Wasserstein Distributionally Robust Optimization with Heterogeneous Data Sources

Yves Rychener,[1)] Adrián Esteban-Pérez,[2)]
Juan M. Morales[2)] and Daniel Kuhn[1)]

[1)] École Polytechnique Fédérale de Lausanne
www.epfl.ch/labs/rao/

[2)] University of Málaga
https://sites.google.com/site/jnmmgo/

SWISS NATIONAL SCIENCE FOUNDATION
NCCR Automation
EPFL

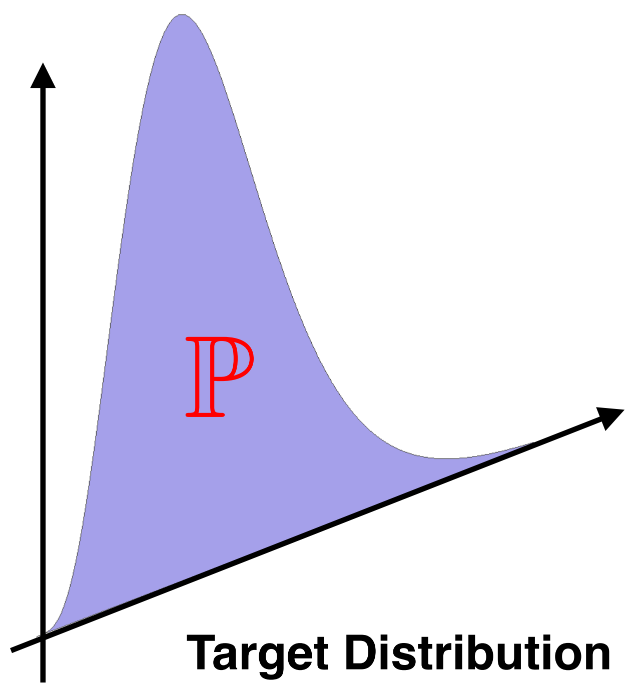
Data-Driven Decision-Making

Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$

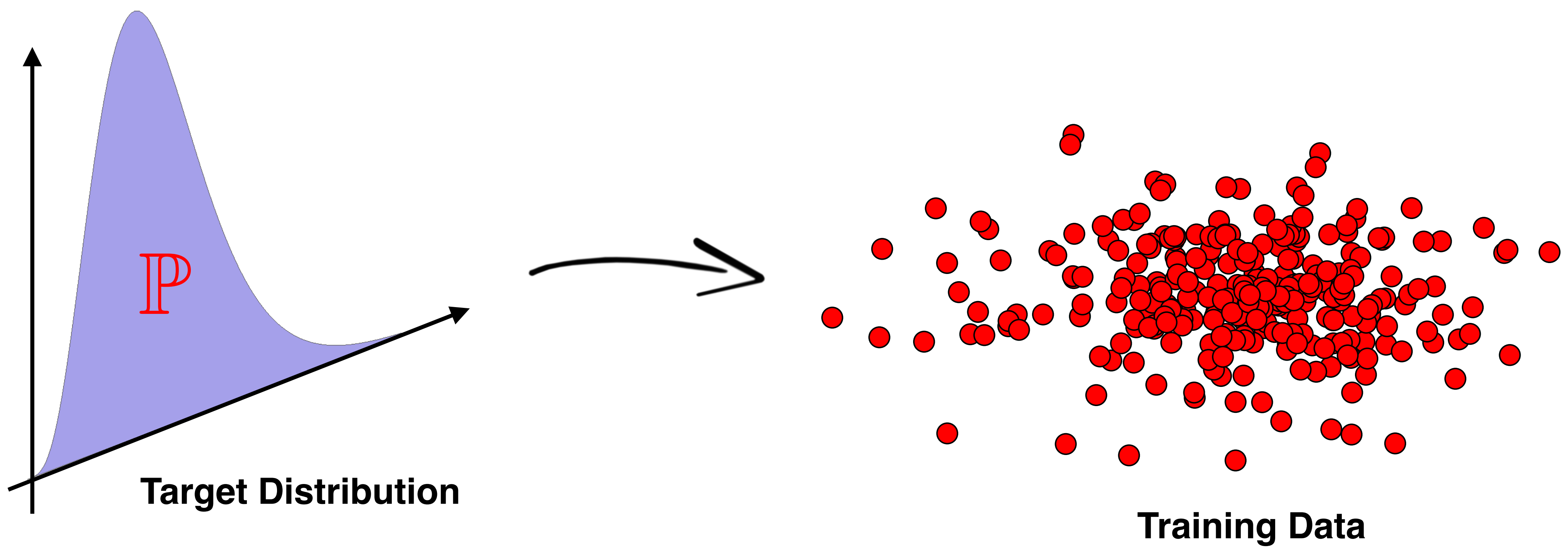
Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \; \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$



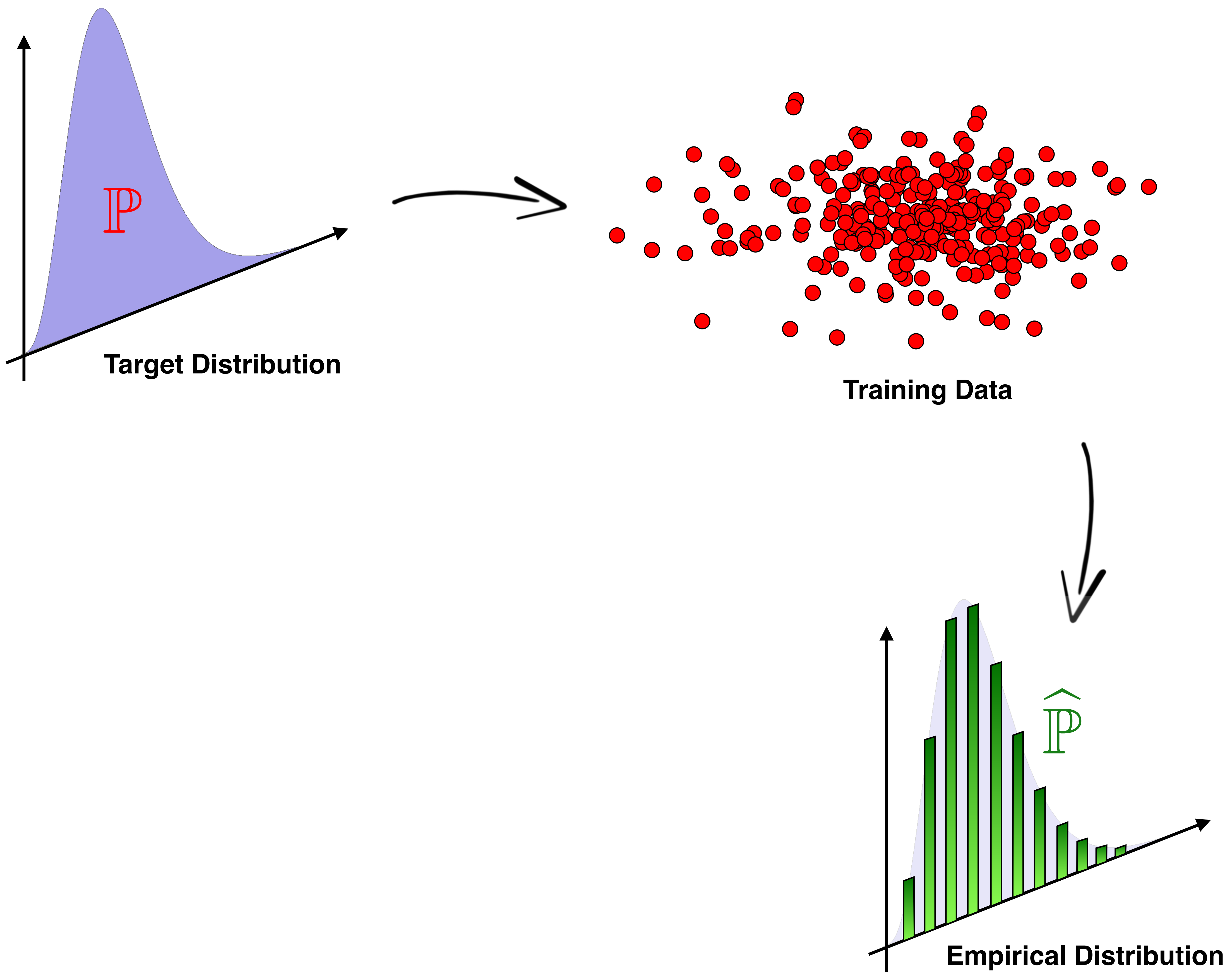
Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$



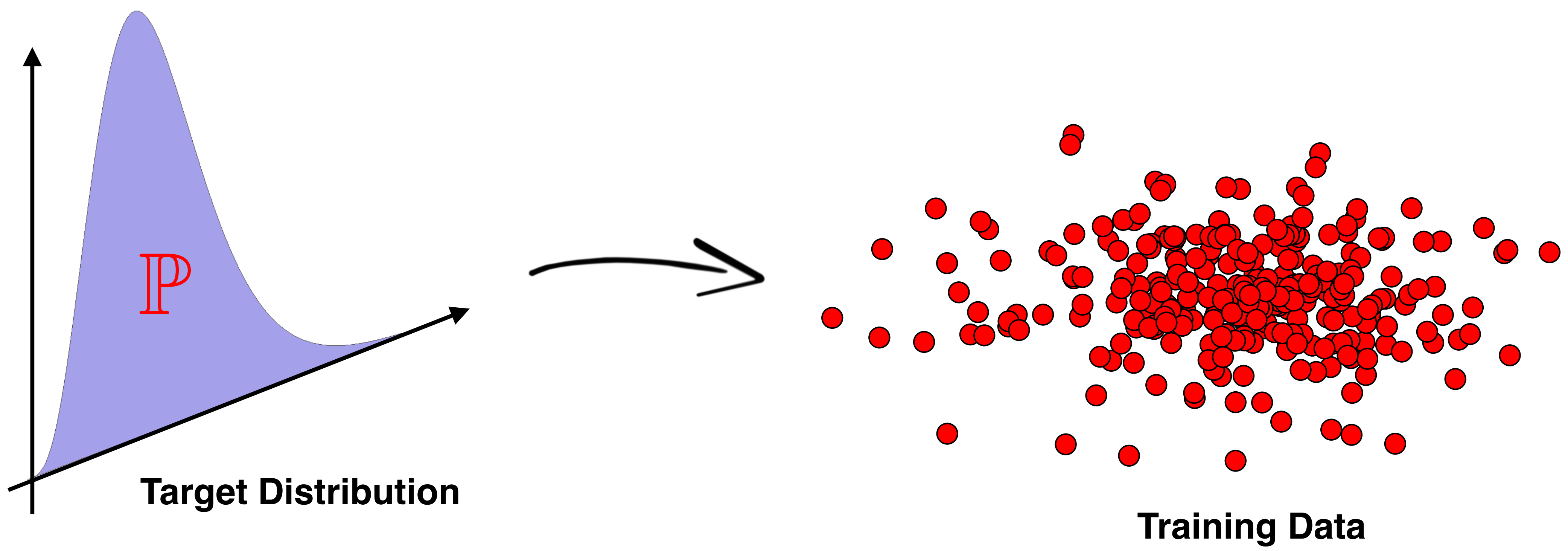
Data-Driven Decision-Making

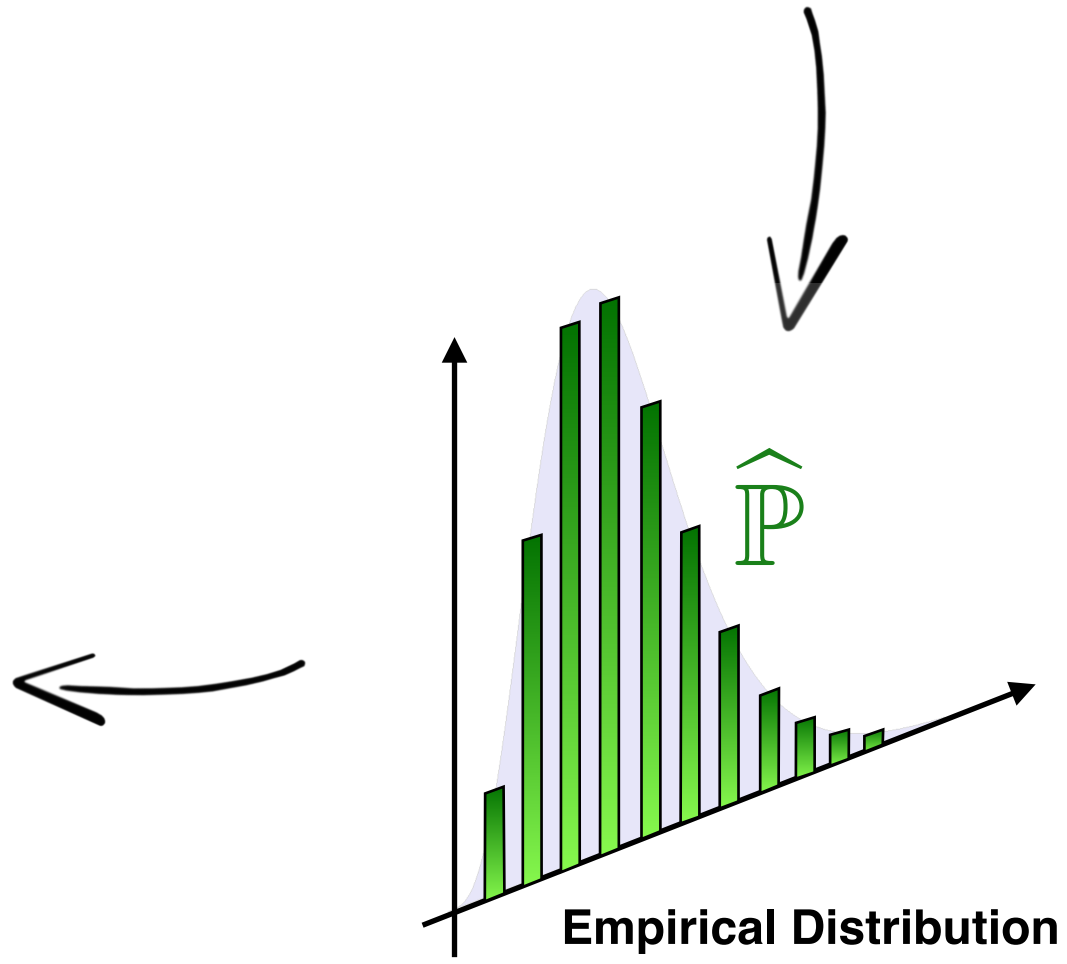
Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$

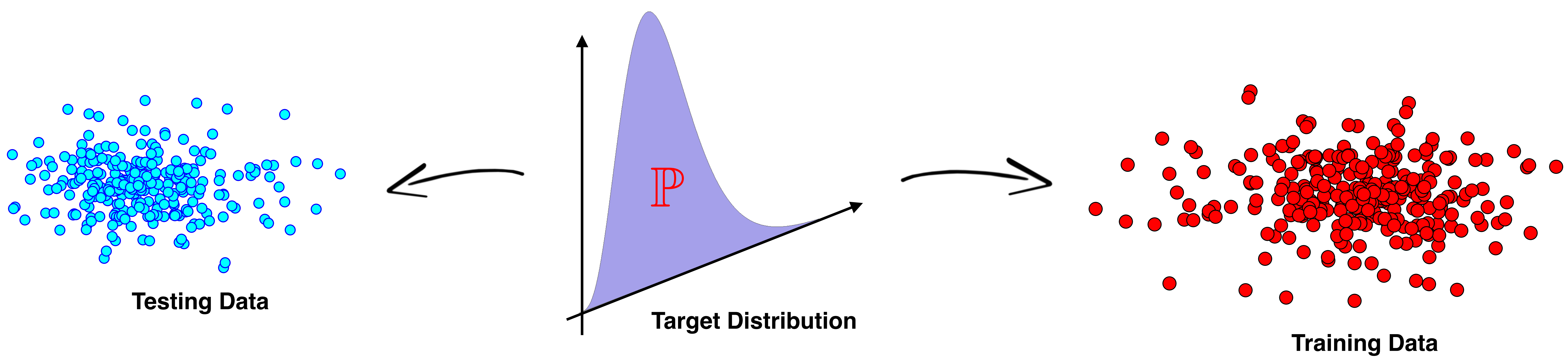




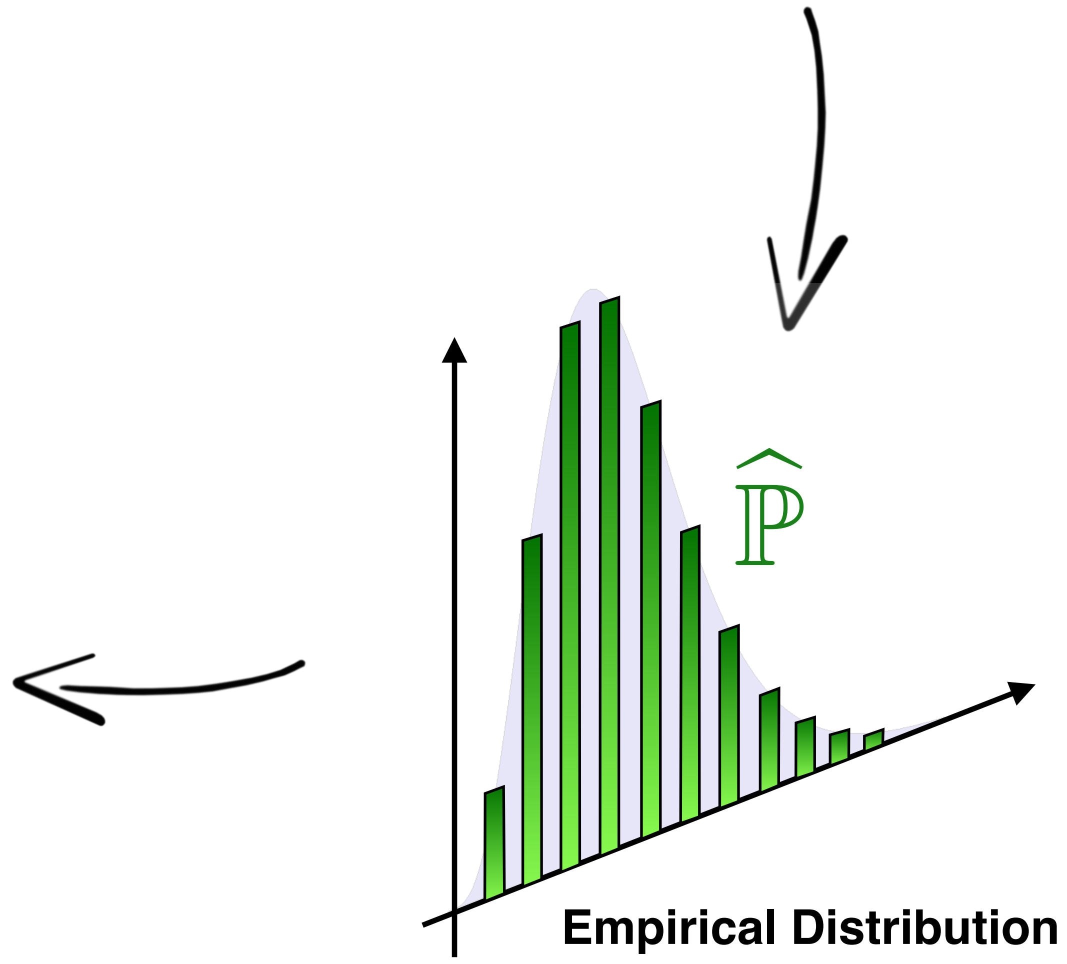
SAA: $\min_{\theta \in \Theta} \mathbb{E}_{\widehat{\mathbb{P}}} [\ell(\theta, \xi)]$

Data-Driven Decision-Making

Stochastic Program: $$\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$$

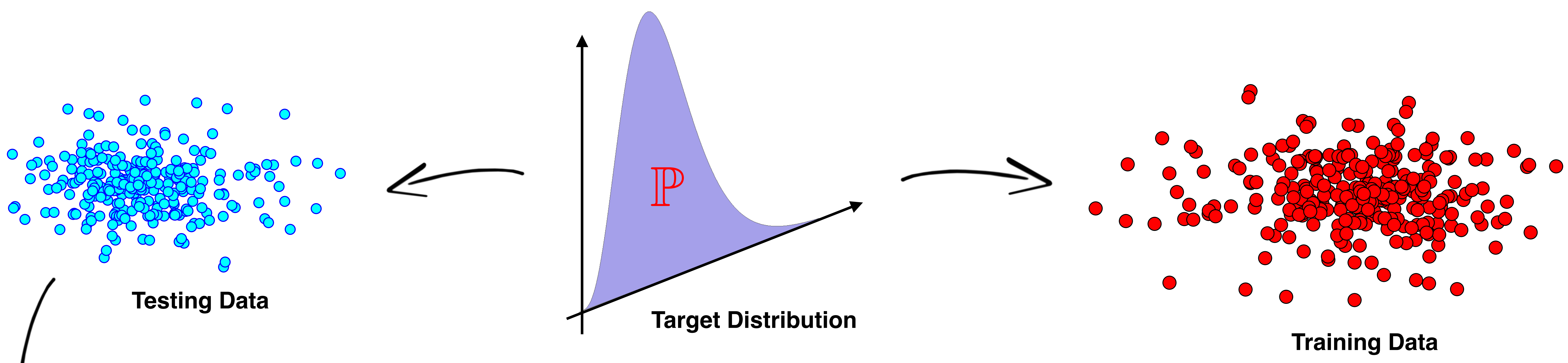


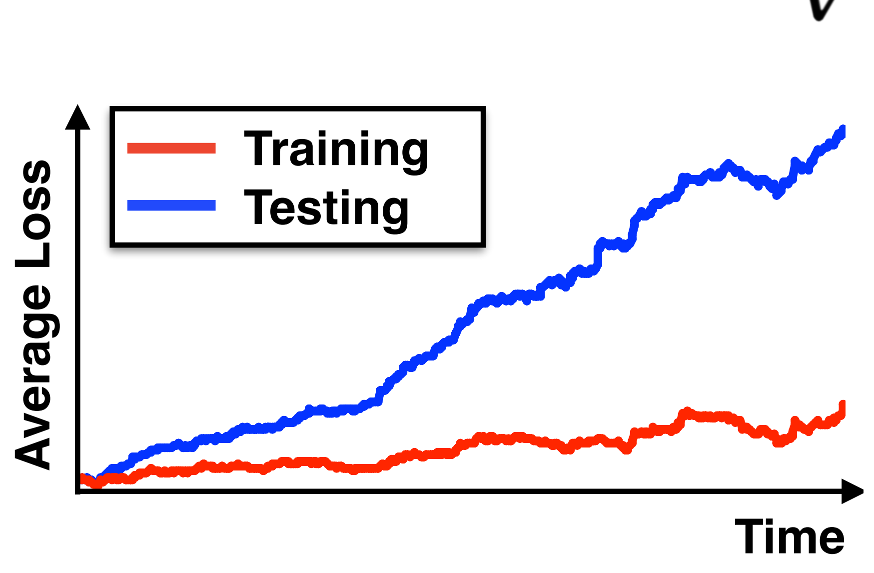
SAA: $$\min_{\theta \in \Theta} \mathbb{E}_{\widehat{\mathbb{P}}} \left[\ell(\theta, \xi) \right]$$



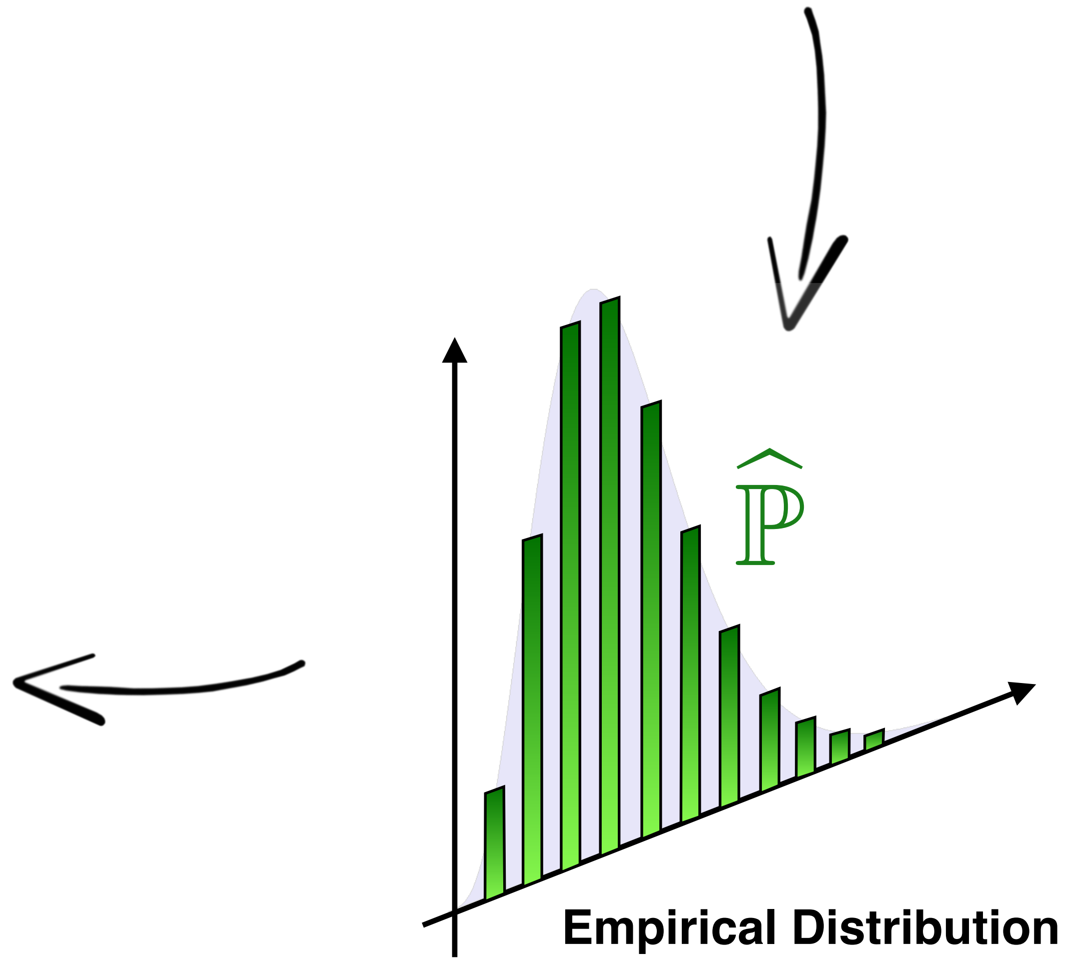
Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



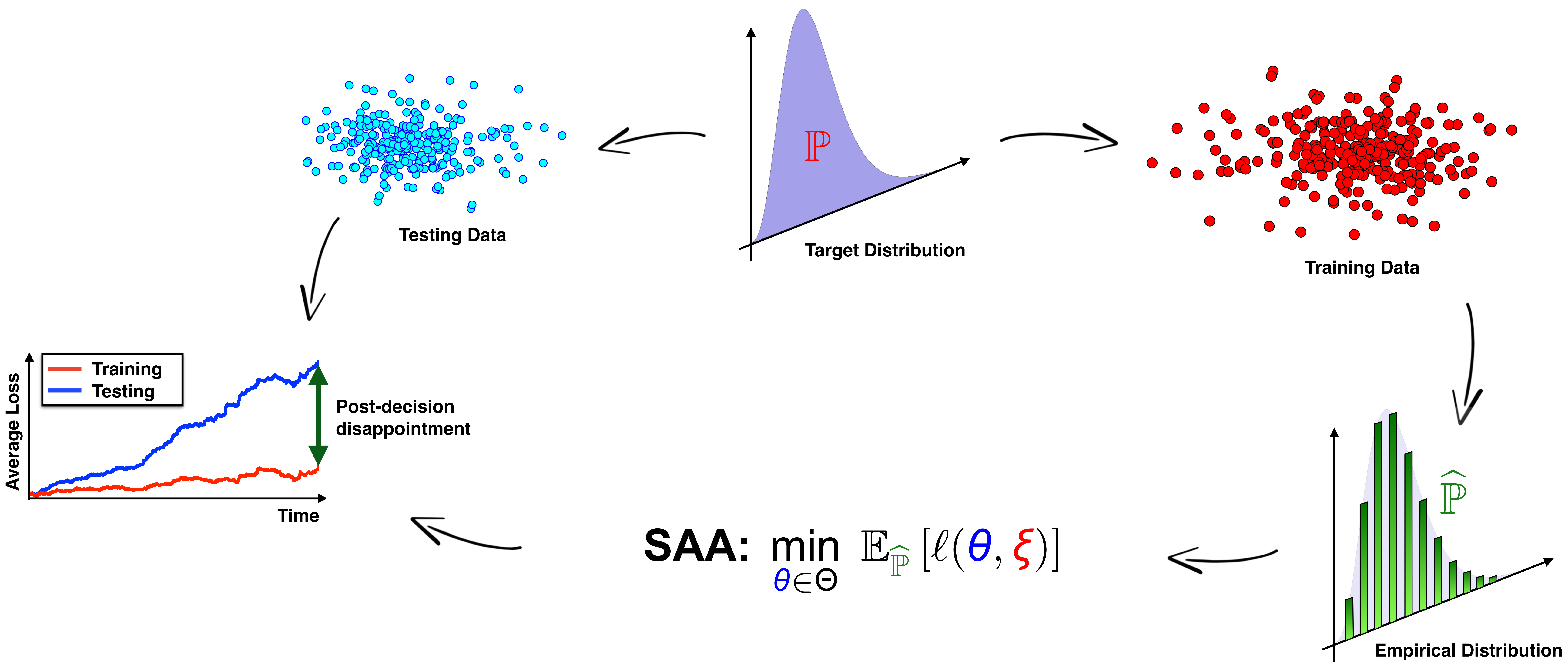


SAA: $\min_{\theta \in \Theta} \mathbb{E}_{\widehat{\mathbb{P}}} [\ell(\theta, \xi)]$



Data-Driven Decision-Making

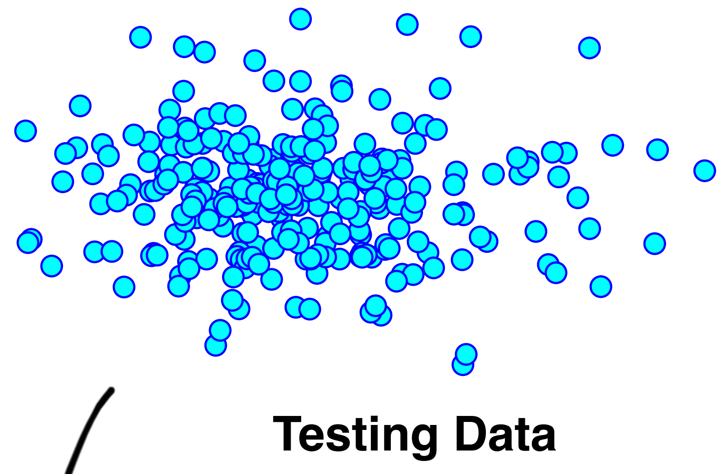
Stochastic Program: $\min_{\theta \in \Theta} \ \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$

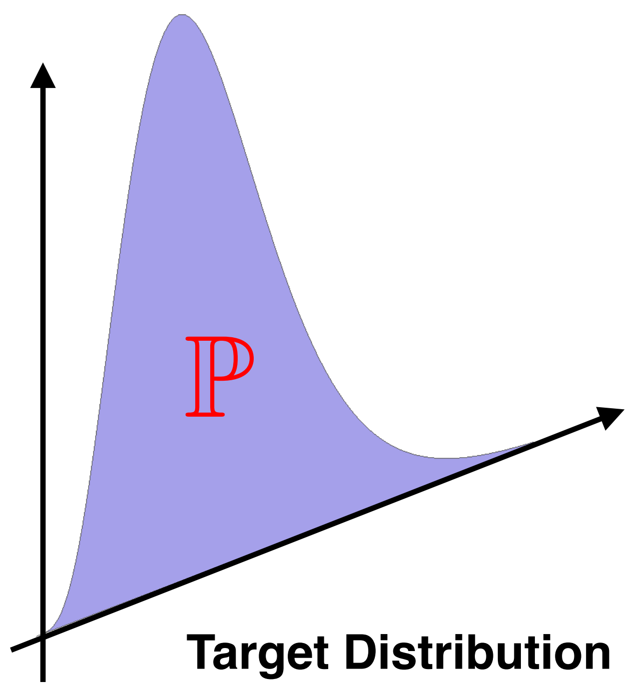


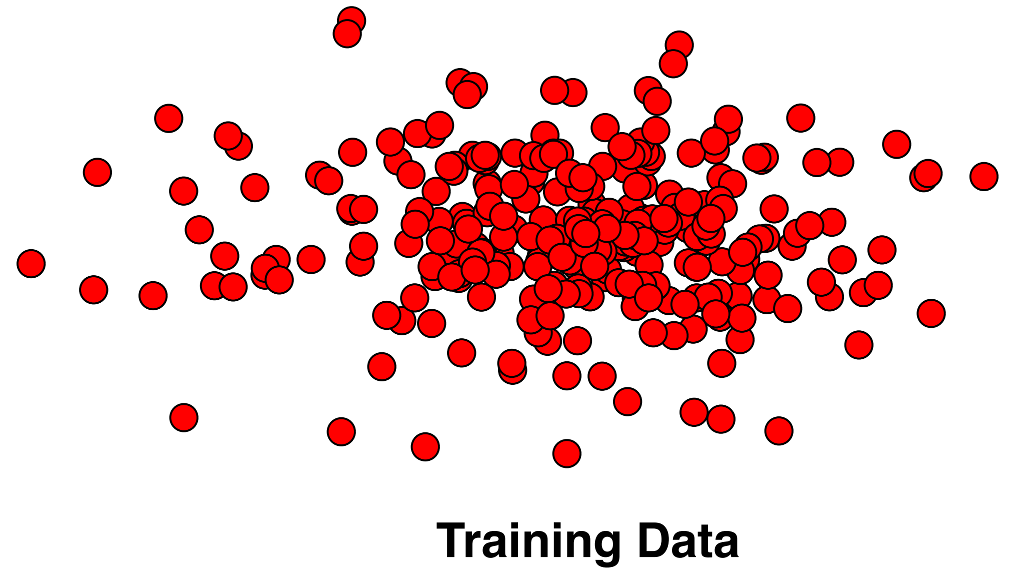
SAA: $\min_{\theta \in \Theta} \ \mathbb{E}_{\widehat{\mathbb{P}}} \left[\ell(\theta, \xi) \right]$

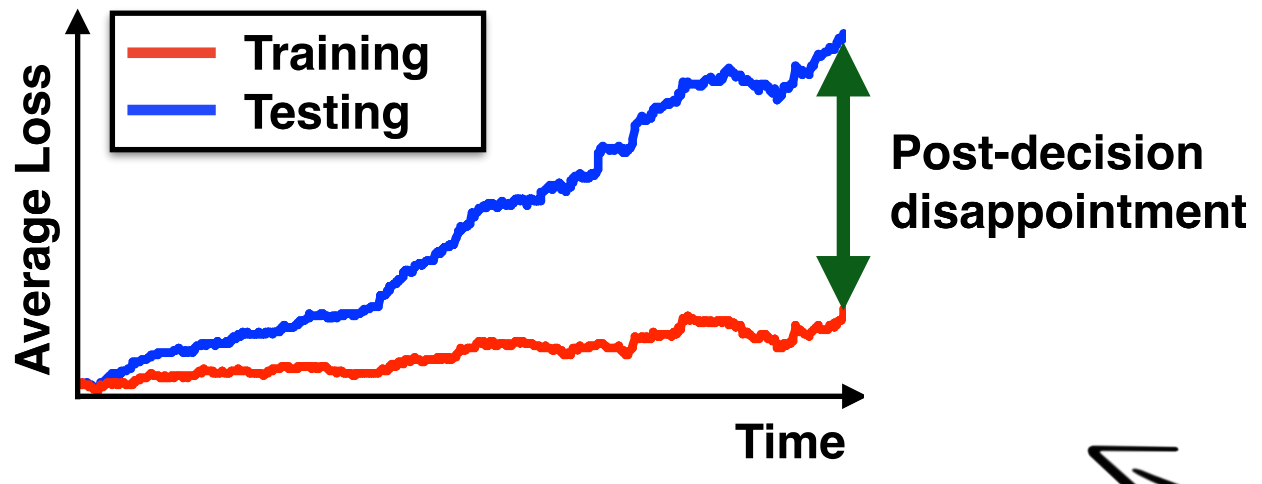
Data-Driven Decision-Making

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$

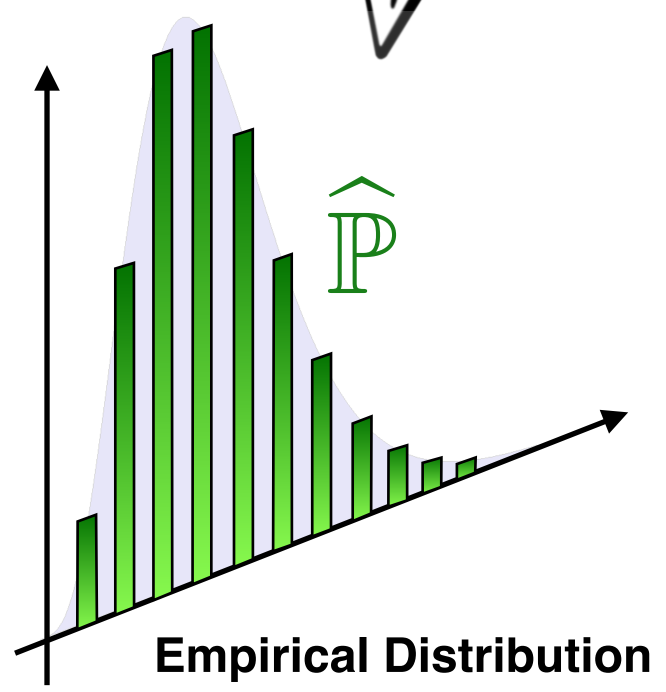






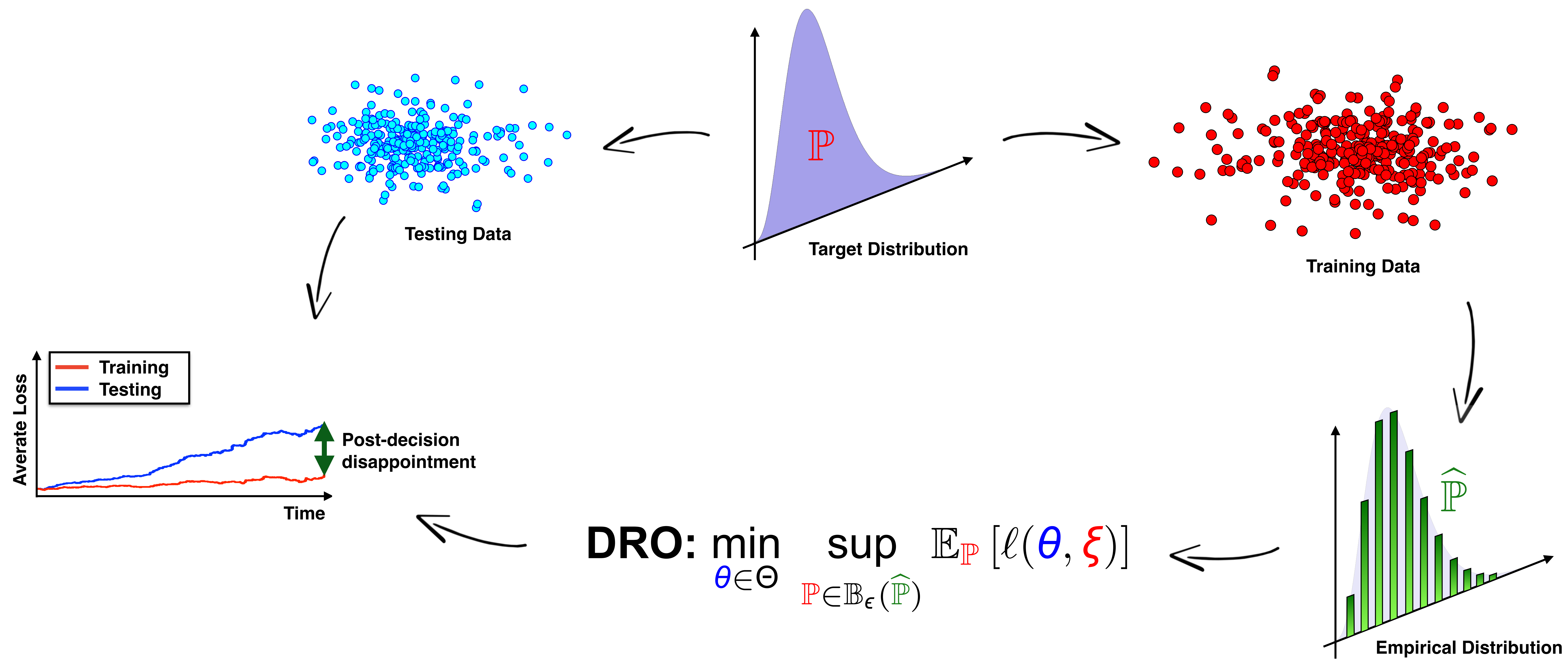


DRO: $\min_{\theta \in \Theta} \sup_{\mathbb{P} \in \mathbb{B}_\epsilon(\widehat{\mathbb{P}})} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



Data-Driven Decision-Making

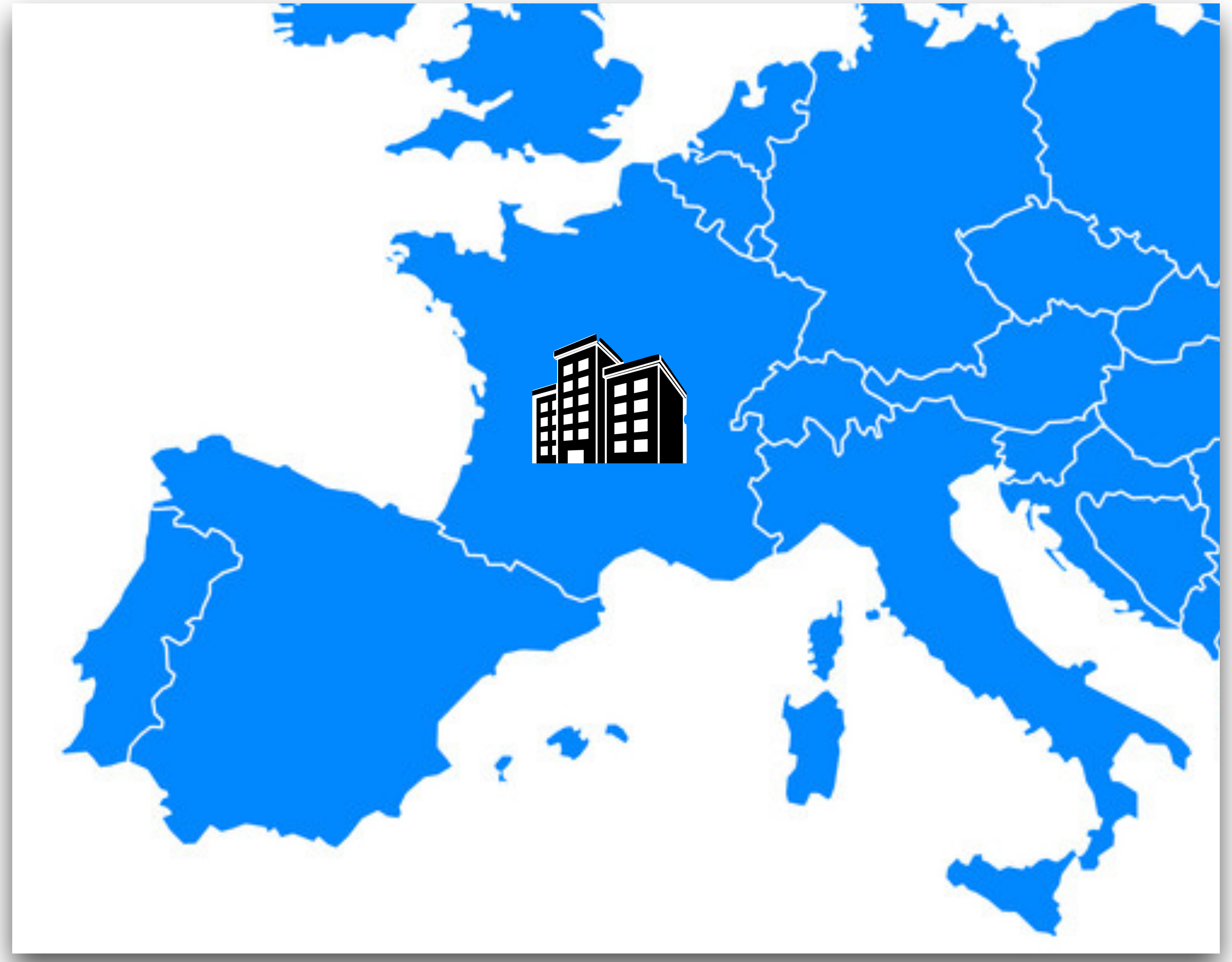
Stochastic Program: $$\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$$



DRO: $$\min_{\theta \in \Theta} \sup_{\mathbb{P} \in \mathbb{B}_{\epsilon}(\widehat{\mathbb{P}})} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$$

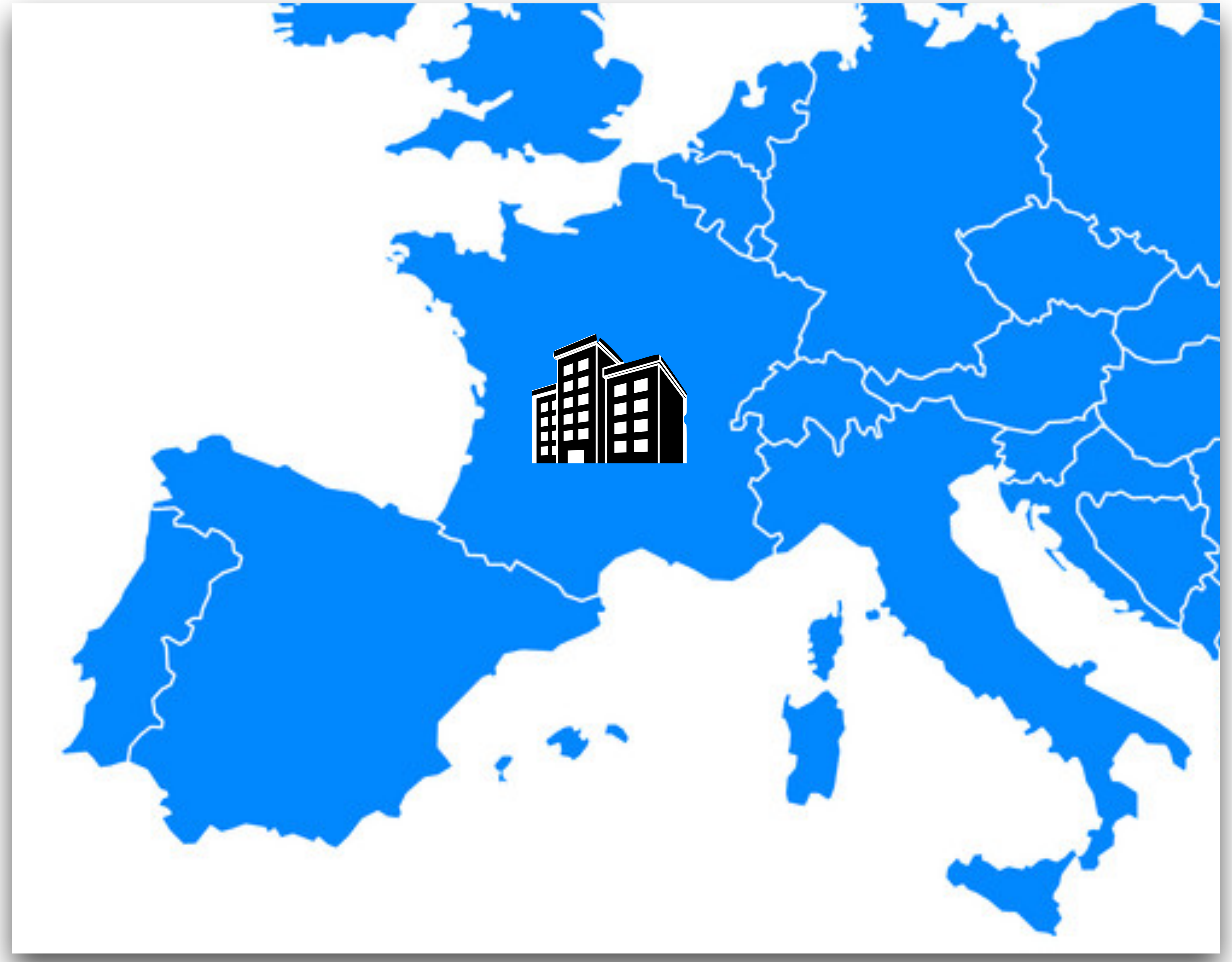
Often, There is No Data!

Expanding into new markets:



Often, There is No Data!

Expanding into new markets:



No historical data on consumer behavior!

Often, There is No Data!

Launching new products:

Smartwatch

Often, There is No Data!

Launching new products:

Smartwatch

No historical
sales data!

Often, There is No Data!

Trading in new securities:

Google Share

Often, There is No Data!

Trading in new securities:

Google Share

No historical return data!

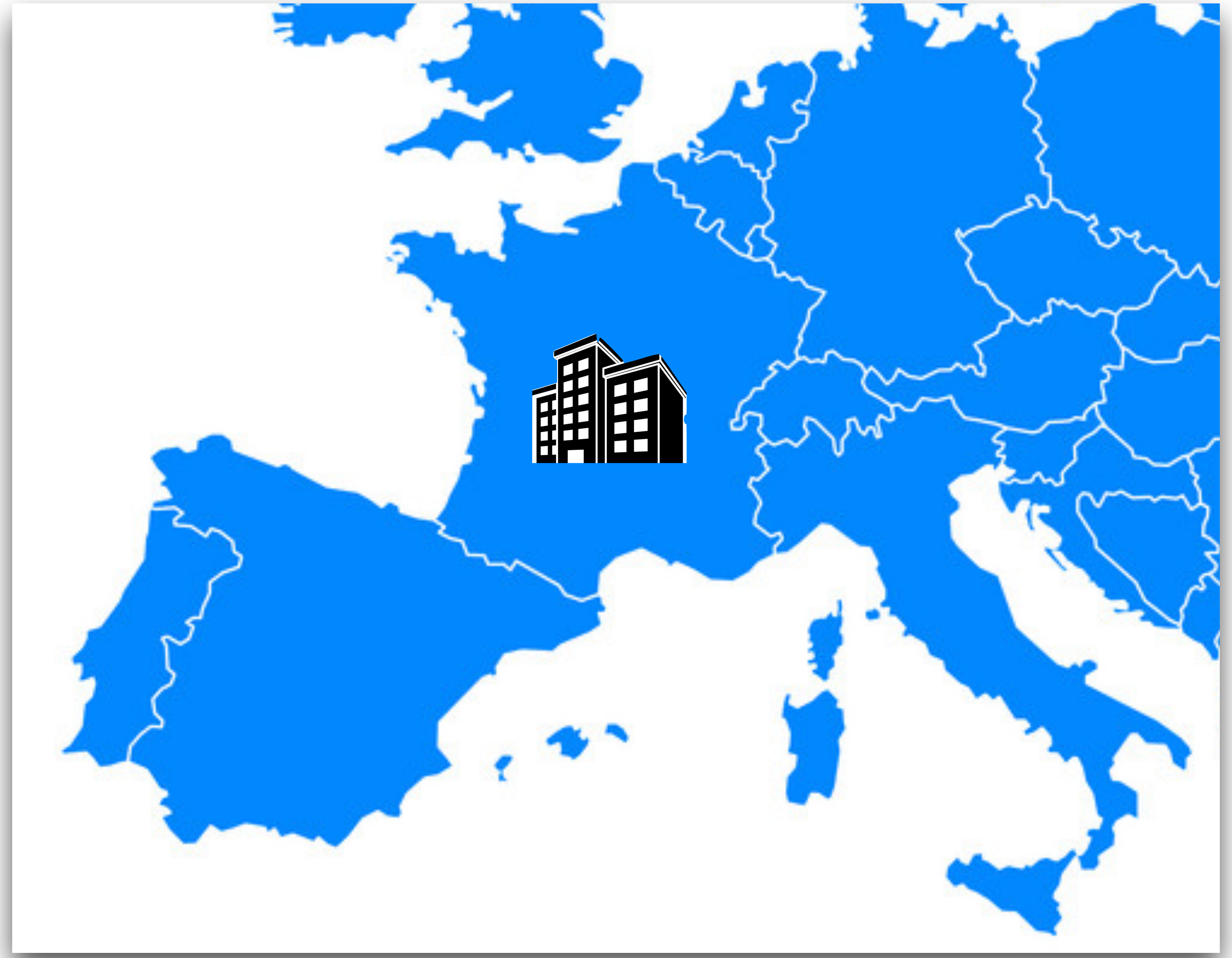
Often, There is No Data!

Trading in new securities:

Structural disruptions / strategic changes $\Longrightarrow$ lack of relevant data

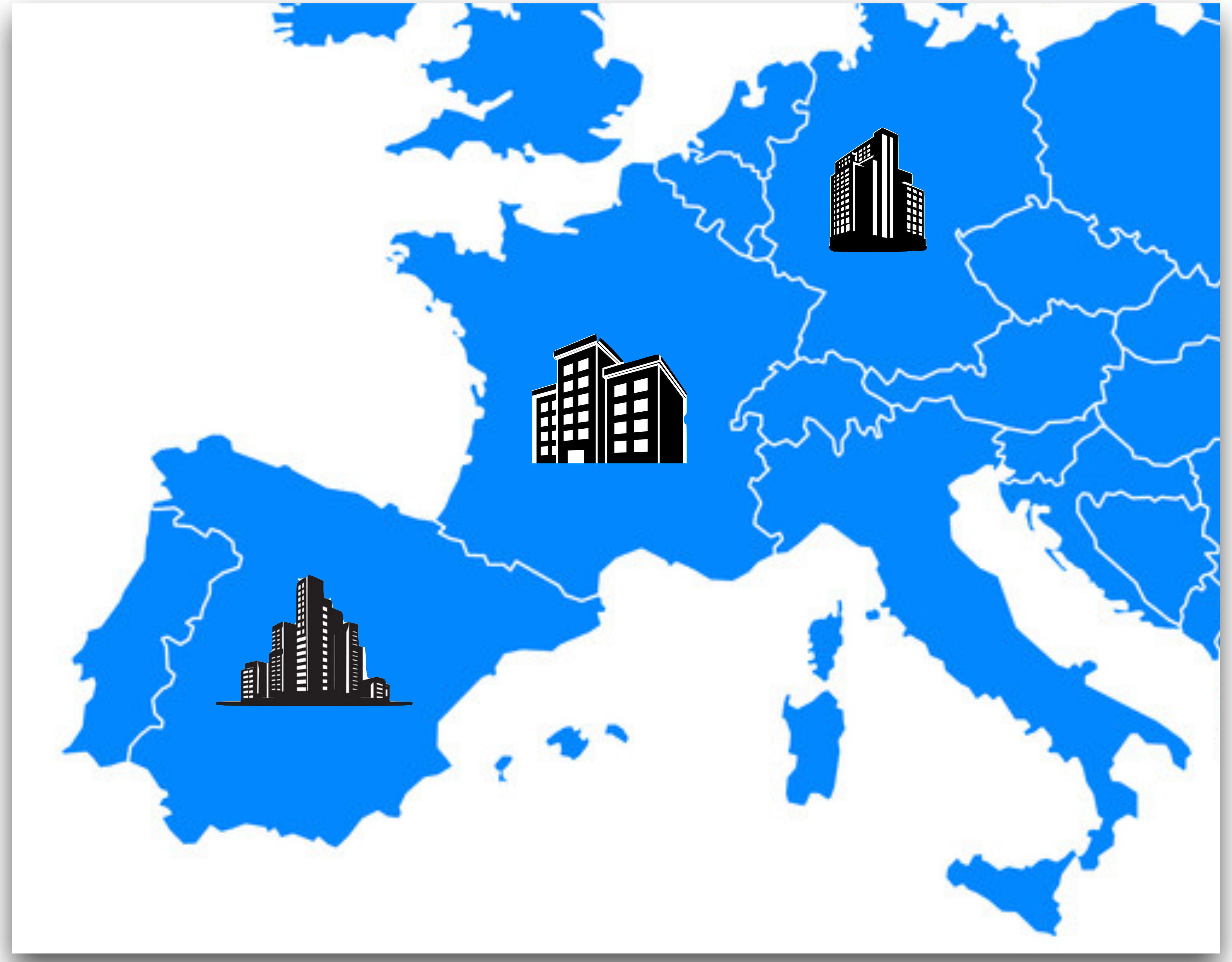
Often, There is No Data!

Expanding into new markets:



Often, There is No Data!

Expanding into new markets:



Demand data from markets with similar customer structure

Often, There is No Data!

Launching new products:

Often, There is No Data!

Launching new products:

Smartphone

Smartwatch

Traditional Watch

Sales data for similar <u>existing</u> products

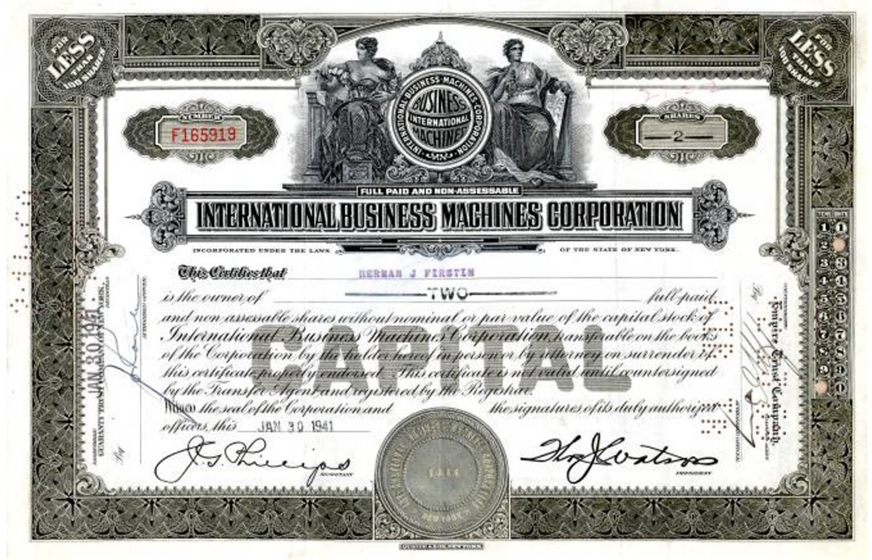
Often, There is No Data!

Trading in new securities:

Google Share

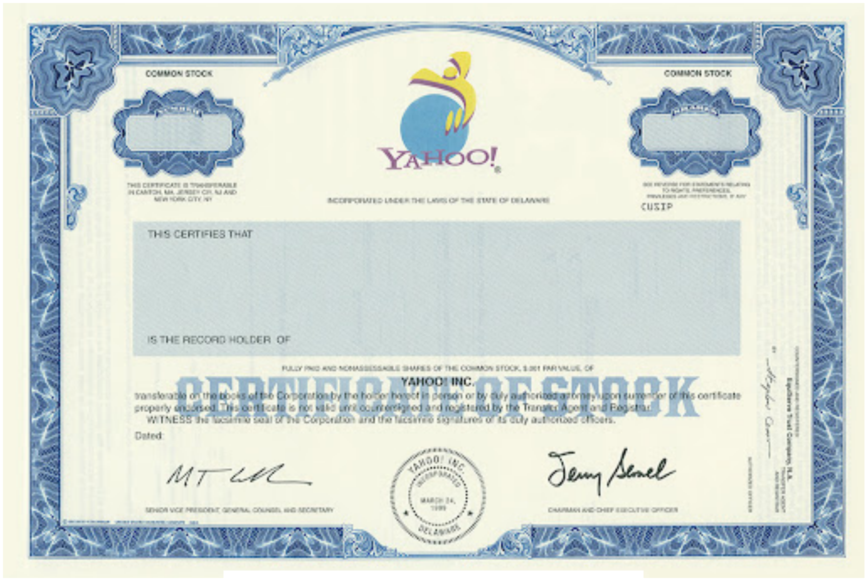
Often, There is No Data!

Trading in new securities:



IBM Share

Google Share

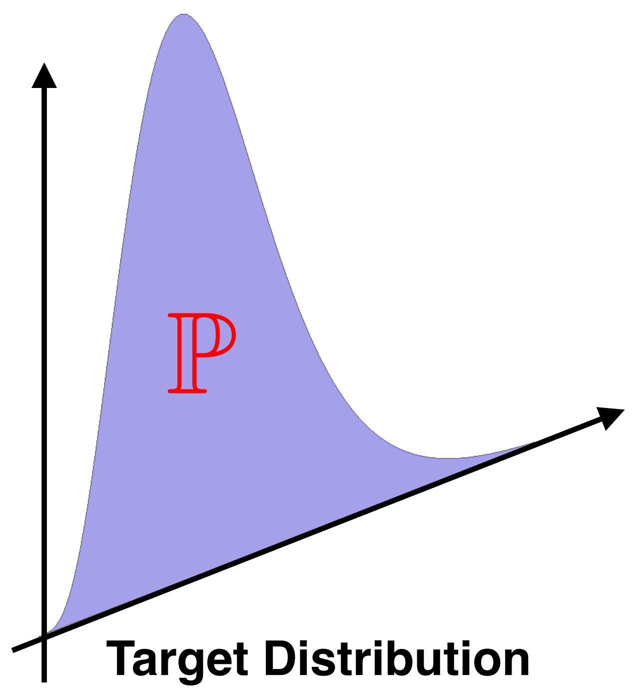


Yahoo Share

Return data for <u>existing</u> securities
issued by similar companies

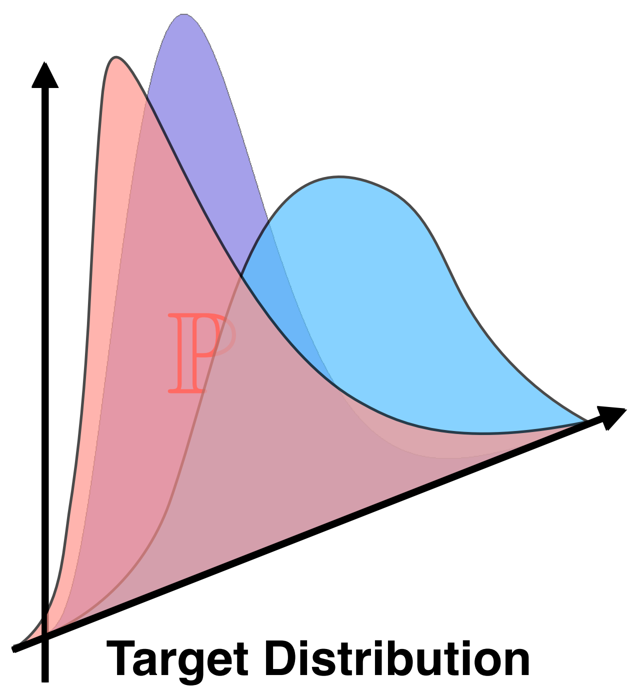
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}}\left[\ell(\theta, \xi)\right]$



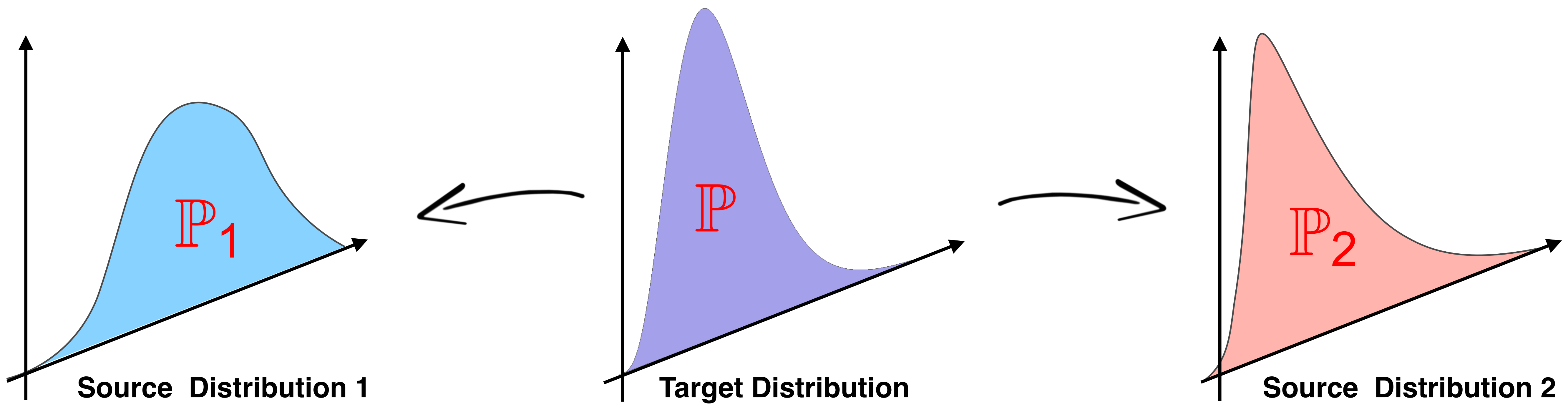
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$



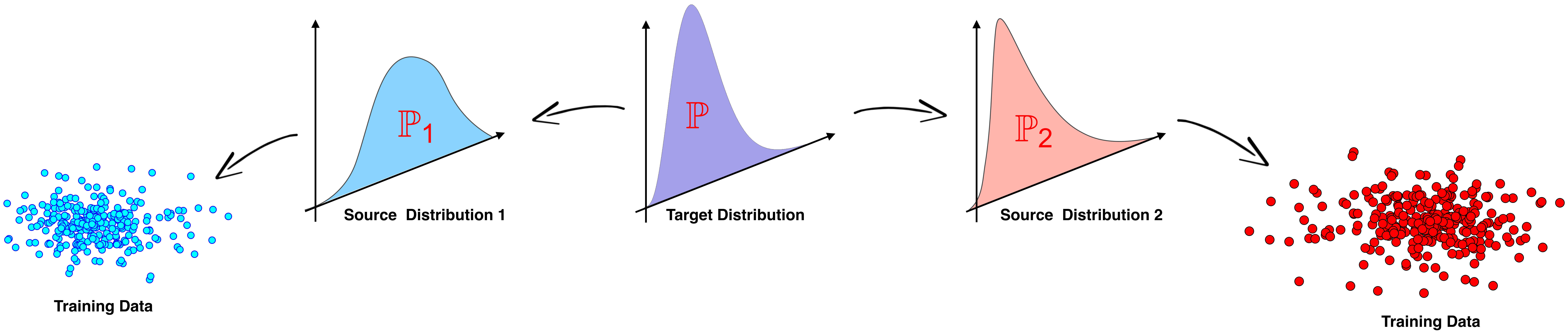
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



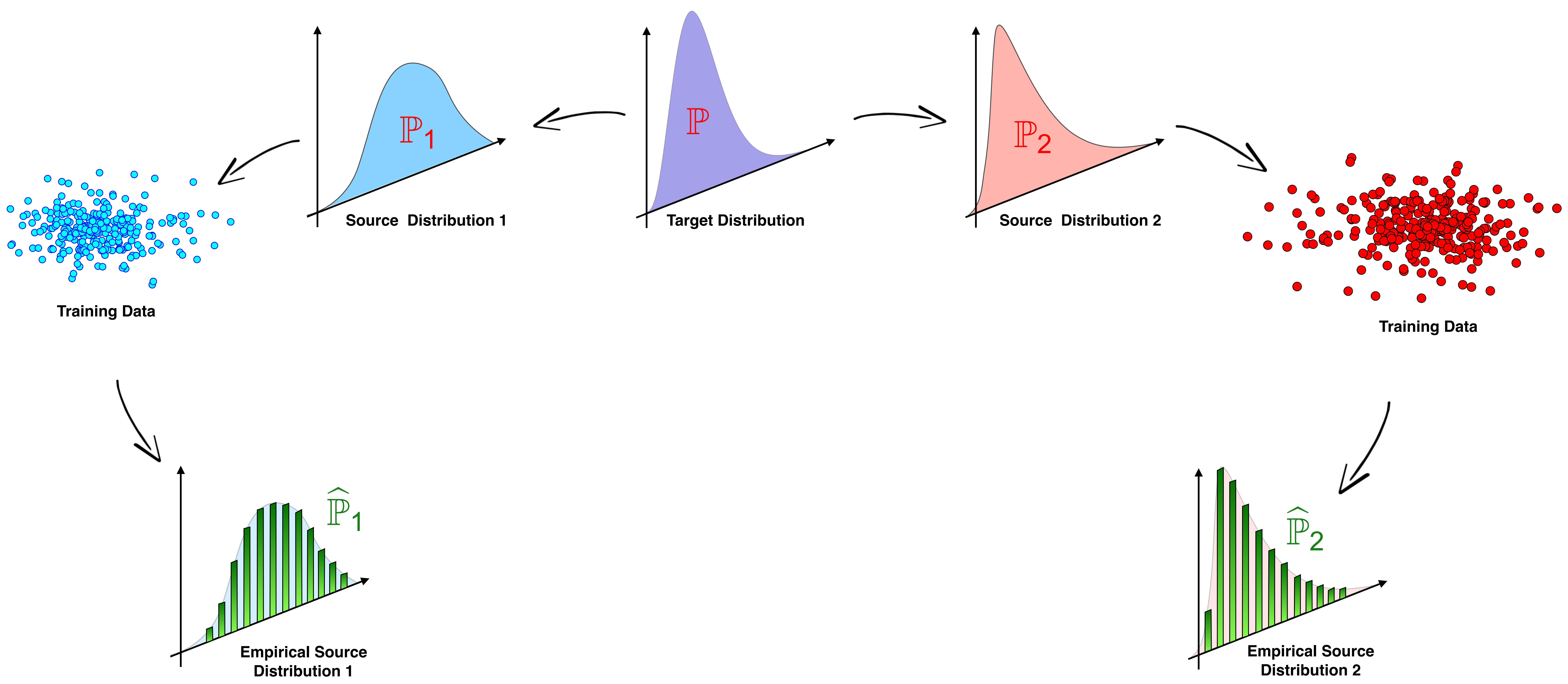
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} \left[\ell(\theta, \xi) \right]$



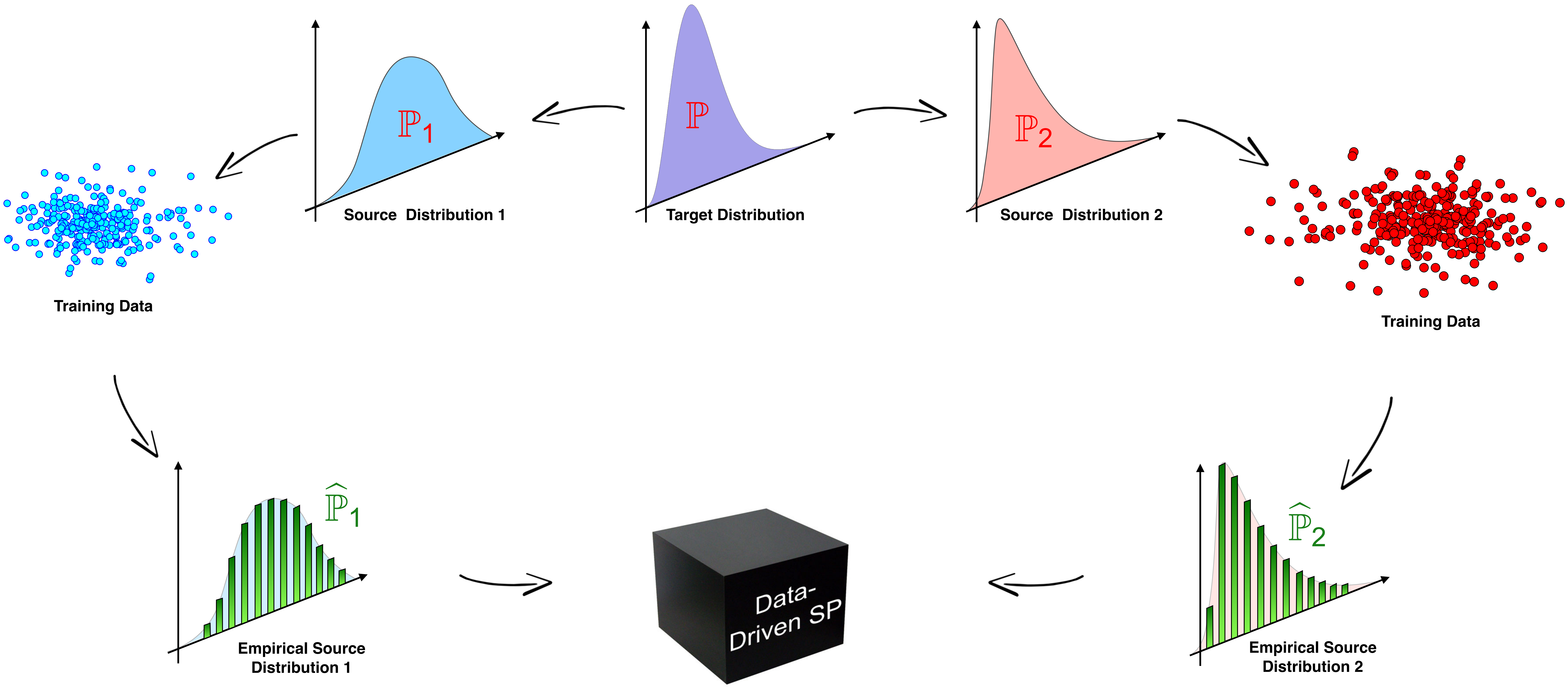
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



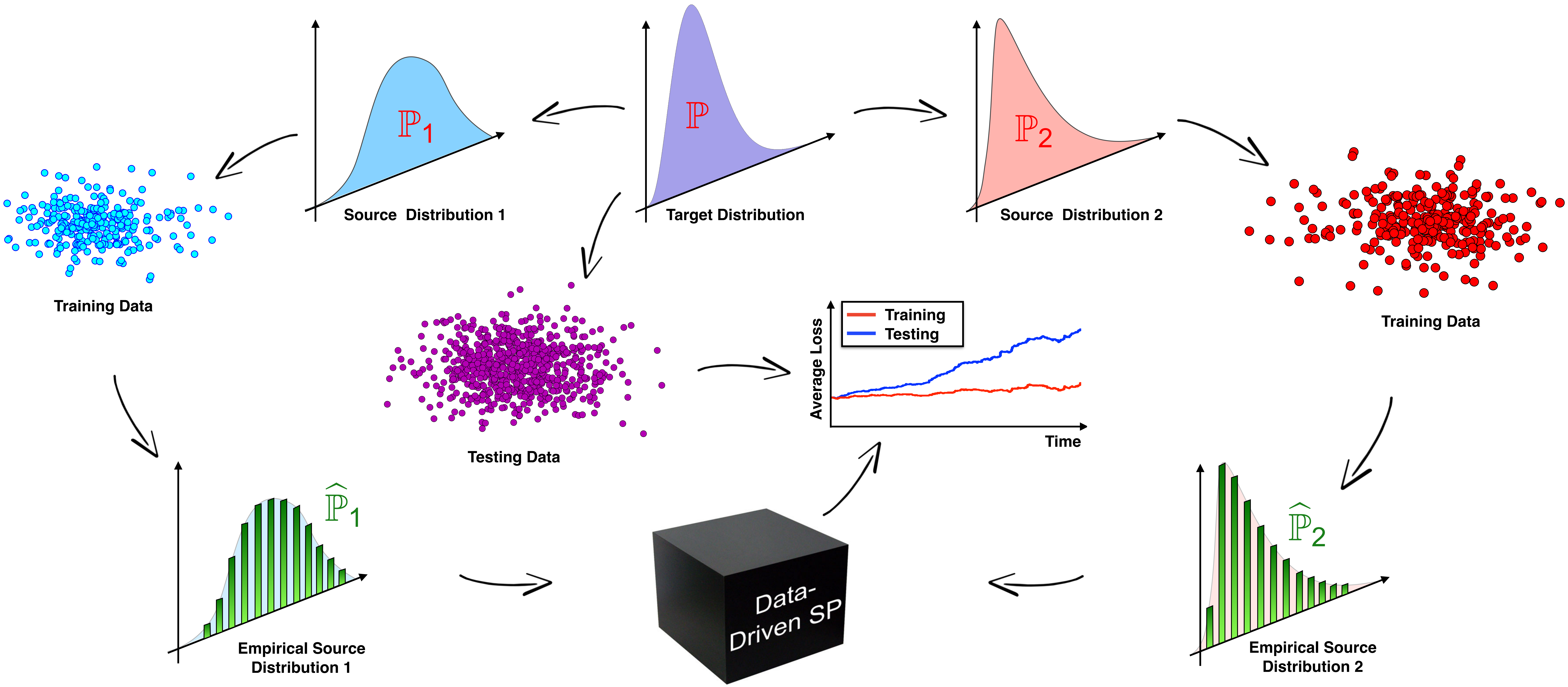
Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



Decision-Making with Multiple Data Sources

Stochastic Program: $\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$



What is in the Black Box?

DRO model: $\min_{\theta \in \Theta} \sup_{\mathbb{P} \in \mathbb{B}_\epsilon(\widehat{\mathbb{P}})} \mathbb{E}_{\mathbb{P}} [\ell(\theta, \xi)]$

Possible choices for $\widehat{\mathbb{P}}$:

- empirical distribution on target data
- empirical distribution on source data
- empirical distribution on pooled data
- barycenter of empirical source distributions[1)]

[1)] Lau & Liu, *arXiv*, 2022.

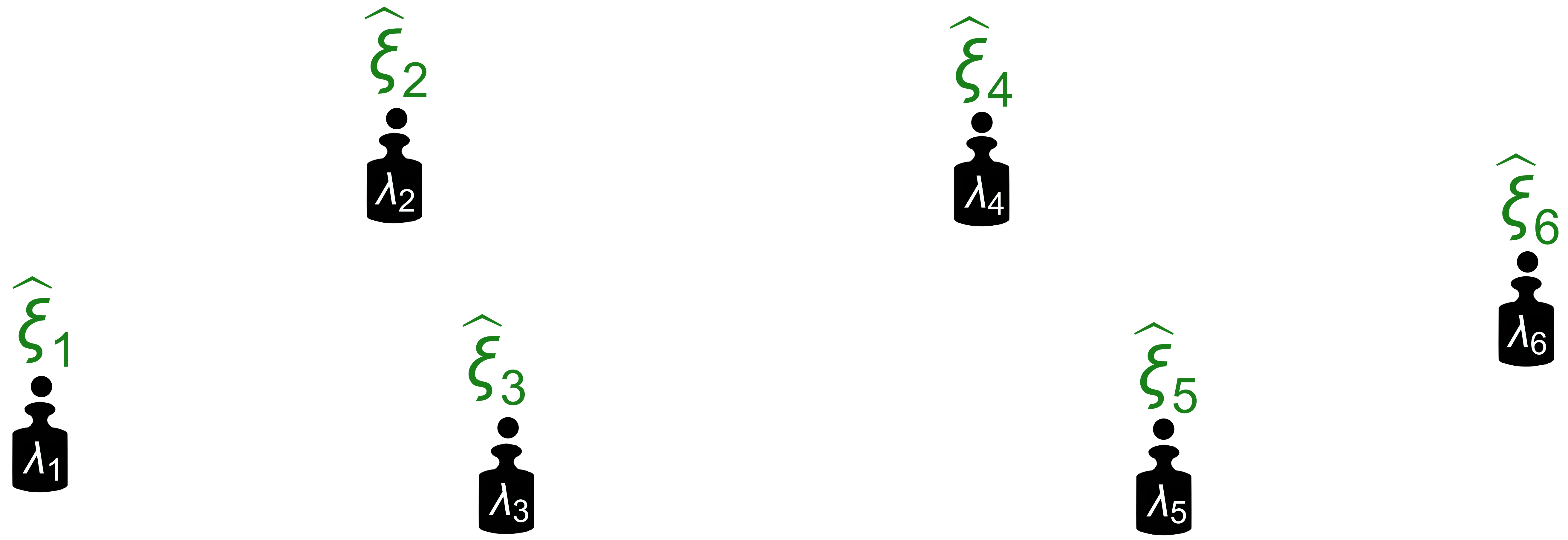
Wasserstein Barycenters

Averaging Points in $\mathbb{R}^d$

Points in $\mathbb{R}^d$:

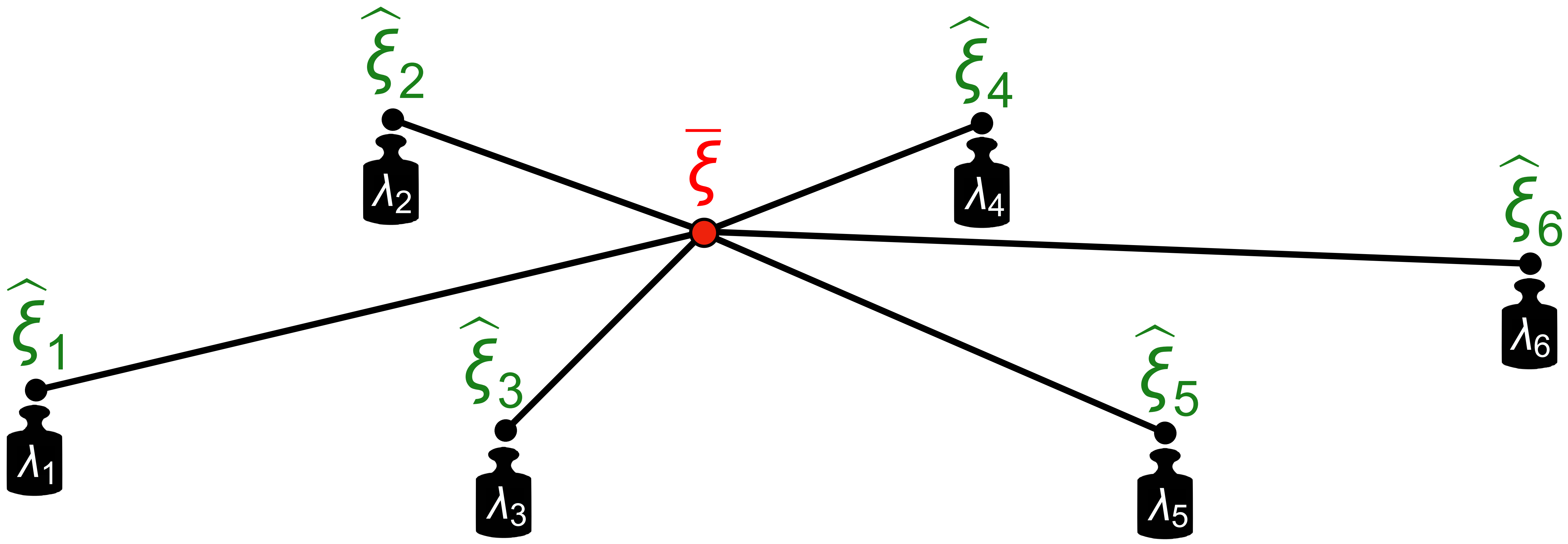
Averaging Points in $\mathbb{R}^d$

Weights in $\mathbb{R}_+$:



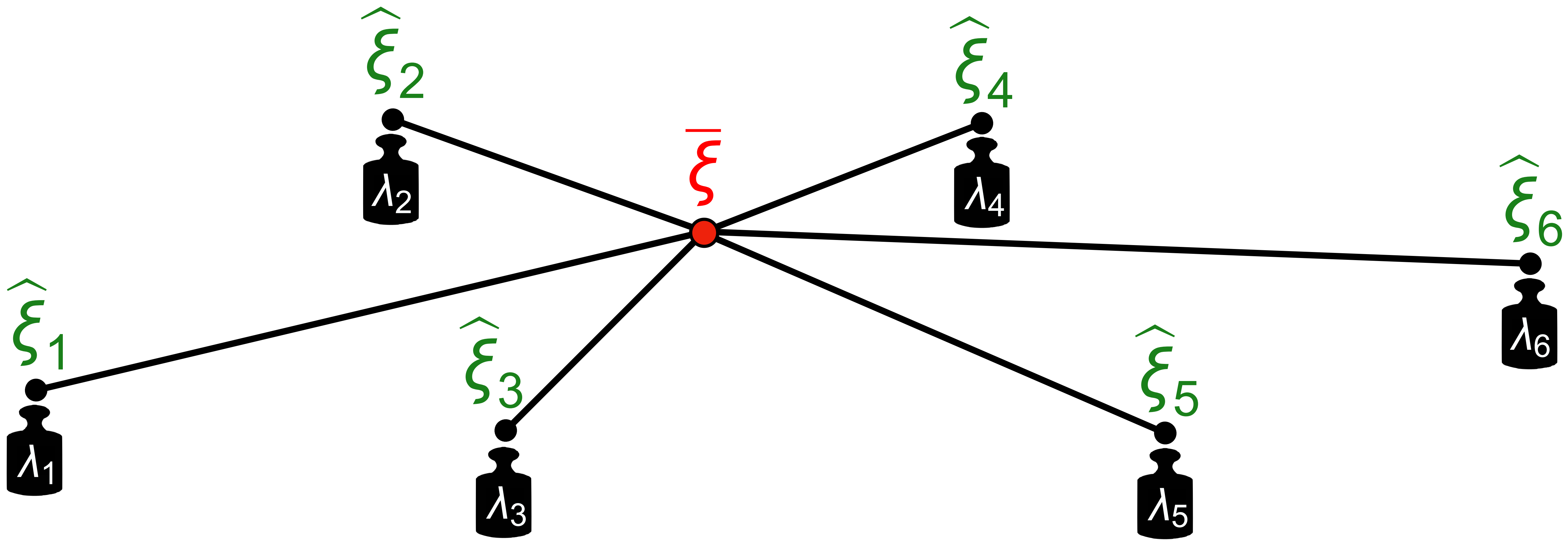
Averaging Points in $\mathbb{R}^d$

Barycenter: $\overline{\xi} = \sum_{k=1}^{K} \lambda_k \widehat{\xi}_k$



Averaging Points in $\mathbb{R}^d$

Barycenter: $\overline{\xi} = \sum_{k=1}^{K} \lambda_k \widehat{\xi}_k$

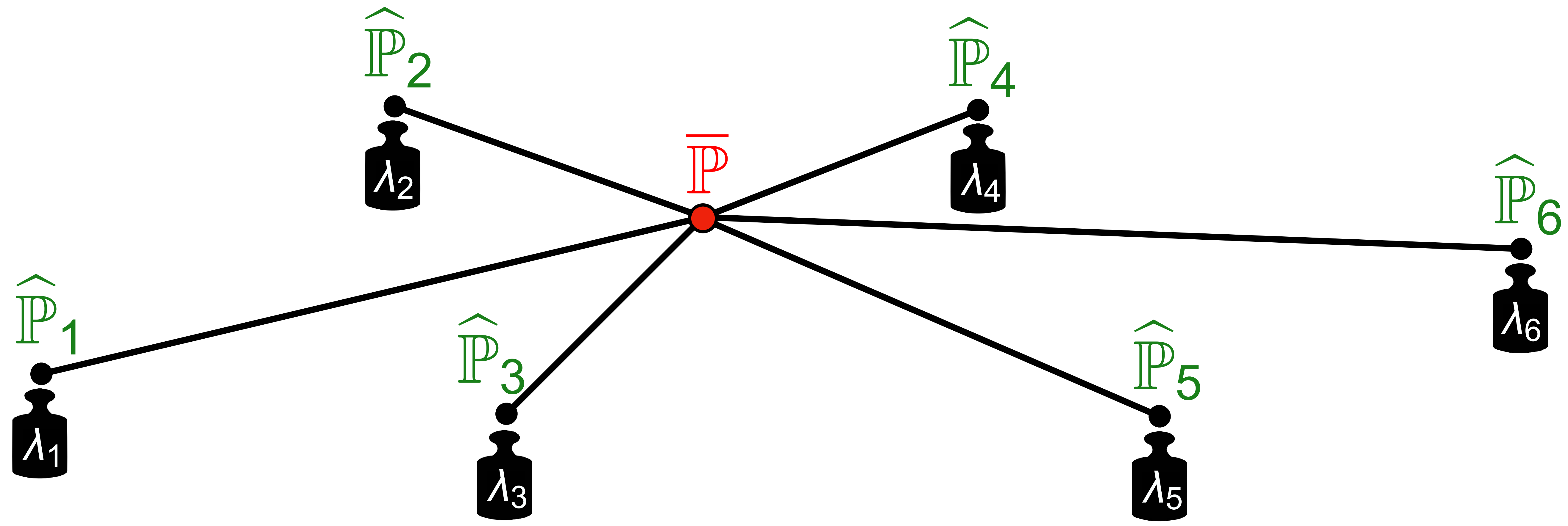


$\overline{\xi}$ can be viewed as solution of convex program

$$\min_{\xi \in \mathbb{R}^d} \sum_{k=1}^{K} \lambda_k \|\xi - \widehat{\xi}_k\|_2^2$$

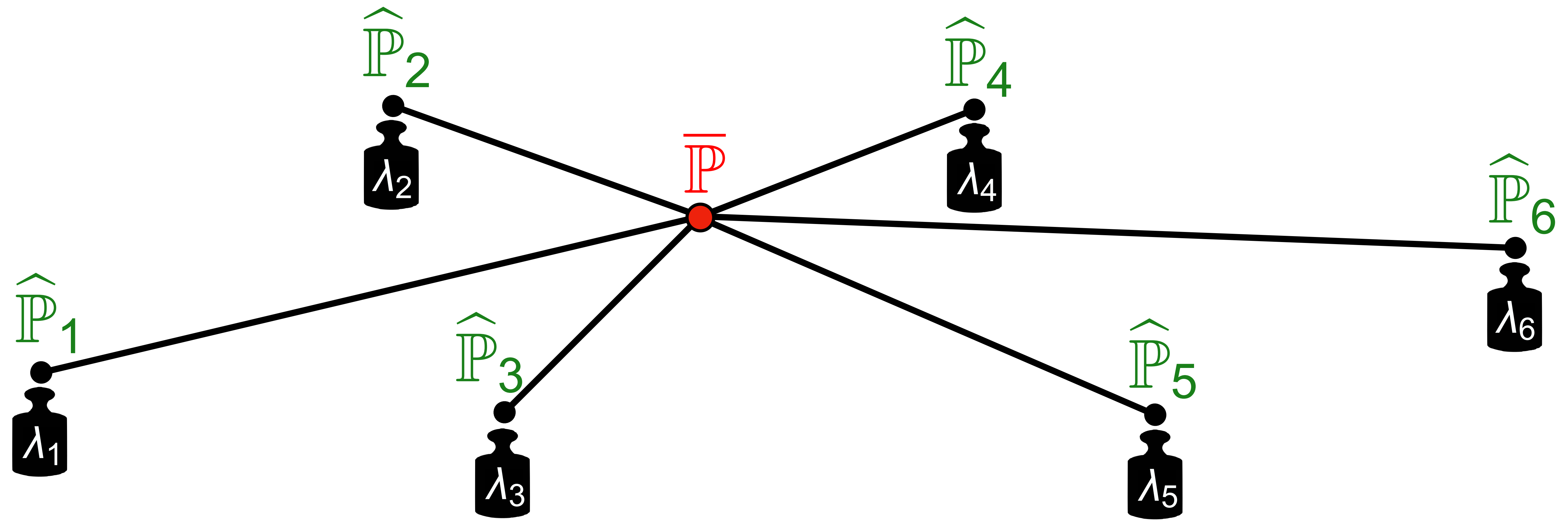
Averaging Probability Distributions $\mathcal{P}(\mathbb{R}^d)$

Barycenter: $\overline{\mathbb{P}} = ?$



Averaging Probability Distributions $\mathcal{P}(\mathbb{R}^d)$

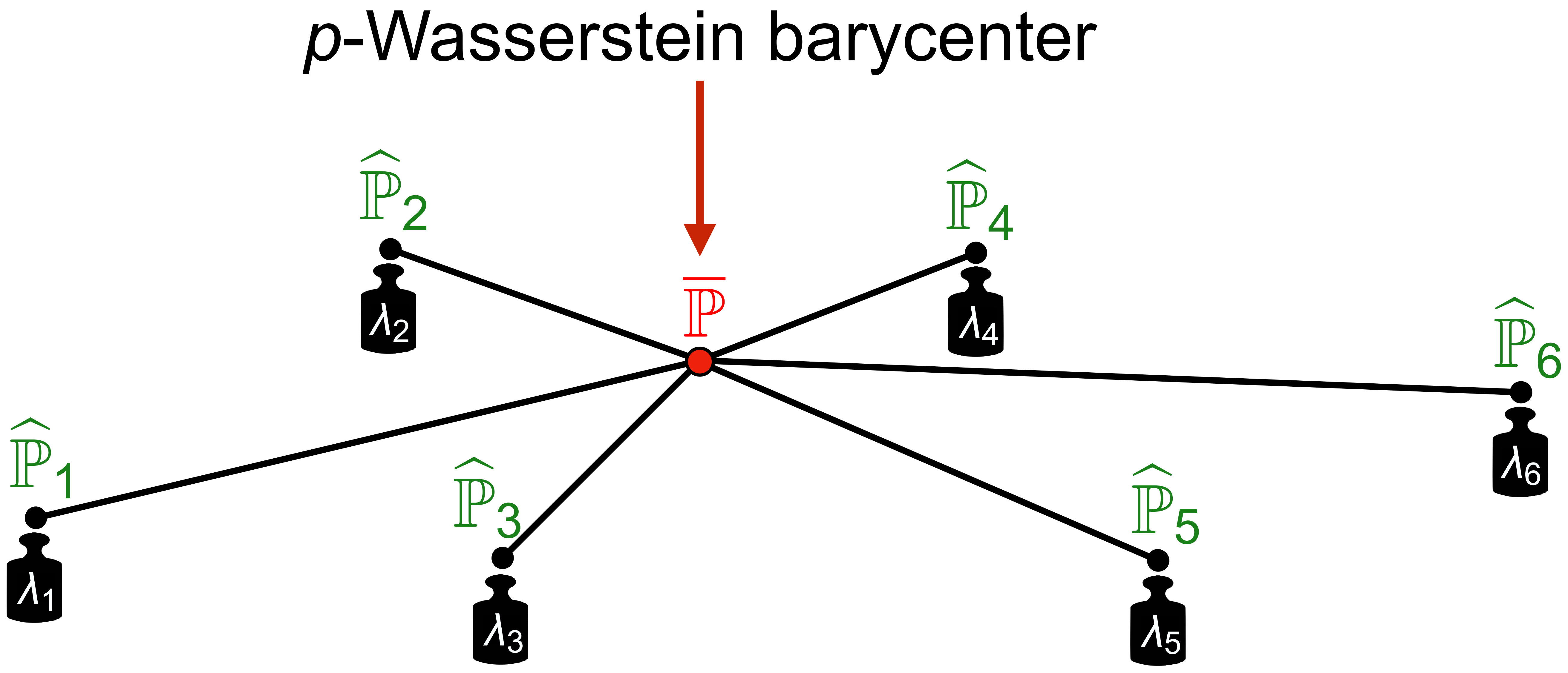
Barycenter: $\overline{\mathbb{P}} = ?$



$\overline{\mathbb{P}}$ can be **defined** as solution of convex program $\min_{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d)} \sum_{k=1}^{K} \lambda_k W_2(\mathbb{P}, \widehat{\mathbb{P}}_k)^2$

Averaging Probability Distributions $\mathcal{P}(\mathbb{R}^d)$

Barycenter: $\overline{\mathbb{P}} = ?$



$\overline{\mathbb{P}}$ can be **defined** as solution of convex program

$$\min_{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d)} \sum_{k=1}^{K} \lambda_k \underbrace{W_p(\mathbb{P}, \widehat{\mathbb{P}}_k)}_{} {}^p$$

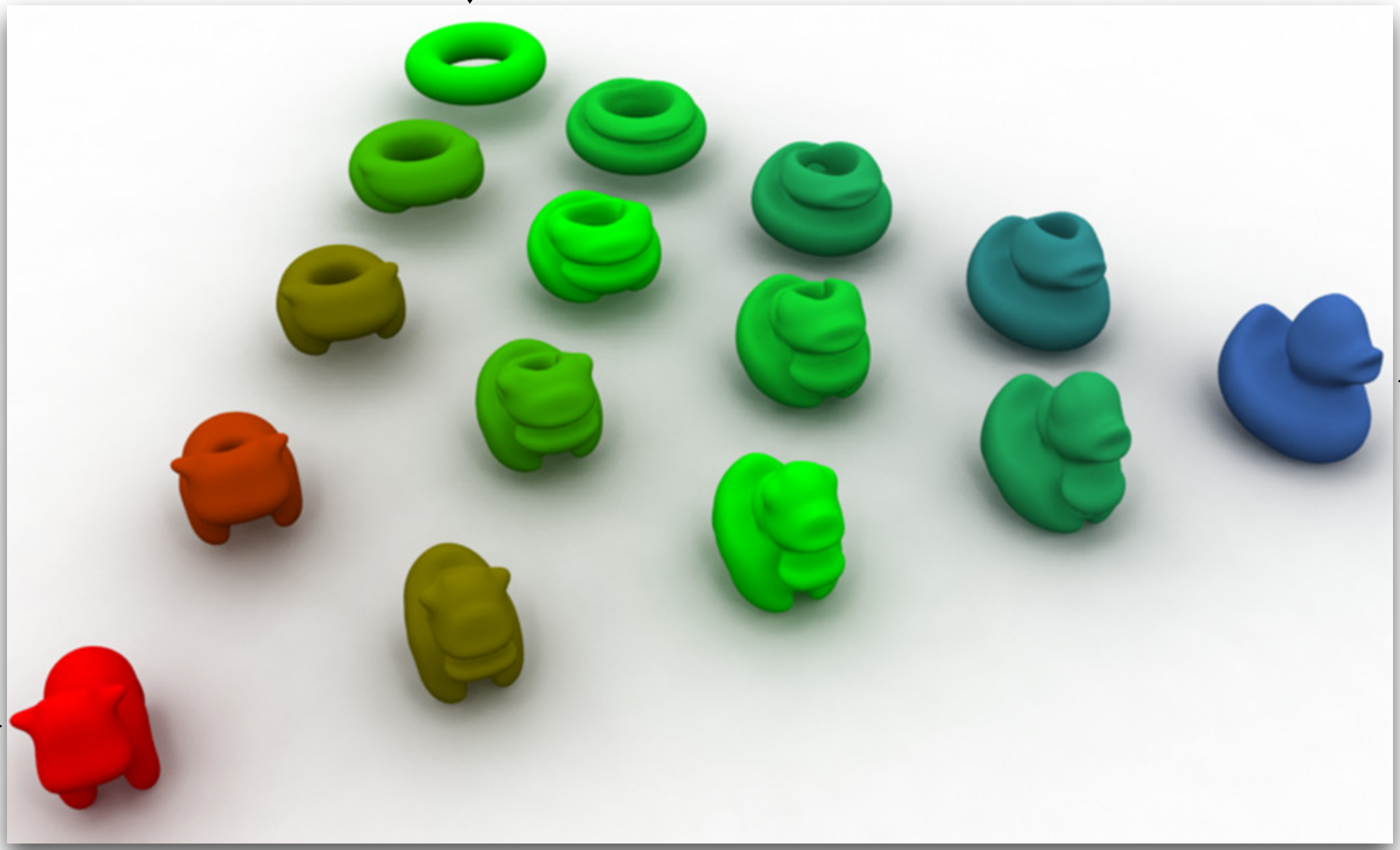
p-Wasserstein distance

2-Wasserstein Barycenters[1)]

$\mathbb{P}_1 \sim$ uniform on donut

$\mathbb{P}_2 \sim$ uniform on cow

$\mathbb{P}_2 \sim$ uniform on duck



1) Solomon et al., *ACM Trans. Graph.*, 2015.

Aggregating Distributions

Source distributions: $\mathbb{P}_k \sim \mathcal{N}(\mu_k, \sigma^2), \quad \lambda_k \geq 0, \quad k \in [K], \quad \sum_{k=1}^{K} \lambda_k = 1$

Aggregating Distributions

Source distributions: $\mathbb{P}_k \sim \mathcal{N}(\mu_k, \sigma^2), \quad \lambda_k \geq 0, \quad k \in [K], \quad \sum_{k=1}^{K} \lambda_k = 1$

	$\overline{\mathbb{P}} = \sum_{k=1}^{K} \lambda_k \cdot \mathbb{P}_k$	$\overline{\mathbb{P}}$ = 2-Wasserstein barycenter
Shape		
Mean		
Variance		

Aggregating Distributions

Source distributions: $\mathbb{P}_k \sim \mathcal{N}(\mu_k, \sigma^2), \quad \lambda_k \geq 0, \quad k \in [K], \quad \sum_{k=1}^{K} \lambda_k = 1$

	$\overline{\mathbb{P}} = \sum_{k=1}^{K} \lambda_k \cdot \mathbb{P}_k$	$\overline{\mathbb{P}}$ = 2-Wasserstein barycenter
Shape	mutimodal	normal
Mean		
Variance		

Aggregating Distributions

Source distributions: $\mathbb{P}_k \sim \mathcal{N}(\mu_k, \sigma^2), \quad \lambda_k \geq 0, \quad k \in [K], \quad \sum_{k=1}^{K} \lambda_k = 1$

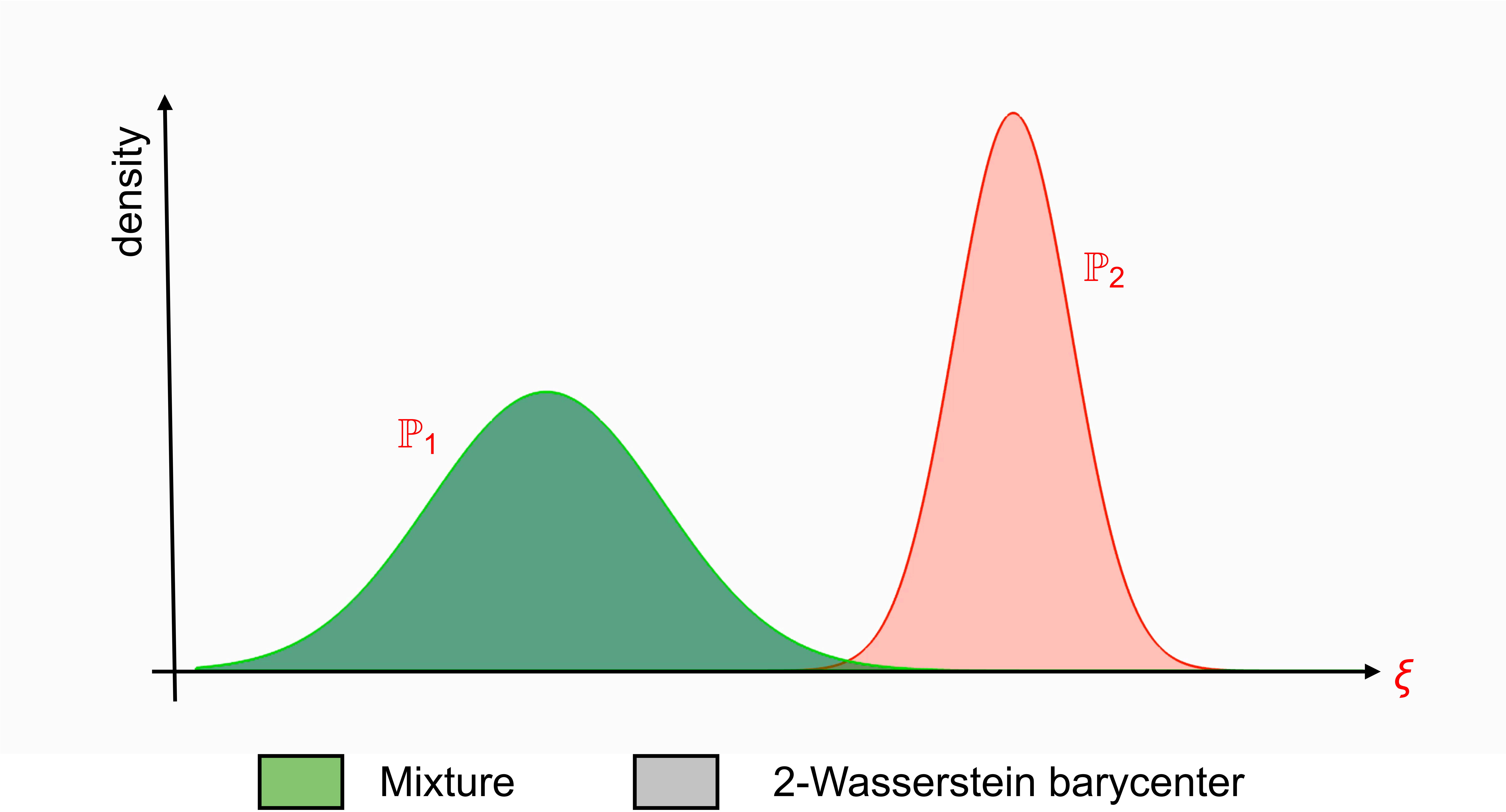
	$\overline{\mathbb{P}} = \sum_{k=1}^{K} \lambda_k \cdot \mathbb{P}_k$	$\overline{\mathbb{P}}$ = 2-Wasserstein barycenter
Shape	mutimodal	normal
Mean	$\overline{\mu} = \sum_{k=1}^{K} \lambda_k \cdot \mu_k$	$\overline{\mu} = \sum_{k=1}^{K} \lambda_k \cdot \mu_k$
Variance		

Aggregating Distributions

Source distributions: $\mathbb{P}_k \sim \mathcal{N}(\mu_k, \sigma^2), \quad \lambda_k \geq 0, \quad k \in [K], \quad \sum_{k=1}^{K} \lambda_k = 1$

	$\overline{\mathbb{P}} = \sum_{k=1}^{K} \lambda_k \cdot \mathbb{P}_k$	$\overline{\mathbb{P}} =$ 2-Wasserstein barycenter
Shape	mutimodal	normal
Mean	$\overline{\mu} = \sum_{k=1}^{K} \lambda_k \cdot \mu_k$	$\overline{\mu} = \sum_{k=1}^{K} \lambda_k \cdot \mu_k$
Variance	$\sigma^2 + \sum_{k=1}^{K} \lambda_k (\mu_k)^2 - (\sum_{k=1}^{K} \lambda_k \mu_k)^2$	σ^2

Aggregating Distributions



Wasserstein Barycenters - Challenges

2-Wasserstein barycenters: Sensitive to perturbations.[1)]

[1)] Zhuang, Chen & Yang, *NeurIPS.*, 2022.

Wasserstein Barycenters - Challenges

2-Wasserstein barycenters: Sensitive to perturbations.[1)]

Theorem: Set $\mathbb{P}^\star$ to the 2-Wasserstein barycenter of $\mathbb{P}_1, \ldots, \mathbb{P}_K$, and set $\widehat{\mathbb{P}}$ to the 2-Wasserstein barycenter of $\widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_K$. Then, we have

$$\mathbb{E}[\widehat{\mathbb{P}}\text{-VAR}(\xi)] < \mathbb{P}^\star\text{-VAR}(\xi).$$

1) Zhuang, Chen & Yang, *NeurIPS.*, 2022.

Wasserstein Barycenters - Challenges

2-Wasserstein barycenters: Sensitive to perturbations.[1)]

Theorem: Set $\mathbb{P}^\star$ to the 2-Wasserstein barycenter of $\mathbb{P}_1, \ldots, \mathbb{P}_K$, and set $\widehat{\mathbb{P}}$ to the 2-Wasserstein barycenter of $\widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_K$. Then, we have

$$\mathbb{E}[\widehat{\mathbb{P}}\text{-VAR}(\xi)] < \mathbb{P}^\star\text{-VAR}(\xi).$$

trace of the covariance matrix

1) Zhuang, Chen & Yang, *NeurIPS.*, 2022.

Wasserstein Barycenters - Challenges

2-Wasserstein barycenters: Sensitive to perturbations.[1)]

Theorem: Set $\mathbb{P}^\star$ to the 2-Wasserstein barycenter of $\mathbb{P}_1, \ldots, \mathbb{P}_K$, and set $\widehat{\mathbb{P}}$ to the 2-Wasserstein barycenter of $\widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_K$. Then, we have

$$\mathbb{E}[\widehat{\mathbb{P}}\text{-VAR}(\xi)] < \mathbb{P}^\star\text{-VAR}(\xi).$$

w.r.t. all training samples

[1)] Zhuang, Chen & Yang, *NeurIPS.*, 2022.

Wasserstein Barycenters - Challenges

2-Wasserstein barycenters: Sensitive to perturbations.[1)]

Theorem: Set $\mathbb{P}^\star$ to the 2-Wasserstein barycenter of $\mathbb{P}_1, \ldots, \mathbb{P}_K$, and set $\widehat{\mathbb{P}}$ to the 2-Wasserstein barycenter of $\widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_K$. Then, we have

$$\mathbb{E}[\widehat{\mathbb{P}}\text{-VAR}(\xi)] < \mathbb{P}^\star\text{-VAR}(\xi).$$

1-Wasserstein barycenters:

Theorem: A 1-Wasserstein barycenter of $\mathbb{P}_1$ and $\mathbb{P}_2$ with weights λ_1 and λ_2 is given by $\mathbb{P}_1$ if $\lambda_1 \geq \lambda_2$ and by $\mathbb{P}_2$ if $\lambda_1 \leq \lambda_2$.

1) Zhuang, Chen & Yang, *NeurIPS.*, 2022.

Decision-Making with Multiple Data Sources

DRO model: $\min_{\theta\in\Theta} \sup_{\mathbb{P}\in\mathbb{B}_\epsilon(\widehat{\mathbb{P}})} \mathbb{E}_{\mathbb{P}}\left[\ell(\theta,\xi)\right]$

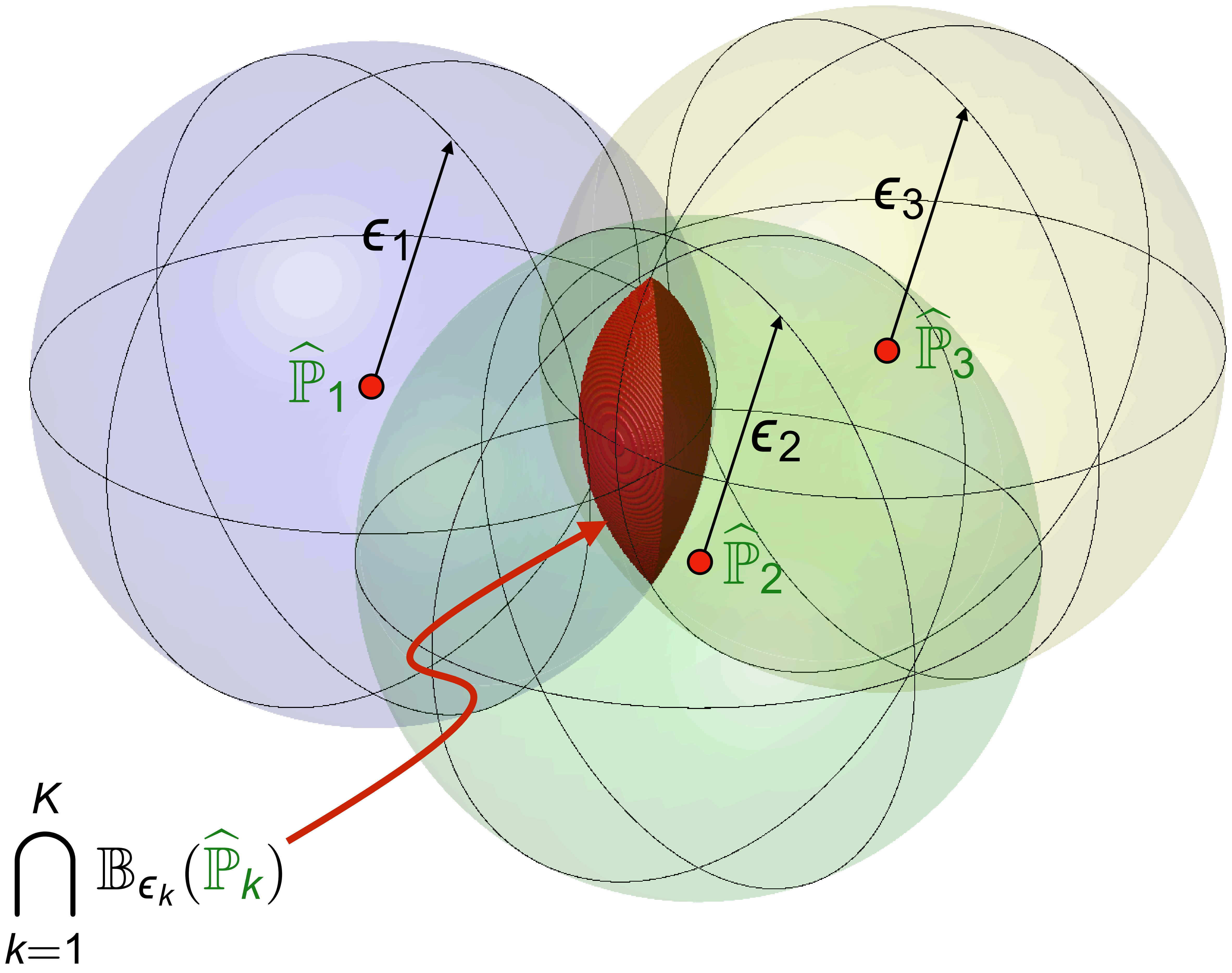
Possible choices for $\widehat{\mathbb{P}}$:

- empirical distribution on target data $\Longrightarrow$ high variance
- empirical distribution on source data $\Longrightarrow$ high bias
- empirical distribution on pooled data $\Longrightarrow$ high bias / misses stylized features
- barycenter of empirical source distributions[1)] $\Longrightarrow$ sensitive to perturbations / biased

1) Lau & Liu, *arXiv*, 2022.

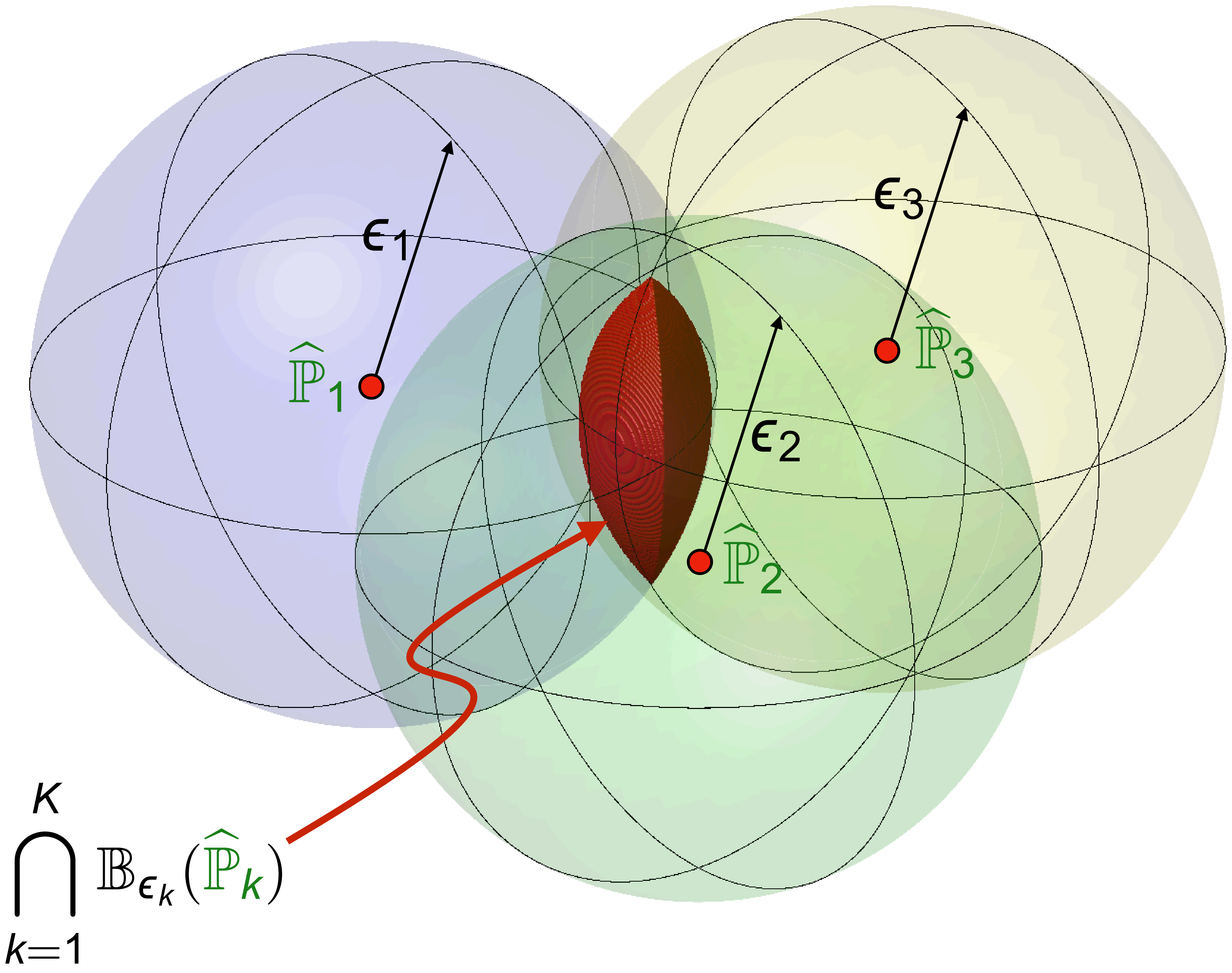
Multi-Source DRO

Multi-Source DRO



$$\min_{\theta \in \Theta} \sup_{\mathbb{P} \in \bigcap_{k=1}^{K} \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k)} \mathbb{E}_{\mathbb{P}}[\ell(\theta, \xi)]$$

Multi-Source DRO



$$\min_{\theta \in \Theta} \sup_{\mathbb{P} \in \bigcap_{k=1}^{K} \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k)} \mathbb{E}_{\mathbb{P}}[\ell(\theta, \xi)]$$

Special cases (all for $K = 2$):

- Linear regression [1)]
- Logistic regression [2)]
- Intersection of a Wasserstein ball and a GoF-based ambiguity set [3)]

---

1) Taskesen, Yue, Blanchet, Kuhn & Nguyen, *ICML*, 2021.
2) Awasthi, Jung & Morgenstern, *arXiv*, 2022; Selvi et al., *arXiv*, 2024.
3) Tanoumand, Bodur & Naoum-Sawaya, *Optimization Online*, 2023.

Nature's Subproblem

$$\sup_{\mathbb{P} \in \bigcap_{k=1}^{K} \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k)} \mathbb{E}_{\mathbb{P}} \left[\ell(\xi) \right]$$

Nature's Subproblem

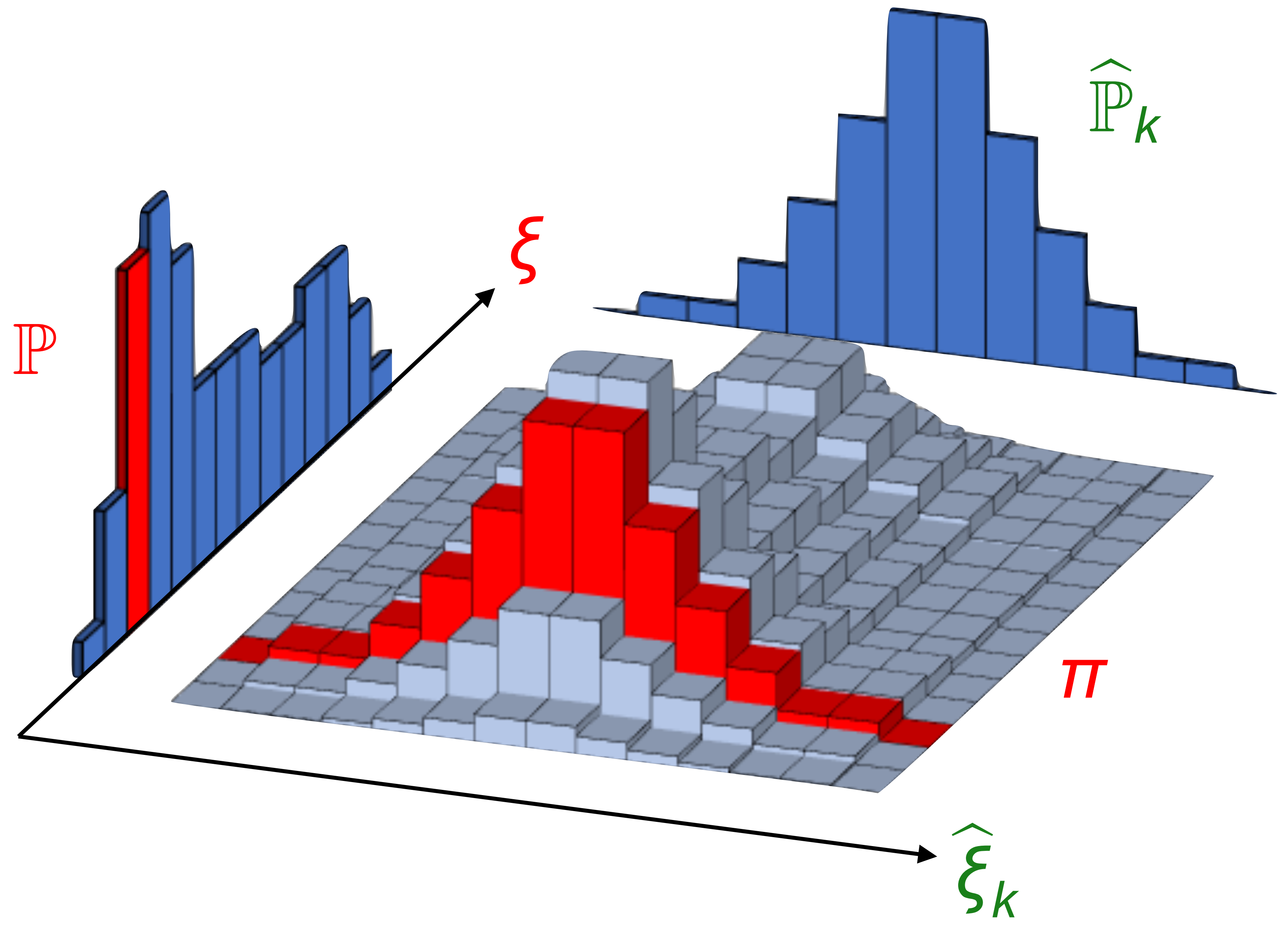
$$\begin{aligned} \sup \quad & \mathbb{E}_{\mathbb{P}}[\ell(\xi)] \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d) \\ & C(\mathbb{P}, \widehat{\mathbb{P}}_k) \leq \epsilon_k \quad \forall k \in [K] \end{aligned}$$

Nature's Subproblem

$$\sup \quad \mathbb{E}_{\mathbb{P}}[\ell(\xi)]$$

$$\text{s.t.} \quad \mathbb{P} \in \mathcal{P}(\mathbb{R}^d)$$

$$C(\mathbb{P}, \widehat{\mathbb{P}}_k) \leq \epsilon_k \quad \forall k \in [K]$$



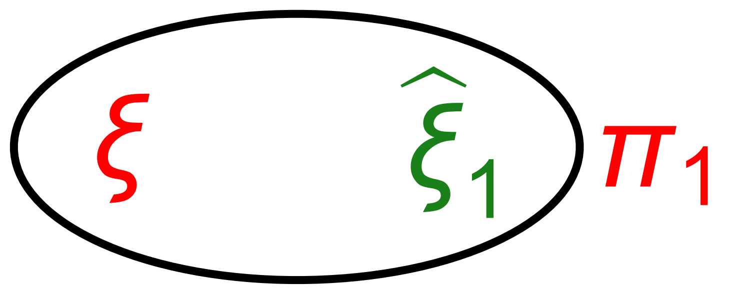
OT discrepancy:

$$C(\mathbb{P}, \widehat{\mathbb{P}}_k) = \min_{\pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k) \, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k)$$

Nature's Subproblem

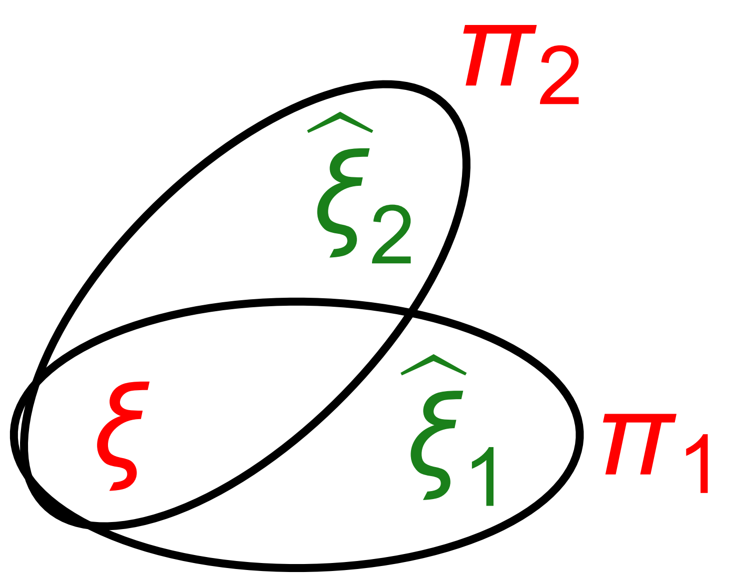
$$\begin{aligned} \sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) & \forall k \in [K] \\ & \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \le \epsilon_k & \forall k \in [K] \end{aligned}$$

Nature's Subproblem



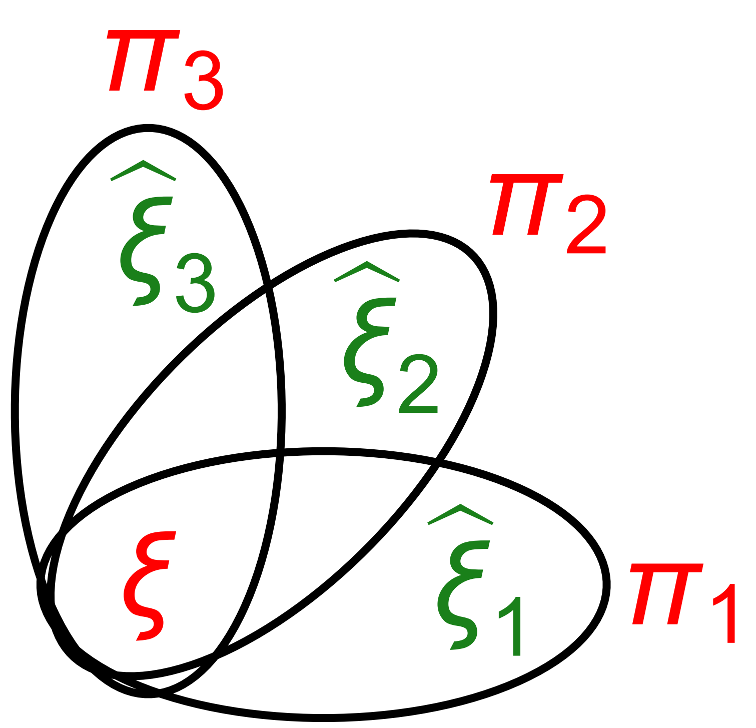
$$
\begin{aligned}
\sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d), \;\; \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) && \forall k \in [K] \\
& \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \le \epsilon_k && \forall k \in [K]
\end{aligned}
$$

Nature's Subproblem



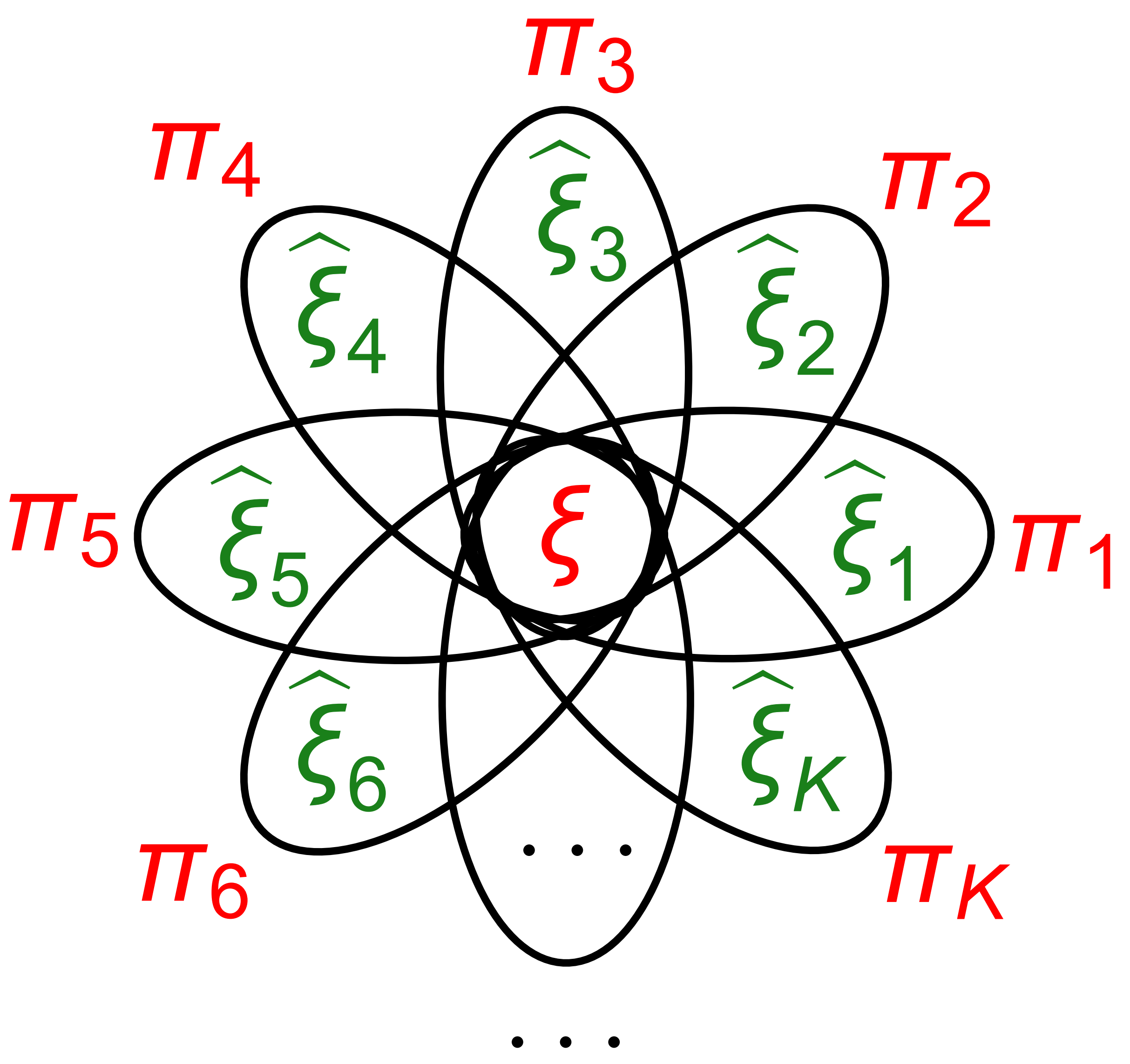
$$
\begin{aligned}
\sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) && \forall k \in [K] \\
& \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \leq \epsilon_k && \forall k \in [K]
\end{aligned}
$$

Nature's Subproblem



$$
\begin{aligned}
\sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) && \forall k \in [K] \\
& \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \leq \epsilon_k && \forall k \in [K]
\end{aligned}
$$

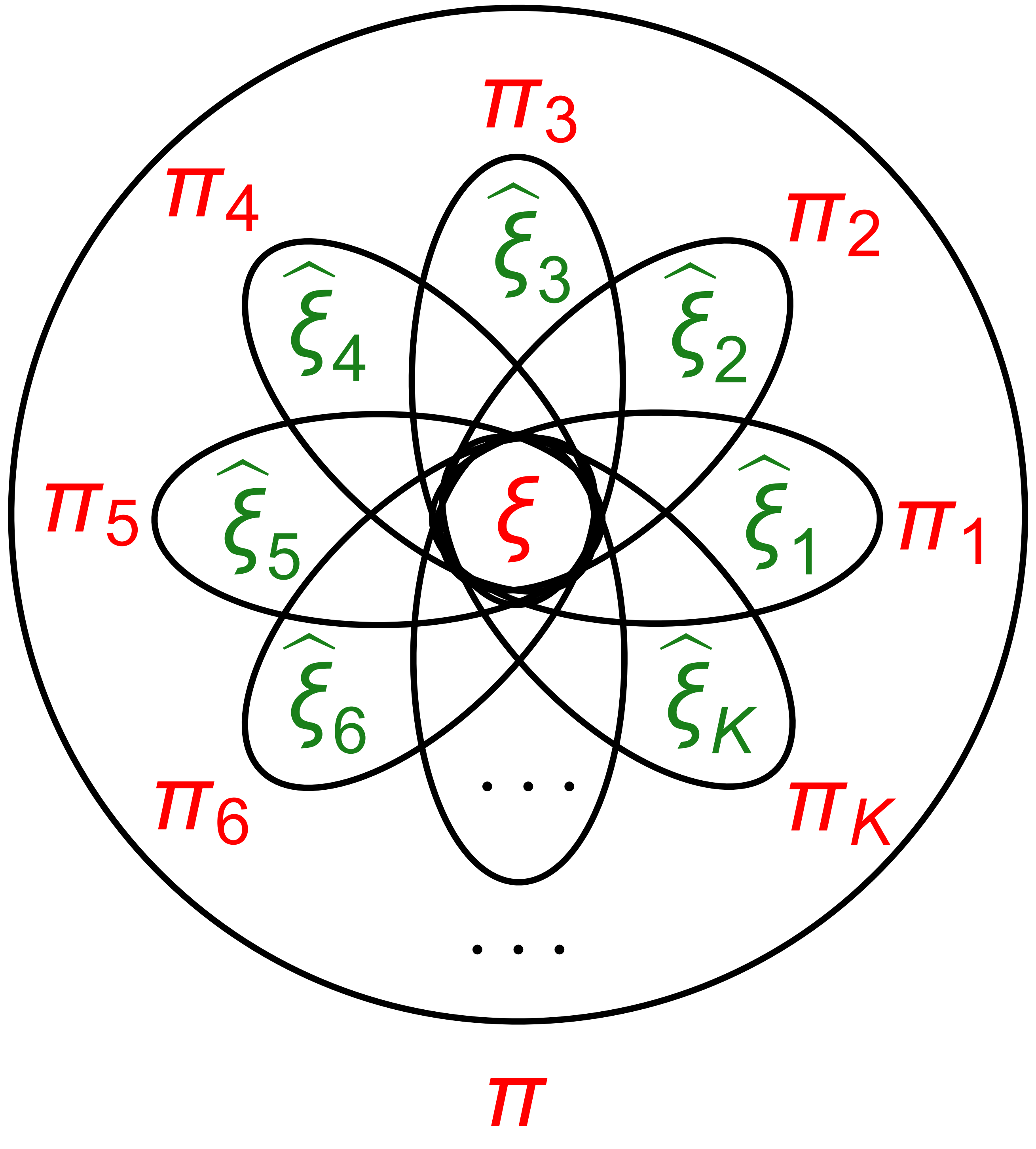
Nature's Subproblem



$$
\begin{aligned}
\sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) && \forall k \in [K] \\
& \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \leq \epsilon_k && \forall k \in [K]
\end{aligned}
$$

Nature's Subproblem

Gluing lemma:[1)]

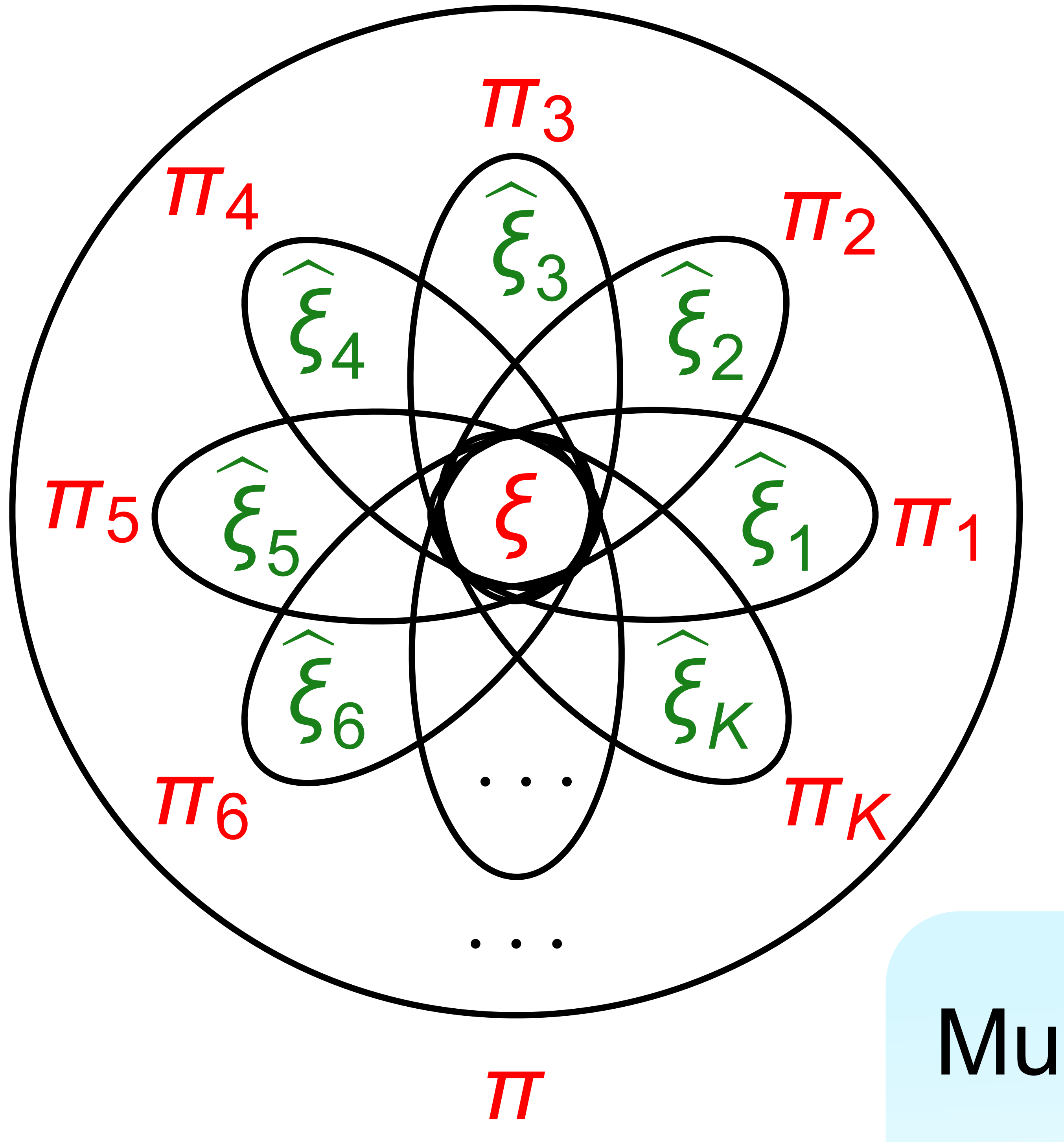


$$
\begin{array}{lll}
\sup & \displaystyle\int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) & \\
\text{s.t.} & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) & \forall k \in [K] \\
& \displaystyle\int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \leq \epsilon_k & \forall k \in [K]
\end{array}
$$

[1)] Villani, *Springer*, 2009.

Nature's Subproblem

Gluing lemma:[1)]



$$\begin{aligned} \sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \pi_k \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_k) && \forall k \in [K] \\ & \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi_k(\xi, \widehat{\xi}_k) \leq \epsilon_k && \forall k \in [K] \end{aligned}$$

Multi-margin transportation plans:

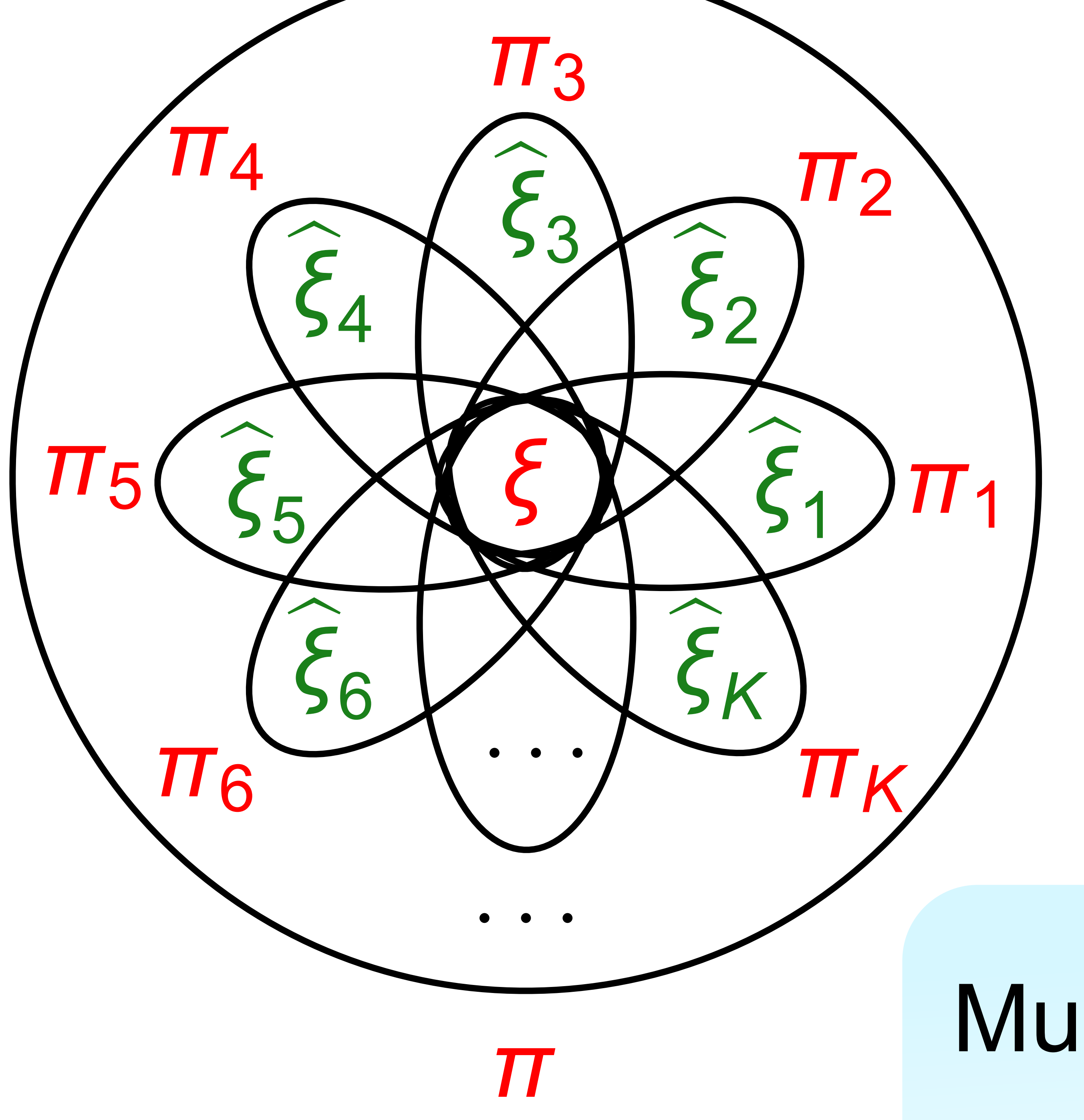
$$\Pi(\mathbb{P}, \widehat{\mathbb{P}}_1 \dots, \widehat{\mathbb{P}}_K) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^{K+1}) \ : \ \begin{array}{l} \xi, \widehat{\xi}_1 \dots, \widehat{\xi}_K \text{ have marginals} \\ \mathbb{P}, \widehat{\mathbb{P}}_1 \dots, \widehat{\mathbb{P}}_K \text{ under } \pi, \text{ resp.} \end{array} \right\}$$

---

1) Villani, *Springer*, 2009.

Nature's Subproblem

Gluing lemma:[1)]



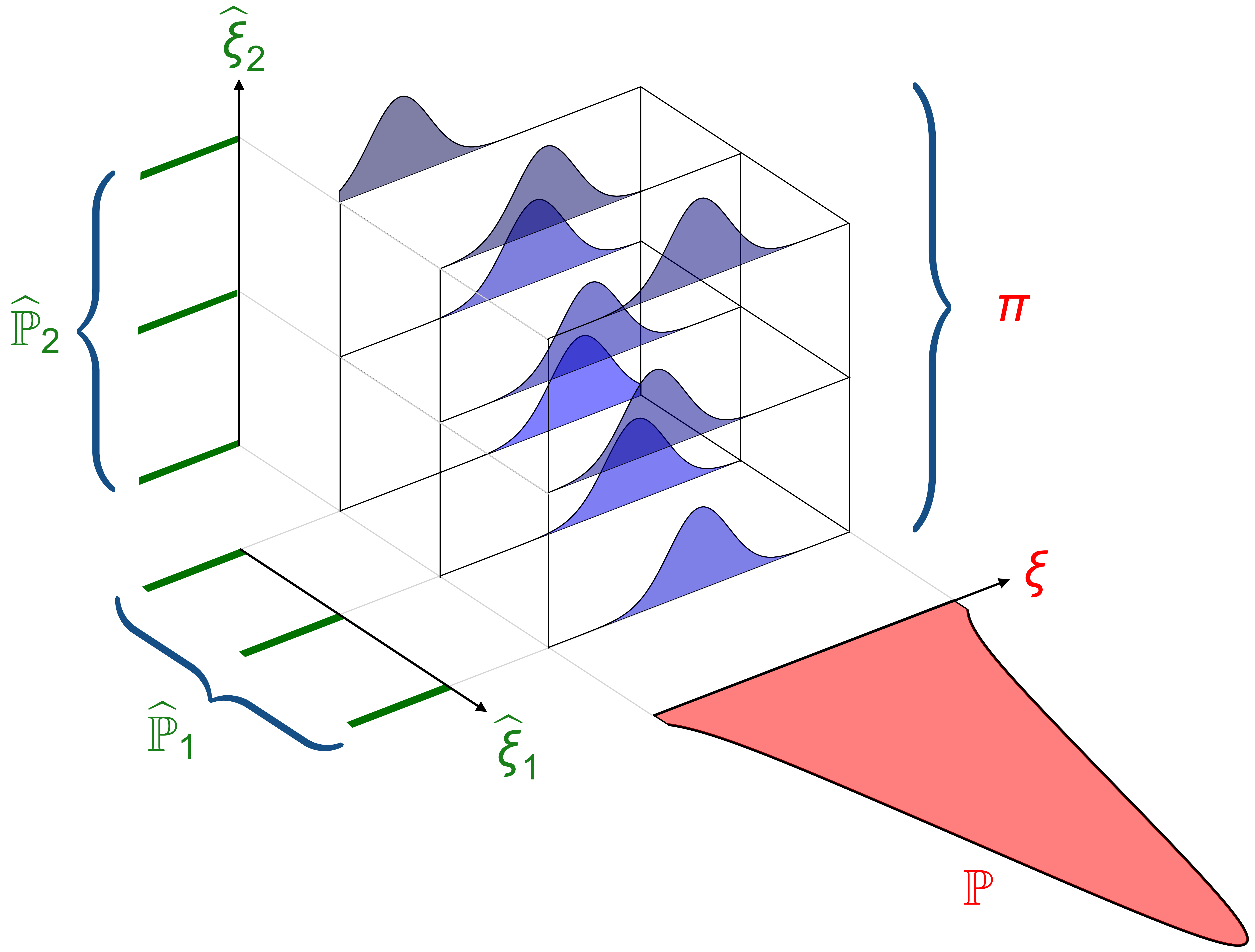
$$\begin{aligned} \sup \quad & \int_{\mathbb{R}^d} \ell(\xi)\, \mathrm{d}\mathbb{P}(\xi) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathbb{R}^d),\ \pi \in \Pi(\mathbb{P}, \widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_K) \\ & \int_{(\mathbb{R}^d)^{K+1}} c(\xi, \widehat{\xi}_k)\, \mathrm{d}\pi(\xi, \widehat{\xi}_1, \ldots, \widehat{\xi}_K) \leq \epsilon_k \quad \forall k \in [K] \end{aligned}$$

Multi-margin transportation plans:

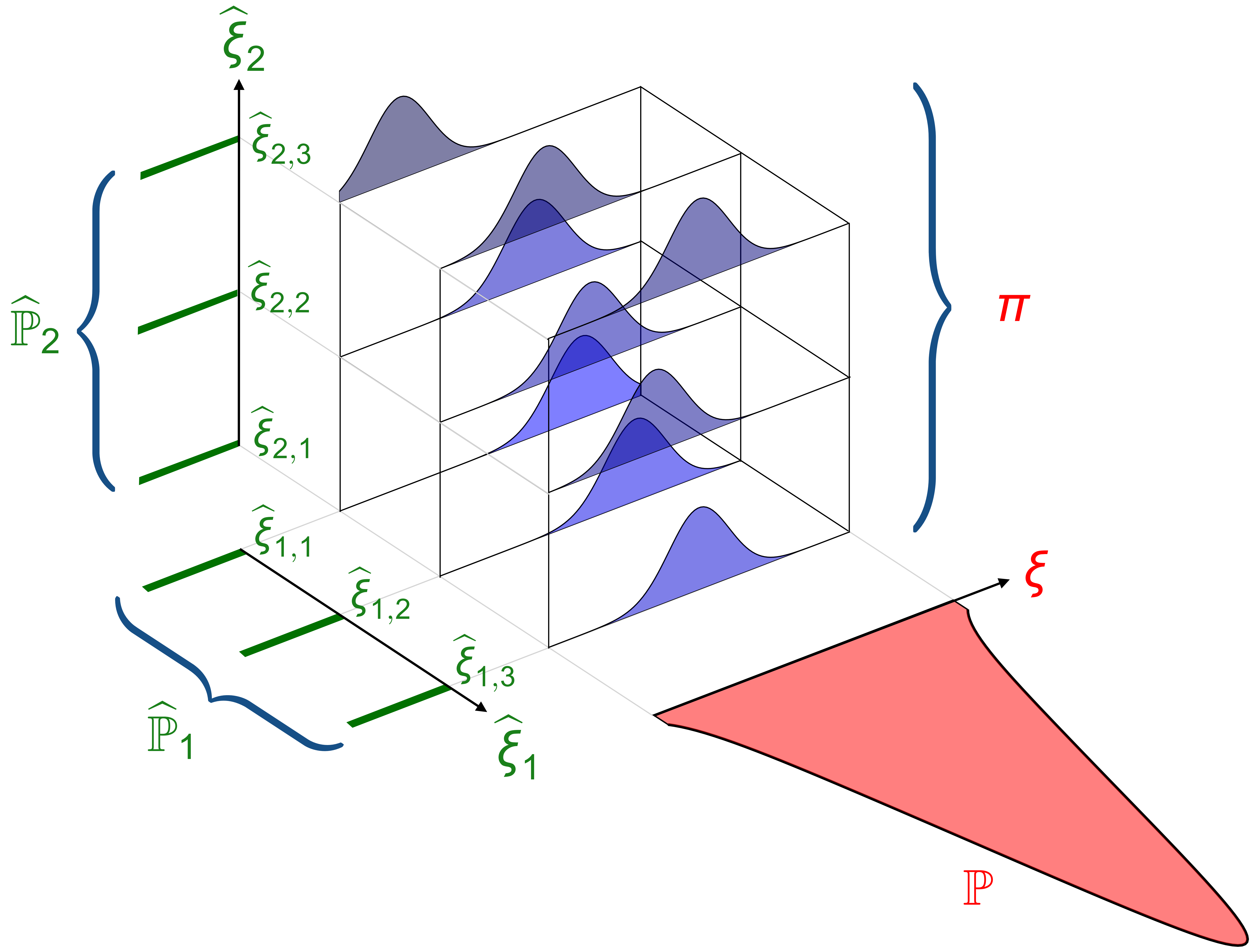
$$\Pi(\mathbb{P}, \widehat{\mathbb{P}}_1 \ldots, \widehat{\mathbb{P}}_K) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^{K+1}) \ : \ \begin{array}{l} \xi, \widehat{\xi}_1 \ldots, \widehat{\xi}_K \text{ have marginals} \\ \mathbb{P}, \widehat{\mathbb{P}}_1 \ldots, \widehat{\mathbb{P}}_K \text{ under } \pi, \text{ resp.} \end{array} \right\}$$

[1)] Villani, *Springer*, 2009.

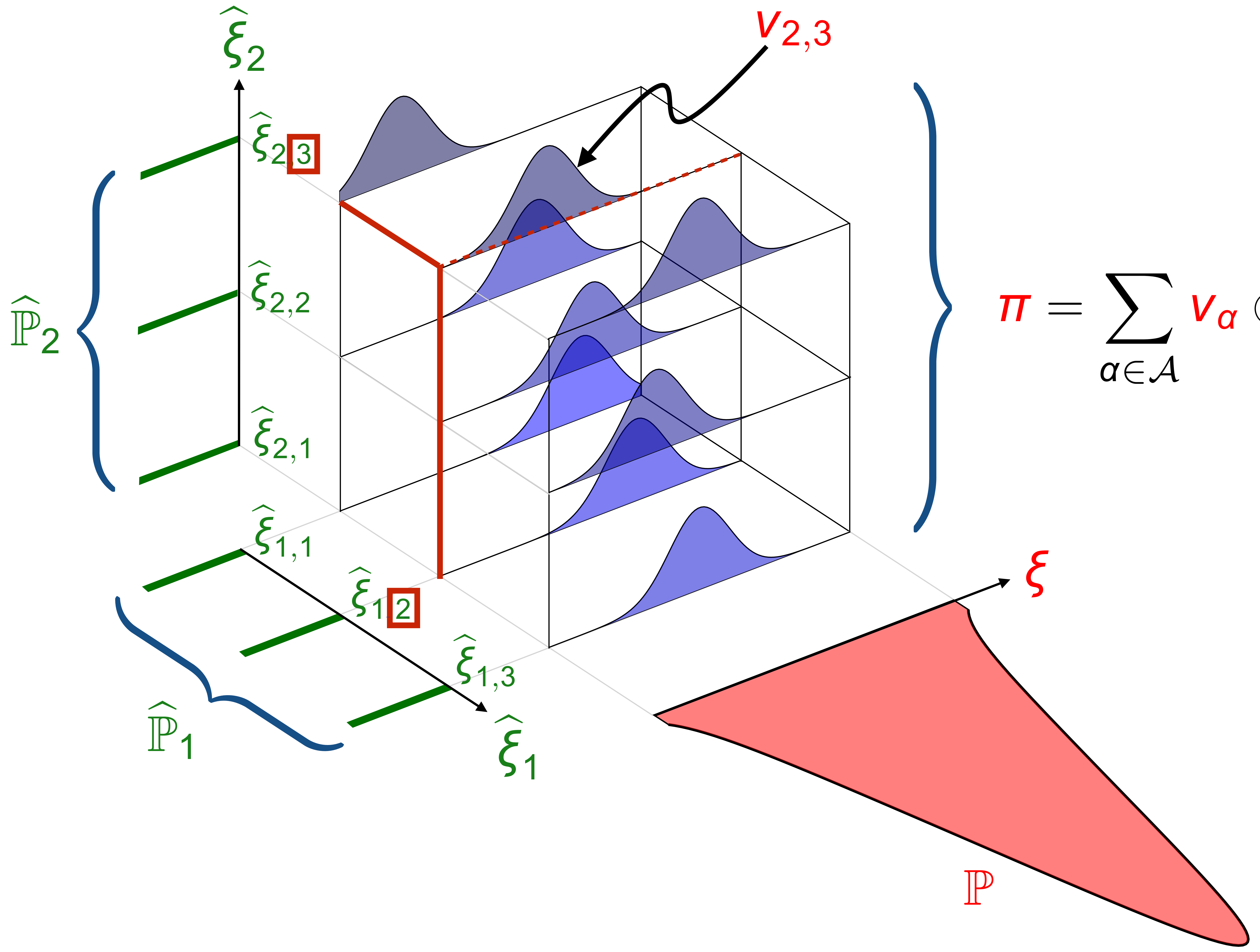
Multi-Margin Transportation Plans



Multi-Margin Transportation Plans



Multi-Margin Transportation Plans



$$\pi = \sum_{\alpha \in \mathcal{A}} v_\alpha \otimes \delta_\alpha$$

Nature's Subproblem

$$
\begin{array}{lll}
\sup & \displaystyle\sum_{\alpha\in\mathcal{A}} \int_{\mathbb{R}^d} \ell(\xi)\,\mathrm{d}v_\alpha(\xi) & \\
\text{s.t.} & v_\alpha \in \mathcal{M}_+(\mathbb{R}^d) & \forall \alpha\in\mathcal{A} \\
& \displaystyle\sum_{\substack{\alpha\in\mathcal{A}:\\ \alpha_k=j}} \int_{\mathbb{R}^d} \mathrm{d}v_\alpha(\xi) = \frac{1}{N_k} & \forall j\in[N_k],\ \forall k\in[K] \\
& \displaystyle\sum_{\alpha\in\mathcal{A}} \int_{\mathbb{R}^d} c(\xi,\widehat{\xi}_{k,\alpha_k})\,\mathrm{d}v_\alpha(\xi) \le \epsilon_k & \forall k\in[K],
\end{array}
$$

Nature's Subproblem

semi-infinite LP

$$
\left\{
\begin{array}{lll}
\sup & \displaystyle\sum_{\alpha\in\mathcal{A}} \int_{\mathbb{R}^d} \ell(\xi)\,\mathrm{d}v_\alpha(\xi) & \\
\text{s.t.} & v_\alpha \in \mathcal{M}_+(\mathbb{R}^d) & \forall \alpha \in \mathcal{A} \\
& \displaystyle\sum_{\substack{\alpha\in\mathcal{A}:\\ \alpha_k=j}} \int_{\mathbb{R}^d} \mathrm{d}v_\alpha(\xi) = \frac{1}{N_k} & \forall j \in [N_k],\ \forall k \in [K] \\
& \displaystyle\sum_{\alpha\in\mathcal{A}} \int_{\mathbb{R}^d} c(\xi, \widehat{\xi}_{k,\alpha_k})\,\mathrm{d}v_\alpha(\xi) \le \epsilon_k & \forall k \in [K],
\end{array}
\right.
$$

Nature's *Dual* Subproblem

$$\begin{aligned} \inf \quad & \sum_{k=1}^{K} \epsilon_k \lambda_k + \sum_{k=1}^{K} \frac{1}{N_k} \sum_{j=1}^{N_k} \gamma_{k,j} \\ \text{s.t.} \quad & \lambda_k \in \mathbb{R}_+, \ \gamma_k \in \mathbb{R}^{N_k} && \forall k \in [K] \\ & \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 && \forall \xi \in \mathbb{R}^d, \ \forall \alpha \in \mathcal{A} \end{aligned}$$

Nature's *Dual* Subproblem

$$
\begin{aligned}
\inf \quad & \sum_{k=1}^{K} \epsilon_k \lambda_k + \sum_{k=1}^{K} \frac{1}{N_k} \sum_{j=1}^{N_k} \gamma_{k,j} \\
\text{s.t.} \quad & \lambda_k \in \mathbb{R}_+, \ \gamma_k \in \mathbb{R}^{N_k} && \forall k \in [K] \\
& \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 && \forall \xi \in \mathbb{R}^d, \ \forall \alpha \in \mathcal{A}
\end{aligned}
$$

Theorem: If $c(\cdot, \widehat{\xi}_{k,\alpha_k})$ is convex and $\ell(\cdot)$ is piecewise concave, then nature's subproblem becomes a convex program with $\mathcal{O}(N^K)$ variables and constraints.

Nature's *Dual* Subproblem

$$\begin{aligned} \inf \quad & \sum_{k=1}^{K} \epsilon_k \lambda_k + \sum_{k=1}^{K} \frac{1}{N_k} \sum_{j=1}^{N_k} \gamma_{k,j} \\ \text{s.t.} \quad & \lambda_k \in \mathbb{R}_+, \ \gamma_k \in \mathbb{R}^{N_k} && \forall k \in [K] \\ & \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 && \forall \xi \in \mathbb{R}^d, \ \forall \alpha \in \mathcal{A} \end{aligned}$$

Theorem: If $c(\cdot, \widehat{\xi}_{k,\alpha_k})$ is convex and $\ell(\cdot)$ is piecewise concave, then nature's subproblem becomes a convex program with $\mathcal{O}(N^K)$ variables and constraints.

Proof: Standard duality tricks from convex analysis.[1)]

[1)] Ben-Tal, Den Hertog & Vial, *Math. Program.*, 2015; Zhen, Kuhn & Wiesemann, *Oper. Res.*, 2023.

Nature's *Dual* Subproblem

$$\begin{aligned} \inf \quad & \sum_{k=1}^{K} \epsilon_k \lambda_k + \sum_{k=1}^{K} \frac{1}{N_k} \sum_{j=1}^{N_k} \gamma_{k,j} \\ \text{s.t.} \quad & \lambda_k \in \mathbb{R}_+, \; \gamma_k \in \mathbb{R}^{N_k} & \forall k \in [K] \\ & \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 & \forall \xi \in \mathbb{R}^d, \; \forall \alpha \in \mathcal{A} \end{aligned}$$

Theorem: The problem is NP-hard even if $c(\xi, \widehat{\xi}_{k,\alpha_k}) = \|\xi - \widehat{\xi}_{k,\alpha_k}\|_2^2$ and $\ell(\xi) = 0$.

Nature's *Dual* Subproblem

$$
\begin{aligned}
\inf \quad & \sum_{k=1}^{K} \epsilon_k \lambda_k + \sum_{k=1}^{K} \frac{1}{N_k} \sum_{j=1}^{N_k} \gamma_{k,j} \\
\text{s.t.} \quad & \lambda_k \in \mathbb{R}_+, \ \gamma_k \in \mathbb{R}^{N_k} && \forall k \in [K] \\
& \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 && \forall \xi \in \mathbb{R}^d, \ \forall \alpha \in \mathcal{A}
\end{aligned}
$$

Theorem: The problem is NP-hard even if $c(\xi, \widehat{\xi}_{k,\alpha_k}) = \|\xi - \widehat{\xi}_{k,\alpha_k}\|_2^2$ and $\ell(\xi) = 0$.

Proof: Reduction from the NP-hard 2-Wasserstein barycenter problem.[1)]

---

1) Altschuler & Boix-Adsera, *J. Mach. Learning Res.*, 2021.

Worst-Case Distribution

Theorem: If $c(\cdot, \widehat{\xi}_{k,\alpha_k})$ grows faster than $\ell(\cdot)$, then nature's subproblem is solved by a discrete distribution $\mathbb{P}^\star$ with at most $1 + NK$ atoms.

$\Longrightarrow$ $\mathbb{P}^\star$ can be written in time poly(d, K, N)

Constraint Elimination

Robust constraints: $\ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d, \ \forall \alpha \in \mathcal{A}$

Constraint Elimination

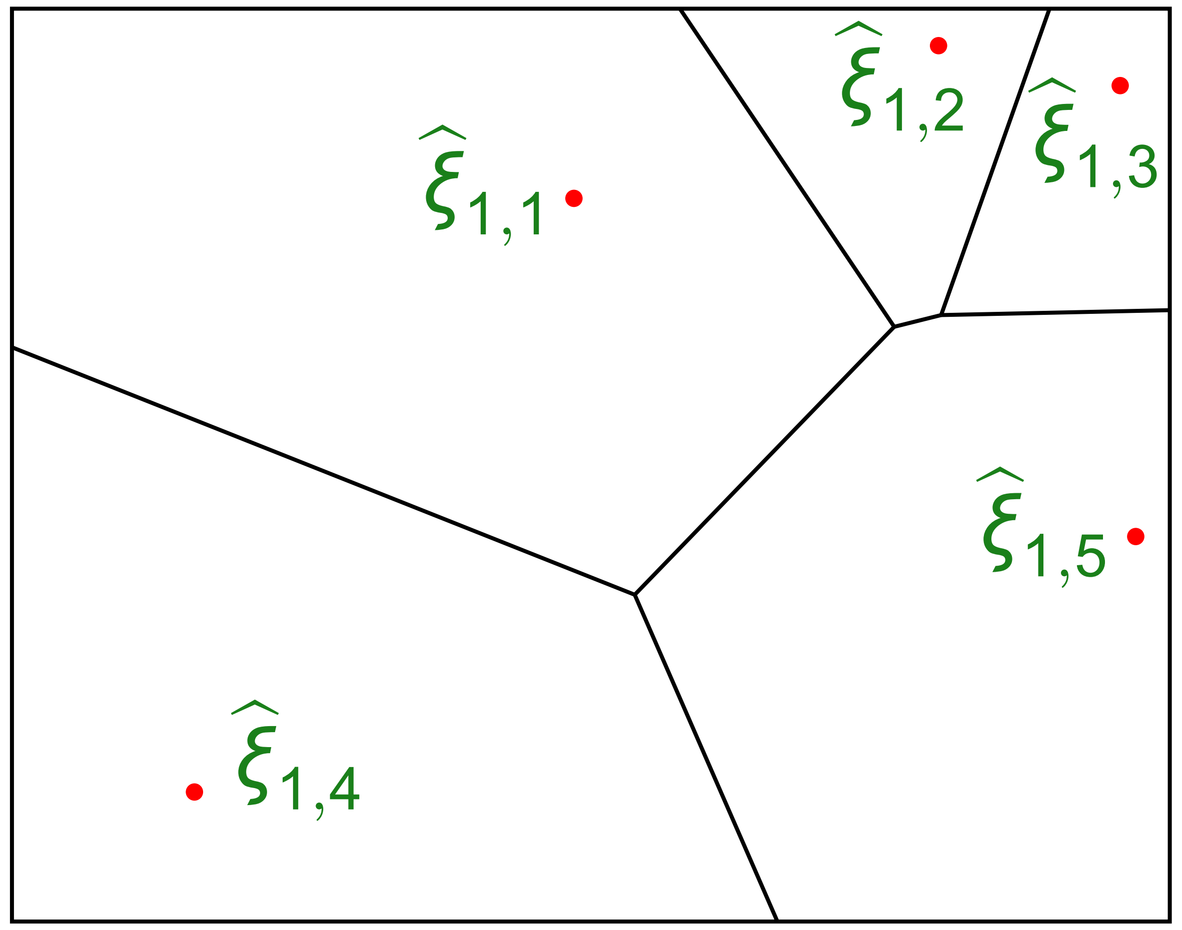
Robust constraints: $$\max_{\alpha \in \mathcal{A}} \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$

Constraint Elimination

Robust constraints: $$\ell(\xi) - \sum_{k=1}^{K} \min_{\alpha_k \in [N]} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$

Constraint Elimination

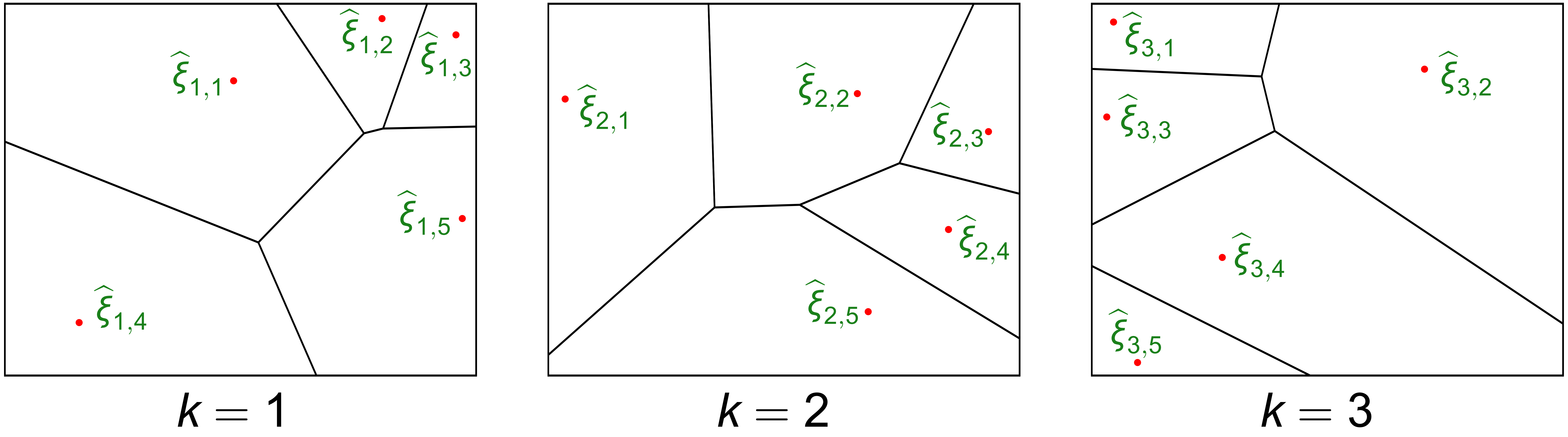
Robust constraints: $\ell(\xi) - \sum_{k=1}^{K} \min_{\alpha_k \in [N]} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$



$k = 1$

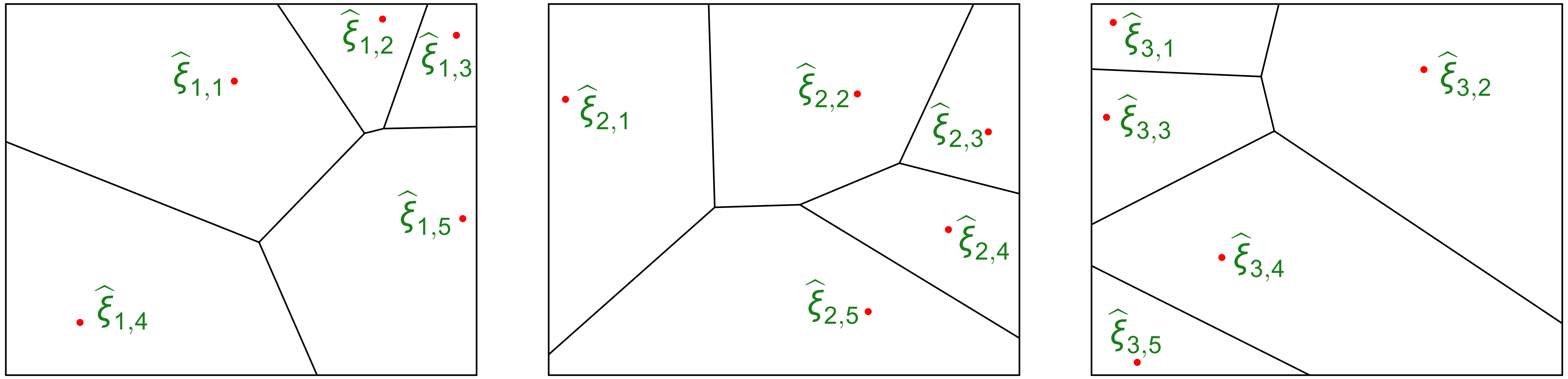
Constraint Elimination

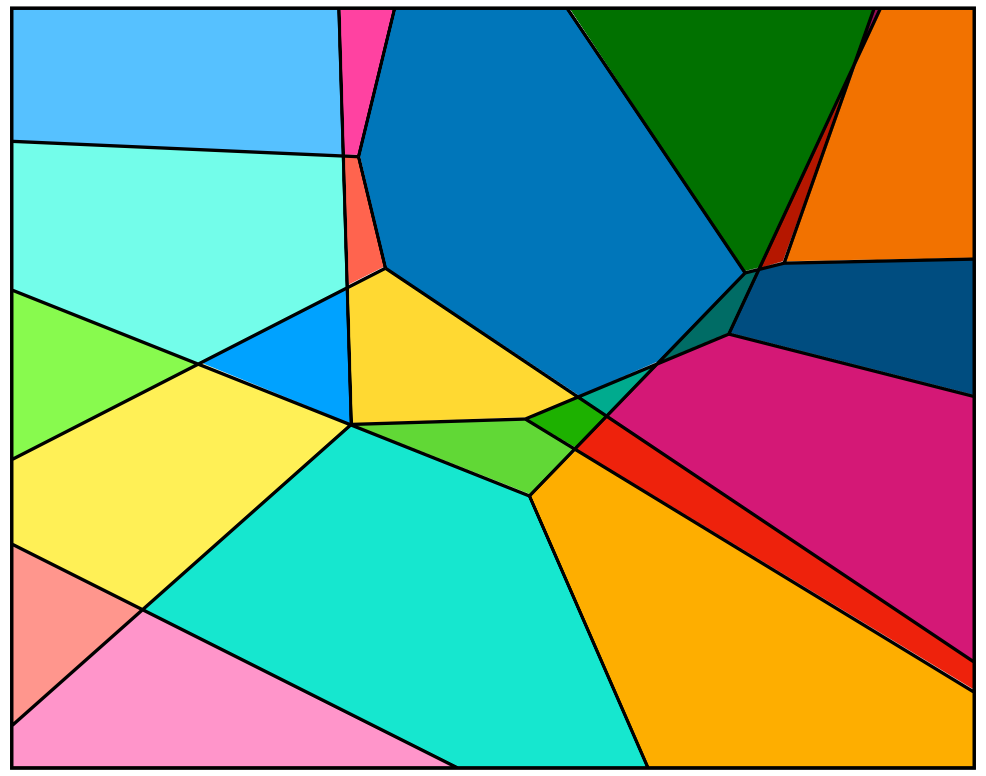
Robust constraints: $\ell(\xi) - \sum_{k=1}^{K} \min_{\alpha_k \in [N]} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$



Constraint Elimination

Robust constraints: $$\ell(\xi) - \sum_{k=1}^{K} \min_{\alpha_k \in [N]} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$

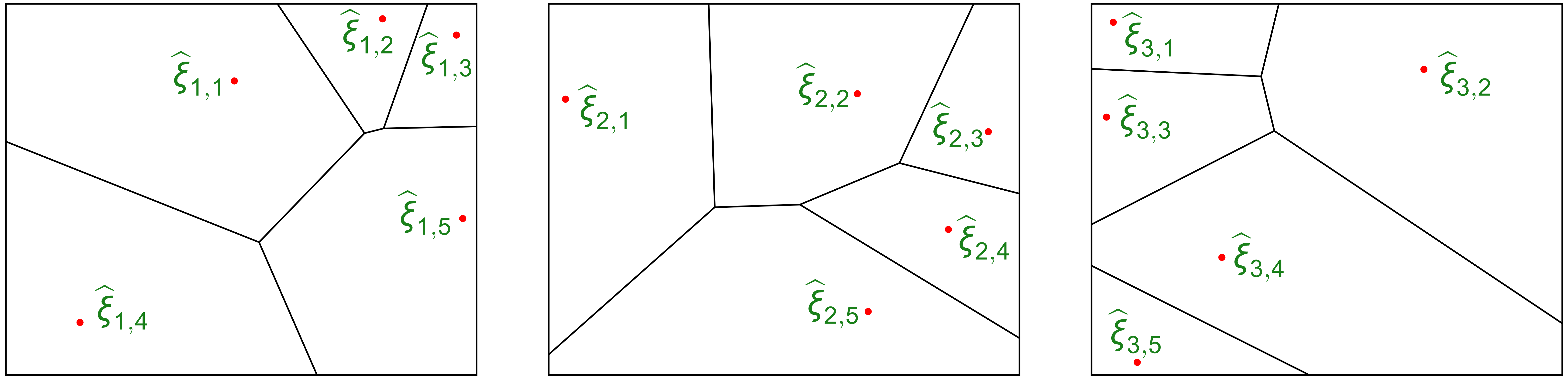


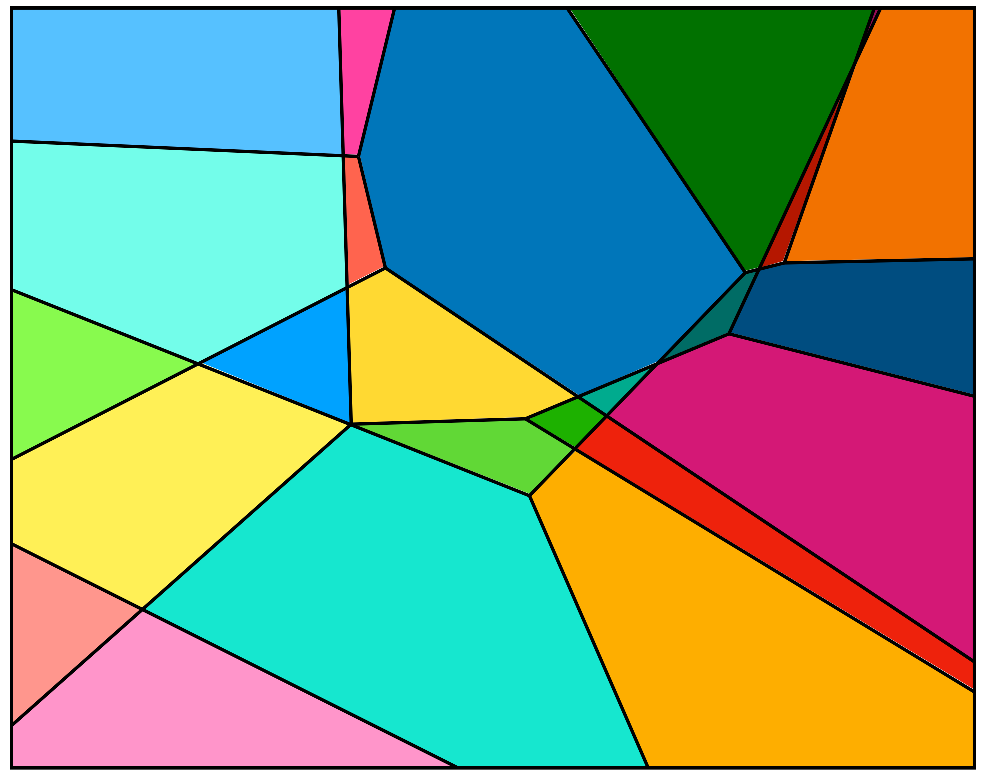


$\mathcal{C}_\alpha, \ \alpha \in \mathcal{A}$

Constraint Elimination

Robust constraints: $$\max_{\alpha \in \mathcal{A}} \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$



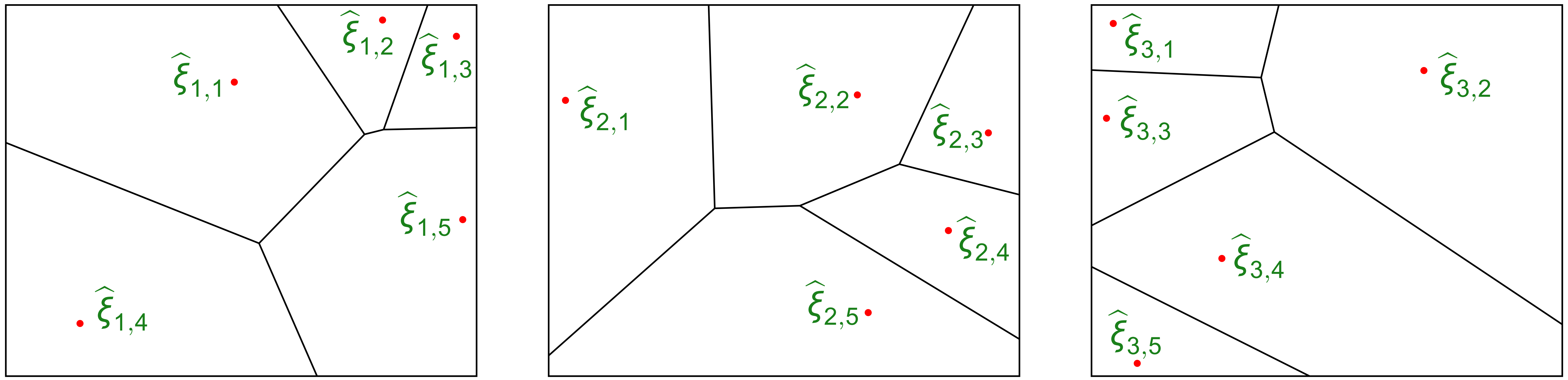


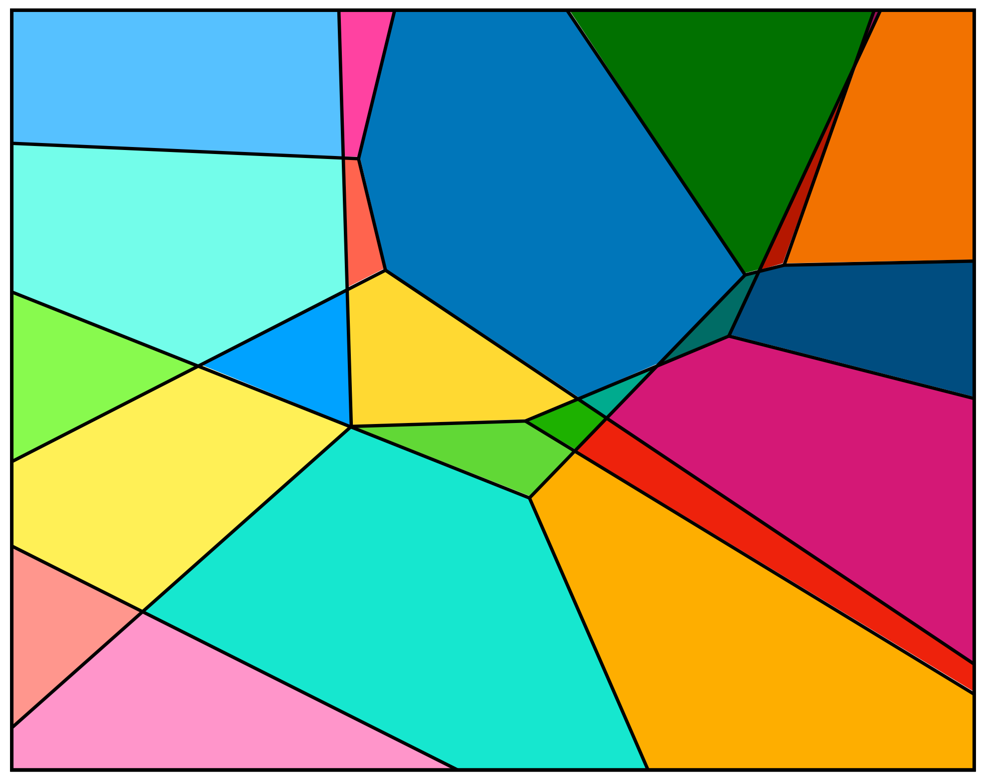
$$\mathcal{C}_\alpha, \ \alpha \in \mathcal{A}$$

$$\mathcal{A}' = \{\alpha \in \mathcal{A} : \mathcal{C}_\alpha \neq \emptyset\}$$

Constraint Elimination

Robust constraints: $$\max_{\alpha \in \mathcal{A}'} \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$





$\mathcal{C}_\alpha, \ \alpha \in \mathcal{A}$

$\mathcal{A}' = \{\alpha \in \mathcal{A} : \mathcal{C}_\alpha \neq \emptyset\}$

$\mathcal{A}'$ can be computed in time[1] poly(N, K)

[1] Altschuler & Boix-Adsera, *J. Mach. Learning Res.*, 2021.

Constraint Elimination

Robust constraints: $$\max_{\alpha \in \mathcal{A}'} \ell(\xi) - \sum_{k=1}^{K} \left(\lambda_k c(\xi, \widehat{\xi}_{k,\alpha_k}) + \gamma_{k,\alpha_k} \right) \leq 0 \quad \forall \xi \in \mathbb{R}^d$$

Theorem: If standard convexity conditions hold, then the optimal value of nature's subproblem can be computed to any accuracy $\delta > 0$

- in time $\text{poly}(N, d, \log 1/\delta)$ (with exponential dependence on K) or
- in time $\text{poly}(N, K, \log 1/\delta)$ (with exponential dependence on d).

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \middle| W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \leq \widehat{\epsilon}_k\right) \geq 1 - \beta_{F_k}(\epsilon_k, N_1, N_k) \quad \forall \epsilon_k \geq \widehat{\epsilon}_k.$$

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \,\middle|\, W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \leq \widehat{\epsilon}_k\right) \geq 1 - \beta_{F_k}(\epsilon_k, N_1, N_k) \quad \forall \epsilon_k \geq \widehat{\epsilon}_k.$$

w.r.t. beliefs over $\mathbb{P}_k$, $k \in [K]$, and
w.r.t. samples from $\mathbb{P}_k$, $k \in [K]$

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \,\middle|\, W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \leq \widehat{\epsilon}_k\right) \geq 1 - \beta_{F_k}(\epsilon_k, N_1, N_k) \quad \forall \epsilon_k \geq \widehat{\epsilon}_k.$$

observed distance between $\widehat{\mathbb{P}}_1$ and $\widehat{\mathbb{P}}_k$

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \,\middle|\, W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \leq \widehat{\epsilon}_k\right) \geq 1 - \underbrace{\beta_{F_k}(\epsilon_k, N_1, N_k)} \quad \forall \epsilon_k \geq \widehat{\epsilon}_k.$$

Bayesian significance level
(decreases with ϵ_k)

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

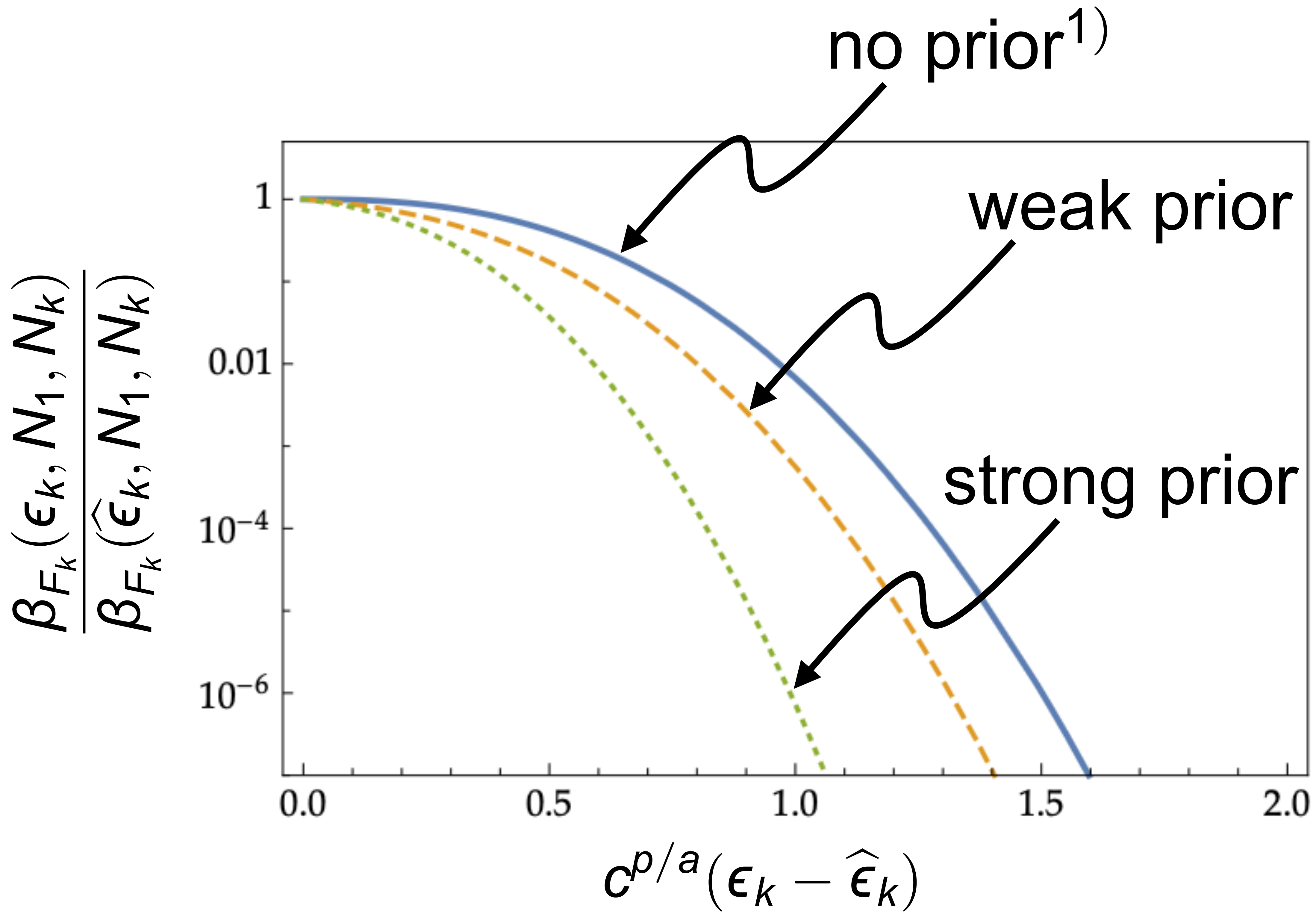
$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \,\middle|\, W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \leq \widehat{\epsilon}_k\right) \geq 1 - \beta_{F_k}(\epsilon_k, N_1, N_k) \quad \forall \epsilon_k \geq \widehat{\epsilon}_k.$$

prior cdf of $W_p(\mathbb{P}_1, \mathbb{P}_k)$

Bayesian Measure Concentration

Theorem: If $\mathbb{P} = \mathbb{P}_1$, then

$$\text{Prob}\left(\mathbb{P}_1 \in \mathbb{B}_{\epsilon_k}(\widehat{\mathbb{P}}_k) \,\middle|\, W_p(\widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_k) \le \widehat{\epsilon}_k \right) \ge 1 - \beta_{F_k}(\epsilon_k, N_1, N_k) \quad \forall \epsilon_k \ge \widehat{\epsilon}_k.$$



---

1) Fournier & Guillin, *Probab. Theory Rel.*, 2014.

Numerical Results

Portfolio Selection

Mean-risk portfolio problem: $\inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}}\left[-\langle \theta, \xi \rangle\right] + \rho \cdot \mathbb{P}\text{-CVaR}_{\eta}(-\langle \theta, \xi \rangle)$

Portfolio Selection

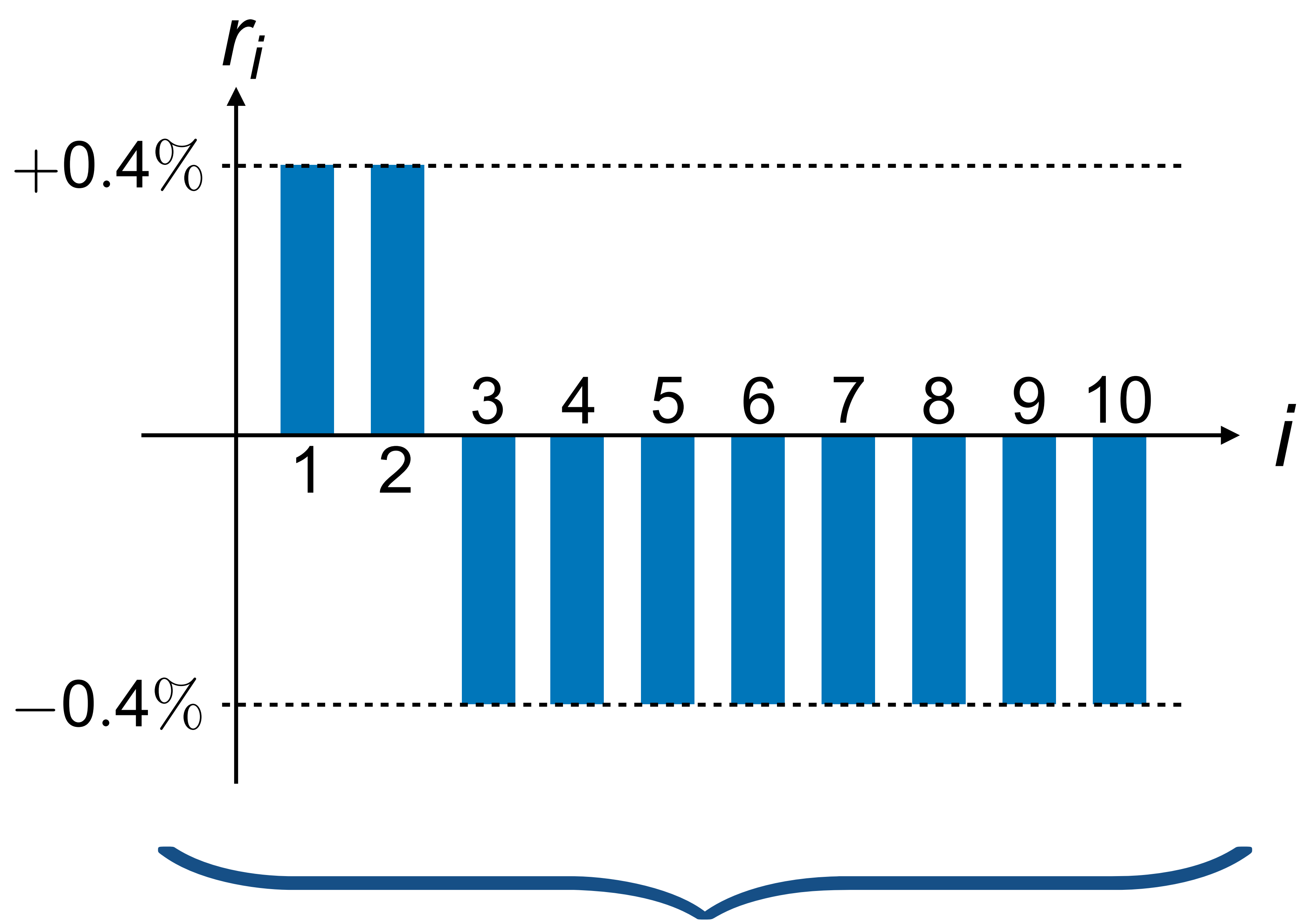
Mean-risk portfolio problem: $\inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}}[-\langle \theta, \xi \rangle] + \rho \cdot \mathbb{P}\text{-CVaR}_{\eta}(-\langle \theta, \xi \rangle)$

Synthetic data: $\xi_i = \psi + \zeta_i$, where $\psi \sim \mathcal{N}(0, 2\%)$ and $\zeta_i \sim \mathcal{N}(r_i, 1\%)$ i.i.d.

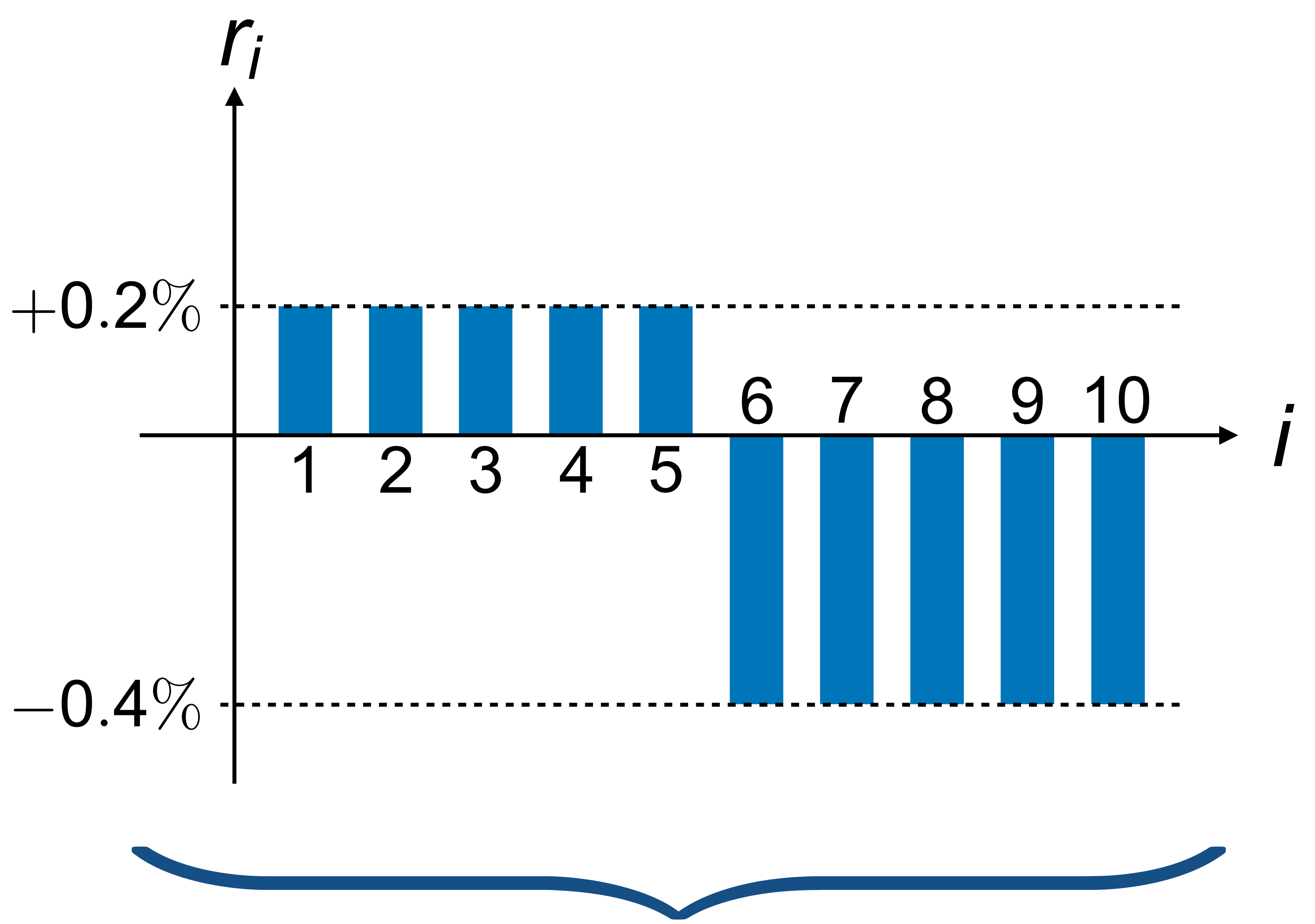
Portfolio Selection

Mean-risk portfolio problem: $\inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}} [-\langle \theta, \xi \rangle] + \rho \cdot \mathbb{P}\text{-CVaR}_{\eta}(-\langle \theta, \xi \rangle)$

Synthetic data: $\xi_i = \psi + \zeta_i$, where $\psi \sim \mathcal{N}(0, 2\%)$ and $\zeta_i \sim \mathcal{N}(r_i, 1\%)$ i.i.d.



$\mathbb{P}_1 =$ target distribution, $N_1 = 5$



$\mathbb{P}_2 =$ source distribution, $N_2 = 30$

Portfolio Selection

Mean-risk portfolio problem: $\inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}}[-\langle \theta, \xi \rangle] + \rho \cdot \mathbb{P}\text{-CVaR}_{\eta}(-\langle \theta, \xi \rangle)$

Synthetic data: $\xi_i = \psi + \zeta_i$, where $\psi \sim \mathcal{N}(0, 2\%)$ and $\zeta_i \sim \mathcal{N}(r_i, 1\%)$ i.i.d.

Table 1 Out-of-sample performance of optimal portfolios on synthetic data (mean (std. error) over 10 replications)

	Single-source DRO on target data	Single-source DRO on source data	Single-source DRO on pooled data	Single-source DRO on barycenter	Multi-source DRO
Sharpe ratio	0.002 (0.017)	−0.024 (0.009)	−0.021 (0.012)	0.007 (0.014)	**0.034 (0.024)**
Expected value	0.01 (0.067)	−0.097 (0.036)	−0.084 (0.047)	0.03 (0.059)	**0.143 (0.1)**
Standard deviation	4.035 (0.018)	**4.018 (0.016)**	4.02 (0.017)	4.034 (0.018)	4.084 (0.022)

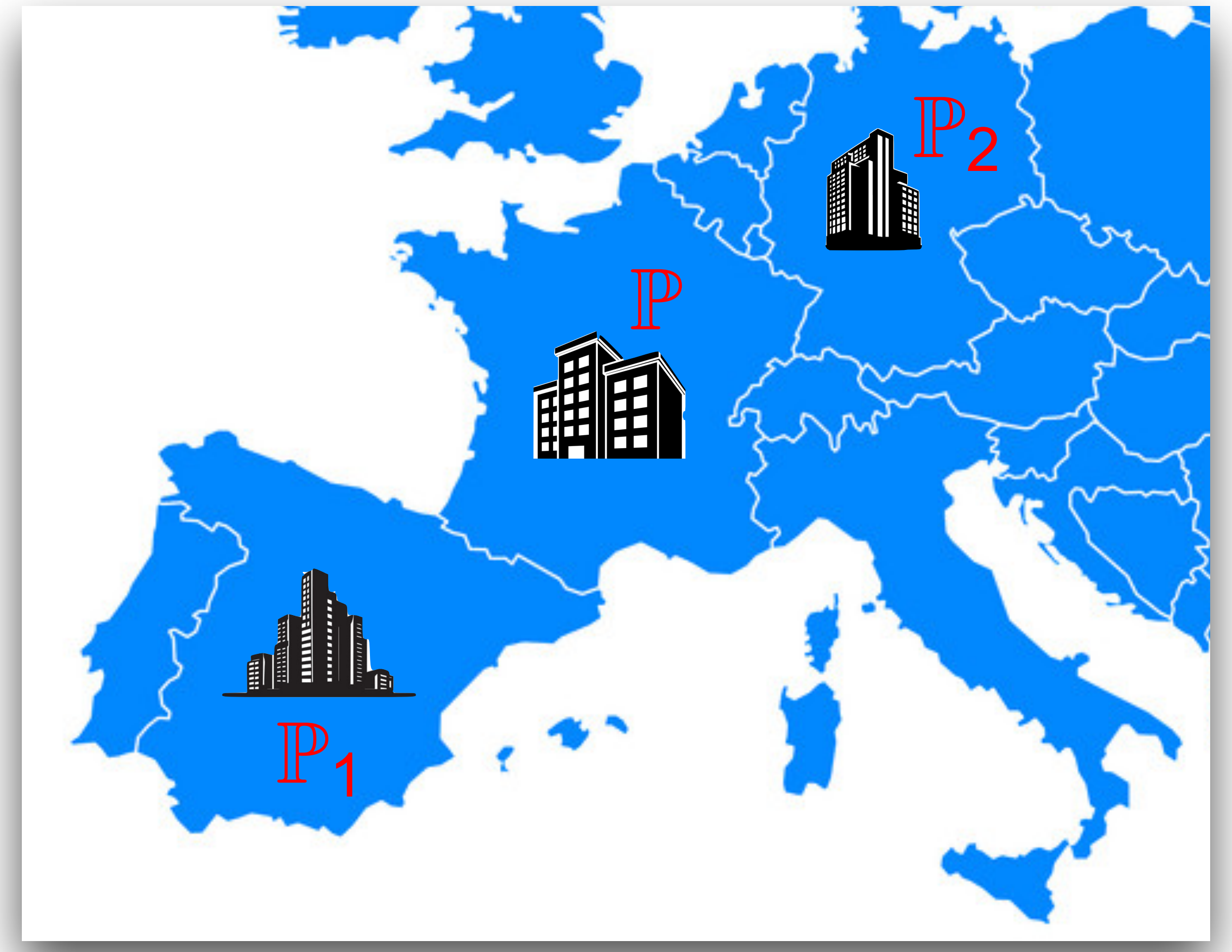
Portfolio Selection with Regional MSCI Data

Table 2 Out-of-sample Sharpe ratios of optimal portfolios on real data (mean (std. error) over 10 replications)

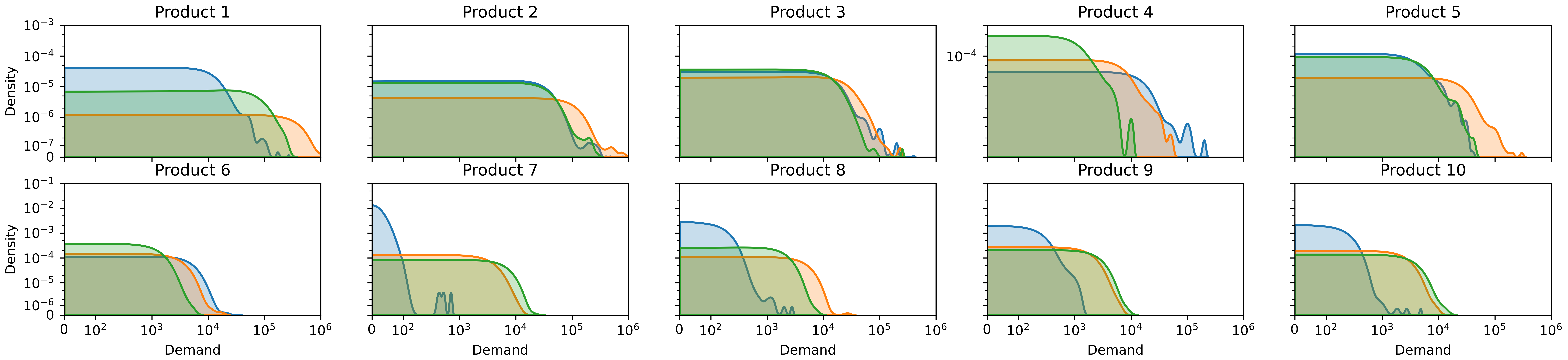
Source	Target	N_1	Single-s. DRO on target distribution	Single-s. DRO on source distribution	Single-s. DRO on pooled data	Single-s. DRO on barycenter	Multi-s. DRO
Europe	Pacific	5	0.059 (0.004)	0.054 (0.006)	0.052 (0.005)	0.052 (0.003)	0.039 (0.006)
		10	0.047 (0.005)	0.04 (0.004)	0.043 (0.0)	0.036 (0.004)	0.035 (0.006)
		20	0.052 (0.004)	0.049 (0.008)	0.049 (0.006)	0.042 (0.006)	0.035 (0.006)
		50	0.051 (0.004)	0.047 (0.005)	0.047 (0.003)	0.045 (0.004)	0.028 (0.004)
	USA	5	0.101 (0.01)	0.107 (0.006)	0.103 (0.007)	0.112 (0.008)	0.113 (0.003)
		10	0.095 (0.007)	0.103 (0.007)	0.104 (0.007)	0.108 (0.006)	0.112 (0.004)
		20	0.094 (0.008)	0.092 (0.009)	0.103 (0.007)	0.107 (0.008)	0.116 (0.004)
		50	0.105 (0.006)	0.095 (0.008)	0.105 (0.006)	0.113 (0.005)	0.115 (0.004)
Pacific	Europe	5	0.111 (0.009)	0.088 (0.0)	0.088 (0.0)	0.096 (0.008)	0.123 (0.014)
		10	0.112 (0.012)	0.088 (0.0)	0.088 (0.0)	0.111 (0.012)	0.126 (0.014)
		20	0.107 (0.01)	0.088 (0.0)	0.088 (0.0)	0.106 (0.01)	0.111 (0.012)
		50	0.113 (0.01)	0.089 (0.0)	0.088 (0.0)	0.101 (0.008)	0.145 (0.011)
	USA	5	0.109 (0.006)	0.115 (0.006)	0.115 (0.006)	0.113 (0.006)	0.116 (0.006)
		10	0.088 (0.006)	0.11 (0.007)	0.11 (0.007)	0.109 (0.008)	0.113 (0.006)
		20	0.099 (0.007)	0.107 (0.007)	0.106 (0.006)	0.114 (0.006)	0.114 (0.006)
		50	0.111 (0.005)	0.119 (0.004)	0.119 (0.004)	0.119 (0.004)	0.119 (0.004)
USA	Europe	5	0.117 (0.014)	0.091 (0.001)	0.09 (0.001)	0.116 (0.014)	0.144 (0.013)
		10	0.117 (0.014)	0.091 (0.001)	0.091 (0.001)	0.117 (0.014)	0.151 (0.014)
		20	0.119 (0.011)	0.09 (0.001)	0.095 (0.005)	0.108 (0.009)	0.15 (0.011)
		50	0.11 (0.01)	0.091 (0.001)	0.091 (0.002)	0.11 (0.01)	0.141 (0.013)
	Pacific	5	0.046 (0.007)	0.049 (0.002)	0.051 (0.004)	0.042 (0.004)	0.046 (0.006)
		10	0.07 (0.004)	0.06 (0.004)	0.058 (0.004)	0.058 (0.005)	0.046 (0.006)
		20	0.048 (0.004)	0.053 (0.003)	0.049 (0.003)	0.047 (0.003)	0.041 (0.004)
		50	0.047 (0.003)	0.052 (0.003)	0.048 (0.003)	0.047 (0.003)	0.046 (0.002)

1

Assortment Optimization with Kaggle Data



$$\max_{\theta \in \{0,1\}^d} \left\{ \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^{d} \theta_i p_i \xi_i \right] : \sum_{i=1}^{d} \theta_i \leq 3 \right\}$$



Assortment Optimization with Kaggle Data

Table 3 Out-of-sample expected revenue of optimal product portfolios (mean (std. error) over 20 replications)

N_1	N_2	Single-s. DRO on distribution $\widehat{\mathbb{P}}_1$	Single-s. DRO on distribution $\widehat{\mathbb{P}}_2$	Single-s. DRO on pooled data	Single-s. DRO on barycenter	Multi-s. DRO
25	25	**504.52 (37.51)**	431.53 (31.17)	481.08 (23.25)	480.16 (30.72)	502.20 (22.90)
	50	481.51 (41.26)	460.61 (29.33)	488.86 (22.87)	476.58 (37.78)	**516.93 (15.66)**
	75	478.79 (42.02)	475.65 (29.29)	470.58 (32.40)	501.66 (30.18)	**515.83 (17.53)**
	100	485.10 (37.71)	443.66 (34.65)	479.74 (29.72)	485.10 (37.71)	**505.23 (22.31)**
50	25	482.26 (37.96)	473.96 (31.94)	489.10 (15.0)	489.93 (30.76)	**515.55 (21.24)**
	50	481.97 (30.82)	442.18 (32.37)	485.77 (14.60)	473.80 (33.48)	**519.13 (10.93)**
	75	**510.30 (21.27)**	450.55 (32.32)	485.77 (14.60)	482.96 (30.69)	506.92 (17.70)
	100	471.92 (26.69)	430.48 (30.53)	482.18 (22.24)	443.68 (29.12)	**519.13 (10.93)**
75	25	490.24 (18.25)	465.64 (27.43)	482.43 (14.03)	486.91 (17.97)	**514.70 (14.54)**
	50	489.10 (15.00)	462.31 (26.60)	489.10 (15.00)	477.40 (18.45)	**499.11 (15.23)**
	75	477.50 (28.54)	450.55 (32.32)	479.10 (13.26)	465.20 (31.46)	**492.44 (15.23)**
	100	479.49 (34.36)	468.98 (28.15)	492.19 (23.03)	486.41 (30.31)	**515.79 (12.25)**
100	25	481.08 (23.25)	487.36 (27.41)	489.10 (15.00)	457.08 (27.38)	**515.79 (12.25)**
	50	461.92 (24.31)	460.61 (29.33)	482.43 (14.03)	462.11 (25.48)	**505.79 (14.60)**
	75	477.50 (28.54)	467.34 (24.41)	485.77 (14.60)	489.10 (15.00)	**509.12 (14.03)**
	100	470.83 (27.45)	465.64 (27.43)	485.77 (14.60)	481.08 (23.25)	**509.12 (14.03)**

Summary

- **Multi-source DRO**
  - tool for data-driven optimization when target data is **scarce** or **unavailable**
  - reminiscent of **domain adaptation** and **transfer learning**
  - dually related to computation of **Wasserstein barycenters**

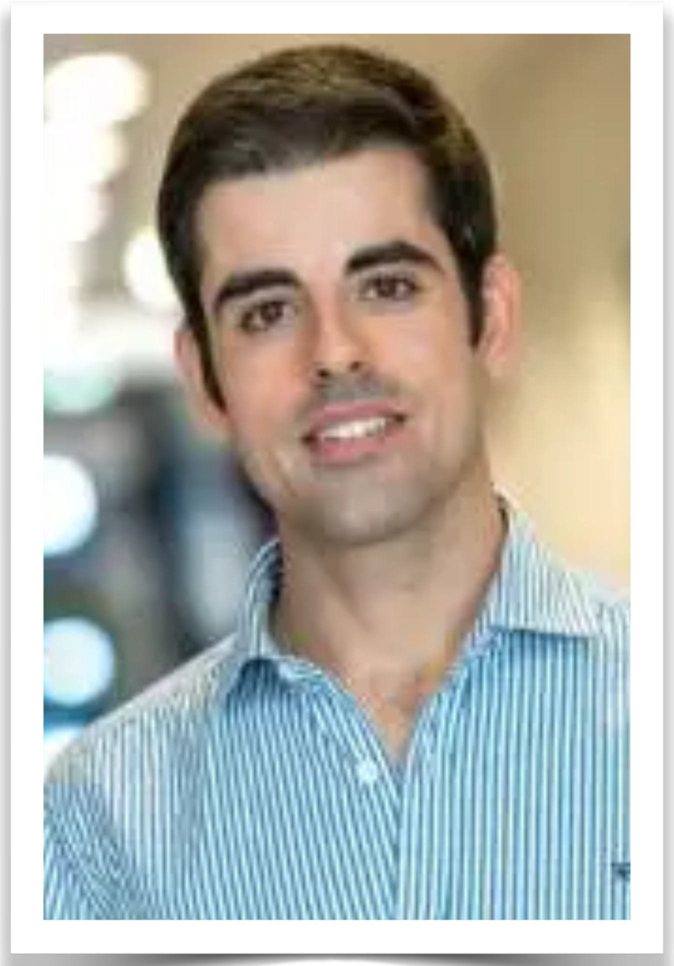
- **Nature's Subproblem**
  - equivalent to **convex program** with $O(N^K)$ variables and constraints
  - **NP-hard** but solved by **discrete distribution** with $1 + NK$ atoms
  - **tractable** if either d or K is fixed

- **Statistical Guarantees**
  - choice of Wasserstein radii based on **Bayesian measure concentration** result
  - **source data is useful** in the presence of strong priors

This Talk is Based on…

Y. Rychener, A. Esteban-Pérez, J. M. Morales & D. Kuhn. **Wasserstein Distributionally Robust Optimization with Heterogeneous Data Sources**. *arXiv preprint*. 2024.

Yves Rychener



Adrián Esteban-Pérez

Juan M. Morales