Quantitative Stability Analysis in Distributionally Robust Optimization

Huifu Xu

School of Mathematical Sciences, University of Southampton, UK.

Presented at BIRS Workshop, Banff, Canada, 5 March 2018

The one-stage DRO model

Consider the following distributionally robust optimization problem:

$$\text{(DRO)} \qquad \begin{array}{cl} \min\limits_{x} \sup\limits_{P\in\mathcal{P}} & \mathbb{E}_P[f(x,\xi)] \\ \text{s.t.} & x \in X, \end{array} \tag{1}$$

where $\mathcal{P}$ is a set of distributions which contains/approximates the true probability distribution of random variable ξ.

References: Scarf (1958), Dupacová (1987), Shapiro and Kleywegt (2002), Prékopa (1995), Bertsimas and Popescu (2005), Zhu and Fukushima (2006), Goh and Sim (2010), Delage and Ye (2010), Wiesemann, Kuhn and Sim (2015), Mohajerin Esfahani and Kuhn (2017) . . .

Part I: Quantifying change of the ambiguity set

Construction of ambiguity sets

- Moment conditions
- Mixture distribution – contamination distribution
- Bayesian inference
- Marginal distributions
- Samples & empirical data
- Maximal likelihood
- Kullback-Leibler divergence
- Wasserstein ball, ζ-ball
- ...

Construction of ambiguity sets

- Moment conditions
- Mixture distribution – contamination distribution
- Bayesian inference
- Marginal distributions
- Samples & empirical data
- Maximal likelihood
- Kullback-Leibler divergence
- Wasserstein ball, ζ-ball
- $\ldots$

Ambiguity set defined through moment conditions

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}, \quad (2)$$

where $\psi_i : \Xi \to \mathbb{R}$, $i = 1, \cdots, q$, are measurable functions.

Ambiguity set defined through moment conditions

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}, \tag{2}$$

where $\psi_i : \Xi \to \mathbb{R}$, $i = 1, \cdots, q$, are measurable functions.

Question: What if the true μ_i is not known?

Ambiguity set defined through moment conditions

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}, \tag{2}$$

where $\psi_i : \Xi \to \mathbb{R}$, $i = 1, \cdots, q$, are measurable functions.

Question: What if the true μ_i is not known?

$$\mathcal{P}_N := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i^N, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i^N, & \text{for } i = p+1, \cdots, q \end{array} \right\} \tag{3}$$

where μ_i^N is often constructed through samples.

Difference between $\mathcal{P}$ and $\mathcal{P}_N$?

Let

$$\langle P, \psi \rangle := \int_{\Xi} \psi(\xi) P(d\xi).$$

We can write $\mathcal{P}$ and $\mathcal{P}_N$ as

$$\mathcal{P} = \{P \in \mathscr{P}(\Xi) : \langle P, \psi_E(\xi) \rangle = \mu_E, \langle P, \psi_I(\xi) \rangle \leq \mu_I\}$$

and

$$\mathcal{P}_N = \{P \in \mathscr{P}(\Xi) : \langle P, \psi_E(\xi) \rangle = \mu_E^N, \langle P, \psi_I(\omega) \rangle \leq \mu_I^N\}.$$

Difference between $\mathcal{P}$ and $\mathcal{P}_N$?

Let

$$\langle P, \psi \rangle := \int_{\Xi} \psi(\xi) P(d\xi).$$

We can write $\mathcal{P}$ and $\mathcal{P}_N$ as

$$\mathcal{P} = \{P \in \mathscr{P}(\Xi) : \langle P, \psi_E(\xi) \rangle = \mu_E, \langle P, \psi_I(\xi) \rangle \le \mu_I\}$$

and

$$\mathcal{P}_N = \{P \in \mathscr{P}(\Xi) : \langle P, \psi_E(\xi) \rangle = \mu_E^N, \langle P, \psi_I(\omega) \rangle \le \mu_I^N\}.$$

Q: Does $\mathcal{P}_N$ approximate $\mathcal{P}$? How?

Measuring the distance between probability measures: Metrics of ζ-structure

Definition (ζ-metric)

Let $\mathscr{P}(\Xi)$ denote the set of all probability distributions/measures over space $(\Xi, \mathscr{B})$. Let $P, Q \in \mathscr{P}(\Xi)$ and $\mathscr{G}$ be a family of real-valued measurable functions on Ξ. Define

$$d_{\mathscr{G}}(P, Q) := \sup_{g \in \mathscr{G}} |\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]| \,.$$

The (semi-) distance defined as such is called a metric with $\zeta-$structure.

Measuring the distance between probability measures: Metrics of ζ-structure

Definition (ζ-metric)

Let $\mathscr{P}(\Xi)$ denote the set of all probability distributions/measures over space $(\Xi, \mathscr{B})$. Let $P, Q \in \mathscr{P}(\Xi)$ and $\mathscr{G}$ be a family of real-valued measurable functions on Ξ. Define

$$d_{\mathscr{G}}(P, Q) := \sup_{g \in \mathscr{G}} \left| \mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)] \right| .$$

The (semi-) distance defined as such is called a metric with $\zeta-$structure.

Qustion: How to choose $\mathscr{G}$?

Metrics of ζ-structure

$$d_{\mathscr{G}}(P,Q) := \sup_{g \in \mathscr{G}} \left| \mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)] \right| .$$

- Total variation metric (denoted by d_{TV}):
  $\mathscr{G} := \left\{ g : \sup_{\xi \in \Xi} |g(\xi)| \leq 1 \right\}.$

Metrics of ζ-structure

$$d_{\mathscr{G}}(P, Q) := \sup_{g \in \mathscr{G}} \left| \mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)] \right| .$$

- Total variation metric (denoted by d_{TV}):
  $\mathscr{G} := \left\{ g : \sup_{\xi \in \Xi} |g(\xi)| \leq 1 \right\}$.
- Kantorovich/ Wasserstein metric (denoted by d_K):
  $\mathscr{G} = \{ g : g \text{ is Lipschitz continuous with } L_1(g) \leq 1 \}$.

Metrics of ζ-structure

$$d_{\mathscr{G}}(P,Q) := \sup_{g\in\mathscr{G}} \left|\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]\right|.$$

- Total variation metric (denoted by d_{TV}):
  $\mathscr{G} := \left\{g : \sup_{\xi\in\Xi} |g(\xi)| \leq 1\right\}.$
- Kantorovich/ Wasserstein metric (denoted by d_K):
  $\mathscr{G} = \{g : g \text{ is Lipschitz continuous with } L_1(g) \leq 1\}.$
- Bounded Lipschitz metric (denoted by d_{BL}):
  $\mathscr{G} := \left\{g : \sup_{\xi\in\Xi} |g(\xi)| \leq 1,\ g \text{ is Lipschitz continuous with } L_1(g) \leq 1\right\}$
  where $L_1(g)$ denotes the Lipschitz modulus.

Hoffman's lemma for moment problem

Hoffman's lemma (Sun and Xu (2015))

There exists a positive constant C depending on ψ such that

$$d_{TV}(Q, \mathcal{P}) \leq C[(\|\mathbb{E}_Q[\psi_I(\xi)] - \mu_I)_+\| + \|\mathbb{E}_Q[\psi_E(\xi)] - \mu_E\|],$$

for any $Q \in \mathscr{P}(\Xi)$, where $\|\cdot\|$ denotes the Euclidean norm, $(a)_+ = \max(0, a)$ and the maximum is taken componentwise.

Quantifying the difference between $\mathcal{P}$ and $\mathcal{P}_N$

Proposition 2.1

There exists a positive constant C depending on ψ such that

$$\begin{aligned}\mathbb{H}_{TV}(\mathcal{P}_N, \mathcal{P}) \quad &\leq \quad C[\max(\|(\mu_I^N - \mu_I)_+\|, \|(\mu_I - \mu_I^N)_+\|) \\ &\quad\quad + \|\mu_E^N - \mu_E\|],\end{aligned}$$

where C is defined as in the Hoffman's lemma and $\mathbb{H}_{TV}$ denotes the Hausdorff distance under the total variation metric.

Proof. Let $Q \in \mathcal{P}_N$. By the Hoffman lemma, there is a constant C such that

$$d_{TV}(Q, \mathcal{P}) \quad \leq \quad C\,[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))] - \mu_I)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))] - \mu_E\|]$$

Proof. Let $Q \in \mathcal{P}_N$. By the Hoffman lemma, there is a constant C such that

$$\begin{aligned} d_{TV}(Q, \mathcal{P}) &\leq C\left[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))] - \mu_I)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))] - \mu_E\|\right] \\ &\leq C\Big[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))] - \mu_I^N)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))] - \mu_E^N\| \\ &\quad + \|(\mu_I^N - \mu_I)_+\| + \|\mu_E^N - \mu_E\|\Big] \end{aligned}$$

Proof. Let $Q \in \mathcal{P}_N$. By the Hoffman lemma, there is a constant C such that

$$\begin{aligned} d_{TV}(Q, \mathcal{P}) &\leq C\,[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))] - \mu_I)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))] - \mu_E\|] \\ &\leq C\,\Big[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))] - \mu_I^N)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))] - \mu_E^N\| \\ &\quad + \|(\mu_I^N - \mu_I)_+\| + \|\mu_E^N - \mu_E\|\Big] \\ &= C[\|(\mu_I^N - \mu_I)_+\| + \|\mu_E^N - \mu_E\|], \end{aligned}$$

because $(a+b)_+ \leq (a)_+ + (b)_+$.

Proof. Let $Q \in \mathcal{P}_N$. By the Hoffman lemma, there is a constant C such that

$$\begin{aligned} d_{TV}(Q,\mathcal{P}) &\le C\,[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))]-\mu_I)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))]-\mu_E\|] \\ &\le C\left[\|(\mathbb{E}_Q[\psi_I(\xi(\omega))]-\mu_I^N)_+\| + \|\mathbb{E}_Q[\psi_E(\xi(\omega))]-\mu_E^N\|\right. \\ &\quad \left. +\|(\mu_I^N-\mu_I)_+\| + \|\mu_E^N-\mu_E\|\right] \\ &= C[\|(\mu_I^N-\mu_I)_+\| + \|\mu_E^N-\mu_E\|], \end{aligned}$$

because $(a+b)_+ \le (a)_+ + (b)_+$. This gives

$$\mathbb{D}_{TV}(\mathcal{P}_N,\mathcal{P}) = \sup_{Q\in\mathcal{P}_N} d_{TV}(Q,\mathcal{P}) \le C(\|(\mu_I^N-\mu_I)_+\| + \|\mu_E^N-\mu_E\|).$$

Likewise,

$$\mathbb{D}_{TV}(\mathcal{P}, \mathcal{P}_N) \leq C(\|(\mu_I - \mu_I^N)_+\| + \|\mu_E - \mu_E^N\|).$$

Likewise,

$$\mathbb{D}_{TV}(\mathcal{P}, \mathcal{P}_N) \leq C(\|(\mu_I - \mu_I^N)_+\| + \|\mu_E - \mu_E^N\|).$$

Combining the inequalities, we have

$$\mathbb{H}_{TV}(\mathcal{P}_N, \mathcal{P}) \leq C\left[\max(\|(\mu_I^N - \mu_I)_+\|, \|(\mu_I - \mu_I^N)_+\|) + \|\mu_E^N - \mu_E\|\right].$$

Generalizations

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}$$

Generalizations

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}$$

$$\Downarrow$$

- $\mathcal{P}$ depends on the decision variable x and other parameter u

$$\mathcal{P}(x, u) := \{P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\Psi(x, u, \xi)] \in \mathcal{K}\},$$

where Ψ is a mapping consisting of matrices and $\mathcal{K}$ is a cone in the respective matrix spaces.

Generalizations

$$\mathcal{P} := \left\{ P : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\psi_i(\xi)] \le \mu_i, & \text{for } i = p+1, \cdots, q \end{array} \right\}$$

$$\Downarrow$$

- $\mathcal{P}$ depends on the decision variable x and other parameter u

$$\mathcal{P}(x, u) := \{P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\Psi(x, u, \xi)] \in \mathcal{K}\},$$

  where Ψ is a mapping consisting of matrices and $\mathcal{K}$ is a cone in the respective matrix spaces.
- Measuring the distance under ζ-metric.

Slater condition

There exist $P_0 \in \mathscr{P}(\Xi)$ and a constant $\alpha > 0$ such that

$$\langle P_0, \Psi(x_0, u_0, \xi)\rangle + \alpha\mathbb{B} \subset \mathcal{K}, \tag{4}$$

where $\mathbb{B}$ is the unit ball in the space that $\mathcal{K}$ is defined.

Hoffman's lemma under ζ-metric

Proposition 2.2 (Liu, Pichler and Xu (2017))

Under the Slater condition (4)

$$d_{\mathscr{G}}(Q, \mathcal{P}(x,u)) \le \frac{\Delta}{\alpha} \inf_{w \in \mathcal{K}} \|w - \langle Q, \Psi(x,u,\xi)\rangle\| \tag{5}$$

for any $Q \in \mathscr{P}(\Xi)$ *and* (x,u) *close to* (x_0, u_0), *where* α *is the positive constant defined in the Slater condition and*

$$\Delta := \max_{P \in \mathscr{P}(\Xi)} d_{\mathscr{G}}(P, P_0). \tag{6}$$

Hoffman's lemma under ζ-metric

Proposition 2.2 (Liu, Pichler and Xu (2017))

Under the Slater condition (4)

$$d_{\mathscr{G}}(Q, \mathcal{P}(x,u)) \le \frac{\Delta}{\alpha} \inf_{w \in \mathcal{K}} \|w - \langle Q, \Psi(x,u,\xi)\rangle\| \tag{5}$$

for any $Q \in \mathscr{P}(\Xi)$ *and* (x,u) *close to* (x_0, u_0)*, where* α *is the positive constant defined in the Slater condition and*

$$\Delta := \max_{P \in \mathscr{P}(\Xi)} d_{\mathscr{G}}(P, P_0). \tag{6}$$

Question:

How to estimate Δ?

$$\Delta := \max_{P\in\mathscr{P}(\Xi)} d_{\mathscr{G}}(P, P_0).$$

- $\Delta \leq 2$ under the total variation metric d_{TV} and the Bounded Lipschitz metric d_{BL};

$$\Delta := \max_{P \in \mathscr{P}(\Xi)} d_{\mathscr{G}}(P, P_0).$$

- $\Delta \leq 2$ under the total variation metric d_{TV} and the Bounded Lipschitz metric d_{BL};
- $\Delta \leq \mathrm{diam}(\Xi)$ under the Kantorovich/ Wasserstein metric when the support set Ξ is bounded.

Recall

$$\mathcal{P}(x,u) := \{P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\Psi(x,u,\xi)] \in \mathcal{K}\}$$

vs

$$\mathcal{P}(x',u') := \left\{P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\Psi(x',u',\xi)] \in \mathcal{K}\right\}.$$

Theorem 2.1

Assume:

(a) *there exist positive constants* $\gamma \in \mathbb{R}_+$ *and* $\nu_1, \nu_2 \in (0, 1]$ *such that*

$$\|\Psi(x, u, \xi) - \Psi(x', u', \xi)\| \leq \gamma(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2})$$

for all $\xi \in \Xi$ *and* $(x, u), (x', u') \in X \times U$ *close to* (x_0, u_0).

Theorem 2.1

Assume:

(a) *there exist positive constants* $\gamma \in \mathbb{R}_+$ *and* $\nu_1, \nu_2 \in (0, 1]$ *such that*

$$\|\Psi(x, u, \xi) - \Psi(x', u', \xi)\| \leq \gamma(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2})$$

for all $\xi \in \Xi$ *and* $(x, u), (x', u') \in X \times U$ *close to* (x_0, u_0).

(b) Ξ *is a compact set.*

Theorem 2.1

Assume:

(a) *there exist positive constants* $\gamma \in \mathbb{R}_+$ *and* $\nu_1, \nu_2 \in (0, 1]$ *such that*

$$\|\Psi(x, u, \xi) - \Psi(x', u', \xi)\| \leq \gamma(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2})$$

for all $\xi \in \Xi$ *and* $(x, u), (x', u') \in X \times U$ *close to* (x_0, u_0).

(b) Ξ *is a compact set.*

(c) *The* **Slater condition** *(4) is fulfilled.*

Theorem 2.1

Assume:

(a) *there exist positive constants* $\gamma \in \mathbb{R}_+$ *and* $\nu_1, \nu_2 \in (0, 1]$ *such that*

$$\|\Psi(x, u, \xi) - \Psi(x', u', \xi)\| \leq \gamma(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2})$$

for all $\xi \in \Xi$ *and* $(x, u), (x', u') \in X \times U$ *close to* (x_0, u_0).

(b) Ξ *is a compact set.*

(c) *The* **Slater condition** *(4) is fulfilled.*

Then there exists a positive constant C *such that*

$$\mathbb{H}_{\mathscr{G}}(\mathcal{P}(x, u), \mathcal{P}(x', u')) \leq C(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2}) \quad (7)$$

for any $(x, u), (x', u') \in X \times U$ *close to* (x_0, u_0).

Ambiguity set constructed through sample information: ζ-ball

Let $P \in \mathscr{P}(\Xi)$ and r be a positive number. We call the following set of probability distributions as ζ-ball:

$$\mathcal{B}(P, r) := \{P' \in \mathscr{P}(\Xi) : d_{\mathscr{G}}(P', P) \leq r\}, \tag{8}$$

where

$$d_{\mathscr{G}}(P, Q) := \sup_{g \in \mathscr{G}} \left|\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]\right|.$$

and P is a nominal distribution which may be an empirical probability distribution.

Quantifying change of the ζ-ball

Theorem 2.2 (Liu, Pichler and Xu (2017))

Let $\mathcal{B}(P, r)$ be the ζ-ball defined as in (8). For every P, $Q \in \mathscr{P}(\Xi)$ and r_1, $r_2 \in \mathbb{R}_+$, it holds that

$$\mathbb{H}_{\mathscr{G}}(\mathcal{B}(P, r_1), \mathcal{B}(Q, r_2)) \leq d_{\mathscr{G}}(P, Q) + |r_1 - r_2|, \tag{9}$$

where $\mathbb{H}_{\mathscr{G}}$ denotes the Hausdorff distance in $\mathscr{P}(\Xi)$ associated with ζ-metric.

Part II: Quantitative stability analysis of the DRO

Distributionally robust formulation

Consider the following distributionally robust problem:

$$(\mathrm{DRO}) \qquad \begin{array}{cl} \min\limits_{x} \sup\limits_{P \in \mathcal{P}} & \mathbb{E}_P[f(x,\xi)] \\ \text{s.t.} & x \in X, \end{array} \tag{10}$$

and its perturbation

$$\begin{array}{cl} \min\limits_{x} \sup\limits_{P \in \tilde{\mathcal{P}}} & \mathbb{E}_P[f(x,\xi)] \\ \text{s.t.} & x \in X. \end{array} \tag{11}$$

Distributionally robust formulation

Consider the following distributionally robust problem:

$$(\mathrm{DRO}) \qquad \begin{array}{ll} \min\limits_{x} \sup\limits_{P\in\mathcal{P}} & \mathbb{E}_P[f(x,\xi)] \\ \text{s.t.} & x \in X, \end{array} \tag{10}$$

and its perturbation

$$\begin{array}{ll} \min\limits_{x} \sup\limits_{P\in\tilde{\mathcal{P}}} & \mathbb{E}_P[f(x,\xi)] \\ \text{s.t.} & x \in X. \end{array} \tag{11}$$

Question

How does perturbation of of the ambiguity set $\mathcal{P}$ affect the optimal value and the optimal solution?

Stability of optimal value

Theorem 3.1 (Liu, Pichler and Xu (2017))

Let $\vartheta(\tilde{\mathcal{P}})$ and $\vartheta(\mathcal{P})$ denote the optimal value of the DRO and its perturbation. Then the following assertions hold:

(i)

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq \mathbb{H}_{\mathscr{G}}(\tilde{\mathcal{P}}, \mathcal{P})$$

where $\mathbb{H}_{\mathscr{G}}$ is the Hausdorff distance under ζ-metric with

$$\mathscr{G} := \{f(x, \cdot) : x \in X\}.$$

In particular, if $\mathcal{P} = \mathcal{B}(P, r), \quad \tilde{\mathcal{P}} = \mathcal{B}(\tilde{P}, \tilde{r})$, then

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq d_{\mathscr{G}}(P, \tilde{P}) + |r - \tilde{r}|. \tag{12}$$

In particular, if $\mathcal{P} = \mathcal{B}(P, r), \quad \tilde{\mathcal{P}} = \mathcal{B}(\tilde{P}, \tilde{r})$, then

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq d_{\mathscr{G}}(P, \tilde{P}) + |r - \tilde{r}|. \tag{12}$$

- If the functions in the set $\mathscr{G}$ are Lipschitz continuous with modulus κ, then

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq \kappa d_K(P, \tilde{P}) + |r - \tilde{r}|. \tag{13}$$

where d_K denotes the Kantorovich/Wasserstein metric.

In particular, if $\mathcal{P} = \mathcal{B}(P, r), \quad \tilde{\mathcal{P}} = \mathcal{B}(\tilde{P}, \tilde{r})$, then

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq d_{\mathscr{G}}(P, \tilde{P}) + |r - \tilde{r}|. \tag{12}$$

- If the functions in the set $\mathscr{G}$ are Lipschitz continuous with modulus κ, then
$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq \kappa d_K(P, \tilde{P}) + |r - \tilde{r}|. \tag{13}$$
where d_K denotes the Kantorovich/Wasserstein metric.
- If the functions in $\mathscr{G}$ are bounded by a positive constant C, then
$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| \leq C d_{TV}(P, \tilde{P}) + |r - \tilde{r}|. \tag{14}$$

Proof. Let

$$v(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

Proof. Let

$$v(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x \in X} v(x) - \sup_{x \in X} \tilde{v}(x) \right| \leq \sup_{x \in X} |v(x) - \tilde{v}(x)|\,.$$

Proof. Let

$$v(x) := \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P\in\tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x\in X} v(x) - \sup_{x\in X} \tilde{v}(x) \right| \leq \sup_{x\in X} \left| v(x) - \tilde{v}(x) \right|.$$

$$v(x) - \tilde{v}(x) \quad = \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] - \sup_{P\in\tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)]$$

Proof. Let

$$v(x) := \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P\in\tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x\in X} v(x) - \sup_{x\in X} \tilde{v}(x) \right| \le \sup_{x\in X} |v(x) - \tilde{v}(x)| .$$

$$\begin{aligned} v(x) - \tilde{v}(x) &= \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] - \sup_{P\in\tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] \\ &= \sup_{P\in\mathcal{P}} \inf_{\tilde{P}\in\tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)] \end{aligned}$$

Proof. Let

$$v(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x \in X} v(x) - \sup_{x \in X} \tilde{v}(x) \right| \le \sup_{x \in X} |v(x) - \tilde{v}(x)| \,.$$

$$\begin{aligned} v(x) - \tilde{v}(x) &= \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)] - \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] \\ &= \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)] \\ &\le \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \sup_{x \in X} |\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)]| \end{aligned}$$

Proof. Let

$$v(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x \in X} v(x) - \sup_{x \in X} \tilde{v}(x) \right| \leq \sup_{x \in X} |v(x) - \tilde{v}(x)|.$$

$$\begin{aligned} v(x) - \tilde{v}(x) &= \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)] - \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] \\ &= \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)] \\ &\leq \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \sup_{x \in X} |\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)]| \\ &= \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} d_{\mathscr{G}}(P, \tilde{P}) \end{aligned}$$

Proof. Let

$$v(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)], \quad \tilde{v}(x) := \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)].$$

$$|\vartheta(\tilde{\mathcal{P}}) - \vartheta(\mathcal{P})| = \left| \sup_{x \in X} v(x) - \sup_{x \in X} \tilde{v}(x) \right| \le \sup_{x \in X} |v(x) - \tilde{v}(x)| .$$

$$\begin{aligned} v(x) - \tilde{v}(x) &= \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)] - \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] \\ &= \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)] \\ &\le \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} \sup_{x \in X} |\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_{\tilde{P}}[f(x,\xi)]| \\ &= \sup_{P \in \mathcal{P}} \inf_{\tilde{P} \in \tilde{\mathcal{P}}} d_{\mathscr{G}}(P, \tilde{P}) \\ &= \mathbb{D}_{\mathscr{G}}(\mathcal{P}, \tilde{\mathcal{P}}), \end{aligned}$$

Stability of the optimal solutions

Theorem 3.1

(ii) If, in addition, $\sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)]$ satisfies the second order growth condition at $X^*(\mathcal{P})$, that is, there exist positive constants C and σ such that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] \geq \vartheta(\mathcal{P}) + \sigma d(x, X^*(\mathcal{P}))^2 \qquad \forall\, x \in X,$$

then

$$\mathbb{D}\big(X^*(\tilde{\mathcal{P}}), X^*(\mathcal{P})\big) \leq \sqrt{\frac{3}{\sigma}\mathbb{H}_{\mathscr{G}}(\tilde{\mathcal{P}}, \mathcal{P})}. \tag{15}$$

Stability of the optimal solutions

Theorem 3.1

(ii) If, in addition, $\sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)]$ satisfies the second order growth condition at $X^*(\mathcal{P})$, that is, there exist positive constants C and σ such that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] \geq \vartheta(\mathcal{P}) + \sigma d(x, X^*(\mathcal{P}))^2 \qquad \forall\, x \in X,$$

then

$$\mathbb{D}\left(X^*(\tilde{\mathcal{P}}), X^*(\mathcal{P})\right) \leq \sqrt{\frac{3}{\sigma}\mathbb{H}_{\mathscr{G}}(\tilde{\mathcal{P}}, \mathcal{P})}. \tag{15}$$

A sufficient condition is that there exists a positive function $\alpha(\xi)$ with $\inf_{P\in\mathcal{P}} \mathbb{E}_P[\alpha(\xi)] > 0$ such that

$$f(x',\xi) \geq f(x,\xi) + \alpha(\xi)\|x' - x\|^2 \qquad \forall\, x' \in X, \xi \in \Xi. \tag{16}$$

Part III: Mathematical program with distributionally robust chance constraint (MPDRCC)

Guo, Xu and Zhang (2017)

MPDRCC

We consider mathematical program with distributionally robust chance constraint:

$$\text{(MPDRCC)} \qquad \begin{array}{ll} \min_{x \in X} & f(x) \\ \text{s.t.} & \inf_{P \in \mathcal{P}} P(g(x,\xi) \leq 0) \geq 1 - \beta, \end{array} \tag{17}$$

where $\mathcal{P}$ is a set of distributions which contains/approximates the true probability distribution of random variable ξ.

MPDRCC

We consider mathematical program with distributionally robust chance constraint:

$$\text{(MPDRCC)} \qquad \begin{array}{ll} \min\limits_{x\in X} & f(x) \\ \text{s.t.} & \inf\limits_{P\in\mathcal{P}} P(g(x,\xi)\leq 0) \geq 1-\beta, \end{array} \tag{17}$$

where $\mathcal{P}$ is a set of distributions which contains/approximates the true probability distribution of random variable ξ.

References: Calafiore and El Ghaoui (2006), Zymler, Kuhn and Rustem (2013), Yang and Xu (2015), Erdoğan and Iyengar (2006), Jiang and Guan (2015), Hu and Hong (2013), Hanasusanto, Roitch, Kuhn and Wiesemann (2015) ...

Approximation of MPDRCC

$$(\mathrm{MPDRCC_N}) \qquad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & \inf\limits_{P \in \mathcal{P}_N} P(g(x,\xi) \le 0) \ge 1 - \beta. \end{array} \tag{18}$$

Approximation of MPDRCC

$$(\mathrm{MPDRCC_N}) \qquad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & \inf\limits_{P \in \mathcal{P}_N} P(g(x,\xi) \le 0) \ge 1 - \beta. \end{array} \tag{18}$$

Question:

What is the impact on the optimal value and the optimal solutions of MPDRCC as N increases?

Reformulation of the chance constraint

For each fixed $x \in X$, let

$$H(x) := \{z \in \Xi : g(x,z) \leq 0\}.$$

Then

$$P(g(x,\xi) \leq 0) \geq 1-\beta \Longleftrightarrow P(H(x)) \geq 1-\beta.$$

Reformulation of the chance constraint

For each fixed $x \in X$, let

$$H(x) := \{z \in \Xi : g(x, z) \leq 0\}.$$

Then

$$P(g(x, \xi) \leq 0) \geq 1 - \beta \Longleftrightarrow P(H(x)) \geq 1 - \beta.$$

Let

$$\mathbb{1}_{H(x)}(z) := \begin{cases} 1 & \text{for } z \in H(x), \\ 0 & \text{for } z \notin H(x), \end{cases}$$

denote the indicator function of $H(x)$. Then

$$P(H(x)) = \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)].$$

$$(\text{MPDRCC}) \quad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & v(x) := \inf\limits_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta, \end{array}$$

vs

$$(\text{MPDRCC}_N) \quad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & v_N(x) := \inf\limits_{P \in \mathcal{P}_N} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta, \end{array}$$

$$(\text{MPDRCC}) \quad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & v(x) := \inf\limits_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta, \end{array}$$

vs

$$(\text{MPDRCC}_N) \quad \begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & v_N(x) := \inf\limits_{P \in \mathcal{P}_N} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta, \end{array}$$

We call $v(x)$ and $v_N(x)$ robust probability function.

$$(\text{MPDRCC}) \qquad \min_{x \in X} \; f(x) \quad \text{s.t.} \quad v(x) := \inf_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta,$$

vs

$$(\text{MPDRCC}_N) \qquad \min_{x \in X} \; f(x) \quad \text{s.t.} \quad v_N(x) := \inf_{P \in \mathcal{P}_N} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \geq 1 - \beta,$$

We call $v(x)$ and $v_N(x)$ robust probability function.

Question: Are $v(x)$, $v_N(x)$ continuous? Does $v_N(x)$ converge to $v(x)$? How?

Pseudo-metric

$$P(H(x)) = \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)].$$

Consider the following set of random indicator functions

$$\mathscr{G} := \{\mathbb{1}_{H(x)}(\xi(\cdot)) : x \in X\}.$$

For $P, Q \in \mathscr{P}(\Xi)$, let

$$\mathscr{D}(P, Q) := \sup_{g \in \mathscr{G}} \left|\mathbb{E}_P[g] - \mathbb{E}_Q[g]\right| = \sup_{x \in X} \left|P(H(x)) - Q(H(x))\right|.$$

We call $\mathscr{D}(P, Q)$ pseudo-metric.

Assumptions

Assumption 4.1 (Convergence of the ambiguity set under the pseudo-metric)

The ambiguity sets $\mathcal{P}$ and $\mathcal{P}_N$ satisfy the following conditions:

(a) $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}_N, \mathcal{P}) = 0,$ (Outer semi-convergence)

(b) $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}, \mathcal{P}_N) = 0.$ (Inner semi-convergence)

Comment: A combination of (a) and (b) implies $\lim_{N\to\infty} \mathscr{H}(\mathcal{P}, \mathcal{P}_N) = 0.$

Uniform approximation of the robust probability function

Theorem 4.1 (Uniform convergence)

Under Assumption 4.1, i.e., $\lim_{N\to\infty} \mathscr{H}(\mathcal{P}, \mathcal{P}_N) = 0$, $v_N(x)$ *converges to* $v(x)$ *uniformly over* X *as* N *tends to* ∞, *i.e.,*

$$\lim_{N\to\infty} \sup_{x\in X} |v_N(x) - v(x)| = 0.$$

Proof.

$$
\begin{aligned}
v_N(x) - v(x) &= \inf_{P_N \in \mathcal{P}_N} P_N(H(x)) - \inf_{P \in \mathcal{P}} P(H(x)) \\
&= \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} P_N(H(x)) - P(H(x)) \\
&\leq \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \left| P(H(x)) - P_N(H(x)) \right| \\
&\leq \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \sup_{x \in X} \left| P(H(x)) - P_N(H(x)) \right| \\
&= \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \sup_{g \in \mathscr{G}} \left| \mathbb{E}_P[g] - \mathbb{E}_{P_N}[g] \right| \\
&= \mathscr{D}(\mathcal{P}, \mathcal{P}_N).
\end{aligned}
$$

Proof.

$$
\begin{aligned}
v_N(x) - v(x) &= \inf_{P_N \in \mathcal{P}_N} P_N(H(x)) - \inf_{P \in \mathcal{P}} P(H(x)) \\
&= \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} P_N(H(x)) - P(H(x)) \\
&\leq \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \left| P(H(x)) - P_N(H(x)) \right| \\
&\leq \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \sup_{x \in X} \left| P(H(x)) - P_N(H(x)) \right| \\
&= \sup_{P \in \mathcal{P}} \inf_{P_N \in \mathcal{P}_N} \sup_{g \in \mathscr{G}} \left| \mathbb{E}_P[g] - \mathbb{E}_{P_N}[g] \right| \\
&= \mathscr{D}(\mathcal{P}, \mathcal{P}_N).
\end{aligned}
$$

This shows

$$
\sup_{x \in X} [v_N(x) - v(x)] \leq \mathscr{D}(\mathcal{P}, \mathcal{P}_N).
$$

Stability/convergence of MPDRCC

- $\mathcal{F}$ and $\mathcal{F}_N$ denote the feasible set,
- $\vartheta := \inf\{f(x) : x \in \mathcal{F}\}$ the optimal value,
- $\vartheta_N := \inf\{f(x) : x \in \mathcal{F}_N\}$ the optimal value,
- $S := \{x \in \mathcal{F} : \vartheta = f(x)\}$ the set of optimal solutions,
- $S_N := \{x \in \mathcal{F}_N : \vartheta_N = f(x)\}$ the set of optimal solutions,
- $\mathcal{F}^s := \{x \in X : v(x) > 1 - \beta\}$ strict feasible solution.

Assumptions

Assumption 4.2 (Continuity of robust probability function)

$v(\cdot) = \inf_{P \in \mathcal{P}} P(H(\cdot))$ *is continuous over* X.

Convergence of the optimal value and optimal solutions

Theorem 4.2

Suppose: (a) $\lim_{N\to\infty} \mathscr{H}(\mathcal{P}, \mathcal{P}_N) = 0$*; (b)* $v(x)$ *is continuous; (c)* $\mathrm{cl}\, \mathcal{F}^s \cap S \neq \emptyset$*. Then*

(i) $\limsup_{N\to\infty} \mathcal{F}_N \subset \mathcal{F}$;

(ii) $\lim_{N\to\infty} \vartheta_N = \vartheta$;

(iii) $\limsup_{N\to\infty} S_N \subset S$.

Sufficient for continuity of the robust probability function

Question

- Under what conditions is the robust probability function

$$v(x) := \inf_{P \in \mathcal{P}} P(H(x))$$

continuous? Recall that $H(x) := \{z \in \Xi : g(x, z) \leq 0\}$.

Sufficient for continuity of the robust probability function

Question

- Under what conditions is the robust probability function

$$v(x) := \inf_{P \in \mathcal{P}} P(H(x))$$

continuous? Recall that $H(x) := \{z \in \Xi : g(x, z) \leq 0\}$.

- When is $P(H(x))$ continuous w.r.t x?

Conditions for continuity of $P(H(x))$

Condition 4.1

For $H(x) := \{z \in \Xi : g(x,z) \leq 0\}$, $K(x) := \{z \in \Xi : g(x,z) = 0\}$ *and* $P \in \mathscr{P}(\Xi)$,

(C1) $P(K(x)) = 0$ *for any* $x \in X$;

(C2) $H(\cdot)$ *is continuous and convex-valued over* X *and for any* $x \in X$,

$$P(\mathrm{bd}\, H(x)) = 0. \tag{19}$$

Note that $K(x) = \mathrm{bd}\, H(x)$ *when* $g(x,\cdot)$ *is strictly convex and* $\Xi = \mathbb{R}^k$.

Conditions for continuity of $P(H(x))$

Condition 4.1

For $H(x) := \{z \in \Xi : g(x,z) \le 0\}$, $K(x) := \{z \in \Xi : g(x,z) = 0\}$ *and* $P \in \mathscr{P}(\Xi)$,

(C1) $P(K(x)) = 0$ *for any* $x \in X$;

(C2) $H(\cdot)$ *is continuous and convex-valued over* X *and for any* $x \in X$,

$$P(\mathrm{bd}\, H(x)) = 0. \tag{19}$$

Note that $K(x) = \mathrm{bd}\, H(x)$ *when* $g(x,\cdot)$ *is strictly convex and* $\Xi = \mathbb{R}^k$.

Theorem 4.3

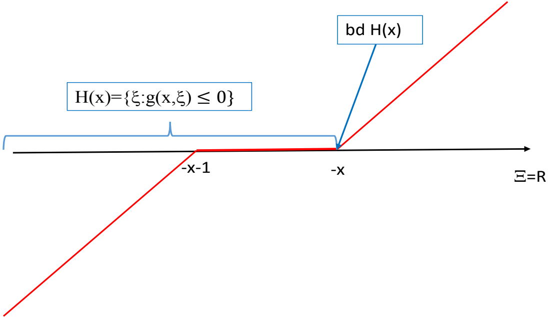
Let $P \in \mathscr{P}(\Xi)$. *Then* $P(H(\cdot))$ *is continuous over* X *when either condition (C1) or (C2) is fulfilled.*

Why do we need C2?

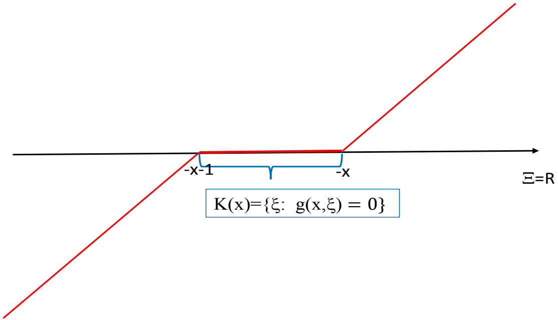
Example 4.1

Let $\xi : \Omega \to \mathbb{R}$ be a random variable with support set $\Xi = \mathbb{R}$. Let

$$g(x,z) := \begin{cases} z + x & \text{for } z \geq -x, \\ 0 & \text{for } z \in [-x-1, -x], \\ z + x + 1 & \text{for } z \leq -x - 1. \end{cases}$$



- $H(x) = \{z \in \mathbb{R} : g(x, z) \leq 0\} = (-\infty, -x]$.
- $H(\cdot)$ is convex set-valued, continuous and $P(\mathrm{bd}\, H(x)) = 0$.
- So (C2) is satisfied!



- $K(x) := \{z \in \mathbb{R} : g(x,z) = 0\} = [-x-1, -x]$.
- $P(K(x)) \neq 0$.
- (C1) is failed!

Pointwise continuity of the robust probability function

Theorem 4.4 (Continuity of $v(x) := \inf_{P \in \mathcal{P}} P(H(x))$)

Suppose $\mathcal{P}$ is weakly compact and one of the following condition holds:

Pointwise continuity of the robust probability function

Theorem 4.4 (Continuity of $v(x) := \inf_{P\in\mathcal{P}} P(H(x))$)

Suppose $\mathcal{P}$ is weakly compact and one of the following condition holds:

(a) *(C1) holds for each $P \in \mathcal{P}$ and for each $x \in X$, $g(\cdot, \xi)$ is continuous at x uniformly w.r.t. $\xi \in \Xi$;*

Pointwise continuity of the robust probability function

Theorem 4.4 (Continuity of $v(x) := \inf_{P\in\mathcal{P}} P(H(x))$)

Suppose $\mathcal{P}$ is weakly compact and one of the following condition holds:

(a) *(C1) holds for each $P \in \mathcal{P}$ and for each $x \in X$, $g(\cdot, \xi)$ is continuous at x uniformly w.r.t. $\xi \in \Xi$;*

(b) *(C2) holds for each $P \in \mathcal{P}$.*

Then $v(\cdot)$ is continuous on X.

Pointwise continuity of the robust probability function

Theorem 4.4 (Continuity of $v(x) := \inf_{P \in \mathcal{P}} P(H(x))$)

Suppose $\mathcal{P}$ is weakly compact and one of the following condition holds:

(a) *(C1) holds for each $P \in \mathcal{P}$ and for each $x \in X$, $g(\cdot, \xi)$ is continuous at x uniformly w.r.t. $\xi \in \Xi$;*

(b) *(C2) holds for each $P \in \mathcal{P}$.*

Then $v(\cdot)$ is continuous on X.

The set $\mathcal{A}$ is said to be weakly compact if every sequence $\{P_N\} \subset \mathcal{A}$ contains a subsequence $\{P_{N'}\}$ and moreover there exists $P \in \mathcal{A}$ such that $P_{N'}$ converges to P weakly.

Sufficient conditions for convergence of $\mathscr{D}(P_N, P)$

Lemma 4.1

Let $\{P_N\} \subset \mathscr{P}$ be a sequence of probability measures and $P \in \mathscr{P}$. Suppose P_N converges to P weakly. Then

$$\lim_{N \to \infty} \mathscr{D}(P_N, P) = 0 \tag{20}$$

under one of the following conditions:

(a) *$g(\cdot, \xi)$ is continuous on X uniformly w.r.t. $\xi \in \Xi$ + condition (C1) for P.*

Sufficient conditions for convergence of $\mathscr{D}(P_N, P)$

Lemma 4.1

Let $\{P_N\} \subset \mathscr{P}$ be a sequence of probability measures and $P \in \mathscr{P}$. Suppose P_N converges to P weakly. Then

$$\lim_{N\to\infty} \mathscr{D}(P_N, P) = 0 \tag{20}$$

under one of the following conditions:

(a) *$g(\cdot, \xi)$ is continuous on X uniformly w.r.t. $\xi \in \Xi$ + condition (C1) for P.*

(b) *Condition (C2) for the P.*

Sufficient conditions for convergence of $\mathscr{D}(\mathcal{P}_N, \mathcal{P})$

Proposition 4.1

- $\mathcal{P}_N$ converges to $\mathcal{P}$ weakly, i.e., for every sequence $\{P_N\} \subseteq \mathcal{P}_N$, $\{P_N\}$ has a subsequence $\{P_{N_k}\}$ converging to P with $P \in \mathcal{P}$.

Sufficient conditions for convergence of $\mathscr{D}(\mathcal{P}_N, \mathcal{P})$

Proposition 4.1

- $\mathcal{P}_N$ converges to $\mathcal{P}$ weakly, i.e., for every sequence $\{P_N\} \subseteq \mathcal{P}_N$, $\{P_N\}$ has a subsequence $\{P_{N_k}\}$ converging to P with $P \in \mathcal{P}$.
- Condition (a) or (b) in Lemma 4.1 holds for any $P \in \mathcal{P}$.

Then

$$\lim_{N\to\infty} \mathscr{D}(\mathcal{P}_N, \mathcal{P}) = 0.$$

Sufficient conditions for convergence of $\mathscr{D}(\mathcal{P}, \mathcal{P}_N)$

Proposition 4.2

Assume:

- *$\mathcal{P}$ is weakly compact, i.e., for any $P \in \mathcal{P}$, there exists a sequence $\{P_N\} \in \mathcal{P}_N$ such that P_N converges to P weakly.*

Sufficient conditions for convergence of $\mathscr{D}(\mathcal{P}, \mathcal{P}_N)$

Proposition 4.2

Assume:

- *$\mathcal{P}$ is weakly compact, i.e., for any $P \in \mathcal{P}$, there exists a sequence $\{P_N\} \in \mathcal{P}_N$ such that P_N converges to P weakly.*
- *Condition (a) or (b) in Lemma 4.1 holds for any $P \in \mathcal{P}$.*

Then

$$\lim_{N\to\infty} \mathscr{D}(\mathcal{P}, \mathcal{P}_N) = 0.$$

- H. Sun and H. Xu, Convergence Analysis for Distributional Robust Optimization and Equilibrium Problems, Mathematics of Operations Research, Vol. 41, pp. 377-401, 2015.
- S. Guo, H. Xu and L. Zhang, Convergence Analysis for Mathematical Programs with Distributionally Robust Chance Constraint, *SIAM J. Optimization*, Vol. 27, pp. 784–816, 2017.
- J. Zhang, H. Xu and L.W. Zhang, Quantitative Stability Analysis for Distributionally Robust Optimization with Moment Constraints. SIAM J. Optimization, Vol. 26, pp. 1855-1882, 2016.
- Y. Liu, A. Pichler and H. Xu, Discrete Approximation and Quatification in Distributionally Robust Optimization, to appear in MOR.
- A. Pichler and H. Xu, Quantitative Stability Analysis for Minimax Distributionally Robust Risk Optimization, Optimization online, 2017.