Distirbutional Robust Optimization Theorem and Application

Lu Sun 228002579

2019.05.08

Content

Content

Robust Optimization (RO)

- "Quite small(just 0.1%) pertubations of 'obviously uncertain' data coefficients can make the 'nominal' optimal solution x^* heavily infeasible and thus practically meaningless."
- Only 2% error in the estimation of the conversion can results in 22% drop for profit.
- The method RO comes out to overcome the dependence on parameters.

Popularity of RO

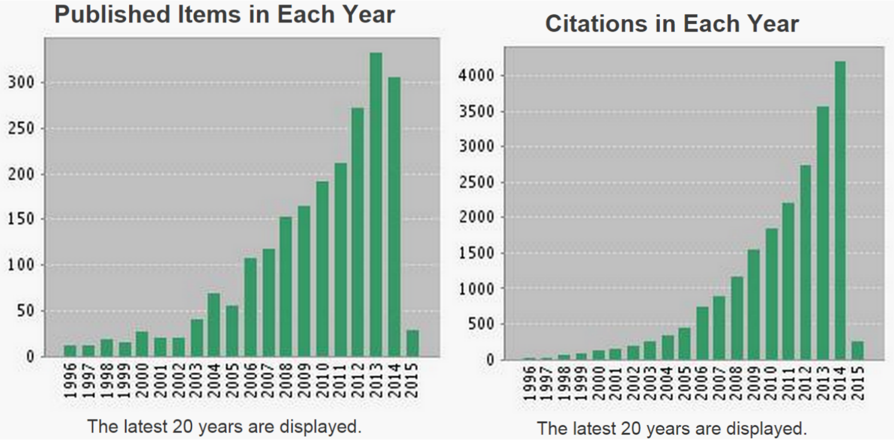


Figura: Rise in popularity of "robust optimization" in the scientific literature

Characteristic of RO

- Uncertainty set of parameters is the most important part of RO.
- The number of values or parameters is large enough
- Eg. semi-infinite dimension problem
  (number of decision variables is infinite and of constraints is finite)

Distributional Robust Optimization (DRO)

- Traditional RO is called Static Robust Optimization(SRO)
- DRO comes from SRO model
- DRO is used for stochastic programming (SP)
- Uncertainty set of distributional functions :
  the most important part of DRO.

Content

SRO

Definition

$$\min_{x\in X} \max_{\epsilon\in D} h(x,\epsilon) \quad or \quad \begin{array}{ll} \min_{x\in X} & h_1(x) \\ s.t. & h_2(x,\epsilon) \leq 0, \forall \epsilon \in D \end{array} \tag{1}$$

Here, D includes any feasible parameter and D,h are not related to probability.

Remark

- We can always choose one in set D to be the worst case
- The solution in worst case may act in-feasible in real life
- SRO is too conservation.

SP

Definition

$$\min_{x \in X} \mathbb{E}_{F_\epsilon}[h(x, \epsilon)] \tag{2}$$

Here, F_ϵ is a fixed distribution.

Remark

- The distribution F_ϵ is not available
- Sub-optimal or meaningless solution can come out
- SP problem need to decide an infinite time's activity
- RO method can be used on SP problem

DRO

Definition

$$\min_{x \in X} \max_{F_\epsilon \in D} \mathbb{E}_{F_\epsilon}[h(x,\epsilon)] \tag{3}$$

Here, D is the uncertainty set of distributions, which only includes part of distributional functions.

Remark

- $D, \mathbb{E}_{F_\epsilon}[h(x,\epsilon)]$ are related to probability
- An example of set D:

$$D := \{p : dist|p - p_{ref}| \leq \sigma\}$$

Here, p_{ref} is based on historical data choosing from basic known distributions(e.g. Guassian); Wasserstein metric is used for probability distance; σ is choosen to suit well for some percentage interval.

Content

Methods to construct D

Definition

$$\min_{x \in X} \sup_{F \in D} \mathbb{E}_F[h(x, \epsilon)] \tag{4}$$

- Moment based models
  - **1.**mean and support models
  - **2.**mean and variance models
  - **3.**moment uncertainty models
- Scenario-based models
- Wasserstein distance based models

Definition

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

Here, M is the set of all probability measures on the measurable space $(\mathbb{R}^m, B)$, B is the Borel $\sigma-$algebra on $\mathbb{R}^m$, and Z is a Borel set in $\mathbb{R}^m$.

Rewrite half part of equation(4)

Definition

$$\begin{array}{cl} \max_{F\in M} & \int_Z h(x,\epsilon)dF(\epsilon) \\ s.t. & \int_Z dF(\epsilon) = 1 \\ & \int_Z \epsilon dF(\epsilon) = \mu \end{array} \tag{5}$$

Here, we further assume that $h(x,.)$ is real-valued measurable in $(\mathbb{R}^m, B)$.

Theorem (3.1)

Let $D(Z,\mu)$ be a distribution set for which there exists a feasible solution $F_0 \in D(Z,\mu)$, then the moment problem (5) is equivalent to the following robust optimization problem:

$$\min_{q} \sup_{z\in Z} h(x,z) + (\mu - z)^T q \tag{6}$$

proof:

Step 1: Construct Lagrangean equation

$$\begin{aligned} L(F,r,q) &= \int_Z h(x,\epsilon)dF(\epsilon) + r(1 - \int_Z dF(\epsilon)) + q^T(\mu - \int_Z \epsilon dF(\epsilon)) \\ &= r + \mu^T q + \int_Z (h(x,\epsilon) - r - q^T\epsilon)dF(\epsilon) \end{aligned}$$

Step 2: Analysis properties of L

$$\begin{array}{rl} & \sup_R \inf_{r,q} L(F,r,q) \\ \leq & \inf_{r,q} \sup_F L(F,r,q) \\ = & \inf_{r,q} \begin{cases} r + \mu^T q & if \;\; h(x,z) - r - q^T z \leq 0, \forall z \in Z \\ \infty & otherwise \end{cases} \end{array}$$

Here, we have the assumption that $D(Z, \mu) \neq \emptyset$

Step 3: Translate the right part into program

$$\begin{array}{cl} \min_{r,q} & \mu^T q + r \\ s.t. & z^T + r \geq h(x,z), \forall z \in Z, \end{array}$$

Step 4: Translate into (6)

$$z^T + r = h(x,z), \forall z \in Z.$$

$$\min_q \sup_{z \in Z} h(x,z) + (\mu - z)^T q$$

Mean and Support Models

Theorem (3.2)

Let $Z \in \mathbb{R}^m$ be a Borel set, and F_0 be some feasible distribution according to $D(Z, \mu)$, then problem (5) is equivalent to the following nite dimensional optimization problem

$$\begin{array}{cl} \max_{p,\{z_i\}_{i=1}^{m+1}} & \sum_{i=1}^{m+1} p_i h(x, z_i) \\ s.t. & \sum_{i=1}^{m+1} p_i = 1 \& p \geq 0 \\ & \sum_{i=1}^{m+1} p_i z_i = \mu \\ & z_i \in Z, \forall i = 1, ..., m+1, \end{array} \tag{7}$$

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

Mean and Support Models

Theorem (3.3)

When Z is a convex set and $h(x,z) := \max_{k=1,\dots,K} h_k(x,z)$ for some K with each $h_k(x,z)$ a concave function of z, then problem (5) is equivalent to

$$\begin{array}{rl} \max_{p,\{z_k\}_{k=1}^K} & \sum_{k=1}^K p_k h_k(x, z_k) \\ s.t. & \sum_{k=1}^K p_k = 1, p \geq 0 \\ & \sum_{k=1}^K p_k z_k = \mu \\ & z_k \in Z, \forall k = 1, \dots, K. \end{array} \tag{8}$$

Corollary

When Z is a convex set and h(x,z) is a concave function of z, then the DRO problem presented in (4) is equivalent to

$$\min_{x \in X} h(x, \mu) \tag{9}$$

Mean and Support Models

Theorem

Let $D(Z, \mu)$ be a distribution set for which there exists a feasible solution $F_0 \in D(Z, \mu)$, the DRO problem presented in (4) is equivalent to the following robust optimization problem:

$$\min_{x \in X, q} \sup_{z \in Z} h(x, z) + (\mu - z)^T q \tag{10}$$

Moreover, the problem can be reformulated as follows when 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:

$$\begin{array}{ll} \min_{x, q, \{v_k\}_k, t} & t \\ s.t. & t \geq \delta^*(v_k | Z) + \mu^T q - h_*^k(x, v_k + q), \forall k \end{array} \tag{11}$$

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

Mean and Variance Models

Definition

$$D(\mu, \sigma^2) := \left\{ F \in M \middle| \begin{array}{l} \mathbb{P}(\epsilon \in \mathbb{R}) = 1 \\ \mathbb{E}[\epsilon] = \mu \\ \mathbb{E}[(\epsilon - \mu)^2] = \sigma^2 \end{array} \right\}$$

Step 1: Assume $Z' := \{z' \in \mathbb{R}^2 | z'_2 = (z'_1 - \mu)^2\}$
Step 2: Equivalent form $D(\mu, \sigma) = D(Z', [\mu, \sigma^2]^T)$
Step 3: Change into non-linear robust optimization model

$$\begin{array}{ll} \min_{x \in X, q_1, q_2 \geq 0, t} & t \\ s.t. & t \geq \sup_{z_1 \in \mathbb{R}} h_k(x, z) + \mu q_1 + (\sigma^2 - \mu^2) q_2 \\ & -(q_1 - 2q_2\mu) z_1 - q_2 z_1^2, \forall k, \end{array} \tag{12}$$

when $h(x, z) := \max_k h_k(x, z)$

Moment Uncertainty Models

Definition

$$\min_{x \in X} \sup_{\mu \in U, F \in D(Z,\mu)} \mathbb{E}_F[h(x,z)] \tag{13}$$

Corollary

Let $D(Z,\mu)$ be a distribution set and $U \in \mathbb{R}^m$ be a bounded and convex uncertainty set for the moment vector μ. Given that for all $\mu \in U$, there exists an $F \in D(Z,\mu)$, the DRO problem presented in (14) is equivalent to the following robust optimization problem:

$$\min_{x \in X, q} \sup_{z \in Z} h(x,z) - z^T q + \delta^*(q|U) \tag{14}$$

Moment Uncertainty Models

Moreover, the problem can be reformulated as follows when 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:

$$\begin{array}{ll} \min_{x \in X, q, \{v_k\}_k, t} & t + \delta^*(q|U) \\ s.t. & t \geq \delta^*(v_k|Z) - h_*^k(x, v_k + q), \forall k \end{array} \tag{15}$$

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

Scenario-Based Models

Basic form

$$\min_{x \in X} \sup_{p \in U} \sum_{k=1}^{K} p_k h(x, z^k),$$

where$Z := \{z^1, z^2, ..., z^K\}$ is a set of scenarios, $p \in \mathbb{R}^K$ is a vector describing the probability of obtaining each of the K scenarios for ϵ while $U \subseteq \{p \in \mathbb{R}^K | p \geq 0, \sum_{k=1}^{K} p_k = 1\}$ is the uncertainty set for the distribution, which can also be calibrated using historical data.

Wassertein Distance Based Models

- **Finite sample guarantee:** The property that the optimal value of the DRO model is guaranteed with high probability to bound from above the expected cost when a finte number of i.i.d. realizations have been observed.
- **Consistency:** The property that the optimal solution will eventually converge to the optimal solution of the stocastic program(2) as more i.i.d. realizations are used to construct the distribution set D.
- **Tractability:** The DRO model can be solved using convex optimization algorithms for a large class of problems.

Content

Example 1

Problem

$$\begin{array}{lll} \min & \max_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}[y(\bar{z})] & \\ s.t. & y(z) \geq z, & \forall z \in Z \\ & y(z) \geq -z, & \forall z \in Z \end{array}$$

where $\mathbb{P}$ is the distribution of random variable $\bar{z}$, Z is the support set of $\bar{z}$, which is set to be $[-2,2]$, and $\mathbb{F}$ is the ambiguity set that characterizes a collection of distributions, which is expressed in the equation below.

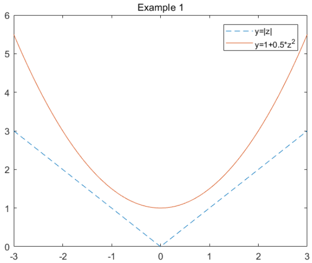
$$\mathbb{F} = \left\{ \mathbb{P} \in P_0(\mathbb{R}) : \begin{array}{l} \bar{z} \in \mathbb{R} \\ \mathbb{E}_{\mathbb{P}}(\bar{z}) = 0 \\ \mathbb{E}_{\mathbb{P}}(\bar{z}^2) \leq 1 \\ \mathbb{P}\{\bar{z} \in\} = \mathbb{P}\{-2 \leq \bar{Z} \leq 2\} = 1 \end{array} \right\}$$

Example 1

Solution

$$\mathbb{G} = \left\{ \mathbb{Q} \in P_0(\mathbb{R}^2) : \begin{array}{l} \bar{z} \in \mathbb{R}, \bar{u} \in \mathbb{R} \\ \mathbb{E}_{\mathbb{Q}}(\bar{z}) = 0 \\ \mathbb{E}_{\mathbb{Q}}(\bar{u}) \leq 1 \\ \mathbb{Q}\left\{ \begin{array}{c} -2 \leq \bar{z} \leq 2 \\ \bar{z}^2 \leq \bar{u} \leq 4 \end{array} \right\} = 1 \end{array} \right\}$$

$\bar{y} = \frac{1+z^2}{2}$



Example 2

Problem

$$\mathbb{F} = \left\{ \mathbb{P} \in P_0(\mathbb{R}) : \begin{array}{l} \bar{z} \in \mathbb{R} \\ \mathbb{E}_{\mathbb{P}}(\bar{z}) = 0 \\ \mathbb{E}_{\mathbb{P}}(\bar{z}^2) \leq 1 \\ \mathbb{P}\{\bar{z} \in Z\} = \mathbb{P}\{-2 \leq \bar{z} \leq 2\} = 1 \\ \mathbb{P}\{\bar{z} \in Z_1\} = \mathbb{P}\{-1 \leq \bar{z} \leq 1\} = 0.9 \\ \mathbb{P}\{\bar{z} \in Z_2\} = \mathbb{P}\{-0.5 \leq \bar{z} \leq 0.5\} \in [0.6, 0.7] \end{array} \right\}$$

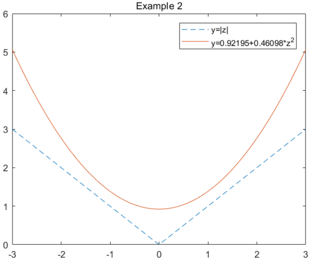
Example 2

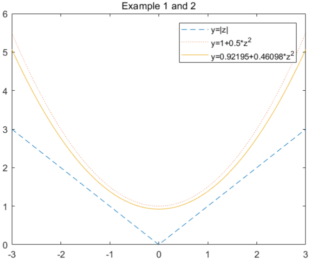
Solution

$$\mathbb{G} = \left\{ \mathbb{Q} \in P_0(\mathbb{R}^2) : \begin{array}{l} \bar{z} \in \mathbb{R}, \bar{u} \in \mathbb{R} \\ \mathbb{E}_{\mathbb{Q}}(\bar{z}) = 0 \\ \mathbb{E}_{\mathbb{Q}}(\bar{u}) \leq 1 \\ \mathbb{Q}\left\{ \begin{array}{c} -2 \leq \bar{z} \leq 2 \\ \bar{z}^2 \leq \bar{u} \leq 4 \end{array} \right\} = 1 \\ \mathbb{Q}\left\{ \begin{array}{c} -1 \leq \bar{z} \leq 1 \\ \bar{z}^2 \leq \bar{u} \leq 1 \end{array} \right\} = 0.9 \\ \mathbb{Q}\left\{ \begin{array}{c} -0.5 \leq \bar{z} \leq 0.5 \\ \bar{z}^2 \leq \bar{u} \leq 0.25 \end{array} \right\} \in [0.6, 0.7] \end{array} \right\}$$

Example 2

$$\bar{y} = 0.9920 + 0.4610 * z^2$$





Reference I

Thank You!