Robust COAML

Fréjus summer school

Solène Delannoy-Pavy (CERMICS)
Axel Parmentier (CERMICS)
Maximilian Schiffer (HEC)
Wolfram Wieseman (Imperial College)

A Data-Driven optimization setting

Contextual Stochastic Optimization (CSO)

Context x, noise ξ, decision $\mathbf{y} \in \mathcal{Y}$. A policy π maps x to a distribution over $\mathcal{Y}$. Find the best policy in a class Π:

$$\min_{\pi \in \Pi} \ \mathbb{E}_{(\mathbf{x},\xi),\, \mathbf{y} \sim \pi(.|x)} \big[c(\mathbf{x}, \mathbf{y}, \xi) \big] \tag{1}$$

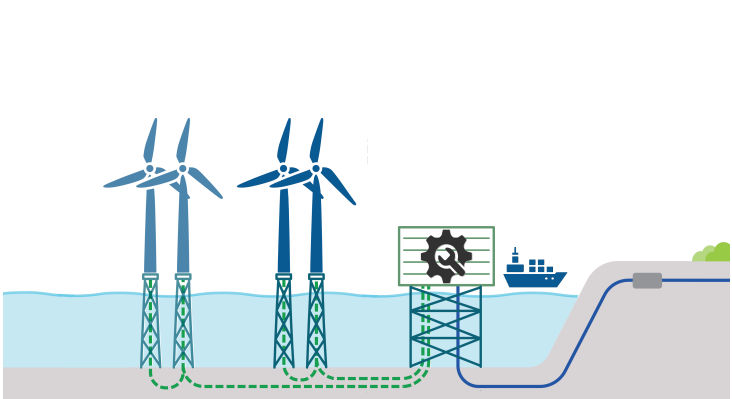
The true distribution over $(\mathbf{x}, \xi)$ is unknown. In a data-driven environment, one may solve ERM instead:

Empirical Risk Minimization (ERM)

Given a training set of N samples (x_i, ξ_i), approximate the true distribution with the empirical distribution:

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_i, \xi_i)} \quad \Longrightarrow \quad \min_{\pi} \ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\mathbf{y} \sim \pi(\cdot|x_i)} \big[c(x_i, \mathbf{y}, \xi_i) \big] \tag{2}$$

Why ERM is (sometimes) not enough: maintenance

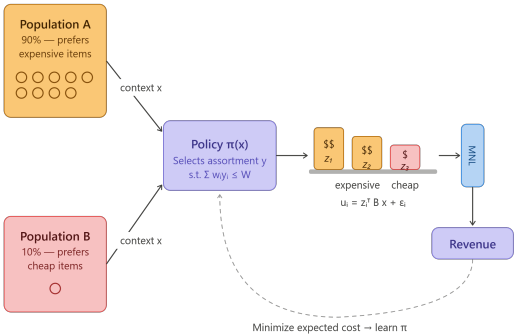


θ **est** \ θ **opt**	0.99	0.67	0.1
0.99	**1195**	1200	1438
0.67	192	126	144
0.1	335	204	194

The cost is highly dependent on the parameter θ of the failure model chosen for optimization and simulation

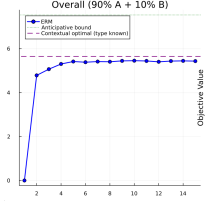
- Small data / high variance: ERM ignores estimation errors
- Distribution shift: no resilience when test $\neq$ train distribution
- Tail risk: rare but costly events

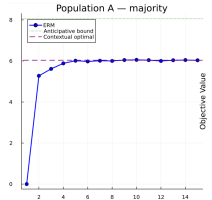
Why ERM is (sometimes) not enough: assortment with heterogenous population

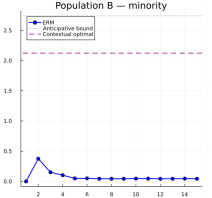


- Heterogeneous population

ERM is dominated by the majority gradient
⇒ always recommends type B items
⇒ near-zero revenue for B.







Distributionally Robust Optimization (DRO)

DRO

Idea : Instead of trusting a single distribution, optimize against the worst distribution in an *ambiguity set* $\mathcal{P}$.

$$\min_{y} \max_{p \in \mathcal{P}} \mathbb{E}_{\xi \sim p}\big[c(y, \xi)\big]$$

Distributionally Robust Optimization (DRO)

DRO

Idea : Instead of trusting a single distribution, optimize against the worst distribution in an *ambiguity set* $\mathcal{P}$.

$$\min_{y} \max_{p \in \mathcal{P}} \mathbb{E}_{\xi \sim p}\big[c(y, \xi)\big]$$

- $\mathcal{P} = \hat{p} \Rightarrow$ stochastic optimization: $\min_y \ \mathbb{E}_{\xi \sim \hat{p}}\big[c(y, \xi)\big]$
- $\mathcal{P} = \mathcal{P}(\Xi) \Rightarrow$ robust optimization optimization: $\min_y \ \max_{\xi \in \Xi} c(y, \xi)\big]$

Ambiguity sets: two main families[1]

Moment-based sets

$$\mathcal{P} = \left\{ P \;\middle|\; \mathbb{E}_P[\phi_k(x, \xi)] \leq b_k,\ \forall k \right\}$$

All distributions whose moments lie in $\mathcal{F}$.

\+ Tractability.

− The ambiguity set may contain unlikely distributions.

Discrepancy-based sets

$$\mathcal{P} = \left\{ P \;\middle|\; D(P, \hat{P}) \leq r \right\}$$

Ball of radius r around a reference $\hat{P}$.

\+ Tunable size parameter.
\+ Statistical guarantees.

Two choices: ϕ-divergences **or** Wasserstein.

[1] Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. *Distributionally Robust Optimization*. May 26, 2025. DOI: 10.48550/arXiv.2411.02549. arXiv: 2411.02549. URL: http://arxiv.org/abs/2411.02549 (visited on 04/13/2026).

We consider Kullback-Leibler ambiguity sets

ϕ-divergence

For $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ convex with $\phi(1) = 0$, and some additional assumptions, the ϕ-divergence of P w.r.t. $\hat{P}$ is

$$D_\phi(P \,\|\, \hat{P}) = \begin{cases} \mathbb{E}_{z\sim\hat{P}}\left[\phi\left(\frac{p(z)}{\hat{p}(z)}\right)\right] & \text{if } P \ll \hat{P} \\ +\infty & \text{otherwise.} \end{cases}$$

We consider Kullback-Leibler ambiguity sets

ϕ-divergence

For $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ convex with $\phi(1) = 0$, and some additional assumptions, the ϕ-divergence of P w.r.t. $\hat{P}$ is

$$D_\phi(P \,\|\, \hat{P}) = \begin{cases} \mathbb{E}_{z\sim\hat{P}}\left[\phi\left(\frac{p(z)}{\hat{p}(z)}\right)\right] & \text{if } P \ll \hat{P} \\ +\infty & \text{otherwise.} \end{cases}$$

Our choice: Kullback–Leibler (KL) divergence ($\phi(s) = s \log s$).

1. With $\hat{P} = \hat{p} = \frac{1}{N}\mathbf{1}$, the distributions in the ambiguity set $\mathcal{P}$ are characterized by vectors $(p_1, \ldots, p_N) \in \Delta_N$.
2. Tractability (see later).

Contextual Distributionally Robust Optimization (Contextual DRO)

- Traditional DRO accounts for uncertainty in the distribution of the noise but ignores the context.

Contextual Distributionally Robust Optimization (Contextual DRO)

- Traditional DRO accounts for uncertainty in the distribution of the noise but ignores the context.

Contextual DRO

- We consider robustness that bears on the **joint** distribution of (x, ξ).

$$\min_{\pi} \max_{p \in \mathcal{P}} \mathbb{E}_{(x,\xi)\sim p,\, y\sim\pi(\cdot|x)} \big[c(x, y, \xi) \big]$$

Contextual Distributionally Robust Optimization (Contextual DRO)

- Traditional DRO accounts for uncertainty in the distribution of the noise but ignores the context.

Contextual DRO

- We consider robustness that bears on the **joint** distribution of (x, ξ).

$$\min_{\pi} \max_{p \in \mathcal{P}} \mathbb{E}_{(x,\xi)\sim p,\, y\sim\pi(\cdot|x)} \big[c(x, y, \xi) \big]$$

Contextual DRO is usually not tractable for large-scale real-world problems. Can we train a COAML policy in a contextual DRO setting ?

Robust COAML equation and relaxation

- **Robust formulation**

$$\min_{w} \max_{\substack{p \\ D_{\mathrm{KL}}(p \,\|\, \hat{p}) \le \varepsilon}} \sum_{i=1}^{N} p_i \, \mathbb{E}_{\mathbf{y} \sim \pi_w(x_i)}[c(x_i, \mathbf{y}, \xi_i)] \,. \tag{3}$$

- **Relaxed problem**

$$\min_{w} \max_{p \in \Delta_N} \sum_{i=1}^{N} p_i \, \mathbb{E}_{\mathbf{y} \sim \pi_w(x_i)}[c(x_i, \mathbf{y}, \xi_i)] \; - \; \gamma \, D_{\mathrm{KL}}(p \,\|\, \hat{p}) \tag{4}$$

$$= \min_{w} \max_{p \in \Delta_N} \sum_{i=1}^{N} p_i \, c_i(w) \; - \; \gamma \, D_{\mathrm{KL}}(p \,\|\, \hat{p}), \tag{5}$$

where $c_i(w) = \mathbb{E}_{\mathbf{y} \sim \pi_w(x_i)}[c(x_i, \mathbf{y}, \xi_i)]$.

The DRO problem is equivalent to minimizing the entropic risk

Inner maximization over p (for fixed w): unique closed-form solution

$$p_i^* = \frac{\exp\big(c_i(w)/\gamma\big)}{\sum_{j=1}^{N} \exp\big(c_j(w)/\gamma\big)}. \tag{6}$$

Substituting p^*:

$$\min_w \max_{p \in \Delta_N} \quad \sum_{i=1}^{N} p_i \, c_i(w) \; - \; \gamma \, D_{\mathrm{KL}}(p \,\|\, \hat{p}) \tag{7}$$

$$= \min_w \boxed{\gamma \log\left(\frac{1}{N} \sum_{i=1}^{N} e^{c_i(w)/\gamma} \right)} \quad \text{Entropic risk} \tag{8}$$

- $\gamma \to +\infty$ recovers ERM while $\gamma \to 0^+$ concentrates on worst-case sample.
- Robustness parameter $\mu = 1/\gamma$.

Alternating minimization for the DRO problem

$$p^{(k)} = \arg\max_{p} \sum_{i=1}^{N} p_i \, \mathbb{E}_{\mathbf{y}\sim\pi_{w^{k-1}}(x_i)}[c(x_i, \mathbf{y}, \xi_i)] \; - \; \gamma \, D_{\mathrm{KL}}(p \,\|\, \hat{p}) \tag{9a}$$

$$w^{(k)} \in \arg\min_{w} \sum_{i=1}^{N} p_i^{(k)} \, \mathbb{E}_{\mathbf{y}\sim\pi_w(x_i)}[c(x_i, \mathbf{y}, \xi_i)] \tag{9b}$$

- Does it converge ?
- How to compute step (9b) ? It looks like a "reweighted" ERM.

Proposition

We can learn w such that π_w minimizes the empirical risk using an Primal dual algorithm, see Louis Bouvier et al. *Primal-dual algorithm for contextual stochastic combinatorial optimization*. 2025. arXiv: 2505.04757 [cs.LG]. URL: https://arxiv.org/abs/2505.04757

Robust algorithm

$$c_i^{(k)} = \mathbb{E}_{\mathbf{y}\sim\pi_{w^{(k-1)}}(\cdot|x_i)}\big[c(x_i,\mathbf{y},\xi_i)\big]$$

$$p^{(k)} = \arg\max_p \left\{ \sum_i p_i c_i^{(k)} + \gamma D_{\mathrm{KL}}(p \,\|\, \hat{p}) \right\} = \mathrm{softmax}\Big(\gamma^{-1}(c_i^{(k)})_{i\in[N]}\Big)$$

(reweighting)

$$q_i^{(k)} = \arg\min_q \left\{ \mathbb{E}_{\mathbf{y}\sim q}[c(x_i,\mathbf{y},\xi_i)] + \kappa\, \ell_{\Omega_\Delta}\big(Y^\top \varphi_{w^{(k-1)}}(x_i); q\big) \right\}$$

(decomposition)

$$w^{(k)} \in \arg\min_w \sum_{i=1}^N p_i^{(k)} \ell_{\Omega_\Delta}\big(Y^\top \varphi_w(x_i); q_i^{(k)}\big)$$

(coordination)

Is this tractable ?

The algorithm is tractable in the sparse perturbation case

In the case of sparse perturbation, these updates have a convenient expression:

$$c_i^{(k)} = \mathbb{E}_{\mathbf{y}\sim\pi_{w^{(k-1)}}(\cdot|x_i)}\left[c(x_i, \mathbf{y}, \xi_i)\right]$$
$$p^{(k)} = \mathrm{softmax}\left(\gamma^{-1}(c_i^{(k)})_{i\in[N]}\right) \qquad \text{(reweighting)}$$

$$\mu_i^{(k)} = \mathbb{E}_{\mathbf{Z}}\left[\arg\min_{y_i\in\mathcal{Y}} c(x_i, y_i, \xi_i) - \kappa(\varphi_{w^{(k-1)}}(x_i) + \epsilon\mathbf{Z})^\top y_i\right] \qquad \text{(decomposition)}$$

$$w^{(k)} \in \arg\min_w \sum_{i=1}^N p_i^{(k)}\left[F_{\epsilon,\mathcal{C}}(\varphi_w(x_i)) - \langle\varphi_w(x_i), \mu_i^{(k)}\rangle\right] \qquad \text{(coordination)}$$

where $\mathbf{Z}$ follows a multivariate standard normal distribution.
We only need a (perturbed) anticipative solver.

The algorithm is tractable in the sparse perturbation case

In the case of sparse perturbation, these updates have a convenient expression:

$$c_i^{(k)} = \mathbb{E}_{\mathbf{y}\sim\pi_{w^{(k-1)}}(\cdot|x_i)}\left[c(x_i, \mathbf{y}, \xi_i)\right]$$

$$p^{(k+\frac{1}{2})} = \operatorname{softmax}\left(\gamma^{-1}(c_i^{(k)})_{i\in[N]}\right) \qquad \text{(reweighting)}$$

$$p^{(k)} = \alpha\, p^{(k-1)} + (1-\alpha)\, p^{(k+\frac{1}{2})}$$

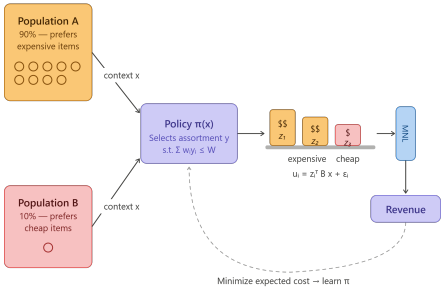
$$\mu_i^{(k)} = \mathbb{E}_{\mathbf{Z}}\left[\arg\min_{y_i\in\mathcal{Y}} c(x_i, y_i, \xi_i) - \kappa(\varphi_{w^{(k-1)}}(x_i) + \epsilon\mathbf{Z})^\top y_i\right] \qquad \text{(decomposition)}$$

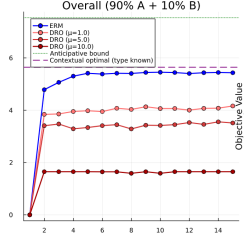
$$w^{(k+\frac{1}{2})} \in \arg\min_w \sum_{i=1}^N p_i^{(k)}\left[F_{\epsilon,\mathcal{C}}(\varphi_w(x_i)) - \langle\varphi_w(x_i), \mu_i^{(k)}\rangle\right] \qquad \text{(coordination)}$$

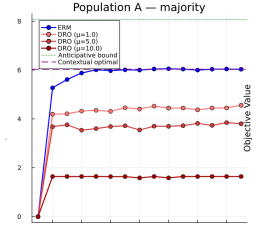
$$w^{(k)} = \beta\, w^{(k-1)} + (1-\beta)\, w^{(k+\frac{1}{2})}$$

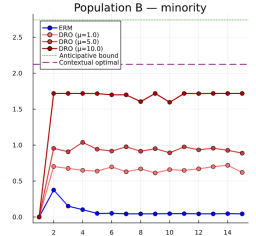
where $\mathbf{Z}$ follows a multivariate standard normal distribution.
We only need a (perturbed) anticipative solver.

Results on assortment with homogenous population









Robust COAML upweights poorly-served samples
⇒ better performance on B

Conclusion and open questions

Our contribution

- A scalable contextual DRO algorithm for combinatorial problems, that does not require strong assumptions on the problem structure.

Open questions

- Do you know any combinatorial problems for which contextual DRO would make sense?
- What convergence guarantees can we obtain?
- Can you extend this learning algorithm to multistage problems? (Drawing inspiration from robust offline RL)?
- Can we extend this learning algorithm to other ambiguity sets?