# Robustness versus Statistical Efficiency: Superiority of Naive Optimization

INFORMS 2024

Arindam Roy Chowdhury Henry Lam

Columbia University

Industrial Engineering and Operations Research

# Background

## Stochastic Optimization

$$\min_{x \in \mathcal{X}} \psi(x, P)$$

$\mathcal{X}$: space of feasible decisions

$P$: probability distribution

| | Machine Learning | Portfolio Optimization | Newsvendor |
|---|---|---|---|
| **Randomness(P)** | features and response | asset returns | demand |
| **Decision (x)** | prediction model | allocation weights | inventory level |
| **Objective($\psi$)** | expected loss | risk-return criterion | expected revenue |

# Background

$P$ is unknown, incurring uncertainties like:

- **Data or Statistical Uncertainty:**
  - ◇ Addressed through data-driven optimization approaches[1]

[1] Shapiro, Darinka, and Andrzej 2016.
[2] Hansen and Sargent 2016.

# Background

$P$ is unknown, incurring uncertainties like:

- **Data or Statistical Uncertainty:**
  - ◇ Addressed through data-driven optimization approaches[1]
- **Distribution Shift:**
  - ◇ Motivates robust optimization methods[2]

[1] Shapiro, Darinka, and Andrzej 2016.
[2] Hansen and Sargent 2016.

# Common Approaches

# Empirical Optimization (EO)

Replace $P$ with it's empirical estimate:

$$x_n^{EO} = \underbrace{x^*(\hat{P}_n)}_{\text{best decision at } \hat{P}_n}$$

[3] Vaart 1998; Kim, Pasupathy, and Henderson 2015.
[4] Lam 2021.

# Empirical Optimization (EO)

Replace $P$ with it's empirical estimate:

$$x_n^{EO} = \underbrace{x^*(\hat{P}_n)}_{\text{best decision at } \hat{P}_n}$$

- Most natural way to handle statistical uncertainty[3]

[3] Vaart 1998; Kim, Pasupathy, and Henderson 2015.
[4] Lam 2021.

# Empirical Optimization (EO)

Replace $P$ with it's empirical estimate:

$$x_n^{EO} = \underbrace{x^*(\hat{P}_n)}_{\text{best decision at } \hat{P}_n}$$

- Most natural way to handle statistical uncertainty[3]
- Result[4]:
  - EO gives statistically best regret

$$R(x, Q) := \underbrace{\psi(x, Q)}_{\text{achieved loss}} - \underbrace{\psi(x^*(Q), Q)}_{\text{oracle loss}}$$

[3] Vaart 1998; Kim, Pasupathy, and Henderson 2015.
[4] Lam 2021.

# Distributionally Robust Optimization (DRO)[5]

DRO tries to protect against worst case objective:

$$x^{DRO} := \arg\min_{x \in \mathcal{X}} \max_{Q \in \mathcal{U}_\lambda} \psi(x, Q)$$

for some ambiguity set $\mathcal{U}_\lambda$.

[5] Delage and Ye 2010; Goh and Sim 2010; Ben-Tal et al. 2013.

# Distributionally Robust Optimization (DRO)[5]

DRO tries to protect against worst case objective:

$$x^{DRO} := \arg\min_{x \in \mathcal{X}} \max_{Q \in \mathcal{U}_\lambda} \psi(x, Q)$$

for some ambiguity set $\mathcal{U}_\lambda$. Example: $\phi$-divergence ambiguity

$$\mathcal{U}_\lambda := \{Q : d(Q, P) \leq \lambda\}$$

where $d(\cdot, \cdot)$ is any $\phi$-divergence (like $\chi^2$ or $KL$ - divergence)

[5] Delage and Ye 2010; Goh and Sim 2010; Ben-Tal et al. 2013.

# Performance of DRO

- RO/DRO are robust in the sense of a worst-case objective guarantee[6]:

$$\mathbb{P}\left[\psi(x^{DRO}, P) \leq \max_{Q \in \mathcal{U}_\lambda} \psi(x^{DRO}, Q)\right] \geq 1 - \delta$$

[6] Zeng and Lam 2022; Delage and Ye 2010; Mohajerin Esfahani and Kuhn 2017.

# Performance of DRO

- RO/DRO are robust in the sense of a worst-case objective guarantee[6]:

$$\mathbb{P}\left[\psi(x^{DRO}, P) \leq \max_{Q \in \mathcal{U}_\lambda} \psi(x^{DRO}, Q)\right] \geq 1 - \delta$$

**However, robustness in regret is unclear**

[6] Zeng and Lam 2022; Delage and Ye 2010; Mohajerin Esfahani and Kuhn 2017.

# Main Result

Under general uncertainty, naive optimization gives the best regret

# Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Under general uncertainty, naive optimization gives the best regret

# Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Under general uncertainty, naive optimization gives the best regret

current best guess

# Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Over any reasonable approach

(includes DRO)

Under general uncertainty, naive optimization gives the best regret

current best guess

# Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Over any reasonable approach

(includes DRO)

Under general uncertainty, naive optimization gives the best regret

current best guess

Asymptotic worst case second order stochastic dominance

# Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Over any reasonable approach

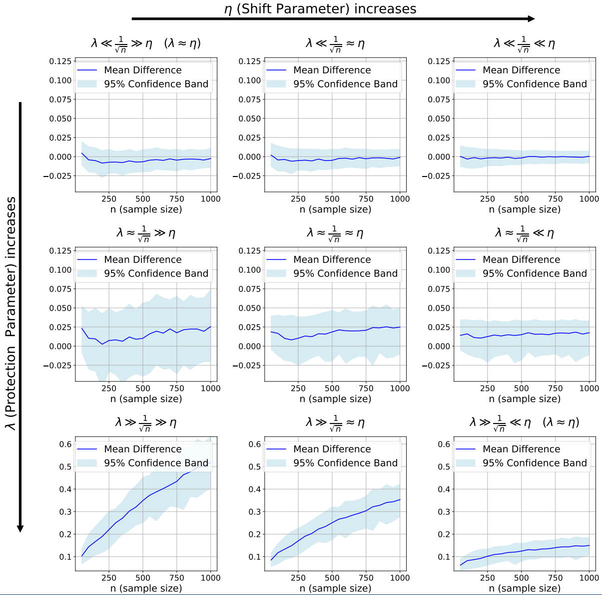
(includes DRO)

Under general uncertainty, naive optimization gives the best regret, due to symmetry.

current best guess

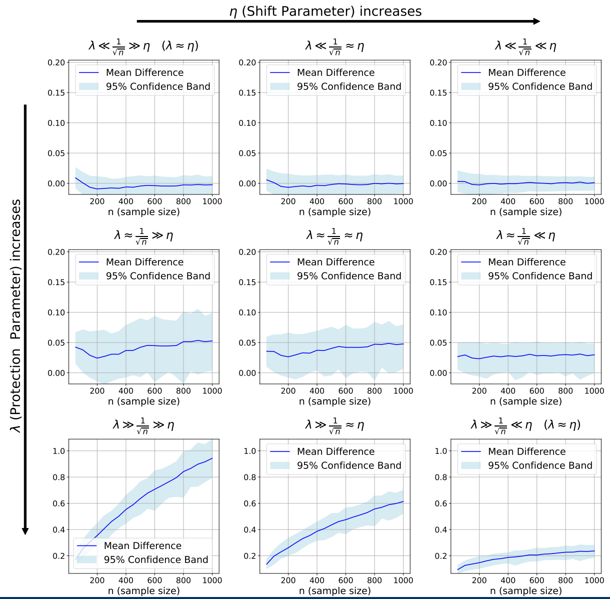
**symmetry of uncertainty**

Asymptotic worst case second order stochastic dominance



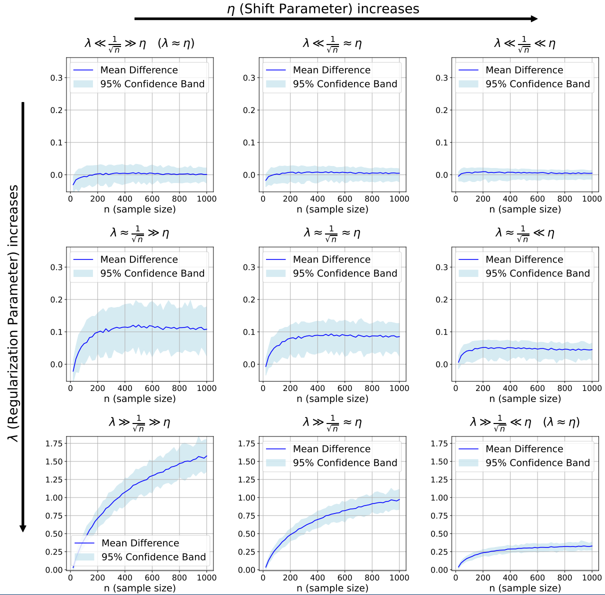

Compare *EO* against $\chi^2 - DRO$

(under $\chi^2-$divergence shift)




Compare $EO$ against $KL - DRO$

(under $\chi^2-$divergence shift)




Compare *EO* against

$L_2$ − *Regularization*

(under $\chi^2$−divergence shift)

# Recall: Main Result

$P$ estimated from data

$P$ shifted to $Q$

Combination of both

Over any reasonable approach

(includes DRO)

Under general uncertainty, naive optimization gives the best regret, due to symmetry.

current best guess

**symmetry of uncertainty**

Asymptotic worst case second order stochastic dominance

# The Competing approaches

We consider solutions of the form:

$$x_n^\lambda \approx x_n^{EO} + \lambda K$$

This includes:

- DRO ($\phi-$ divergence, Wasserstein) with protection parameter $\lambda$
- Regularization with parameter $\lambda$

# Performance Criterion: Worst Case Second Order Stochastic Dominance

- With suitable scaling,

$$[s \cdot R(x^{EO}, Q)] \xrightarrow{d} Y \quad \text{and} \quad [s \cdot R(x^{\lambda}, Q)] \xrightarrow{d} Z$$

- $Z$ dominates $Y$ if for any non-decreasing convex function $u$:

$$\mathbb{E}[u(Y)] \leq \mathbb{E}[u(Z)]$$

# Performance Criterion: Worst Case Second Order Stochastic Dominance

- With suitable scaling,

$$[s \cdot R(x^{EO}, Q)] \xrightarrow{d} Y \quad \text{and} \quad [s \cdot R(x^{\lambda}, Q)] \xrightarrow{d} Z$$

- $Z$ dominates $Y$ if for any non-decreasing convex function $u$:

$$\mathbb{E}[u(Y)] \leq \mathbb{E}[u(Z)]$$

For example: $\mathbb{E}[Y] \leq \mathbb{E}[Z]$ and $\mathbb{E}[Y^2] \leq \mathbb{E}[Z^2]$

# Performance Criterion: Worst Case Second Order Stochastic Dominance

- With suitable scaling,

$$[s \cdot R(x^{EO}, Q)] \xrightarrow{d} Y \quad \text{and} \quad [s \cdot R(x^{\lambda}, Q)] \xrightarrow{d} Z$$

- $Z$ dominates $Y$ if for any non-decreasing convex function $u$:

$$\mathbb{E}[u(Y)] \leq \mathbb{E}[u(Z)]$$

For example: $\mathbb{E}[Y] \leq \mathbb{E}[Z]$ and $\mathbb{E}[Y^2] \leq \mathbb{E}[Z^2]$

### Worst Case Dominance

$$\sup_{Q \in \mathcal{U}_\eta} \mathbb{E}\left[u\left(s \cdot R(x_n^{EO}, Q)\right)\right] \leq \sup_{Q \in \mathcal{U}_\eta} \mathbb{E}\left[u\left(s \cdot R(x_n^{\lambda}, Q)\right)\right]$$

# Main Result

- $n$: Sample size
- $\eta$: Distribution shift parameter
- $\lambda$: Protection/Regularization parameter

**Theorem**

As $(n, \eta, \lambda) \to (\infty, 0, 0)$ at any rate,

$$\lim_{(n,\eta)\to(\infty,0)} \sup_{Q\in\mathcal{U}_\eta} \mathbb{E}\left[u\left(s \cdot R(x_n^{EO}, Q)\right)\right] \leq \lim_{(n,\eta,\lambda)\to(\infty,0,0)} \sup_{Q\in\mathcal{U}_\eta} \mathbb{E}\left[u\left(s \cdot R(x_n^{\lambda}, Q)\right)\right]$$

# Symmetry of Uncertainty

- No Distribution Shift: $[x_n^{EO} - x^*(P)]$ is symmetric
- Under Distribution Shift: We introduce a notion of symmetry:

$$Q \in \mathcal{U}_\eta(P) \implies \exists \tilde{Q} \in \mathcal{U}_\eta(P) \text{ s.t } [x^*(P) - x^*(Q)] \approx [x^*(P) + x^*(Q)]$$

Asymptotically satisfied by most common choices of $\mathcal{U}_\eta$

- $\theta_0$ shifts to $\theta \in \mathcal{U}_\eta = (\theta_0 - \eta, \theta_0 + \eta)$
- $P$ shifts to $Q \in \mathcal{U}_\eta = \{Q : d_\phi(Q, P_0) \leq \eta\}$

# Exploiting Symmetry

Notice that:

$$\begin{aligned} Regret(\hat{x}, Q) &\approx \frac{1}{2}(\hat{x} - x^*(Q))^T \nabla^2 (\hat{x} - x^*(Q)) \\ &\approx ||\hat{x} - x^*(Q)||_D^2 \\ &\approx \left|\left|(\hat{x} - x_n^{EO}) + \underbrace{(x_n^{EO} - x^*(P)) + (x^*(P) - x^*(Q))}_{symmetric}\right|\right|_D^2 \end{aligned}$$

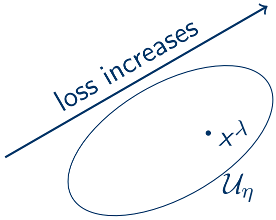
We exploit this idea to provide a unified treatment for all the three cases

# Regret is symmetric, objective is not

- Loss is not symmetric

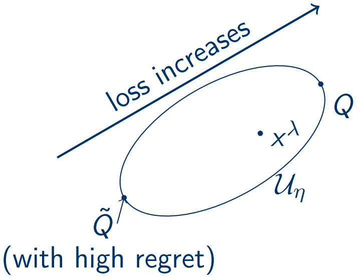
# Regret is symmetric, objective is not

- Loss is not symmetric
  - DRO protects against $Q$ with high objective



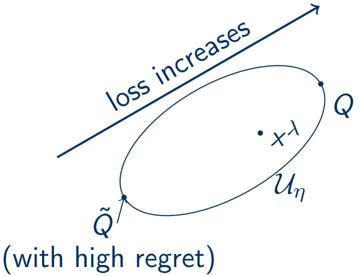
# Regret is symmetric, objective is not

- Loss is not symmetric
  - DRO protects against $Q$ with high objective
- Symmetric counterpart $\tilde{Q}$ has high regret



# Regret is symmetric, objective is not

- Loss is not symmetric
  - DRO protects against $Q$ with high objective
- Symmetric counterpart $\tilde{Q}$ has high regret




**Conclusion**

If the goal is to minimize regret, central estimates work the best

# References


- Ben-Tal, Aharon et al. (2013). **“Robust Solutions of Optimization Problems Affected by Uncertain Probabilities”.** In: *Management Science* 59.2, pp. 341–357.
- Delage, Erick and Yinyu Ye (2010). **“Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems”.** In: *Operations Research* 58.3, pp. 595–612. ISSN: 0030364X, 15265463.
- Goh, Joel and Melvyn Sim (2010). **“Distributionally Robust Optimization and Its Tractable Approximations”.** In: *Operations Research* 58.4, pp. 902–917. ISSN: 0030364X, 15265463.

- Hansen, Lars Peter and Thomas J Sargent (June 2016). ***Robustness.*** en. Princeton, NJ: Princeton University Press.
- Kim, Sujin, Raghu Pasupathy, and Shane Henderson (Jan. 2015). **"A Guide to Sample Average Approximation".** In: vol. 216, pp. 207–243. ISBN: 978-1-4939-1383-1. DOI: 10.1007/978-1-4939-1384-8_8.
- Lam, Henry (2021). ***On the Impossibility of Statistically Improving Empirical Optimization: A Second-Order Stochastic Dominance Perspective.*** arXiv: 2105.13419 [math.OC].

# Bibliography iii


- Mohajerin Esfahani, Peyman and Daniel Kuhn (July 2017). **"Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations".** In: *Mathematical Programming* 171.1–2, pp. 115–166. ISSN: 1436-4646. DOI: 10.1007/s10107-017-1172-1.
- Shapiro, Alexander, Dentcheva Darinka, and Ruszczynski Andrzej (2016). ***Lectures on stochastic programming : modeling and theory*.** en. MPS-SIAM series on optimization. Cambridge, TAS, Australia: Cambridge University Press.
- Vaart, A. W. van der (Oct. 1998). ***Asymptotic Statistics*.** Cambridge University Press. ISBN: 9780521784504. DOI: 10.1017/cbo9780511802256.

Zeng, Yibo and Henry Lam (2022). **"Generalization Bounds with Minimal Dependency on Hypothesis Class via Distributionally Robust Optimization".** In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., pp. 27576–27590.

# Thank You

## Questions?