Distributionally Robust Optimization

Stochastic Programming

- Consider the following stochastic program:

$$\underset{x \in \mathcal{X}}{\text{maximize}} \quad \mathbb{E}\left[g_0(x, Z)\right] \tag{8.1a}$$

$$\text{subject to} \quad \mathbb{E}\left[g_j(x, Z)\right] \leq b_j \;, \; \forall j = 1, \ldots, J \tag{8.1b}$$

- This model is quite flexible: bounds on probability, expected utility models, risk measures, etc.

- DRO questions the assumption that the distribution of Z is known

Ellsberg's urn game

- Consider a game in which two urns are presented to you

- Urn #1 has an equal amount of blue and red balls inside
- Urn #2 also has red and blue balls but of unknown proportion

- You are asked to choose between urn #1 & #2.

- I will draw a ball from the chosen urn
  - If you chose urn #1 and a red ball is drawn, you win 1000$
  - If you chose urn #2 and a red ball is drawn, you win 1100$

- What wou

A strict preference for urn #1 demonstrates ambiguity aversion

Distributionally Robust Optimization

- Assume that one only knows that $F \in \mathcal{D}$
  - E.g. 1: normal distrib. with mean and covariance in some confidence region
  - E.g. 2: distribution supported on some region with known mean

- Instead of maximizing expected value, maximize the worst-case expected value (similarly for constraints)

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{maximize}} \quad & \inf_{F\in\mathcal{D}} \mathbb{E}_F[g_0(x,Z)] & \text{(8.2a)} \\ \text{subject to} \quad & \mathbb{E}_F[g_j(x,Z)] \leq b_j \,,\ \forall j=1,\ldots,J \,,\ \forall\, F \in \mathcal{D}\,. & \text{(8.2b)} \end{aligned}$$

- In this chapter, we focus on :

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{F\in\mathcal{D}} \mathbb{E}_F[h(x,\xi)]\,, \qquad \text{(8.3a)}$$

Scenario based models

Scenario based models

- An alternative to moment based models consists of using predefined scenarios: $$\mathcal{Z} := \{z^1, z^2, \ldots, z^K\}$$
- The DRO model takes the form:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sup_{p \in \mathcal{U}} \sum_{k=1}^{K} p_k h(x, z^k)$$

- Pros : If all scenarios are covered, then DRO model can be asymptotically consistent in data-driven context (see Bayraksan & Love [5] for survey)

$$\sup_{p \in \mathcal{U}(\{\xi^i\}_{i=1}^M)} \sum_{k=1}^{K} p_k h(x, z^k) \quad \xrightarrow[M \to \infty]{} \quad E[h(x, \xi)]$$

- Cons: If some scenarios are missing, then there is no protection against them

Moment based models

Mean and support models

- We would like to solve:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sup_{F \in \mathcal{D}} \mathbb{E}_F[h(x,\xi)], \tag{8.3a}$$

where the distribution set takes the form

$$\mathcal{D}(\mathcal{Z},\mu) = \left\{ F \in \mathcal{M} \;\middle|\; \begin{array}{l} \mathbb{P}(\xi \in \mathcal{Z}) = 1 \\ \mathbb{E}[\xi] = \mu \end{array} \right\},$$

- E.g. : Markov inequality

$$P(\xi \geq a) \leq \sup_{F \in \mathcal{D}([0,\infty[,\mu)} E[\mathbf{1}\{\xi \geq a\}] = \begin{cases} 1 & \text{if } \mu \geq a \\ \mu/a & \text{otherwise.} \end{cases}$$

Semi-infinite linear programming duality

- The worst-case expected value problem looks like:

$$\begin{aligned}
\underset{F \in \mathcal{M}}{\text{maximize}} \quad & \int_{\mathcal{Z}} h(x,\xi) dF(\xi) && (8.4a)\\
\text{subject to} \quad & \int_{\mathcal{Z}} dF(\xi) = 1 && (8.4b)\\
& \int_{\mathcal{Z}} \xi dF(\xi) = \mu \,, && (8.4c)
\end{aligned}$$

- Duality theory for semi-infinite linear program states that if there exists a feasible distribution then dual problem is equivalent:

$$\begin{aligned}
\underset{r,q}{\text{minimize}} \quad & \mu^T q + r && (8.5a)\\
\text{subject to} \quad & z^T q + r \geq h(x,z)\,,\ \forall\, z \in \mathcal{Z}\,, && (8.5b)
\end{aligned}$$

The main reformulation

Theorem 8.6. : *Let $\mathcal{D}(\mathcal{Z}, \mu)$ be a distribution set for which there exists a feasible solution $F_0 \in \mathcal{D}(\mathcal{Z}, \mu)$, the DRO problem presented in* (8.3) *is equivalent to the following robust optimization problem:*

$$\underset{x \in \mathcal{X}, q}{\text{minimize}} \quad \sup_{z \in \mathcal{Z}} h(x, z) + (\mu - z)^T q \,. \tag{8.11}$$

Moreover, the problem can be reformulated as follows when $\mathcal{Z}$ is a convex set and $h(x, z) := \max_k h_k(x, z)$ where each $h_k(x, z)$ is a concave function of z:

$$\begin{array}{ll} \underset{x \in \mathcal{X}, q, \{v_k\}_k, t}{\text{minimize}} & t + \mu^T q \\ \text{subject to} & t \geq \delta^*(v_k \,|\, \mathcal{Z}) - h_*^k(x, v_k + q) \,, \; \forall\, k \,, \end{array}$$

where for each k, $v_k \in \mathbb{R}^m$, while $\delta^*(v|\mathcal{Z})$ is the support function of $\mathcal{Z}$ and $h_*^k(x, v)$ is the partial concave conjugate function of $h_k(x, z)$.

Example : Generalized Markov Inequality

- Consider trying to bound the following probability with respect to probabilities supported in the non-negative orthant with a mean of μ with $\mathcal{U}$ as a convex set:

$$\mathbb{P}(\xi \in \mathcal{U})$$

- Based on Theorem 8.6, this can be measured using:

$$\begin{aligned}
\underset{t,q,w}{\text{minimize}} \quad & t + q^T \mu \\
\text{subject to} \quad & t \geq \delta^*(w|\mathcal{U}) + 1 \\
& t \geq 0 \\
& q \geq -w \\
& q \geq 0
\end{aligned}$$

Some intuition about the worst-case distribution

- We showed that the worst-case expected value problem is equivalent to:

$$\begin{aligned} \underset{r,q}{\text{minimize}} \quad & \mu^T q + r && (8.5\text{a}) \\ \text{subject to} \quad & z^T q + r \geq h(x, z)\,,\ \forall\, z \in \mathcal{Z}\,, && (8.5\text{b}) \end{aligned}$$

- For finite dimensional LP, it is well known that only m+1 constraints are needed to get an optimal solution

- There should therefore exist a set $\mathcal{Z}^* := \left\{z_1^*, z_2^*, \ldots, z_{m+1}^*\right\}$ for which 8.5 becomes equivalent to

$$\begin{aligned} \underset{r,q}{\text{minimize}} \quad & \mu^T q + r && (8.7\text{a}) \\ \text{subject to} \quad & \xi^T q + r \geq h(x, z)\,,\ \forall\, z \in \mathcal{Z}^*\,, && (8.7\text{b}) \end{aligned}$$

Some intuition about the worst-case distribution

- The finite dimensional LP

$$\underset{r,q}{\text{minimize}} \quad \mu^T q + r \tag{8.7a}$$

$$\text{subject to} \quad \xi^T q + r \geq h(x, z)\,,\, \forall\, z \in \mathcal{Z}^*\,, \tag{8.7b}$$

is equivalent to

$$\underset{p \in \mathbb{R}^{m+1}}{\text{maximize}} \quad \sum_{i=1}^{m+1} p_i h(x, z_i^*) \tag{8.8a}$$

$$\text{subject to} \quad \sum_i p_i = 1 \tag{8.8b}$$

$$\sum_{i=1}^{m+1} p_i z_i^* = \mu\,. \tag{8.8c}$$

Some intuition about the worst-case distribution

Theorem 8.2. : *Let $\mathcal{Z} \in \mathbb{R}^m$ be a Borel set, and F_0 be some feasible distribution according to $\mathcal{D}(\mathcal{Z}, \mu)$, then problem (8.4) is equivalent to the following finite dimensional optimization problem*

$$\begin{aligned}
\underset{p,\{z_i\}_{i=1}^{m+1}}{\text{maximize}} \quad & \sum_{i=1}^{m+1} p_i h(x, z_i) && \text{(8.9a)}\\
\text{subject to} \quad & \sum_{i=1}^{m+1} p_i = 1 \;\;\&\;\; p \geq 0 && \text{(8.9b)}\\
& \sum_{i=1}^{m+1} p_i z_i = \mu && \text{(8.9c)}\\
& z_i \in \mathcal{Z}\,,\, \forall\, i = 1, \ldots, m+1\,, && \text{(8.9d)}
\end{aligned}$$

where $p \in \mathbb{R}^{m+1}$ and each $z_i \in \mathbb{R}^m$.

Example : Mean-variance models

Example 8.7. : Consider that ξ is a random variable known to have a mean μ, and a variance of $\mathbb{E}\left[(\xi-\mu)^2\right]=\sigma^2$. This gives rise to the following DRO problem :

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{F\in\mathcal{D}(\mu,\sigma^2)} \mathbb{E}_F[h(x,\xi)] \ ,$$

where

$$\mathcal{D}(\mu,\sigma^2) := \left\{F \,|\, \mathbb{P}(\xi\in\mathbb{R})=1,\ \mathbb{E}[\xi]=\mu,\ \mathbb{E}\left[(\xi-\mu)^2\right]=\sigma^2\right\} .$$

- Applying Theorem 8.6 we get:

$$\underset{x\in\mathcal{X},q\in\mathbb{R}^2}{\text{minimize}} \ \sup_{z_1\in\mathbb{R}} h(x,z_1)+(\mu-z_1)q_1+(\sigma^2-(z_1-\mu)^2)q_2$$

- If h(x,z) is bounded below this reduces to:

$$\underset{x\in\mathcal{X},q_1,q_2\geq 0}{\text{minimize}} \quad \sup_{z_1\in\mathbb{R}} h(x,z_1)+\boxed{(\mu-z_1)q_1+(\sigma^2-(z_1-\mu)^2)q_2}$$

concave in z_1

Example: Support-mean-bounded covariance model

Example 8.8. : Consider that one has information about the support $\mathcal{Z}$, the mean μ, and an upper bound on the second order moment matrix of the type $\mathbb{E}[\xi\xi^T] \preceq \Sigma$ where $A \preceq B$ refers to the fact that $B - A$ is positive semi-definite, i.e. $z^T(B-A)z \geq 0$ for all $z \in \mathbb{R}^m$. This gives rise to the following DRO problem :

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{F\in\mathcal{D}(\mathcal{Z},\mu,\Sigma)} \mathbb{E}_F[h(x,\xi)] \ ,$$

where

$$\mathcal{D}(\mathcal{Z},\mu,\Sigma) := \left\{F \,\middle|\, \mathbb{P}(\xi\in\mathcal{Z}) = 1,\ \mathbb{E}[\xi] = \mu,\ \mathbb{E}[\xi\xi^T] \preceq \Sigma\right\} .$$

- Solution:

$$\underset{x\in\mathcal{X},q,Q\succeq 0}{\text{minimize}} \ \sup_{z\in\mathcal{Z}} h(x,z) + (\mu - z)^T q + \Sigma \bullet Q - z^T Q z$$

- See Wiesemann et al. [42] for many more moment models

Accounting for moment uncertainty

- Data-driven moment estimation leads to moment uncertainty
- DRO problem might instead take the form:

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{\mu\in\mathcal{U},F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[h(x,z)]\,. \tag{8.13a}$$

Corollary 8.9. : *Let $\mathcal{D}(\mathcal{Z},\mu)$ be a distribution set and $\mathcal{U}\in\mathbb{R}^m$ be a bounded uncertainty set for the moment vector μ. Given that there exists a feasible pair (μ_0,F_0) for which $\mu_0\in\mathcal{U}$ and $F_0\in\mathcal{D}(\mathcal{Z},\mu_0)$, the DRO problem presented in* (8.13) *is equivalent to the following robust optimization problem:*

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}} h(x,z)-z^Tq\boxed{+\,\delta^*(q\,|\,\mathcal{U})} \tag{8.14a}$$

Mean-Covariance Uncertainty

Example 8.10. : In [24], the authors explain how independently and identically distributed samples $\{\xi_i\}_{i=1}^M$ from F can be used to construct the following uncertainty set:

$$\mathcal{D}(\mathcal{Z}, \hat{\mu}, \hat{\Sigma}, \gamma_1, \gamma_2) = \left\{ F \in \mathcal{M} \;\middle|\; \begin{array}{l} \mathbb{P}(\xi \in \mathcal{Z}) = 1 \\ (\mathbb{E}[\xi] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}[\xi] - \hat{\mu}) \leq \gamma_1 \\ \mathbb{E}[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \preceq (1 + \gamma_2)\hat{\Sigma} \end{array} \right\} ,$$

- The parameters can be chosen such that this set has high probability of containing the true distribution

$$\begin{aligned} \underset{x \in \mathcal{X}, q, Q \succeq 0}{\text{minimize}} \quad & \sup_{z \in \mathcal{Z}} h(x, z) - z^T q - z^T Q z \\ & + ((1 + \gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T) \bullet Q + \hat{\mu}^T q + \sqrt{\gamma_1} \|\hat{\Sigma}^{1/2}(q + 2Q\hat{\mu})\|_2 \end{aligned}$$

We will exploit : $\delta^*([v_1^T \;\; v_2^T] \,|\, \mathcal{Z}_1 \times \mathcal{Z}_2) = \delta^*(v_1 \,|\, \mathcal{Z}_1) + \delta^*(v_2 \,|\, \mathcal{Z}_2)$

Exercise 8.1 + 8.2

Consider the following DRO problem:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sup_{F \in \mathcal{D}_1} \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\ x^T C\xi)]\,, \tag{8.17}$$

where $\mathcal{X} := \{x \in \mathbb{R}^n_+ \,|\, Ax \leq b\}$ for some $A \in \mathbb{R}^{p\times n}$ and $b \in \mathbb{R}^p$, $Q(x) := \sum_{i=1}^n Q_i x_i$ with each $Q_i \in \mathbb{R}^{m\times m}$ such that $Q_i \succ 0$, and $C \in \mathbb{R}^{n\times m}$ such that each $C_{ij} \geq 0$.

Exercise 8.1. *Mean-support DRO problem*
Derive an explicit finite dimensional representation for the DRO problem (8.17) when

$$\mathcal{D}_1 := \{F \,|\, \mathbb{P}_F(\xi \in \mathcal{Z}) \geq 1,\ \mathbb{E}_F[\xi] = \bar{\mu}\}\,,$$

where $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, Wz \leq v\}$, with $W \in \mathbb{R}^{p\times m}$, $v \in \mathbb{R}^p$, and $\bar{\mu} \in \mathbb{R}^n$.

Exercise 8.2. *DRO with moment uncertainty*
Derive an explicit finite dimensional representation for problem (8.17) when the distribution ambiguity set takes the form:

$$\mathcal{D}_2(\Gamma) := \{F \,|\, \mathbb{P}_F(\xi \in \mathcal{Z}) = 1,\ \mathbb{E}_F[\xi] \geq \bar{\mu},\ \sum_i \mathbb{E}_F[\xi_i] - \bar{\mu}_i \leq \Gamma\}\,.$$

- Hint: Use Theorem 8.6 and Corollary 8.9

Theorem 8.6. : *Let $\mathcal{D}(\mathcal{Z}, \mu)$ be a distribution set for which there exists a feasible solution $F_0 \in \mathcal{D}(\mathcal{Z}, \mu)$, the DRO problem presented in (8.3) is equivalent to the following robust optimization problem:*

$$\underset{x \in \mathcal{X}, q}{\text{minimize}} \quad \sup_{z \in \mathcal{Z}} h(x, z) - z^T q + \mu^T q \,. \tag{8.11}$$

Moreover, the problem can be reformulated as follows when $\mathcal{Z}$ is a convex set and $h(x, z) := \max_k h_k(x, z)$ where each $h_k(x, z)$ is a concave function of z:

$$\begin{aligned} \underset{x \in \mathcal{X}, q, \{v_k\}_k, t}{\text{minimize}} \quad & t + \mu^T q \\ \text{subject to} \quad & t \geq \delta^*(v_k \,|\, \mathcal{Z}) - h_*^k(x, v_k + q) \,, \, \forall\, k \,, \end{aligned}$$

where for each k, $v_k \in \mathbb{R}^m$, while $\delta^*(v|\mathcal{Z})$ is the support function of $\mathcal{Z}$ and $h_*^k(x, v)$ is the partial concave conjugate function of $h_k(x, z)$.

Corollary 8.9. : *Let $\mathcal{D}(\mathcal{Z},\mu)$ be a distribution set and $\mathcal{U} \in \mathbb{R}^m$ be a bounded and convex uncertainty set for the moment vector μ. Given that for all $\mu \in \mathcal{U}$, there exists an $F \in \mathcal{D}(\mathcal{Z},\mu)$, the DRO problem presented in (8.14) is equivalent to the following robust optimization problem:*

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}} h(x,z) - z^T q + \delta^*(q\,|\,\mathcal{U})\,. \tag{8.15a}$$

Moreover, the problem can be reformulated as follows when $\mathcal{Z}$ is a convex set and $h(x,z) := \max_k h_k(x,z)$ where each $h_k(x,z)$ is a concave function:

$$\begin{aligned} \underset{x\in\mathcal{X},q,\{v_k\}_k,t}{\text{minimize}} \quad & t + \delta^*(q\,|\,\mathcal{U}) \\ \text{subject to} \quad & t \geq \delta^*(v_k\,|\,\mathcal{Z}) - h_*^k(x, v_k + q)\,,\ \forall\, k\,, \end{aligned}$$

where for each k, $v_k \in \mathbb{R}^m$, while $\delta^*(v|\mathcal{Z})$ is the support function of $\mathcal{Z}$ and $h_*^k(x,v)$ is the partial concave conjugate function of $h_k(x,z)$.

Table 6.1: Table of reformulations for uncertainty sets (Table 1 in [8])

Uncertainty region	$\mathcal{Z}$	Support function $\delta^*(v\|\mathcal{Z})$
Box	$\\|z\\|_\infty \leq \rho$	$\rho\\|v\\|_1$
Ball	$\\|z\\|_2 \leq \rho$	$\rho\\|v\\|_2$
Polyhedral	$b - Bz \geq 0$	$\inf_{w\geq 0: B^T w = v} b^T w$
Cone	$b - Bz \in C$	$\inf_{w\in C^*: B^T w = v} b^T w$
KL-Divergence	$\sum_l z_l \ln\left(\frac{z_l}{z_l^0}\right) \leq \rho$	$\inf_{u\geq 0} \sum_l z_l^0 u e^{(v_l/u)-1} + \rho u$
Geometric prog.	$\sum_i \alpha_i e^{(d_i)^T z} \leq \rho$	$\inf_{u\geq 0,\, w\geq 0: \sum_i d_i w_i = v} \sum_i \{w_i \ln\left(\frac{w_i}{\alpha_i u}\right) - w_i\} + \rho u$
Intersection	$\mathcal{Z} = \cap_i \mathcal{Z}_i$	$\inf_{\{w_i\}: \sum_i w_i = v} \sum_i \delta^*(w^i\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_k = \{z\|\\|z\\|_k \leq \rho_k\}$ $k = 1, 2$	$\inf_{(w^1,w^2): w^1+w^2=v} \rho_1\\|w^1\\|_\infty + \rho_2\\|w^2\\|_2$
Minkowski sum	$\mathcal{Z} = \mathcal{Z}_1 + \cdots + \mathcal{Z}_K$	$\sum_i \delta^*(v\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_1 = \{z \mid \\|z\\|_\infty \leq \rho_\infty\}$ $\mathcal{Z}_2 = \{z \mid \\|z\\|_2 \leq \rho_2\}$	$\rho_\infty\\|v\\|_1 + \rho_2\\|v\\|_2$
Convex hull	$\mathcal{Z} = \mathrm{conv}(\mathcal{Z}_1, \ldots, \mathcal{Z}_K)$	$\max_i \delta^*(v\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_1 = \{z \mid \\|z\\|_\infty \leq \rho_\infty\}$ $\mathcal{Z}_2 = \{z \mid \|z - z^0\\|_2 \leq \rho_2\}$	$\max\{\rho_\infty\\|v\\|_1, (z^0)^T v + \rho_2\\|v\|\\|_2\}$

Table 6.2: Table of reformulations for constraint functions (Table 2 in [8])

Constraint function	$g(x,z)$	Partial concave conjugate $g_*(x,v)$
Linear in z	$z^T g(x)$	$\begin{cases} 0 & \text{if } v = g(x) \\ -\infty & \text{otherwise} \end{cases}$
Concave in z, separable in z and x	$g(z)^T x$	$\sup_{\{s^i\}_{i=1}^n : \sum_{i=1}^n s^i = v} \sum_i x_i (g_i)_*(s^i/x_i)$
Example	$-\sum_i \frac{1}{2}(z^T Q_i z) x_i$	$\sup_{\{s^i\}_{i=1}^n : \sum_{i=1}^n s^i = v} -\frac{1}{2} \sum_{i=1}^n \frac{(s^i)^T Q_i^{-1} s^i}{x_i}$
Sum of functions	$\sum_i g_i(x,z)$	$\sup_{\{s^i\}_{i=1}^n : \sum_i s^i = v} \sum_i (g_i)_*(x, s^i)$
Sum of separable functions	$\sum_i g_i(x, z_i)$	$\sum_{i=1}^n (g_i)_*(x, v_i)$
Example	$-\sum_{i=1}^m x_i^{z_i}$, $x_i > 1, 0 \leq z \leq 1$	$\begin{cases} \sum_{i=1}^m \left(\frac{v_i}{\ln x_i} \ln \frac{-v_i}{\ln x_i} - \frac{v_i}{\ln x_i} \right) & \text{if } v \leq 0 \\ -\infty & \text{otherwise} \end{cases}$

Wasserstein distance based models

(see separate set of slides by D. Kuhn)

(see an example of implementation in
RSOME documentation)