Bulk-Calibrated Credal Ambiguity Sets:

Fast, Tractable Decision Making under Out-of-Sample Contamination

LAI Reading Group, 24th June 2026

Mengqi Chen



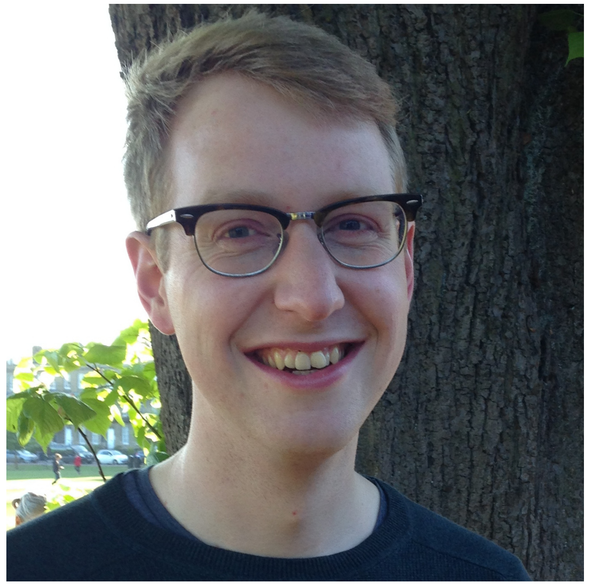
ICML
International Conference
On Machine Learning



WARWICK
THE UNIVERSITY OF WARWICK



Collaborators



Tom Berrett
@Warwick



Theo Damoulas
@Warwick



Michele Caprio
@Manchester

Robust decision-making

Data $\mathscr{D} = \{\xi_{1:n}\}$, $\xi_i \in \Xi \subseteq \mathbb{R}^d \overset{i.i.d.}{\sim} \mathbb{P}^*$, decision $x \in \mathscr{X}$, loss function $f_x(\xi)$

Traditional decision making:

Goal: find x that solves $\min_{x \in \mathscr{X}} \mathbb{E}_{\xi \sim \mathbb{P}^*} \left[f_x(\xi) \right]$

Worst-case protected decision making under distributional uncertainty: we only know that test-time law lies in a plausible set $\mathscr{A}$.

Goal: find x that solves $\min_{x \in \mathscr{X}} \sup_{Q \in \mathscr{A}} \mathbb{E}_{\xi \sim Q} \left[f_x(\xi) \right]$

Ambiguity Sets

Robust decision-making

Data $\mathscr{D} = \{\xi_{1:n}\}$, $\xi_i \in \Xi \subseteq \mathbb{R}^d \overset{i.i.d.}{\sim} \mathbb{P}^*$, decision $x \in \mathscr{X}$, loss function $f_x(\xi)$

Traditional decision making:

Goal: find x that solves $\min_{x \in \mathscr{X}} \mathbb{E}_{\xi \sim \mathbb{P}^*} \left[f_x(\xi) \right]$

Worst-case protected decision making under distributional uncertainty: we only know that test-time law lies in a plausible set $\mathscr{A}$.

Goal: find x that solves $\min_{x \in \mathscr{X}} \sup_{Q \in \mathscr{A}} \mathbb{E}_{\xi \sim Q} \left[f_x(\xi) \right]$

Ambiguity Sets

Kuhn et al. (2025)

- Moment-based, support only, …
  - Isii (1960) Ben-Tal & Nemirovski (1998)
- Fixed **divergence** around a nominal centre
  - Wasserstein, KL, Chi^2, TV, MMD, …
  - Mohajerin Esfahani & Kuhn (2018) Hu & Hong (2013) Namkoong and Duchi (2016)

Robust decision-making

Data $\mathscr{D} = \{\xi_{1:n}\}$, $\xi_i \in \Xi \subseteq \mathbb{R}^d \overset{i.i.d.}{\sim} \mathbb{P}^*$, decision $x \in \mathscr{X}$, loss function $f_x(\xi)$

Traditional decision making:

Goal: find x that solves $\min_{x \in \mathscr{X}} \mathbb{E}_{\xi \sim \mathbb{P}^*} \left[f_x(\xi) \right]$

Worst-case protected decision making under distributional uncertainty: we only know that test-time law lies in a plausible set $\mathscr{A}$.

Goal: find x that solves $\min_{x \in \mathscr{X}} \sup_{Q \in \mathscr{A}} \mathbb{E}_{\xi \sim Q} \left[f_x(\xi) \right]$

Ambiguity Sets

Kuhn et al. (2025)

- Moment-based, support only, …
  - Isii (1960)
  - Ben-Tal & Nemirovski (1998)
- Fixed **divergence** around a nominal centre
  - Wasserstein, KL, Chi^2, TV, MMD, …
  - Mohajerin Esfahani & Kuhn (2018)
  - Hu & Hong (2013)
  - Namkoong and Duchi (2016)

💡 Divergence choice <-> type of shift expected

💡 Computable (dual) forms for optimisation

Credal

Imprecise Probability

$\mathscr{M}$: *set of probabilities (convex and closed)*

Upper expectation

Ambiguity Sets

Distributionally Robust Optimisation

$\mathscr{A}$: distributional ambiguity set

Worst-case risk

$$\bar{\mathbb{E}}_{\mathscr{M}}[f_x(\xi)] \overset{\mathscr{M}=\mathscr{A}}{:=} \sup_{Q\in\mathscr{A}} \mathbb{E}_Q[f_x(\xi)] =: \mathscr{R}_{\mathscr{A}}(f_x(\xi))$$

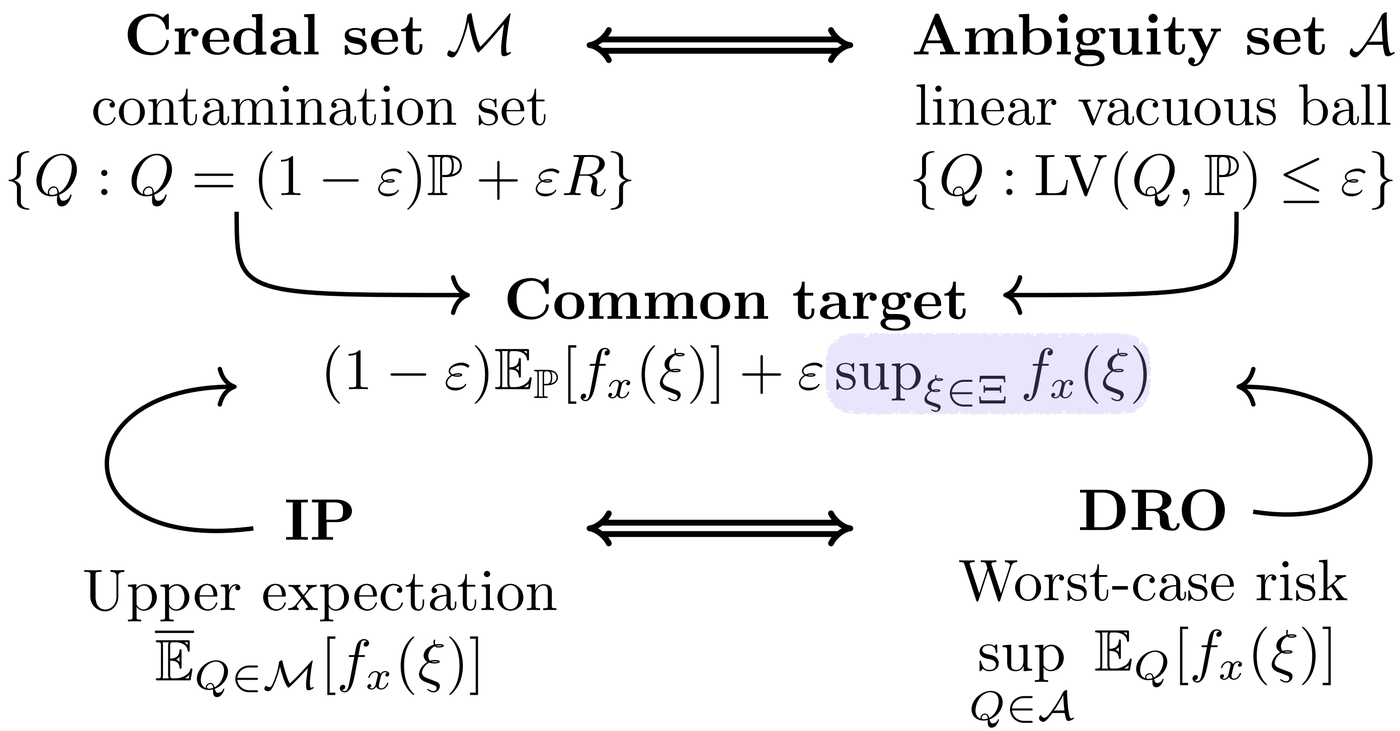
Common Optimisation target

Data $\mathscr{D} = \{\xi_{1:n}\}$, $\xi_i \in \Xi \subseteq \mathbb{R}^d \overset{i.i.d.}{\sim} \mathbb{P}^*$, decision $x \in \mathscr{X}$, loss function $f_x(\xi)$

Out-of-sample environment: **Huber (linear vacuous) contamination** $\mathscr{A} = \{\tilde{\mathbb{P}} : \tilde{\mathbb{P}} = (1-\varepsilon)\mathbb{P}^* + \varepsilon R,\ R \in \mathscr{P}(\Xi)\}$

Goal: find x that solves $\min_{x\in\mathscr{X}} \sup_{Q\in\mathscr{A}} \mathbb{E}_{\xi\sim Q}\left[f_x(\xi)\right]$

Credal Ambiguity Sets



Challenge: $\sup_{\xi \in \Xi} f_x(\xi)$ is **infinite** if f_x and Ξ are unbounded

e.g. linear regression with Gaussian samples

Credal Ambiguity Sets

Jiang & Guan, (2018) Staib & Jegelka, (2019)

Credal set $\mathcal{M}$ $\Longleftrightarrow$ **Ambiguity set** $\mathcal{A}$ $\subseteq$ TV ball, MMD ball, …

contamination set — linear vacuous ball

$\{Q : Q = (1-\varepsilon)\mathbb{P} + \varepsilon R\}$ — $\{Q : \mathrm{LV}(Q, \mathbb{P}) \leq \varepsilon\}$

Bounded Ξ Restricted f

Common target

$$(1-\varepsilon)\mathbb{E}_{\mathbb{P}}[f_x(\xi)] + \varepsilon \sup_{\xi\in\Xi} f_x(\xi)$$

IP $\Longleftrightarrow$ **DRO**

Upper expectation

$\overline{\mathbb{E}}_{Q\in\mathcal{M}}[f_x(\xi)]$

Worst-case risk

$\sup_{Q\in\mathcal{A}} \mathbb{E}_Q[f_x(\xi)]$

Challenge: $\sup_{\xi\in\Xi} f_x(\xi)$ is **infinite** if f_x and Ξ are unbounded

e.g. linear regression with Gaussian samples

Outline for this talk

- Bulk-restricted credal ambiguity set: definition, properties, links to other contamination neighbourhoods
- Bulk set calibration: procedure and guarantees
- Experiments: insights and extensions

Bulk-restricted credal ambiguity sets

Definition

Arbitrary contamination on the **bulk** + separate tail conditions

↓

A bounded set $\Xi_0(\mathscr{D})$ learned from data such that

$$\Pr\left(\mathbb{P}^*\left(\Xi_0\right) \geq 1-\gamma\right) \geq 1-\delta$$

Chosen centre $\mathbb{P}_c$: Bayesian posterior predictive, frequentist plug-in, empirical, …

Define our bulk-restricted credit ambiguity sets, given Ξ_0, as

$$\mathscr{A}^{\mathrm{LV}}_{\varepsilon,\Xi_0} = \left\{(1-\varepsilon)\mathbb{P}_{c,\Xi_0} + \varepsilon R : R \in \mathscr{P}(\Xi_0)\right\} = \left\{Q \in \mathscr{P}(\Xi_0) : \mathrm{LV}\left(Q, \mathbb{P}_{c,\Xi_0}\right) \leq \varepsilon\right\}$$

$\mathbb{P}_{c,\Xi_0}(A) := \dfrac{\mathbb{P}_c(A \cap \Xi_0)}{\mathbb{P}_c(\Xi_0)}$, restricted centre

linear-vacuous distortion

$$\mathrm{LV}(Q, \mathbb{P}) := \sup_{A:\mathbb{P}(A)>0} \frac{\mathbb{P}(A) - Q(A)}{\mathbb{P}(A)}$$

Bulk-restricted credal ambiguity sets

Properties

Worst-case **risk:** $(1-\varepsilon)\mathbb{E}_{\xi\sim\mathbb{P}_{c,\Xi_0}}\left[f_x(\xi)\right] + \varepsilon \sup_{\xi\in\Xi_0} f_x(\xi)$

Approximated via Rejection sampling

Calibrate Ξ_0 to make this tractable common f_x

Bulk-restricted credal ambiguity sets

Properties

Worst-case **risk:** $(1-\varepsilon)\mathbb{E}_{\xi\sim\mathbb{P}_{c,\Xi_0}}\left[f_x(\xi)\right] + \varepsilon \sup_{\xi\in\Xi_0} f_x(\xi)$

Approximated via Rejection sampling

Calibrate Ξ_0 to make this tractable common f_x

Worst-case **distribution:** $(1-\varepsilon)\mathbb{P}_{c,\Xi_0} + \varepsilon\delta_{M_x}$

Set (singleton) attaining supremum

Bulk-restricted credal ambiguity sets

Properties

Worst-case **risk:** $(1-\varepsilon)\mathbb{E}_{\xi\sim\mathbb{P}_{c,\Xi_0}}\left[f_x(\xi)\right] + \varepsilon \sup_{\xi\in\Xi_0} f_x(\xi)$

Approximated via Rejection sampling

Calibrate Ξ_0 to make this tractable common f_x

Worst-case **distribution:** $(1-\varepsilon)\mathbb{P}_{c,\Xi_0} + \varepsilon\delta_{M_x}$

Set (singleton) attaining supremum

Risk certificate: for any $\tilde{\mathbb{P}} = (1-\varepsilon^*)\mathbb{P}^* + \varepsilon^* R,\ R \in \mathscr{P}(\Xi)$

$$\mathbb{E}_{\tilde{\mathbb{P}}}[f_x(\xi)] \le (1-\varepsilon_{\text{eff}})\mathbb{E}_{\mathbb{P}_{c,\Xi_0}}[f_x(\xi)] + \varepsilon_{\text{eff}} \sup_{\xi\in\Xi_0} f_x(\xi) + \text{tail term}$$

Depends on:

ε_c: centre distortion, ε^*: true contamination ratio, ρ_{Ξ_0}: how much mass R places on Ξ_0

Bulk-restricted credal ambiguity sets

Relationship with other contamination neighbourhoods

Forward LV ball: $\{Q : \mathrm{LV}(Q, \mathbb{P}) \leq \varepsilon\} = \{Q : Q = (1-\varepsilon)\mathbb{P} + \varepsilon R\}$

Worst-case risk: $(1-\varepsilon)\mathbb{E} + \varepsilon \sup$

Reverse LV ball: $\{Q : \mathrm{LV}(\mathbb{P}, Q) \leq \varepsilon\} = \{Q : \mathbb{P} = (1-\varepsilon)Q + \varepsilon R\}$

Worst-case risk: $\mathrm{CVaR}_{1-\varepsilon}$

$\subseteq$ TV ball $\{Q : \mathrm{TV}(\mathbb{P}, Q) \leq \varepsilon\}$

Worst-case risk: $(1-\varepsilon)\mathrm{CVaR}_{1-\varepsilon} + \varepsilon \sup$

Bulk-restricted credal ambiguity sets

Relationship with other contamination neighbourhoods

Forward LV ball: $\{Q : \mathrm{LV}(Q, \mathbb{P}) \leq \varepsilon\} = \{Q : Q = (1-\varepsilon)\mathbb{P} + \varepsilon R\}$

Worst-case risk: $(1-\varepsilon)\mathbb{E} + \varepsilon \sup$

Reverse LV ball: $\{Q : \mathrm{LV}(\mathbb{P}, Q) \leq \varepsilon\} = \{Q : \mathbb{P} = (1-\varepsilon)Q + \varepsilon R\}$

Worst-case risk: $\mathrm{CVaR}_{1-\varepsilon}$

$\subseteq$ TV ball $\{Q : \mathrm{TV}(\mathbb{P}, Q) \leq \varepsilon\}$

Worst-case risk: $(1-\varepsilon)\mathrm{CVaR}_{1-\varepsilon} + \varepsilon \sup$

"Similarity decomposition": $\exists C$ s.t.

$$\mathbb{P} = (1-\varepsilon)C + \varepsilon R_1$$

$$Q = (1-\varepsilon)C + \varepsilon R_2$$

Bulk-restricted credal ambiguity sets

Relationship with other contamination neighbourhoods

Forward LV ball: $\{Q : \mathrm{LV}(Q, \mathbb{P}) \le \varepsilon\} = \{Q : Q = (1-\varepsilon)\mathbb{P} + \varepsilon R\}$

Worst-case risk: $(1-\varepsilon)\mathbb{E} + \varepsilon \sup$

Adding ε worst points

Reverse LV ball: $\{Q : \mathrm{LV}(\mathbb{P}, Q) \le \varepsilon\} = \{Q : \mathbb{P} = (1-\varepsilon)Q + \varepsilon R\}$

Worst-case risk: $\mathrm{CVaR}_{1-\varepsilon}$

Deleting ε best part

$\subseteq$ TV ball $\{Q : \mathrm{TV}(\mathbb{P}, Q) \le \varepsilon\}$

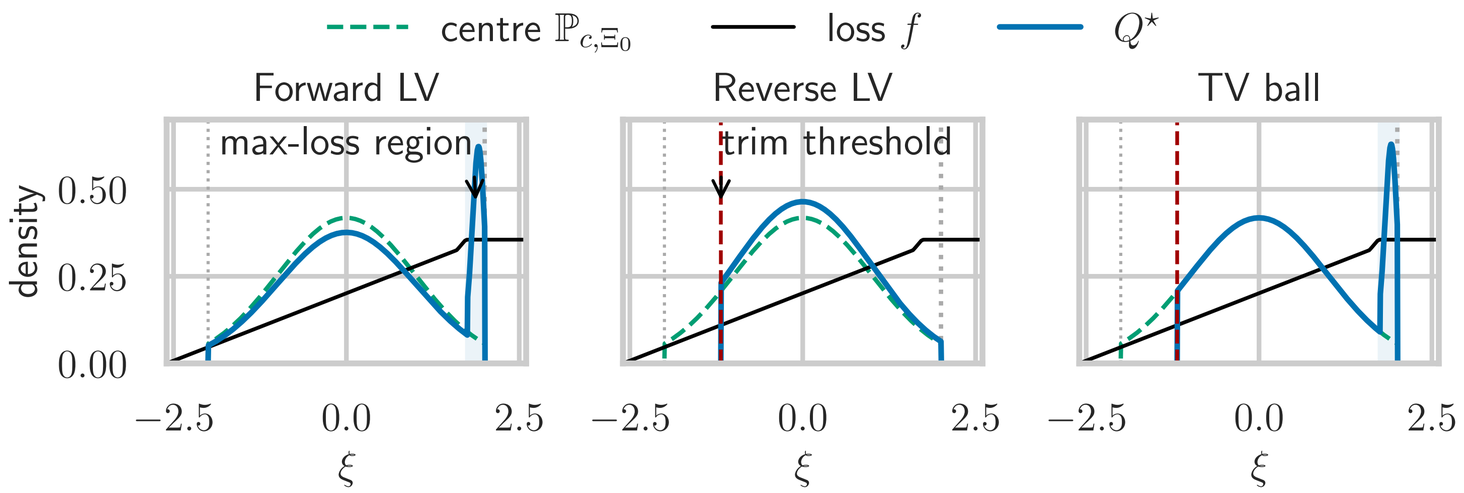
Worst-case risk: $(1-\varepsilon)\mathrm{CVaR}_{1-\varepsilon} + \varepsilon \sup$

"Similarity decomposition": $\exists C$ s.t.

$\mathbb{P} = (1-\varepsilon)C + \varepsilon R_1$

$Q = (1-\varepsilon)C + \varepsilon R_2$

Allows both



Bulk-*calibrated* credal ambiguity sets

Bulk-set calibration

Goal: find $\Xi_0(\mathscr{D})$ such that

1. $\Pr\left(\mathbb{P}^*\left(\Xi_0\right) \geq 1-\gamma\right) \geq 1-\delta$
2. $\sup_{\xi \in \Xi_0} f_x(\xi)$ is tractable for common losses f_x

Bulk-*calibrated* credal ambiguity sets

Bulk-set calibration

Goal: find $\Xi_0(\mathscr{D})$ such that

1. $\Pr\left(\mathbb{P}^*\left(\Xi_0\right) \geq 1-\gamma\right) \geq 1-\delta$
2. $\sup_{\xi \in \Xi_0} f_x(\xi)$ is tractable for common losses f_x

Split $\mathscr{D}$ into score **fitting** set + score **selection** set:

$\mathscr{D}_{\text{fit}}$

↓ Defines a score function

$$s(\xi) = \begin{cases} \|\Sigma_{\text{fit}}^{-1/2}(\xi - \mu_{\text{fit}})\|_2, & \text{ellip.} \\ \max_i |\xi - \mu_{\text{fit},i}|/w_i, & \text{box} \end{cases}$$

$\mathscr{D}_{\text{select}}$

↓ $\tilde{\xi}_{1:m}$ samples

$s(\tilde{\xi}_1), \ldots, s(\tilde{\xi}_m)$ scores

Bulk-*calibrated* credal ambiguity sets

Bulk-set calibration

Goal: find $\Xi_0(\mathscr{D})$ such that

1. $\Pr\left(\mathbb{P}^*\left(\Xi_0\right) \geq 1-\gamma\right) \geq 1-\delta$
2. $\sup_{\xi \in \Xi_0} f_x(\xi)$ is tractable for common losses f_x

Split $\mathscr{D}$ into score **fitting** set + score **selection** set:

$\mathscr{D}_{\text{fit}}$

↓ Defines a score function

$$s(\xi) = \begin{cases} \|\Sigma_{\text{fit}}^{-1/2}(\xi - \mu_{\text{fit}})\|_2, & \text{ellip.} \\ \max_i |\xi - \mu_{\text{fit},i}|/w_i, & \text{box} \end{cases}$$

$\mathscr{D}_{\text{select}}$

↓ $\tilde{\xi}_{1:m}$ samples

$s(\tilde{\xi}_1), \ldots, s(\tilde{\xi}_m)$ scores

DKW inequality:

$$\Pr\left\{\sup_{t \in \mathbb{R}} \left| \underbrace{F_m(t)}_{\text{Empirical cdf}} - \underbrace{F(t)}_{\text{Theoretical cdf}} \leq r_{m,\delta} \right|\right\} \geq 1-\delta$$

$\sqrt{\frac{1}{2m}\log\left(\frac{2}{\delta}\right)}$ minimum certifiable γ

$\hat{t}_{\text{DKW}} := (1-\gamma+r_{m,\delta})$-quantile of selection scores

$$\Pr\left(\mathbb{P}^*\left(s(\xi) \leq \hat{t}_{\text{DKW}}\right) \geq 1-\gamma\right) \geq 1-\delta$$

↓ Defines a bulk set Ξ_0 that satisfies the coverage condition

Bulk-*calibrated* credal ambiguity sets

Bulk-set calibration

Goal: find $\Xi_0(\mathscr{D})$ such that

1. $\Pr\left(\mathbb{P}^*\left(\Xi_0\right) \geq 1-\gamma\right) \geq 1-\delta$
2. $\sup_{\xi \in \Xi_0} f_x(\xi)$ is tractable for common losses f_x

💡This can be extended to **separate dimension-blocks!**

Split $\mathscr{D}$ into score **fitting** set + score **selection** set:

$\mathscr{D}_{\text{fit}}$

↓ Defines a score function

$$s(\xi) = \begin{cases} \|\Sigma_{\text{fit}}^{-1/2}(\xi - \mu_{\text{fit}})\|_2, & \text{ellip.} \\ \max_i |\xi - \mu_{\text{fit},i}|/w_i, & \text{box} \end{cases}$$

$\mathscr{D}_{\text{select}}$

↓ $\tilde{\xi}_{1:m}$ samples

$s(\tilde{\xi}_1), \ldots, s(\tilde{\xi}_m)$ scores

DKW inequality:

$$\Pr\left\{\sup_{t \in \mathbb{R}} \left| \underbrace{F_m(t)}_{\text{Empirical cdf}} - \underbrace{F(t)}_{\text{Theoretical cdf}} \leq r_{m,\delta} \right|\right\} \geq 1-\delta$$

$\sqrt{\frac{1}{2m}\log\left(\frac{2}{\delta}\right)}$ minimum certifiable γ

$\hat{t}_{\text{DKW}} := (1-\gamma+r_{m,\delta})$-quantile of selection scores

$$\Pr\left(\mathbb{P}^*\left(s(\xi) \leq \hat{t}_{\text{DKW}}\right) \geq 1-\gamma\right) \geq 1-\delta$$

↓ Defines a bulk set Ξ_0 that satisfies the coverage condition

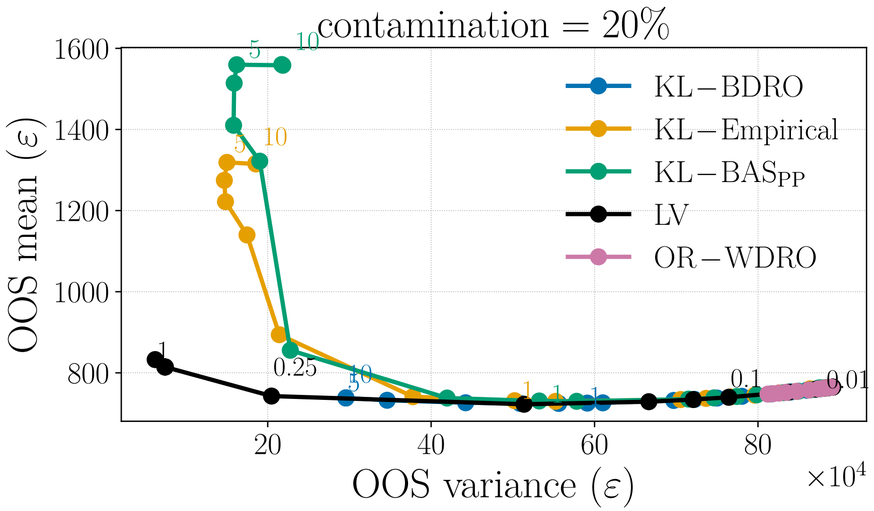
Experiments

Bayesian Centre: Heavy-tailed Synthetic Newsvendor + Demand Spikes

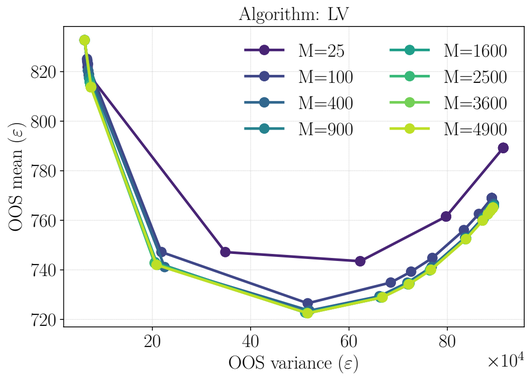
Train: multivariate Student-t

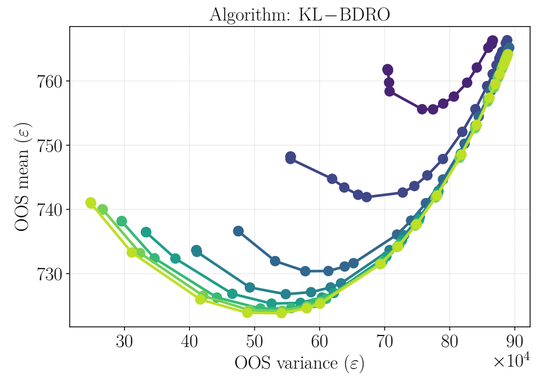
Test: 80% clean Student-t + 20% Gaussian spike

Loss: $f_x(\xi) = \sum_{j=1}^{d} h[x_j - \xi_j]_+ + b[\xi_j - x_j]_+$



OOS frontiers





Sample-efficiency: LV stabilises with smaller SAA budgets

Experiments

Bayesian Centre: Heavy-tailed Synthetic Newsvendor + Demand Spikes

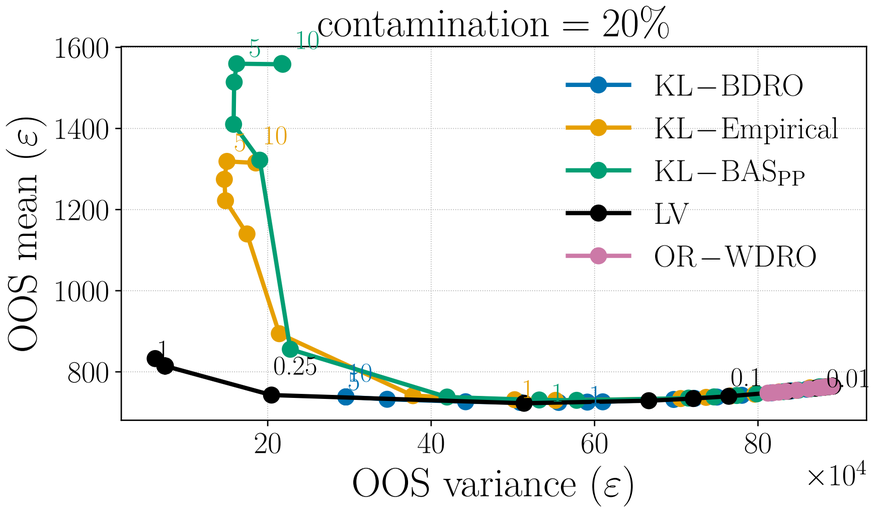
Train: multivariate Student-t

Test: 80% clean Student-t + 20% Gaussian spike

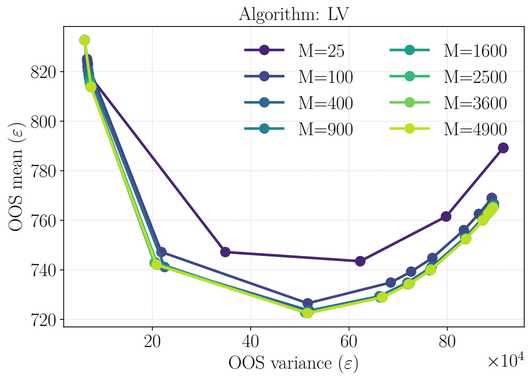
Loss: $f_x(\xi) = \sum_{j=1}^{d} h[x_j - \xi_j]_+ + b[\xi_j - x_j]_+$

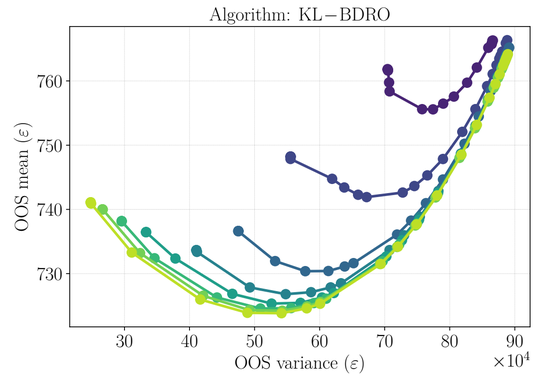
LV: $(1-\varepsilon)\mathbb{E} + \varepsilon \sup$

KL: $\inf_{\lambda>0}\left\{\lambda\varepsilon + \lambda\log\left(\mathbb{E}\left[\exp(f_x(\xi)/\lambda)\right]\right)\right\}$



OOS frontiers





Sample-efficiency: LV stabilises with smaller SAA budgets

Experiments

Frequentist Centre: California Housing Regression + Geographic Shift

Train: eastern 50% **Test:** western 20%

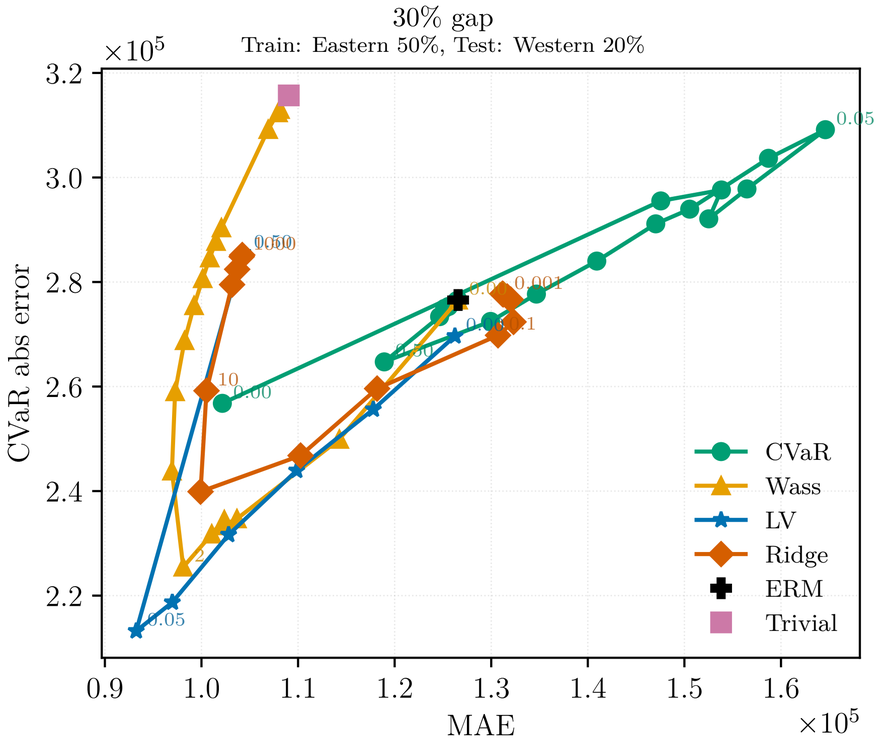
Loss: linear regression MAE

Product bulk-set construction: $\Xi_0 = \Xi_{0,X} \times \Xi_{0,Y}$ to avoid hard-coding $\mathbb{P}_{Y|X}$

Ellipsoid Interval

Runtime comparison

Method	CV	Likelihood	Solve	Total
LV	1.43 (0.09)	0.01 (0.00)	0.14 (0.01)	1.58 (0.10)
CVaR	5.58 (0.14)	—	0.37 (0.02)	5.95 (0.15)
Wass	6.62 (0.47)	—	0.41 (0.03)	7.03 (0.50)



Parameter sweeps

Experiments

Empirical Centre: CivilComments Classification

Train&Test: WILDS split (demographic shift)
Loss: Cross-entropy loss

Sagawa, et al. (2020)

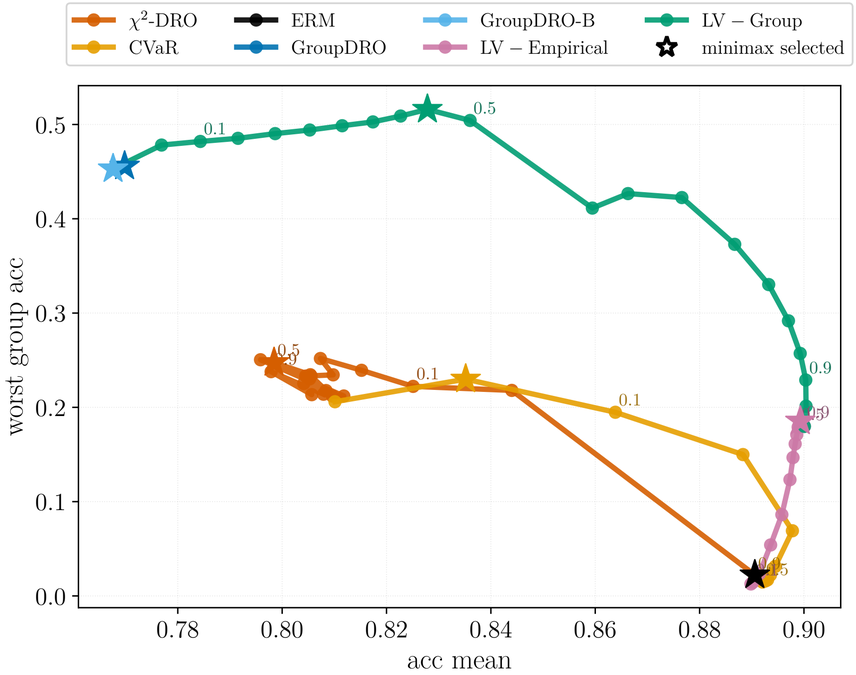
$$\left\{ \sum_{g=1}^{G} q_g \hat{\mathbb{P}}_g : q \in \Delta_G \right\}$$

LV-Group: **discounting** GroupDRO

$$\mathscr{C}_{\mathrm{LVG}}(\varepsilon) = \left\{ (1-\varepsilon)\mathbb{P}_{\Xi_0} + \varepsilon R : \mathbb{P} \in \mathscr{C}_{\mathrm{G}},\ R \in \mathscr{P}(\Xi_0) \right\}$$

Accuracies

Method	Mean acc	Worst-group acc	Gap
LV-Group	0.828 (0.005)	0.516 (0.010)	0.312
GroupDRO	0.770 (0.002)	0.456 (0.005)	0.314
ERM	0.891 (0.000)	0.022 (0.001)	0.869



Parameter sweeps

Conclusions

✓**Bridge** between the IP and DRO communities

✓**Make forward Huber contamination meaningful** under unbounded sample spaces + losses

✓**Tractable** optimisation objective and efficient bulk calibration with **finite-sample guarantees** and risk certificates

Future work

- Bulk-set calibration techniques that require fewer selection samples, or allow online updates
- Unified theory for the bulk-set geometry, choice of centre, and loss function

References

Ben-Tal A. and Nemirovski A. (1998), Robust convex optimization, Mathematics of Operations Research.

Hu, Z. and Hong, L. J. (2013), Kullback-Leibler divergence constrained distributionally robust optimization. Available from Optimization Online.

Isii K. (1960), The extrema of probability determined by generalized moments (I) Bounded random variables, Annals of the Institute of Statistical Mathematics.

Jiang, R. and Guan, Y. (2018), Risk-averse two-stage stochastic program with distributional ambiguity. Operations Research.

Kuhn, D., Shafiee, S., and Wiesemann, W. (2025), Distributionally robust optimization. Acta Numerica.

Mohajerin Esfahani, P. and Kuhn, D. (2018), Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Mathematical Programming.

Namkoong, H. and Duchi, J. C. (2016), Stochastic gradient methods for distributionally robust optimization with fdivergences. In Advances in Neural Information Processing Systems.

Sagawa, S., Koh, P. W., Hashimoto, T. B., and Liang, P. (2020), Distributionally robust neural networks. In International Conference on Learning Representations.

Staib, M. and Jegelka, (2019), S. Distributionally robust optimization and generalization in kernel methods. NeurIPS.

Thank you!

Paper

Code