From Data to Decisions: Distributionally Robust Optimization is Optimal

Bart Van Parys, Peyman Mohajerin Esfahani, Daniel Kuhn

Risk Analytics and Optimization Chair
École Polytechnique Fédérale de Lausanne
rao.epfl.ch

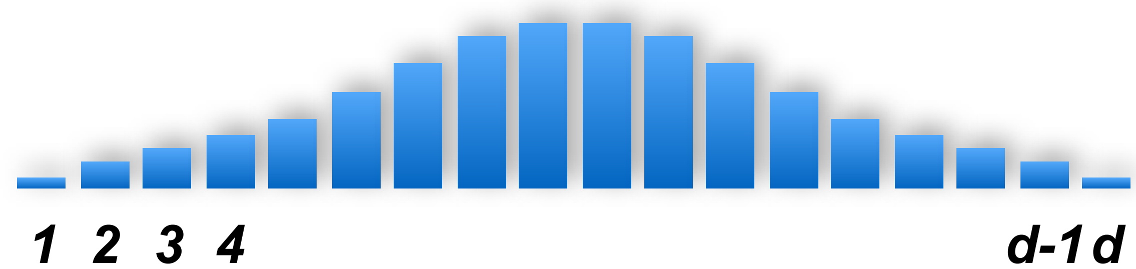
Swiss National Science Foundation

École Polytechnique Fédérale de Lausanne

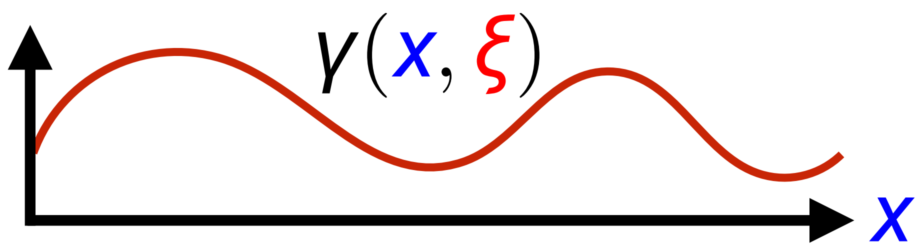
Stochastic Programming

$$\underset{x \in X}{\text{minimize}} \quad \mathbb{E}_{\mathbb{P}}\left[\gamma(x, \xi)\right]$$

- $\xi \in \{1, \ldots, d\}$



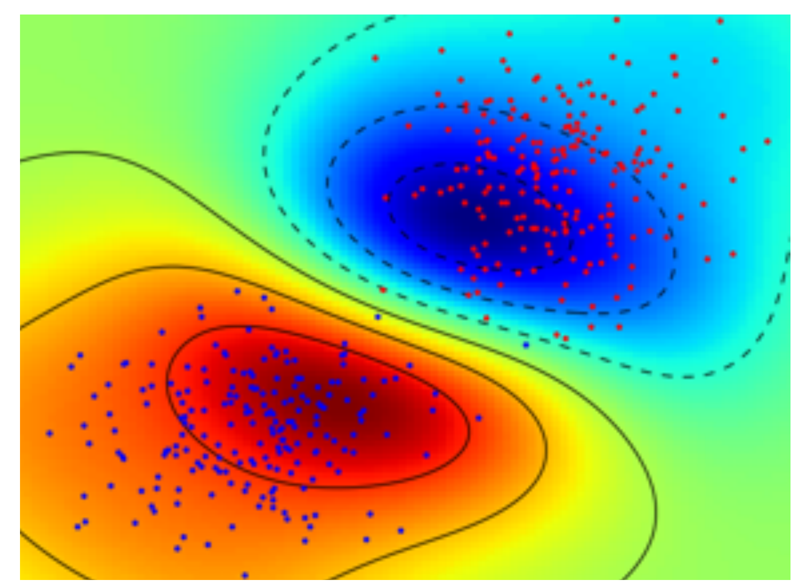
- $\gamma(x, \xi)$ continuous in x



Applications:

Supply Chain Mgmt.

Portfolio Mgmt.



Machine Learning

Stochastic Programming

$$\underset{x \in X}{\text{minimize}} \quad c(x, \mathbb{P})$$

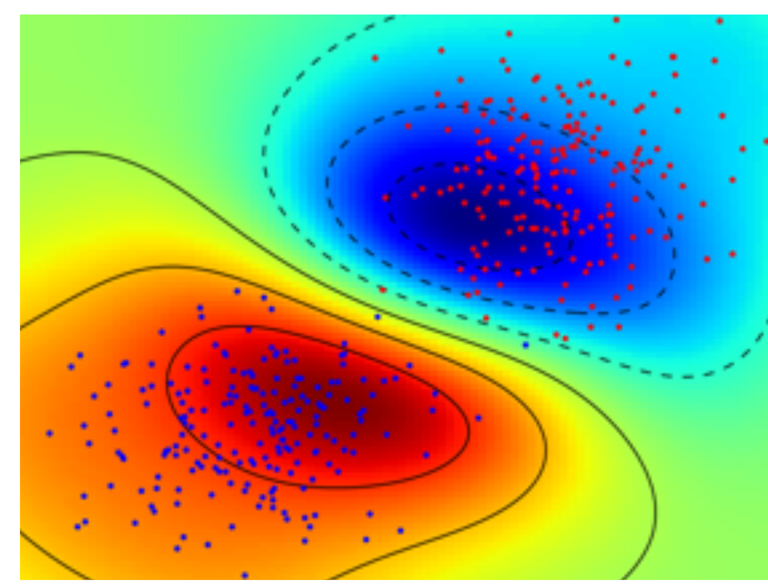
Objective: $$c(x, \mathbb{P}) = \sum_{i=1}^{d} \mathbb{P}(i)\, \gamma(x, i)$$

Optimizer: $$x^\star(\mathbb{P}) \in \underset{x \in X}{\text{argmin}} \; c(x, \mathbb{P})$$

Applications:

Supply Chain Mgmt.

Portfolio Mgmt.



Machine Learning

Separation of Estimation and Optimization

Phase 1:

$\xi_1, \xi_2, \ldots, \xi_T$ → 

→ estimator $\mathbb{P}_{\text{est}}$

training data $\sim \mathbb{P}^T$

Phase 2:

estimator $\mathbb{P}_{\text{est}}$ → 

→
- objective $c(x, \mathbb{P}_{\text{est}})$
- optimizer $x^\star(\mathbb{P}_{\text{est}})$

$\Longrightarrow$ **Estimator not tailored to optimization problem!**

Predictors & Prescriptors

Idea:

- Predictor: $\hat{c}(x, \xi_1, \ldots, \xi_T)$
- Prescriptor: $\hat{x}(\xi_1, \ldots, \xi_T) \in \operatorname{argmin}_{x \in X} \hat{c}(x, \xi_1, \ldots, \xi_T)$

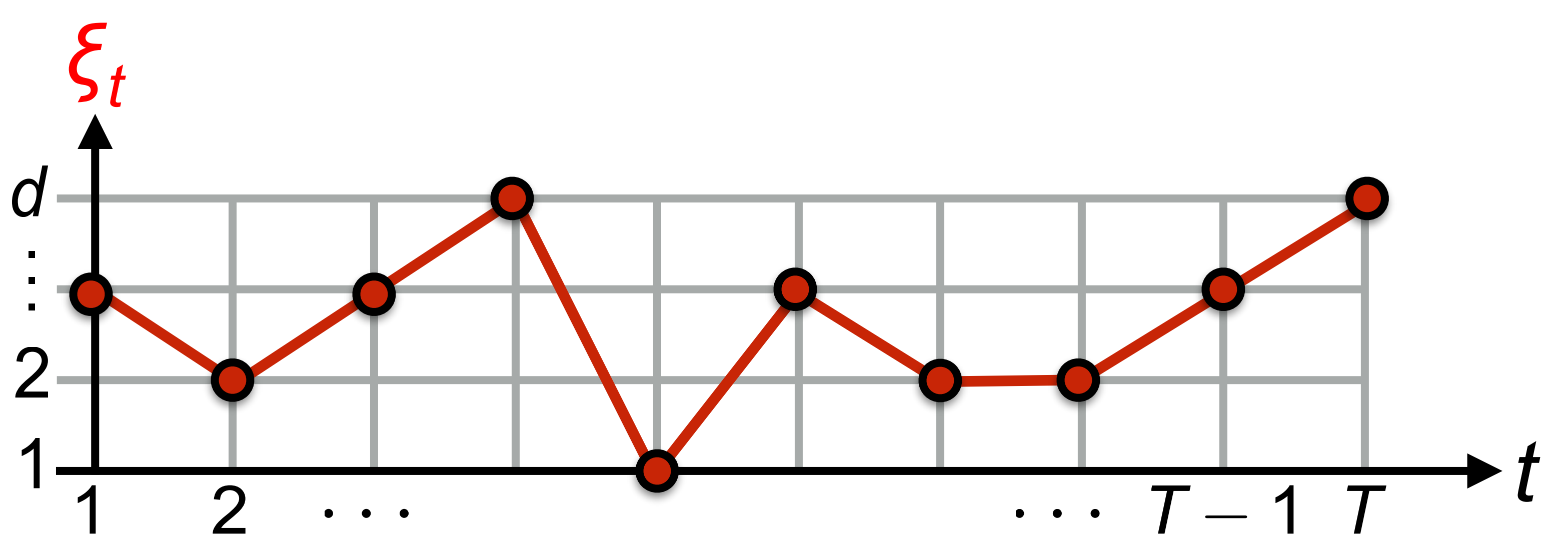
Predictors & Prescriptors

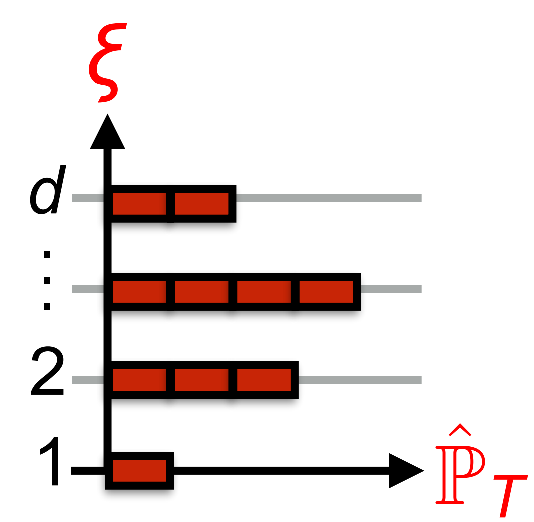
Idea:

- Predictor: $\hat{c}(x, \xi_1, \ldots, \xi_T)$
- Prescriptor: $\hat{x}(\xi_1, \ldots, \xi_T) \in \operatorname{argmin}_{x \in X} \hat{c}(x, \xi_1, \ldots, \xi_T)$

“Compressing” the statistical information:

$$\left.\begin{array}{c}\text{training samples}\\ \xi_1, \xi_2, \ldots, \xi_T\end{array}\right\} \quad \leftrightsquigarrow \quad \left\{\begin{array}{l}\text{empirical distribution}\\ \hat{\mathbb{P}}_T(i) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\xi_t = i}\end{array}\right.$$





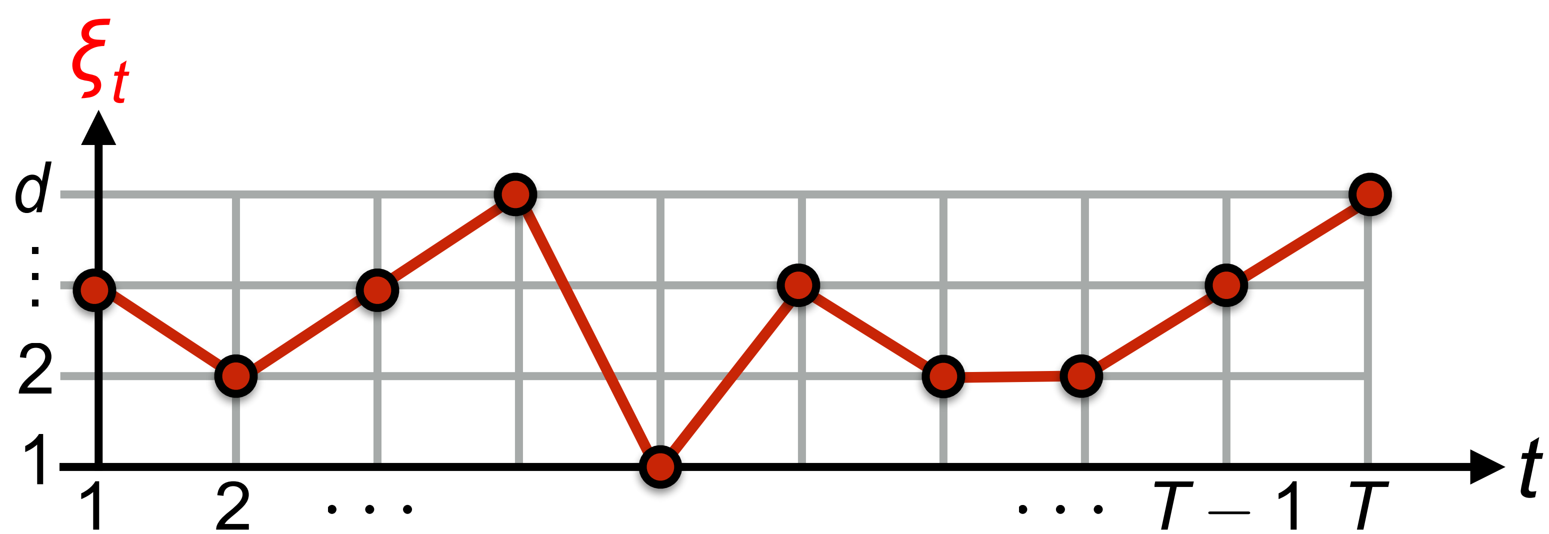
Data-Driven Predictors & Prescriptors

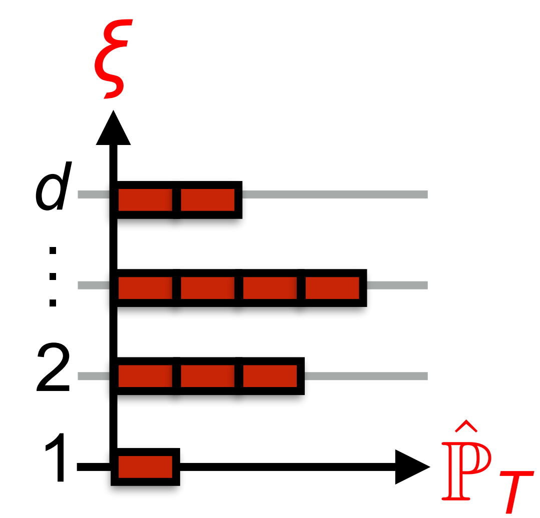
Definition:

- Predictor: $\hat{c}(x, \hat{\mathbb{P}}_T)$
- Prescriptor: $\hat{x}(\hat{\mathbb{P}}_T) \in \underset{x \in X}{\operatorname{argmin}}\, \hat{c}(x, \hat{\mathbb{P}}_T)$

"Compressing" the statistical information:

training samples $\xi_1, \xi_2, \ldots, \xi_T$ $\rightsquigarrow$ empirical distribution $\hat{\mathbb{P}}_T(i) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\xi_t = i}$





Data-Driven Stochastic Programming

$\xi_1, \xi_2, \ldots, \xi_T$ → 

→ $\begin{cases} \text{predictor } \hat{c}(x, \hat{\mathbb{P}}_T) \\ \text{prescriptor } \hat{x}(\hat{\mathbb{P}}_T) \end{cases}$

training data $\sim \mathbb{P}^N$

Examples:

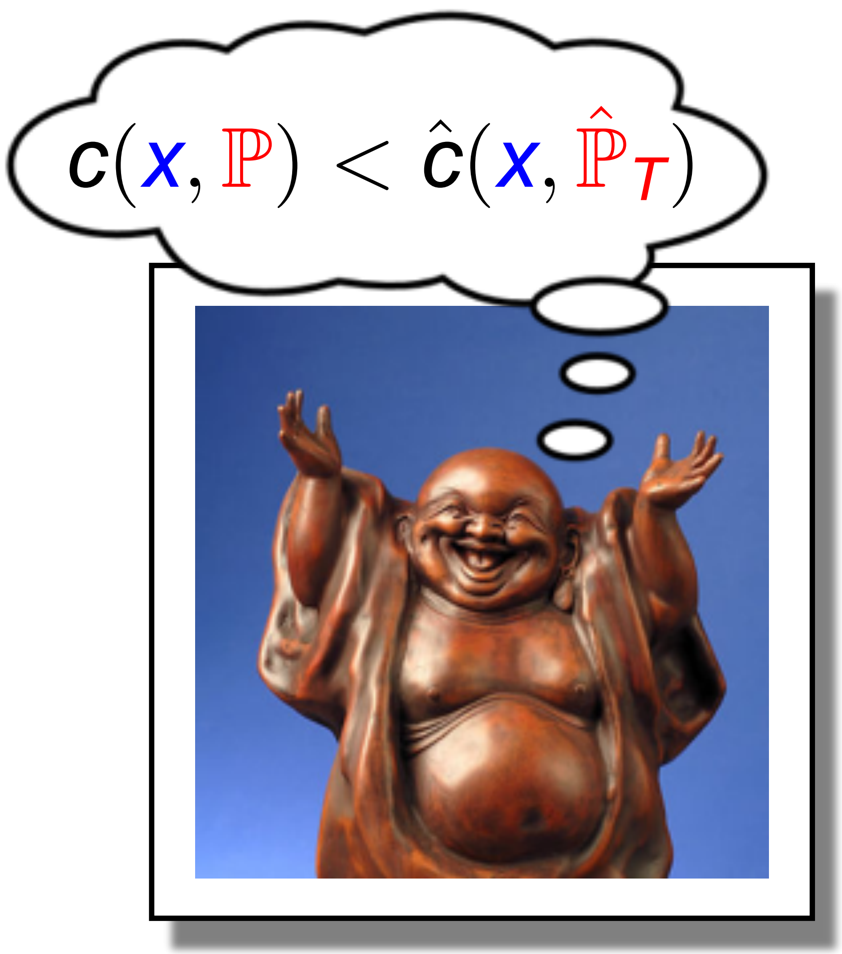
- SAA predictor $\quad \hat{c}(x, \hat{\mathbb{P}}_T) = c(x, \hat{\mathbb{P}}_T) = \frac{1}{T}\sum_{t=1}^{T} \gamma(x, \xi_t)$
- Plug-in predictor $\quad \hat{c}(x, \hat{\mathbb{P}}_T) = c(x, \mathbb{P}_{\text{est}}(\hat{\mathbb{P}}_T))$
- DRO predictor $\quad \hat{c}(x, \hat{\mathbb{P}}_T) = \max_{\mathbb{P} \in \mathcal{P}(\hat{\mathbb{P}}_T)} c(x, \mathbb{P})$
- etc.

Out-of-Sample Disappointment

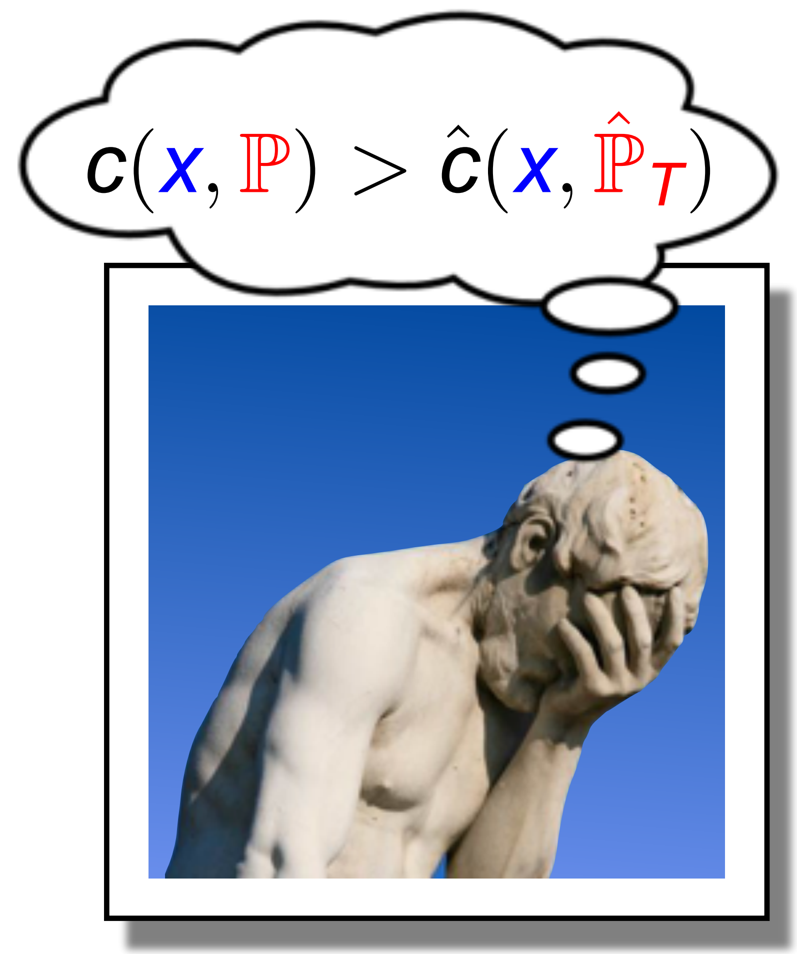
Mean-squared error: $\mathbb{E}_{\mathbb{P}^\infty}\left[\left|c(x,\mathbb{P}) - \hat{c}(x,\hat{\mathbb{P}}_T)\right|^2\right]$

Out-of-Sample Disappointment

Mean-squared error: $\mathbb{E}_{\mathbb{P}^\infty}\left[\left|c(x,\mathbb{P}) - \hat{c}(x,\hat{\mathbb{P}}_T)\right|^2\right]$



Positive surprise



Disappointment

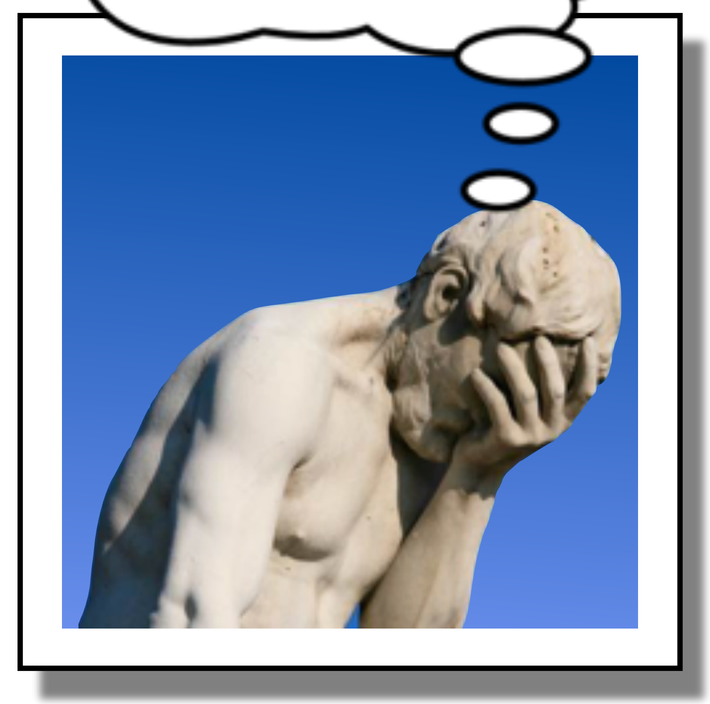
Out-of-Sample Disappointment

Mean-squared error: $\mathbb{E}_{\mathbb{P}^\infty}\left[\left|c(x,\mathbb{P}) - \hat{c}(x,\hat{\mathbb{P}}_T)\right|^2\right]$

$c(x,\mathbb{P}) < \hat{c}(x,\hat{\mathbb{P}}_T)$

Positive surprise

$c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T)$



Disappointment

Out-of-sample disappointment: $\mathbb{P}^\infty\left[c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T)\right]$

Partial Orders for Predictors & Prescriptors

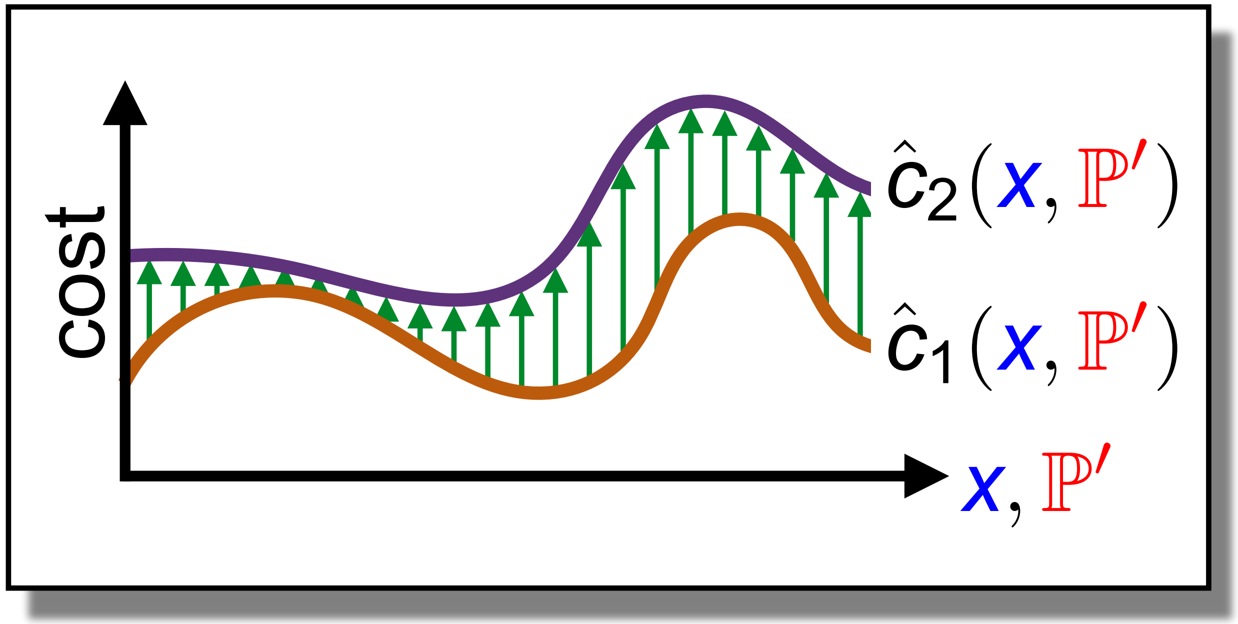
$\hat{c}_1$ **less conservative than** $\hat{c}_2$**:**

$$\hat{c}_1 \preceq_{\mathcal{C}} \hat{c}_2 \quad \Longleftrightarrow \quad \hat{c}_1(x, \mathbb{P}') \leq \hat{c}_2(x, \mathbb{P}') \quad \forall x, \mathbb{P}'$$

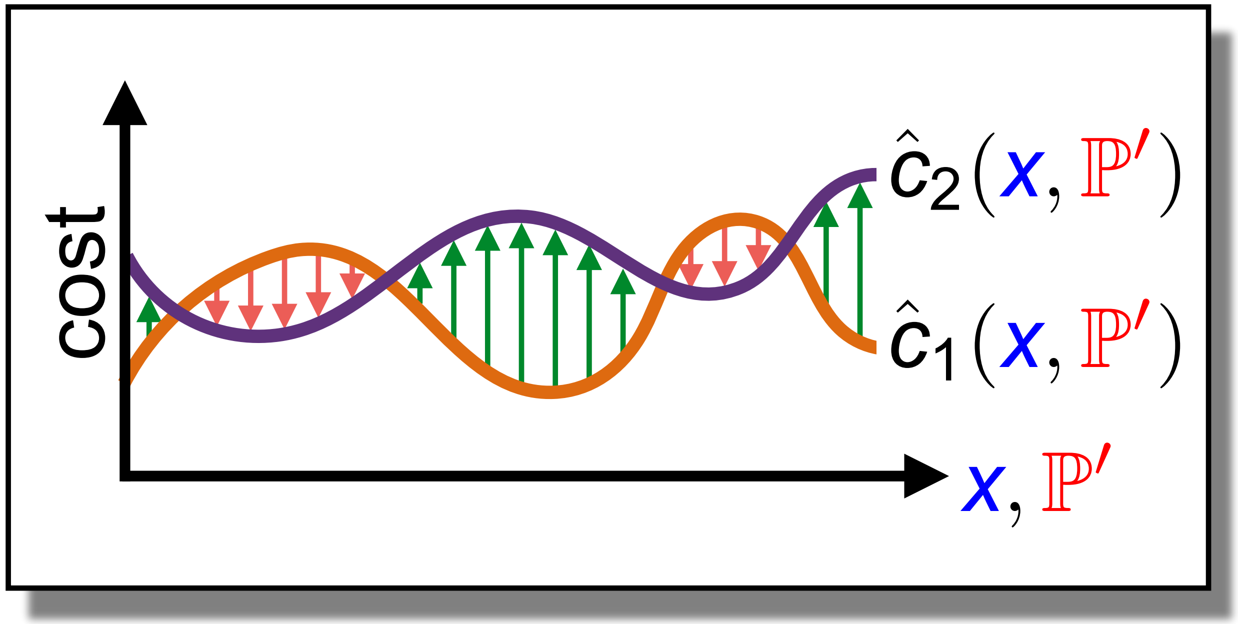
Partial Orders for Predictors & Prescriptors

$\hat{c}_1$ **less conservative than** $\hat{c}_2$**:**

$$\hat{c}_1 \preceq_{\mathcal{C}} \hat{c}_2 \iff \hat{c}_1(x, \mathbb{P}') \leq \hat{c}_2(x, \mathbb{P}') \quad \forall x, \mathbb{P}'$$



$\hat{c}_1 \preceq_{\mathcal{C}} \hat{c}_2$

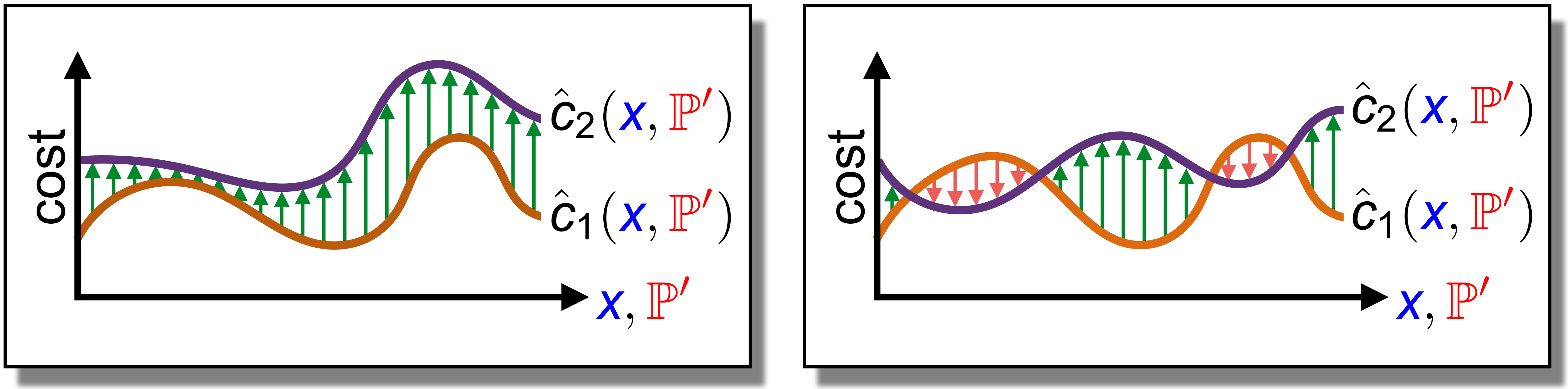


$\hat{c}_1, \hat{c}_2$ incomparable

Partial Orders for Predictors & Prescriptors

$\hat{c}_1$ less conservative than $\hat{c}_2$:

$$\hat{c}_1 \preceq_{\mathcal{C}} \hat{c}_2 \iff \hat{c}_1(x, \mathbb{P}') \leq \hat{c}_2(x, \mathbb{P}') \quad \forall x, \mathbb{P}'$$



$\hat{c}_1 \preceq_{\mathcal{C}} \hat{c}_2$ $\qquad$ $\hat{c}_1, \hat{c}_2$ incomparable

$(\hat{c}_1, \hat{x}_1)$ less conservative than $(\hat{c}_2, \hat{x}_2)$:

$$(\hat{c}_1, \hat{x}_1) \preceq_{\mathcal{X}} (\hat{c}_2, \hat{x}_2) \iff \hat{c}_1(\hat{x}_1(\mathbb{P}'), \mathbb{P}') \leq \hat{c}_2(\hat{x}_2(\mathbb{P}'), \mathbb{P}') \quad \forall \mathbb{P}'$$

Optimizing over Optimization Problems

The best predictor:

$$\begin{aligned} &\underset{\hat{c} \in \mathcal{C}}{\text{minimize}}_{\preceq_{\mathcal{C}}} \; \hat{c} \\ &\text{subject to} \quad \limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}^{\infty} \left(c(x, \mathbb{P}) > \hat{c}(x, \hat{\mathbb{P}}_T) \right) \leq -r \;\; \forall x, \mathbb{P} \end{aligned}$$

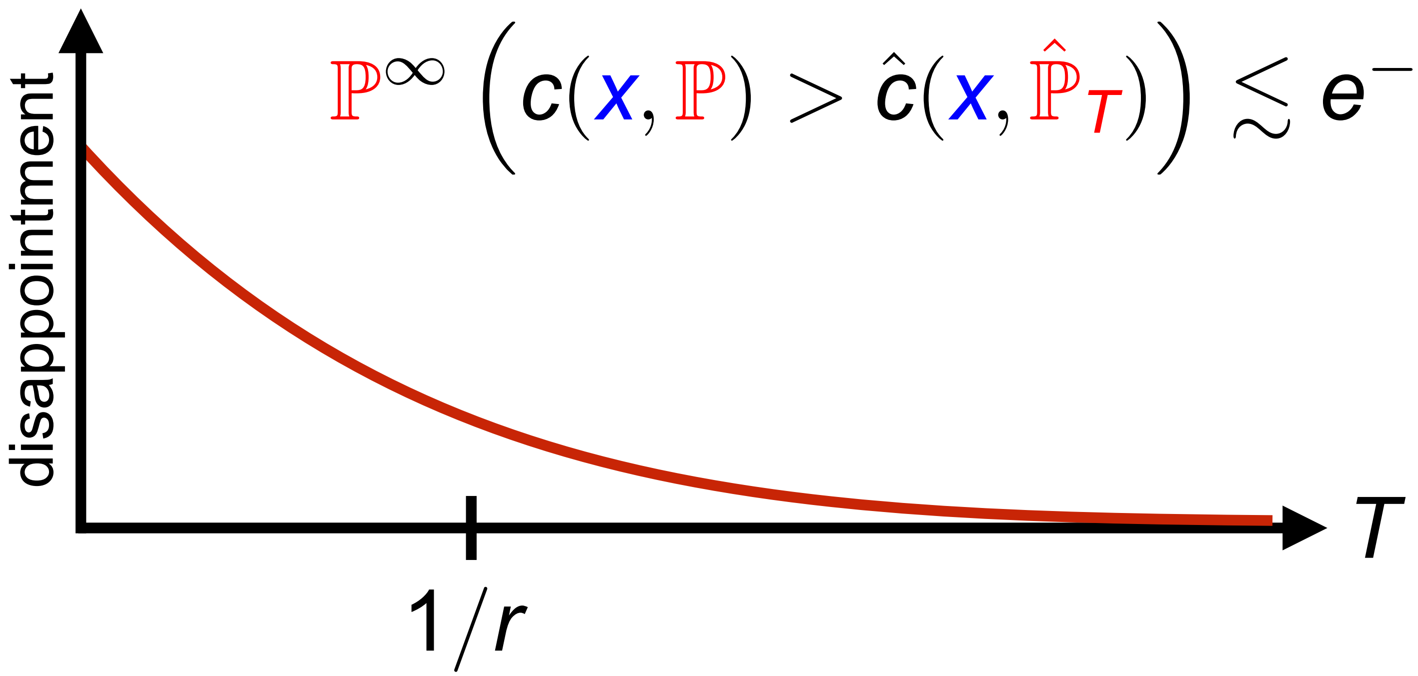
Optimizing over Optimization Problems

The best predictor:

$$\text{minimize}_{\hat{c} \in \mathcal{C}} \;{}_{\preceq_{\mathcal{C}}}\; \hat{c}$$

$$\text{subject to} \quad \limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}^{\infty}\left(c(x, \mathbb{P}) > \hat{c}(x, \hat{\mathbb{P}}_T) \right) \leq -r \;\; \forall x, \mathbb{P}$$

$$\mathbb{P}^{\infty}\left(c(x, \mathbb{P}) > \hat{c}(x, \hat{\mathbb{P}}_T) \right) \lesssim e^{-rT} \text{ for } T \to \infty$$



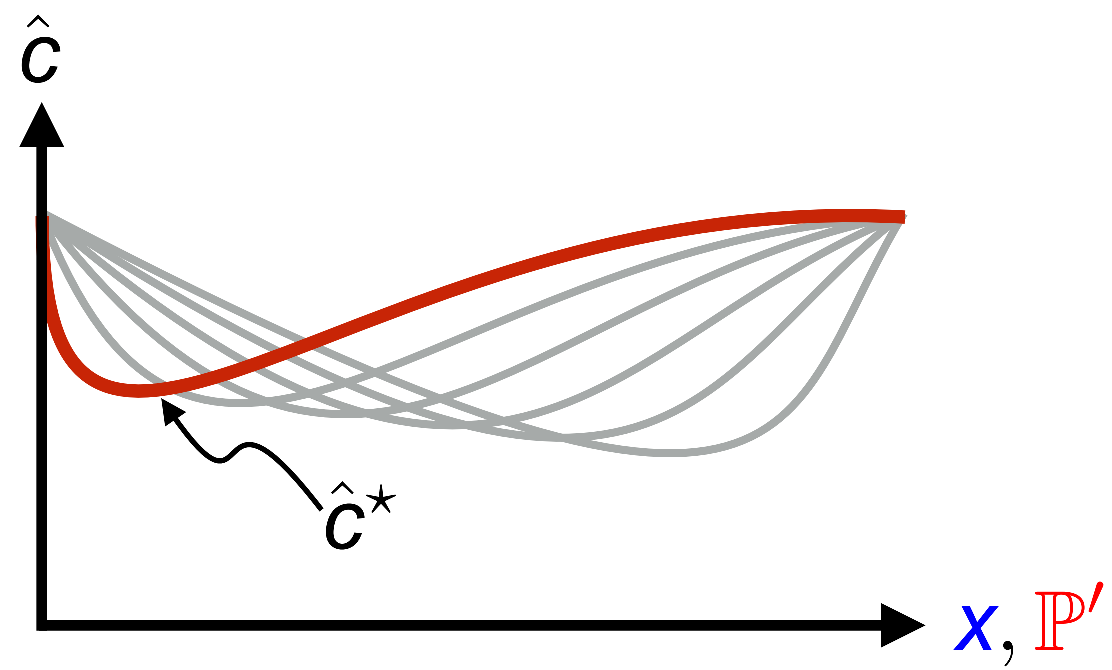
Optimizing over Optimization Problems

The best predictor:

$$\underset{\hat{c}\in\mathcal{C}}{\text{minimize}}_{\preceq_{\mathcal{C}}} \ \hat{c}$$

$$\text{subject to} \quad \limsup_{T\to\infty} \frac{1}{T} \log \mathbb{P}^{\infty}\left(c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T)\right) \le -r \ \ \forall x, \mathbb{P}$$

Weak solution:



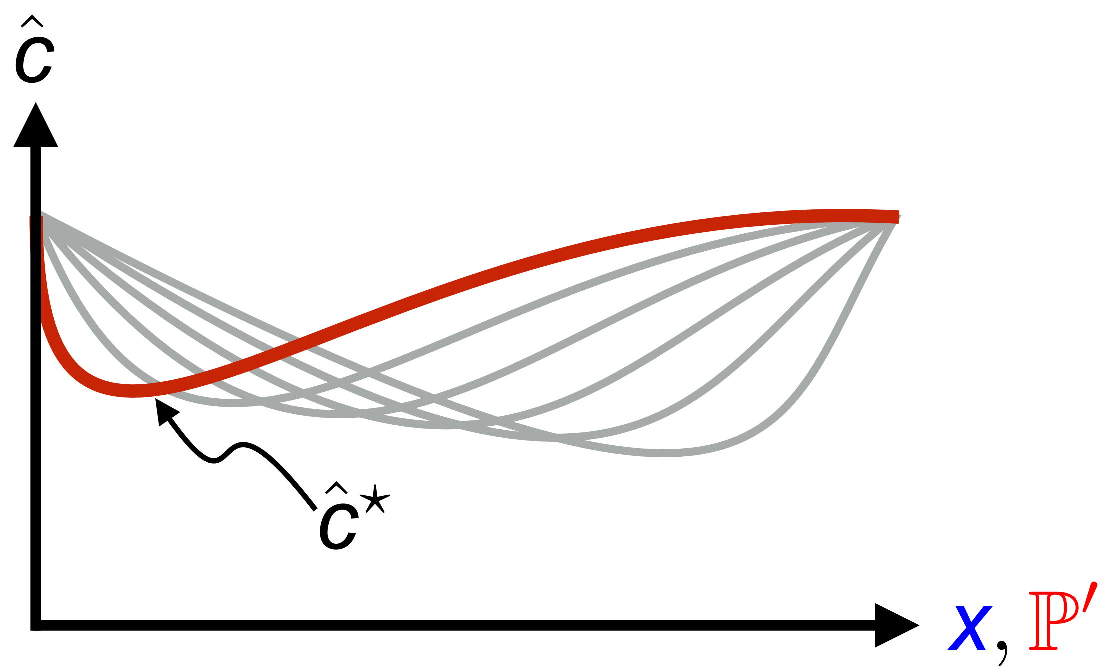
$$\hat{c} \preceq_{\mathcal{C}} \hat{c}^{\star} \implies \hat{c} \text{ infeasible}$$

Optimizing over Optimization Problems

The best predictor:

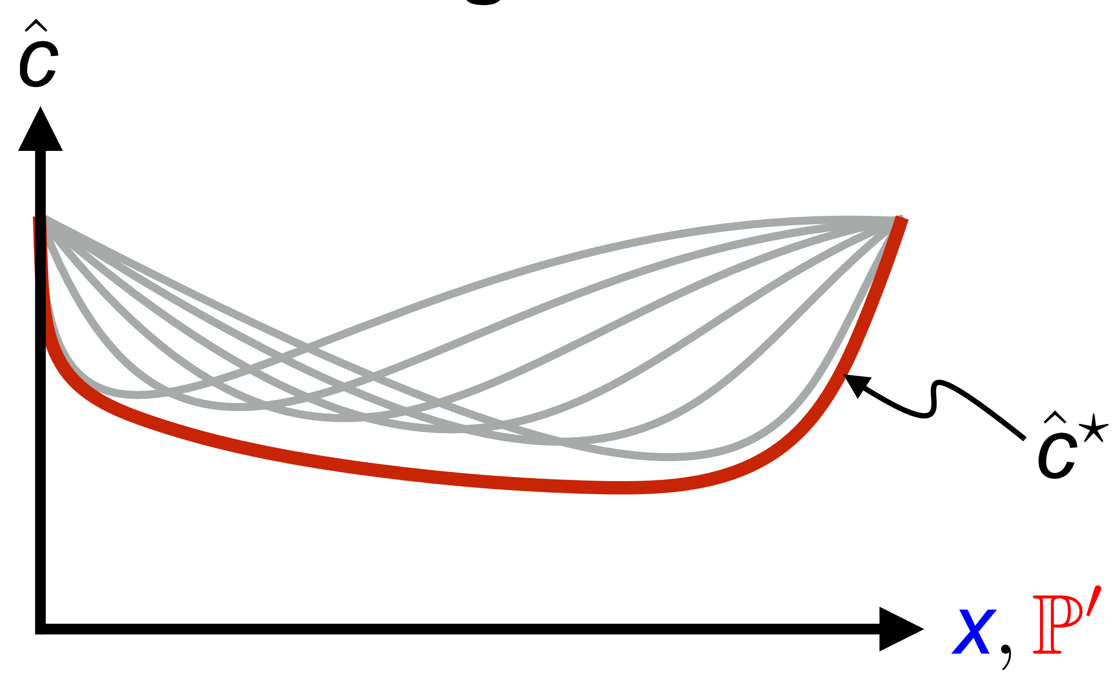
$$\begin{aligned} &\underset{\hat{c}\in\mathcal{C}}{\text{minimize}}_{\preceq_{\mathcal{C}}} \ \hat{c} \\ &\text{subject to} \quad \limsup_{T\to\infty} \frac{1}{T} \log \mathbb{P}^{\infty}\left(c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T) \right) \leq -r \ \ \forall x, \mathbb{P} \end{aligned}$$

Weak solution:



$\hat{c} \preceq_{\mathcal{C}} \hat{c}^{\star} \implies \hat{c}$ infeasible

Strong solution:



$\hat{c}$ feasible $\implies \hat{c}^{\star} \preceq_{\mathcal{C}} \hat{c}$

Optimizing over Optimization Problems

The best predictor:

$$\begin{aligned} &\underset{\hat{c}\in\mathcal{C}}{\text{minimize}}_{\preceq_{\mathcal{C}}} \ \hat{c} \\ &\text{subject to} \quad \limsup_{T\to\infty} \frac{1}{T} \log \mathbb{P}^{\infty}\left(c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T) \right) \leq -r \ \ \forall x, \mathbb{P} \end{aligned}$$

The best predictor-prescriptor pair:

$$\begin{aligned} &\underset{(\hat{c},\hat{x})\in\mathcal{X}}{\text{minimize}}_{\preceq_{\mathcal{X}}} \ (\hat{c},\hat{x}) \\ &\text{subject to} \quad \limsup_{T\to\infty} \frac{1}{T} \log \mathbb{P}^{\infty}\left(c(\hat{x}(\hat{\mathbb{P}}_T),\mathbb{P}) > \hat{c}(\hat{x}(\hat{\mathbb{P}}_T),\hat{\mathbb{P}}_T) \right) \leq -r \ \ \forall \mathbb{P} \end{aligned}$$

Large Deviation Principles

Relative entropy: $$\mathrm{I}(\mathbb{P}', \mathbb{P}) = \sum_{i=1}^{d} \mathbb{P}'(i) \log \left(\frac{\mathbb{P}'(i)}{\mathbb{P}(i)} \right)$$

Large Deviation Principles

Relative entropy: $\mathrm{I}(\mathbb{P}', \mathbb{P}) = \sum_{i=1}^{d} \mathbb{P}'(i) \log\left(\frac{\mathbb{P}'(i)}{\mathbb{P}(i)}\right)$

estimator realization ($\mathbb{P}'$)

data-generating distribution ($\mathbb{P}$)

Large Deviation Principles

Relative entropy: $\mathrm{I}(\mathbb{P}', \mathbb{P}) = \sum_{i=1}^{d} \mathbb{P}'(i) \log \left(\frac{\mathbb{P}'(i)}{\mathbb{P}(i)} \right)$

Weak LDP: If $\xi_1, \xi_2, \ldots \sim \mathbb{P}$ i.i.d. and $\mathcal{D} \subseteq \mathcal{P}$, then:

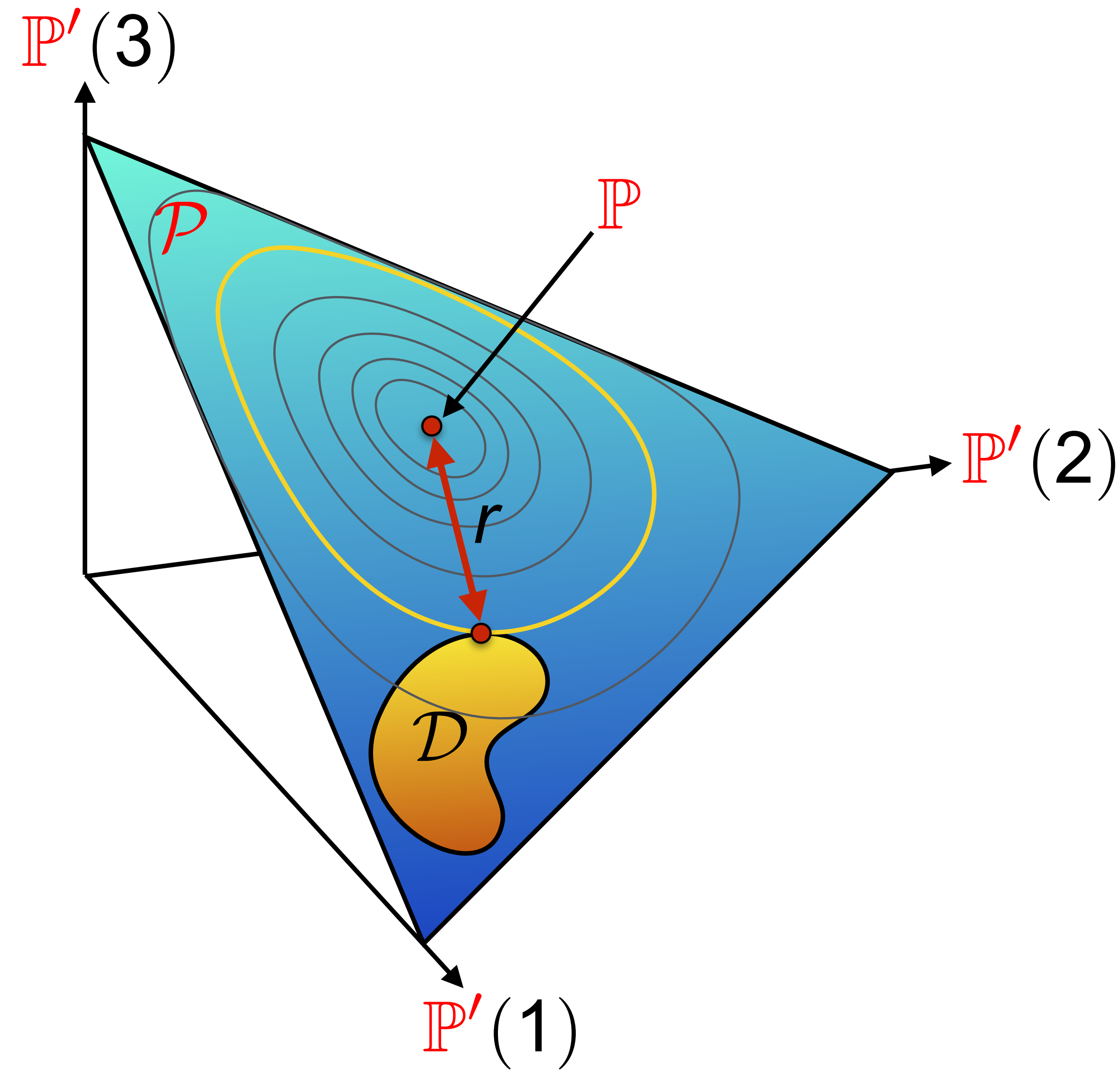
$$\limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}^\infty(\hat{\mathbb{P}}_T \in \mathcal{D}) \leq - \inf_{\mathbb{P}' \in \mathcal{D}} \mathrm{I}(\mathbb{P}', \mathbb{P})$$

$$\liminf_{T \to \infty} \frac{1}{T} \log \mathbb{P}^\infty(\hat{\mathbb{P}}_T \in \mathcal{D}) \geq - \inf_{\mathbb{P}' \in \operatorname{int} \mathcal{D}} \mathrm{I}(\mathbb{P}', \mathbb{P})$$

Large Deviation Principles

$$\mathbb{P}^{\infty}(\hat{\mathbb{P}}_T \in \mathcal{D}) \approx e^{-rT}$$

$$r = \inf_{\mathbb{P}' \in \mathcal{D}} \mathrm{I}(\mathbb{P}', \mathbb{P})$$



Distributionally Robust Predictors/Prescriptors

Distributionally robust predictor:

$$\hat{c}_r(x, \mathbb{P}') = \max_{\mathbb{P} \in \mathcal{P}} \quad c(x, \mathbb{P}) \quad \text{s.t.} \quad \mathrm{I}(\mathbb{P}', \mathbb{P}) \leq r$$

$$\hat{x}_r(\mathbb{P}') \in \operatorname*{argmin}_{x \in X} \hat{c}_r(x, \mathbb{P}')$$

Distributionally Robust Predictors/Prescriptors

Distributionally robust predictor:

$$\hat{c}_r(x, \mathbb{P}') = \max_{\mathbb{P} \in \mathcal{P}} \; c(x, \mathbb{P}) \quad \text{s.t.} \quad \mathrm{I}(\mathbb{P}', \mathbb{P}) \leq r$$

$$\hat{x}_r(\mathbb{P}') \in \operatorname*{argmin}_{x \in X} \hat{c}_r(x, \mathbb{P}')$$

estimator
realization

Distributionally Robust Predictors/Prescriptors

Distributionally robust predictor:

$$\hat{c}_r(x, \mathbb{P}') = \max_{\mathbb{P} \in \mathcal{P}} \; c(x, \mathbb{P}) \quad \text{s.t.} \quad \mathrm{I}(\mathbb{P}', \mathbb{P}) \leq r$$

$$\hat{x}_r(\mathbb{P}') \in \operatorname*{argmin}_{x \in X} \hat{c}_r(x, \mathbb{P}')$$

Reverse **distributionally robust predictor:**

$$\check{c}_r(x, \mathbb{P}') = \max_{\mathbb{P} \in \mathcal{P}} \; c(x, \mathbb{P}) \quad \text{s.t.} \quad \mathrm{I}(\mathbb{P}, \mathbb{P}') \leq r$$

$$\check{x}_r(\hat{\mathbb{P}}') \in \operatorname*{argmin}_{x \in X} \check{c}_r(x, \hat{\mathbb{P}}')$$

Distributionally Robust Predictors/Prescriptors

Distributionally robust predictor:

$$\hat{c}_r(x, \mathbb{P}') = \begin{array}{ll} \max\limits_{\mathbb{P}\in\mathcal{P}} & c(x, \mathbb{P}) \\ \text{s.t.} & \mathrm{I}(\mathbb{P}', \mathbb{P}) \leq r \end{array}$$

$$\hat{x}_r(\mathbb{P}') \in \operatorname*{argmin}_{x\in X} \hat{c}_r(x, \mathbb{P}')$$

Reverse distributionally robust predictor:

$$\check{c}_r(x, \mathbb{P}') = \begin{array}{ll} \max\limits_{\mathbb{P}\in\mathcal{P}} & c(x, \mathbb{P}) \\ \text{s.t.} & \mathrm{I}(\mathbb{P}, \mathbb{P}') \leq r \end{array}$$

$$\check{x}_r(\hat{\mathbb{P}}') \in \operatorname*{argmin}_{x\in X} \check{c}_r(x, \hat{\mathbb{P}}')$$

estimator

realization

Optimizing over Optimization Problems

Meta optimization problem (MOP):

$$\begin{aligned} &\underset{\hat{c}\in\mathcal{C}}{\text{minimize}}_{\preceq_{\mathcal{C}}}\ \hat{c} \\ &\text{subject to} \quad \limsup_{T\to\infty} \frac{1}{T}\log \mathbb{P}^{\infty}\Big(c(x,\mathbb{P}) > \hat{c}(x,\hat{\mathbb{P}}_T)\Big) \leq -r \ \ \forall x,\mathbb{P} \end{aligned}$$

Feasibility

Theorem: If $r \geq 0$, then $\hat{c}_r$ is feasible in (MOP).

Feasibility

Theorem: If $r \geq 0$, then $\hat{c}_r$ is feasible in (MOP).

$$c(x, \mathbb{P}) > \hat{c}_r(x, \mathbb{P}') \implies c(x, \mathbb{P}) > \max_{\mathbb{Q} \in \mathcal{P}} \{c(x, \mathbb{Q}) : \mathrm{I}(\mathbb{P}', \mathbb{Q}) \leq r\}$$

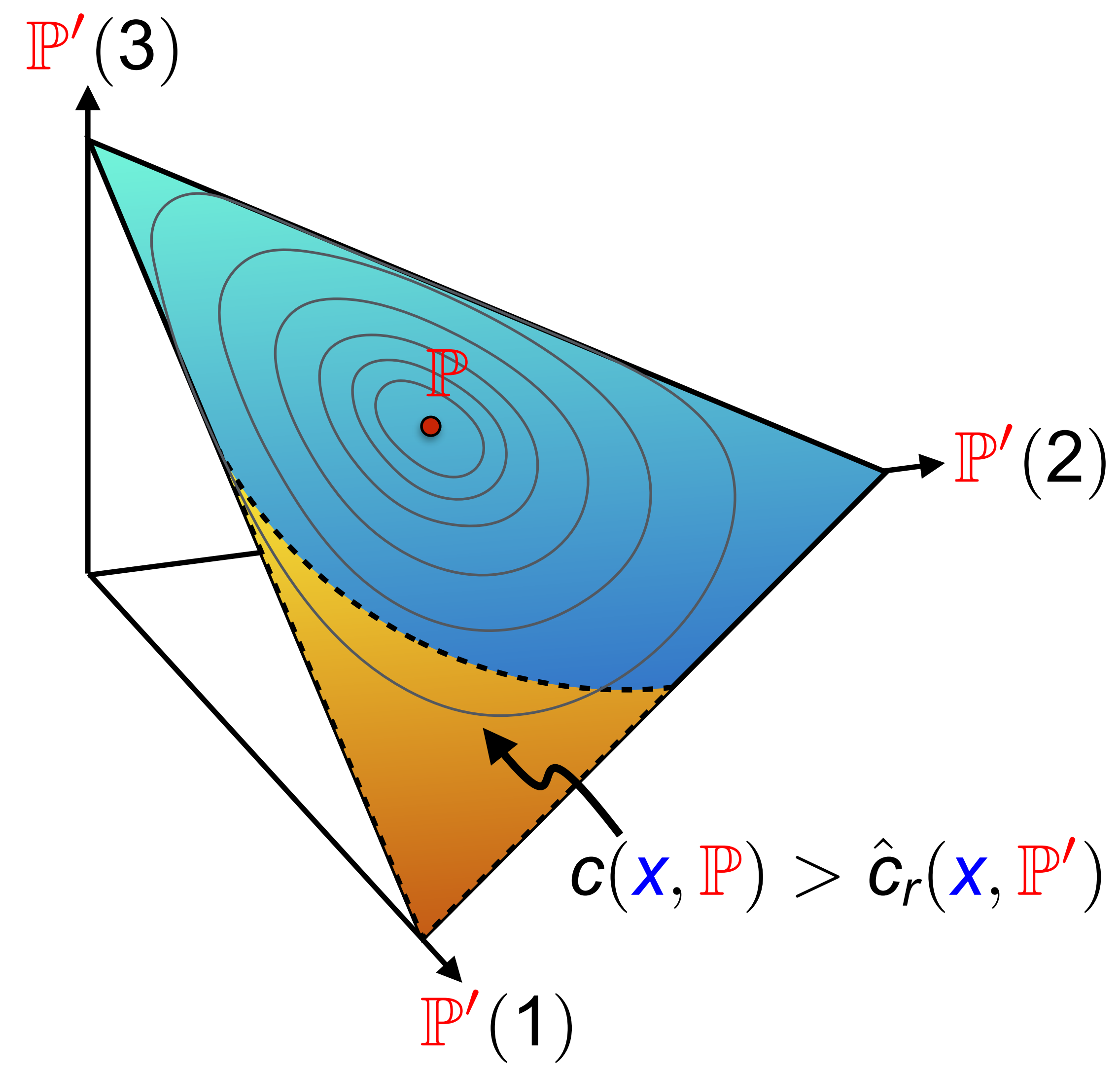
$$\implies \mathrm{I}(\mathbb{P}', \mathbb{P}) > r$$

Feasibility

Theorem: If $r \geq 0$, then $\hat{c}_r$ is feasible in (MOP).

$$c(x, \mathbb{P}) > \hat{c}_r(x, \mathbb{P}') \quad \Longrightarrow \quad c(x, \mathbb{P}) > \max_{\mathbb{Q} \in \mathcal{P}} \{ c(x, \mathbb{Q}) : \mathrm{I}(\mathbb{P}', \mathbb{Q}) \leq r \}$$

$$\Longrightarrow \quad \mathrm{I}(\mathbb{P}', \mathbb{P}) > r$$

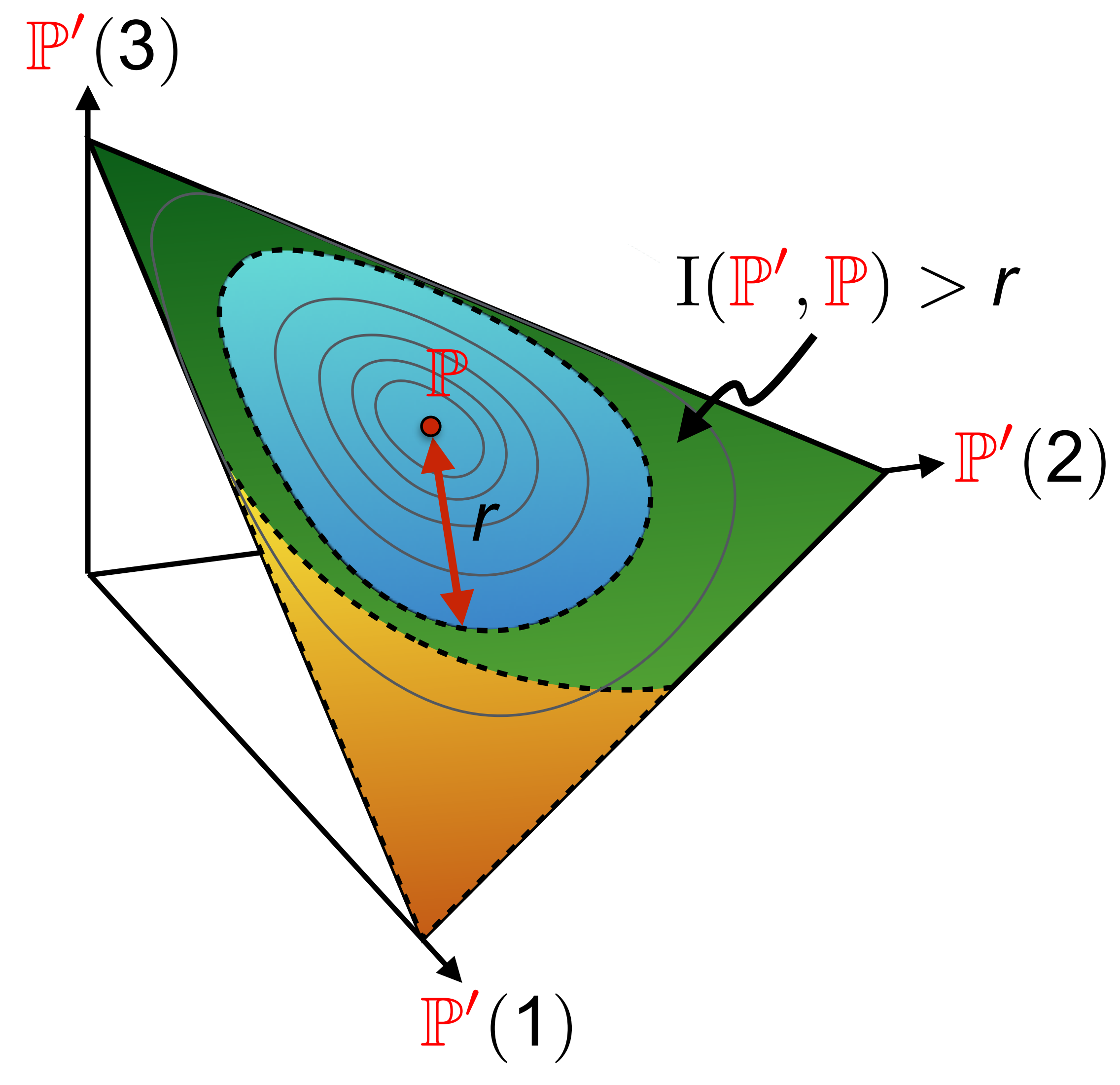


Feasibility

Theorem: If $r \geq 0$, then $\hat{c}_r$ is feasible in (MOP).

$$c(x, \mathbb{P}) > \hat{c}_r(x, \mathbb{P}') \quad \Longrightarrow \quad c(x, \mathbb{P}) > \max_{\mathbb{Q} \in \mathcal{P}} \{c(x, \mathbb{Q}) : \mathrm{I}(\mathbb{P}', \mathbb{Q}) \leq r\}$$

$$\Longrightarrow \quad \mathrm{I}(\mathbb{P}', \mathbb{P}) > r$$



Feasibility

Theorem: If $r \geq 0$, then $\hat{c}_r$ is feasible in (MOP).

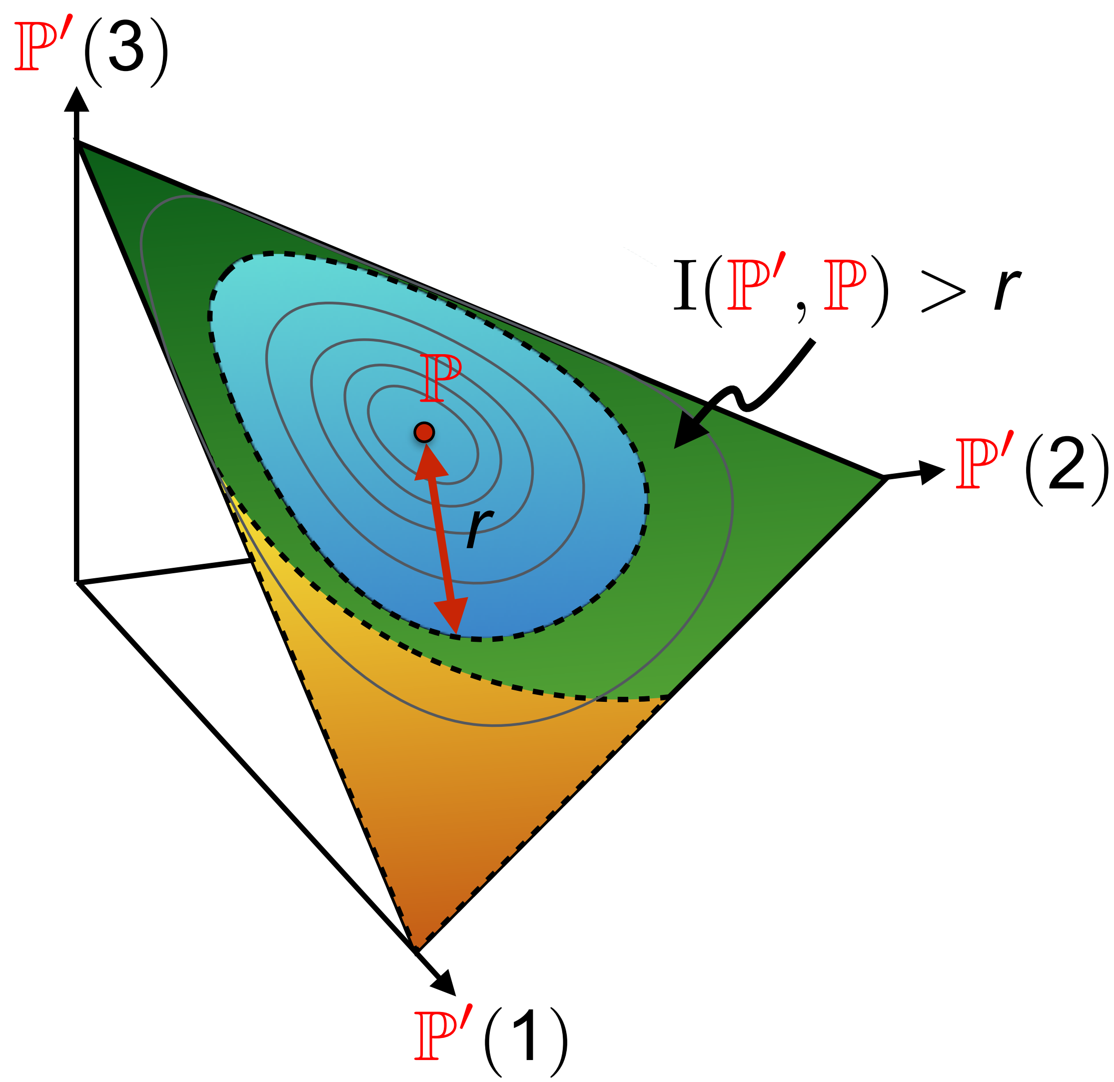
$$c(x, \mathbb{P}) > \hat{c}_r(x, \mathbb{P}') \implies c(x, \mathbb{P}) > \max_{\mathbb{Q} \in \mathcal{P}} \{c(x, \mathbb{Q}) : \mathrm{I}(\mathbb{P}', \mathbb{Q}) \leq r\}$$

$$\implies \mathrm{I}(\mathbb{P}', \mathbb{P}) > r$$

$$\limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}^\infty \left(c(x, \mathbb{P}) > \hat{c}_r(x, \hat{\mathbb{P}}_T) \right)$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}^\infty \left(\mathrm{I}(\hat{\mathbb{P}}_T, \mathbb{P}) > r \right)$$

$$\leq - \inf_{\mathbb{P}' \in \mathcal{P}} \left\{ \mathrm{I}(\mathbb{P}', \mathbb{P}) : \mathrm{I}(\mathbb{P}', \mathbb{P}) > r \right\} \leq -r$$



Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

$$\hat{c}_r \not\preceq_{\mathcal{C}} \hat{c} \quad \implies \quad \exists x, \mathbb{P}'_0 : \; \hat{c}(x, \mathbb{P}'_0) < \hat{c}_r(x, \mathbb{P}'_0)$$

Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

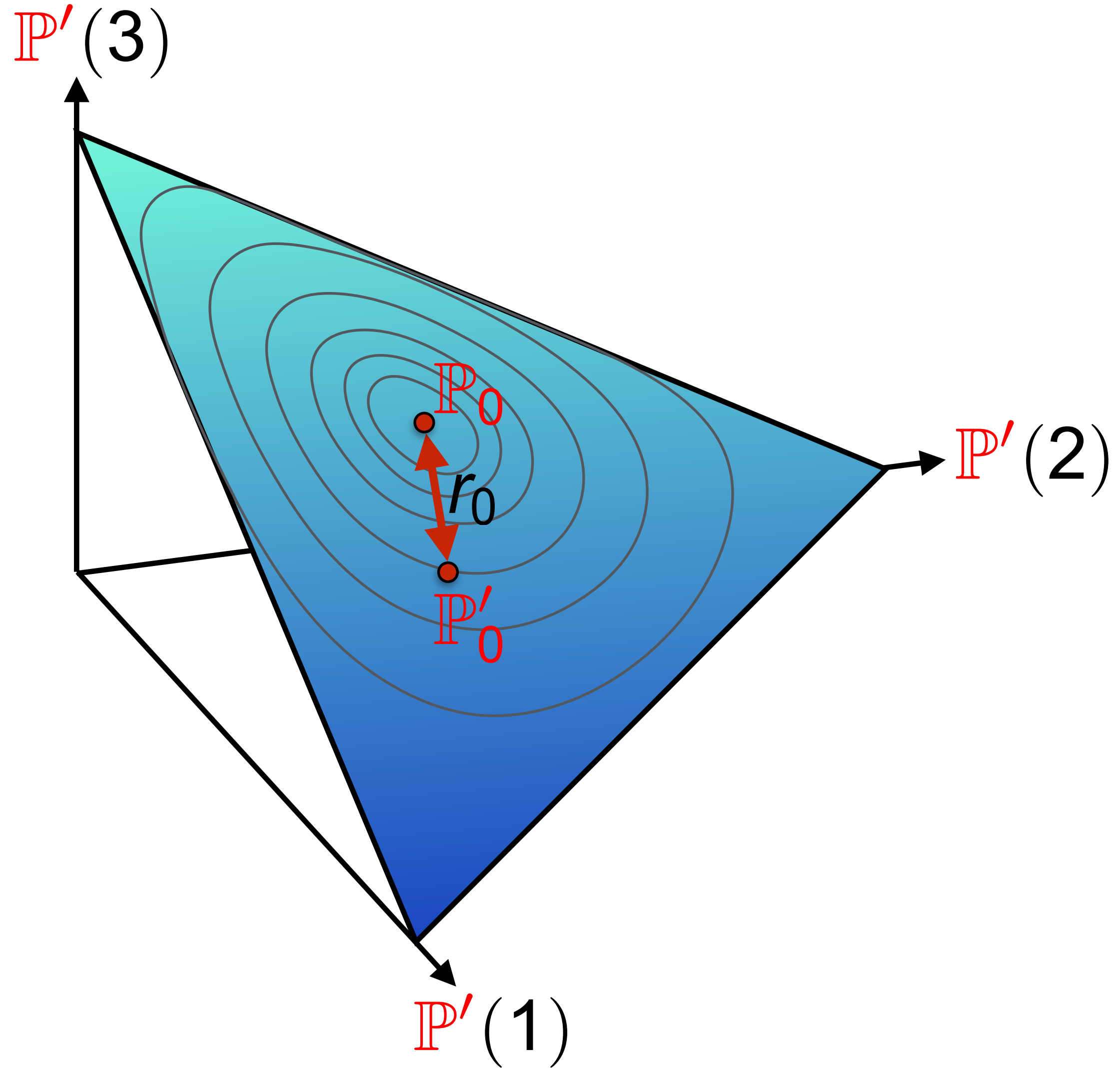
$$\hat{c}_r \not\preceq_{\mathcal{C}} \hat{c} \quad \Longrightarrow \quad \exists x, \mathbb{P}_0' : \quad \hat{c}(x, \mathbb{P}_0') < \max_{\mathbb{P} \in \mathcal{P}} \{ c(x, \mathbb{P}) : \mathrm{I}(\mathbb{P}_0', \mathbb{P}) \le r \}$$

Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

$$\hat{c}_r \not\preceq_{\mathcal{C}} \hat{c} \quad \Longrightarrow \quad \exists x, \mathbb{P}'_0 : \quad \hat{c}(x, \mathbb{P}'_0) < \max_{\mathbb{P} \in \mathcal{P}} \{c(x, \mathbb{P}) : \mathrm{I}(\mathbb{P}'_0, \mathbb{P}) \leq r\}$$

$$\Longrightarrow \quad \exists \mathbb{P}_0 : \quad \hat{c}(x, \mathbb{P}'_0) < c(x, \mathbb{P}_0) \text{ and } \mathrm{I}(\mathbb{P}'_0, \mathbb{P}_0) = r_0 < r$$



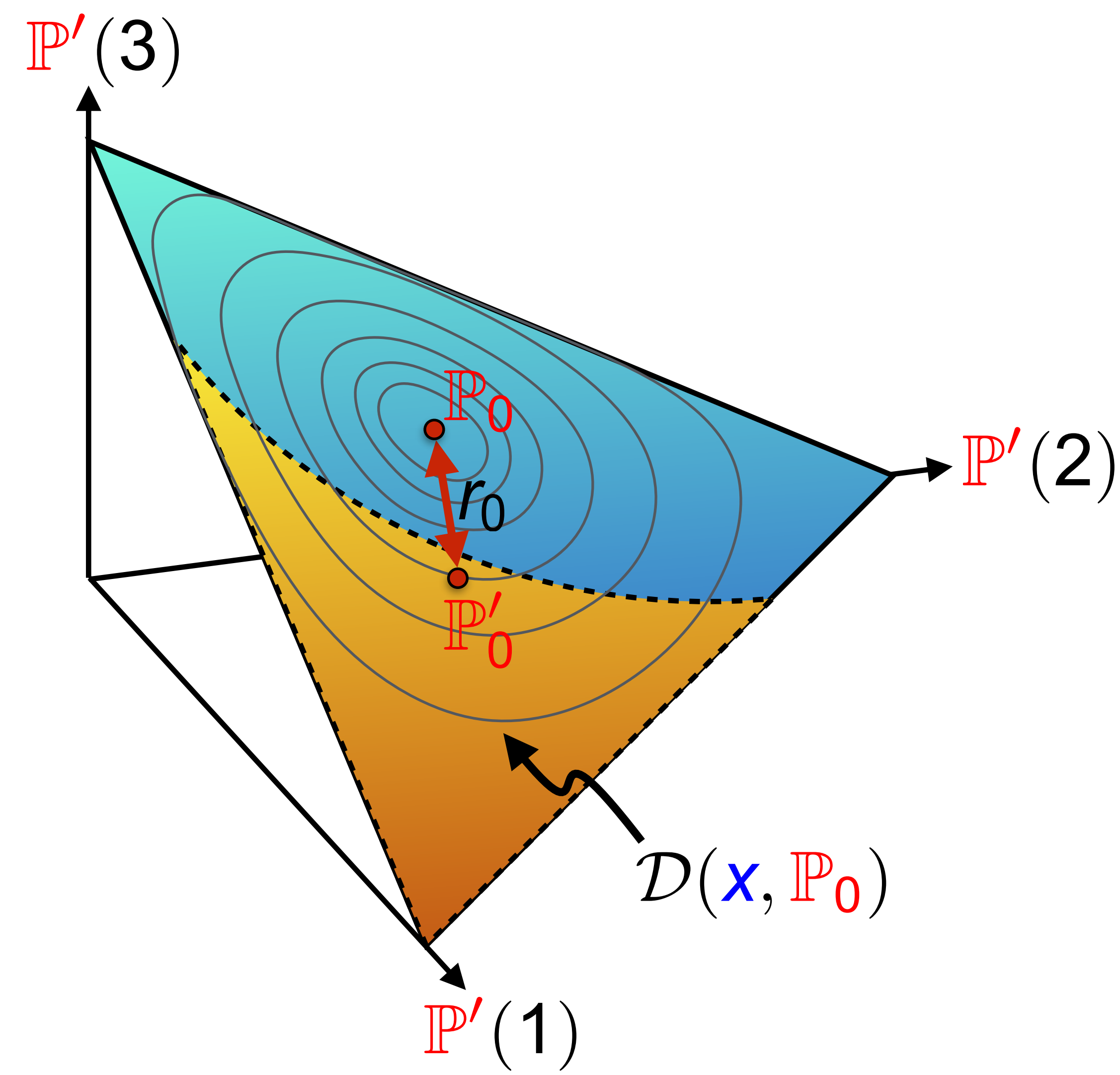
Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

$$\hat{c}_r \not\preceq_{\mathcal{C}} \hat{c} \quad \Longrightarrow \quad \exists x, \mathbb{P}'_0 : \quad \hat{c}(x, \mathbb{P}'_0) < \max_{\mathbb{P} \in \mathcal{P}} \{c(x, \mathbb{P}) : \mathrm{I}(\mathbb{P}'_0, \mathbb{P}) \leq r\}$$

$$\Longrightarrow \quad \exists \mathbb{P}_0 : \quad \hat{c}(x, \mathbb{P}'_0) < c(x, \mathbb{P}_0) \text{ and } \mathrm{I}(\mathbb{P}'_0, \mathbb{P}_0) = r_0 < r$$

$$\mathcal{D}(x, \mathbb{P}_0) = \left\{ \mathbb{P}' \in \mathcal{P} : c(x, \mathbb{P}_0) > \hat{c}(x, \mathbb{P}') \right\}$$



Optimality

Theorem: If $r > 0$, then $\hat{c}_r$ is strongly optimal in (MOP).

$$\hat{c}_r \not\preceq_C \hat{c} \quad \Longrightarrow \quad \exists x, \mathbb{P}_0' : \quad \hat{c}(x, \mathbb{P}_0') < \max_{\mathbb{P} \in \mathcal{P}} \{c(x, \mathbb{P}) : \mathrm{I}(\mathbb{P}_0', \mathbb{P}) \leq r\}$$

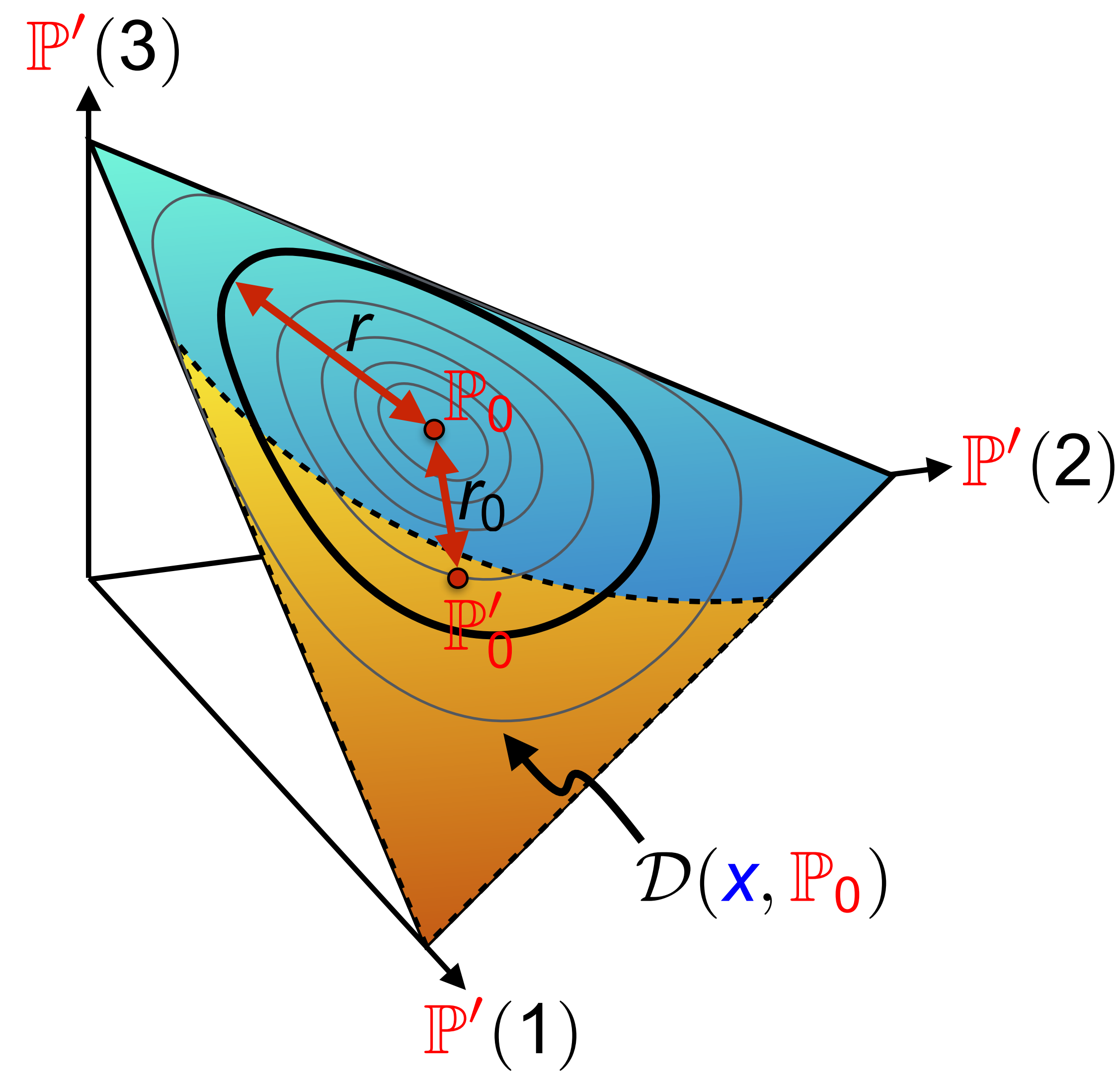
$$\Longrightarrow \quad \exists \mathbb{P}_0 : \quad \hat{c}(x, \mathbb{P}_0') < c(x, \mathbb{P}_0) \ \text{ and } \ \mathrm{I}(\mathbb{P}_0', \mathbb{P}_0) = r_0 < r$$

$$\mathcal{D}(x, \mathbb{P}_0) = \left\{ \mathbb{P}' \in \mathcal{P} : c(x, \mathbb{P}_0) > \hat{c}(x, \mathbb{P}') \right\}$$

$$-r < -r_0 \leq - \inf_{\mathbb{P}' \in \operatorname{int} \mathcal{D}(x, \mathbb{P}_0)} \mathrm{I}(\mathbb{P}', \mathbb{P}_0)$$

$$\leq \liminf_{T \to \infty} \frac{1}{T} \log \mathbb{P}_0^\infty \left(\hat{\mathbb{P}}_T \in \mathcal{D}(x, \mathbb{P}_0) \right)$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} \log \mathbb{P}_0^\infty \left(\hat{\mathbb{P}}_T \in \mathcal{D}(x, \mathbb{P}_0) \right)$$



Properties of Optimal Predictor

- **Unique strong** solution of (MOP)
- Has **distributionally robust** interpretation
  - Worst-case expectation over **relative entropy ball**
  - $r =$ decay rate of **out-of-sample disappointment**
- Tractability
  - **Convex program** for generic $\mathbb{P}'$
  - **SOCP** with $\mathcal{O}(T)$ hyperbolic constraints for empirical $\mathbb{P}'$
- Explicit **finite sample guarantee** (no unknown constants)

This Talk is Based on…

[1] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2016.

[2] V. Gupta. **Near-Optimal Bayesian Ambiguity Sets for Distributionally Robust Optimization**. *Optimization Online*, 2015.

[3] H. Lam. **Recovering Best Statistical Guarantees via the Empirical Divergence-based Distributionally Robust Optimization**. *arXiv*, 2016.

[4] B. Van Parys, P. Mohajerin Esfahani and D. Kuhn. **From Data to Decisions: Distributionally Robust Optimization is Optimal**. *Optimization Online*, 2017.