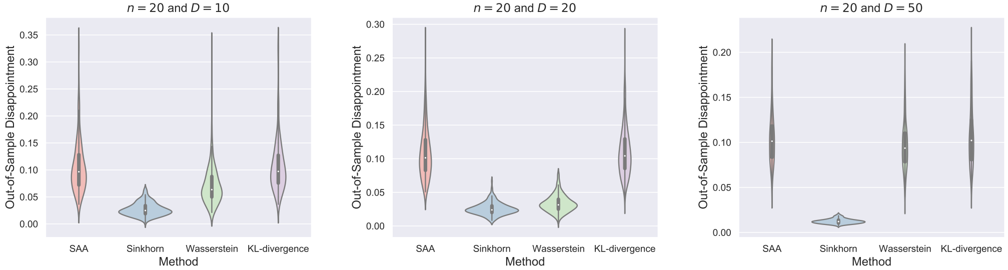


Numerical Results

Portfolio Optimization:

$$\inf_{x} \quad \mathbb{E}_{\mathbb{P}_*}[-\langle x, \zeta \rangle] + \rho \cdot \mathbb{P}_*\text{-CVaR}_\alpha(-\langle x, \zeta \rangle)$$
$$\text{s.t.} \quad x \in \mathscr{X} = \{x \in \mathbb{R}_+^D : \; x^{\mathrm{T}} 1 = 1\}.$$

Dimension $D \in \{10, 20, 50\}$ with sample size $n = 20$.



Numerical Simulation Results

Semi-supervised Learning:

- Train classifiers based on data with labels and without labels;
- Two performance measures:
  - Training error for data without labels;
  - Testing error.

	SAA	Sinkhorn	Wasserstein	KL-divergence
	$.20 \pm .068$	$\mathbf{.12 \pm .068}$	$.17 \pm .073$	$.19 \pm .038$
Breast Cancer	$.19 \pm .073$	$\mathbf{.11 \pm .067}$	$.17 \pm .075$	$.19 \pm .073$
	$.28 \pm .082$	$\mathbf{.25 \pm .091}$	$.27 \pm .077$	$.26 \pm .078$
Magic	$.28 \pm .064$	$\mathbf{.25 \pm .074}$	$.27 \pm .058$	$.27 \pm .066$
	$.25 \pm .057$	$\mathbf{.22 \pm .063}$	$.23 \pm .073$	$.25 \pm .037$
QSAR Bio	$.25 \pm .062$	$\mathbf{.22 \pm .065}$	$.23 \pm .079$	$.25 \pm .042$
	$.19 \pm .038$	$\mathbf{.14 \pm .046}$	$.16 \pm .036$	$.18 \pm .034$
Spambase	$.19 \pm .032$	$\mathbf{.14 \pm .036}$	$.16 \pm .028$	$.18 \pm .042$

Take Home Message

Sinkhorn DRO is a great notion of DRO models:

- Inherit geometric properties from optimal transport;
- Absolutely continuous worst-case distribution thanks to entropic regularization;
- Improve the out-of-sample performance of Wasserstein DRO;
- Optimization by Monte Carlo approximation and first order method;
- More applications in operations research with Sinkhorn DRO can be explored!

Sinkhorn Distributionally Robust Optimization

(Submitted to Operations Research – INFORMS PUBs)

Online Available: arxiv.org/abs/2109.11926

SCAN ME