

2022/08/30

Wasserstein距離に基づく分布的ロバスト最適化の凸最適化問題への帰着

hdk

目次

1. 1. はじめに
2. 2. Wasserstein距離とは
3. 3. 分布的ロバスト最適化とは
4. 4. 最悪ケースの期待コスト問題の凸最適化問題への帰着

4-1. 4.1 定理 1 で使う道具の紹介

4-2. 4.2 定理1の証明
5. 参考文献

1. 1. はじめに

この記事は、文献[2]の式変形の備忘録として残したものです。式番号は論文にあわせておりますのでご注意ください。

ルベーク積分、関数解析、凸解析および測度論的確率論を勉強しはじめた段階なので、間違った理解をしているところがあると思います。見つけた際は、間違いをご指摘頂ければ幸いです。

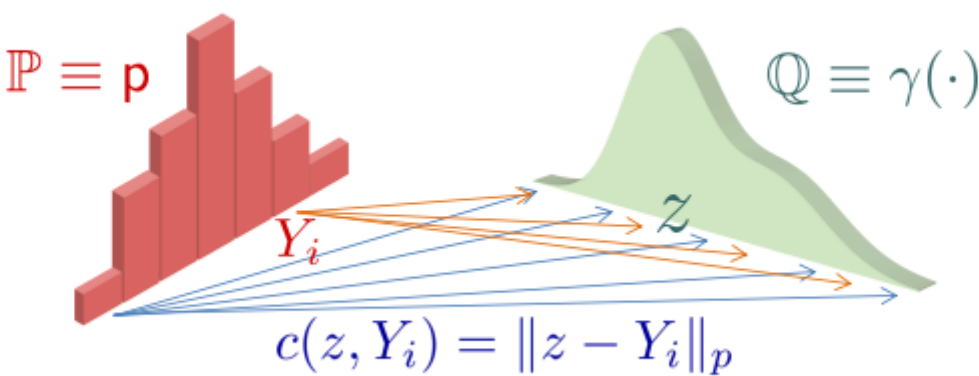
2. 2. Wasserstein距離とは

Wasserstein距離は、輸送問題に着想を得た分布間の距離尺度のことである。連続型分布 $\mathbb{Q} := \gamma(\cdot)$ と、離散型分布である経験分布 $\mathbb{P} := p$ の間に限定して考える。

[トピック] Wasserstein 距離に基づく DRO

$$d(\mathbb{Q}|\mathbb{P}) = \left| \begin{array}{l} \min_{\gamma} \sum_{i=1}^n \int_z \|z - Y_i\|_p \gamma_i(dz) \\ \text{s. t. } \int_z \gamma_i(dz) = p_i, i = 1, \dots, n, \\ \gamma_i(dz) \geq 0 \end{array} \right|$$

- \mathbb{Q} と \mathbb{P} の輸送問題（単位輸送費用が $\|z - Y_i\|_p$ ）
- 連続型の確率分布も許す



82 / 120

図1：Wasserstein距離のイメージ図

図1のように、 \mathbb{P} から \mathbb{Q} へと確率を移動する輸送問題を考え、単位輸送費用が $\|z - Y_i\|_p$ で与えられた場合の最小費用がダイバージェンス $d(\mathbb{Q}|\mathbb{P})$ を与える。

3. 3. 分布的ロバスト最適化とは

事前には認知が難しい事後での不整合に対し、最悪ケースのずれがあったとしても最適解の実行可能性やある種の最適性を担保しようとする最適化モデリングをロバスト最適化という。

ロバスト最適化の比較を図に示す。

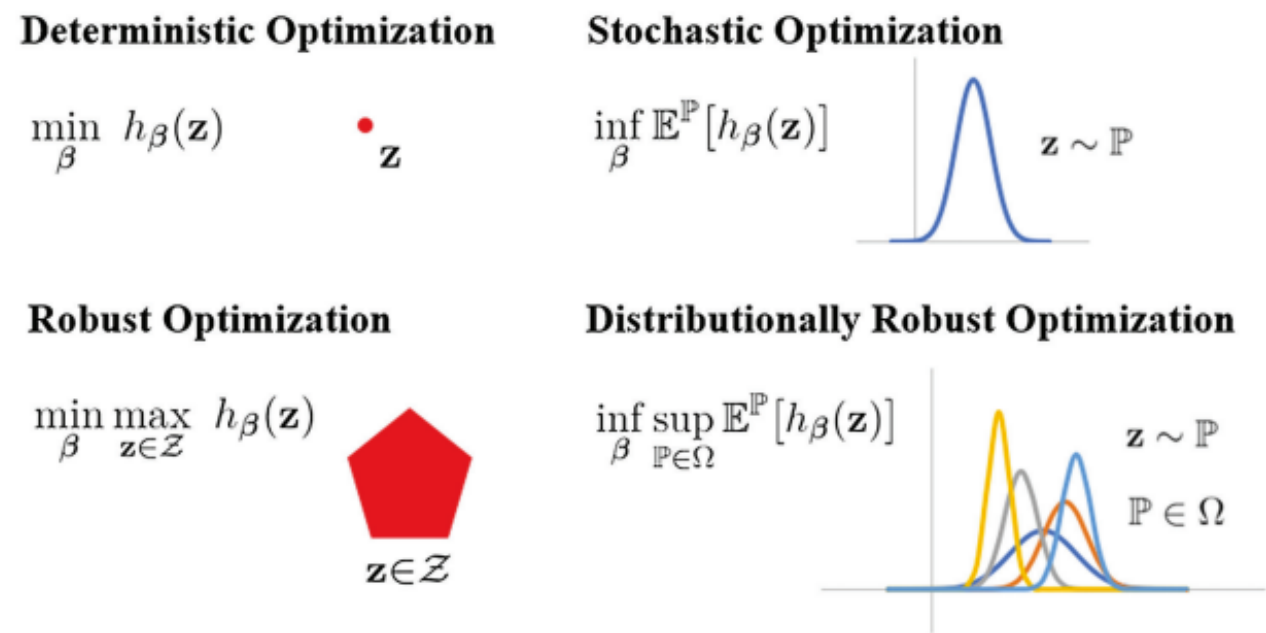


Figure 1.2: Comparison of robust optimization with distributionally robust optimization.

図2：ロバスト最適化の比較

ロバスト最適化の中でも想定する確率分布の不確実性を考慮したものを分布的ロバスト最適化という。

分布的ロバスト最適化は、経験分布 $\hat{\mathbb{P}}_N$ の不確実性を集合で表し、その集合の中で最悪ケースの期待コストの最小化を目指す。

$$\inf_x \sup_{\mathbb{Q} \in \mathbb{B}_\epsilon(\hat{\mathbb{P}}_N)} \mathbb{E}^\mathbb{Q}[l(\xi)]$$

不確実性集合 $\mathbb{B}_\epsilon(\hat{\mathbb{P}}_N)$ は、下式で与えられる。

$$\mathbb{B}_\epsilon(\hat{\mathbb{P}}_N) := \{\mathbb{Q} \in \mathcal{M}(\Xi) : d_W(\hat{\mathbb{P}}_N, \mathbb{Q}) \leq \epsilon\}. \tag{6}$$

経験分布 $\hat{\mathbb{P}}_N$ を中心とした半径 ϵ のWasserstein球とみることができる。

4. 4. 最悪ケースの期待コスト問題の凸最適化問題への帰着

式(10)の最悪ケースの期待コスト問題が凸最適化問題に帰着することができることをみていく。

$$\sup_{\mathbb{Q} \in \mathbb{B}_\epsilon(\hat{\mathbb{P}}_N)} \mathbb{E}^\mathbb{Q}[l(\xi)] \tag{10}$$

仮定4.1 (Convexity)

不確実性集合 $\Xi \subset \mathbb{R}^m$ は凸で閉であるとする。
 l_k はproperであり、凸であり、下半連続であるとする。
さらに、 l_k は $-\infty$ ではないとする。

定理 1 (Convex reduction)

仮説4.1において、任意の $\epsilon \geq 0$ において、式(10)の最悪ケースの期待コスト問題は式(11)の凸最適化問題の最適値に等しい。

$$\left| \begin{array}{ll} \inf_{\lambda, s_i, z_{ik}, v_{ik}} & \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t.} & [-l_k]^*(z_{ik} - v_{ik}) + \sigma_\Xi(v_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \\ & \|z_{ik}\|_* \leq \lambda \end{array} \right. \tag{11}$$

ここで、 $\|z_{ik}\|_*$ は z_{ik} の双対ノルムである。

4-1. 4.1 定理1で使う道具の紹介

この節では、定理1の証明に入る前に、証明でつかう定義や補題を紹介する。
関数解析、凸解析および測度論的確率論はまだ勉強しはじめたばかりなので間違いがあればご指摘ください。

補題 2

$$\sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_\Xi \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) = \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right)$$

証明

まず、以下の不等式の証明をする。

$$\begin{aligned} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) &\geq \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \\ &= \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) \\ &= \sup \left\{ \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) \mid \mathbb{Q}_i \in \mathcal{M}(\Xi) \text{ for } i = 1, 2, \dots, N \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \sup \left\{ \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) \mid \mathbb{Q}_i \in \mathcal{M}(\Xi) \right\} \\ &\geq \frac{1}{N} \sum_{i=1}^N \sup \left\{ \int_{\Xi} \left(l(\zeta) - \lambda \|\zeta - \hat{\xi}_i\| \right) \delta_{\xi}(d\zeta) \mid \xi \in \Xi \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \end{aligned}$$

次に以下の不等式を証明する。

$$\begin{aligned} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) &\leq \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \\ &= \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) \\ &\leq \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \sup_{\xi' \in \Xi} \left(l(\xi') - \lambda \|\xi' - \hat{\xi}_i\| \right) \mathbb{Q}_i(d\xi) \\ &= \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \sup_{\xi' \in \Xi} \left(l(\xi') - \lambda \|\xi' - \hat{\xi}_i\| \right) \int_{\Xi} \mathbb{Q}_i(d\xi) \\ &= \frac{1}{N} \sum_{i=1}^N \sup_{\xi' \in \Xi} \left(l(\xi') - \lambda \|\xi' - \hat{\xi}_i\| \right) \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \int_{\Xi} \mathbb{Q}_i(d\xi) \\ &= \frac{1}{N} \sum_{i=1}^N \sup_{\xi' \in \Xi} \left(l(\xi) - \lambda \|\xi - \hat{\xi}_i\| \right) \end{aligned}$$

よって、等号が成立する。

定義 1（双対ノルム）

x のノルム $\|x\|$ の双対ノルムを以下で定義する。

$$\|x\|_* := \max_{\|y\| \leq 1} \langle y, x \rangle$$

例 1

l^p の共役空間はノルム空間 l^q に等しい距離同型である。

ただし、 $1 < p < \infty$ で、 q は p の共役指数、すなわち、

$$\frac{1}{p} + \frac{1}{q} = 1$$

を満たすものである。

f がノルム空間 l^p 上の有界な線形汎関数であるための必要十分条件は、 $a \in l^q$ が存在して、

$$f(x) = \langle x, a \rangle$$

が成り立つことである。

しかも、 $\|f\| = \|a\|_q$ が成り立つことである。

簡単に書くと

$$\|x\|_{p^*} = \max_{\|y\|_p \leq 1} \langle y, x \rangle = \|x\|_q$$

定義 2（共役関数）

H を Hilbert 空間とし、関数 $f: H \rightarrow (-\infty, \infty]$ を真であるとする。このとき、 H 上の f^* を

$$f^*(x^*) = \sup_{x \in H} \{\langle x, x^* \rangle - f(x)\}$$

で定義する。 f^* は f の共役関数といわれる。

定義 3（epi-sum, inf-convolution）

E を線形空間とし、 $f, g: E \rightarrow (-\infty, \infty]$ を sublinear とする。

$$(f \boxplus g)(x) = \inf_{x \in E} \{f(x) + g(x - z)\}, \forall x \in E$$

で定義される $(f \boxplus g)$ を f と g による epi-sum または inf-convolution という。

定理 3

E を線形区間とし、 $f, g: E \rightarrow (-\infty, \infty]$ を proper な凸関数とする。また、

$$(f \boxplus g)(x) = \inf_{x \in E} \{f(x) + g(x - z)\} > -\infty, \forall x \in E$$

とする。

このとき、 $(f \boxplus g)(x): E \rightarrow (-\infty, \infty]$

は proper な凸関数となる。さらに、 $(f \boxplus g)^* = f^* + g^*$ である。

証明

凸解析と不動点近似のp156 定理4.4.2を参照してください。

補題 4

$\sigma_{\Xi}(v_{ik}) := [\chi_{\Xi}]^*(v_{ik})$ とおく。

$$\begin{aligned} [-l_k + \chi_{\Xi}]^*(z_{ik}) &= \inf_{v_{ik}} ([-l_k]^*(z_{ik} - v_{ik}) + [\chi_{\Xi}]^*(v_{ik})) \\ &= \inf_{v_{ik}} ([-l_k]^*(z_{ik} - v_{ik}) + \sigma_{\Xi}(v_{ik})) \end{aligned}$$

証明

$$\begin{aligned} & \left(\inf_{v_{ik}} ([-l_k]^*(z_{ik} - v_{ik}) + [\chi_{\Xi}]^*(v_{ik})) \right)^* \\ &= ([\chi_{\Xi}^* \boxplus (-l_k^*)](z_{ik}))^* \\ &= (\chi_{\Xi}^*(z_{ik}))^* + (-l_k^*(z_{ik}))^* \\ &= \chi_{\Xi}^{**}(x) - l_k^{**}(z_{ik}) \\ &= \chi_{\Xi}(x) - l_k(z_{ik}) \\ &= -l_k(z_{ik}) + \chi_{\Xi}(z_{ik}) \\ &= (-l_k + \chi_{\Xi})(z_{ik}) \text{ (この行は正しいかわからないです、ごめんなさい)} \end{aligned}$$

4-2. 4.2 定理1の証明

$$\begin{aligned} \sup_{\mathbb{Q} \in \mathbb{B}_{\epsilon}(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[l(\xi)] &= \left| \begin{array}{l} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} l(\xi) \mathbb{Q}_i(d\xi) \\ \text{s.t.} \quad \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \|\xi - \hat{\xi}_i\| \mathbb{Q}_i(d\xi) \leq \epsilon \end{array} \right. \\ &= \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \inf_{\lambda \geq 0} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} l(\xi) \mathbb{Q}_i(d\xi) + \lambda \left(\epsilon - \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \|\xi - \hat{\xi}_i\| \mathbb{Q}_i(d\xi) \right) \\ &\leq \inf_{\lambda \geq 0} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} l(\xi) \mathbb{Q}_i(d\xi) + \lambda \left(\epsilon - \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \|\xi - \hat{\xi}_i\| \mathbb{Q}_i(d\xi) \right) \quad (12a) \\ &= \inf_{\lambda \geq 0} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N \int_{\Xi} (l(\xi) - \lambda \|\xi - \hat{\xi}_i\|) \mathbb{Q}_i(d\xi) \\ &= \inf_{\lambda \geq 0} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} (l(\xi) - \lambda \|\xi - \hat{\xi}_i\|) \quad (\because \text{補題2}) \\ &= \left| \begin{array}{l} \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t.} \quad \sup_{\xi \in \Xi} (l(\xi) - \lambda \|\xi - \hat{\xi}_i\|) \leq s_i, \forall i \leq N \\ \lambda \geq 0 \end{array} \right. \\ &= \left| \begin{array}{l} \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t.} \quad \sup_{\xi \in \Xi} \left(l_k(\xi) - \max_{\|z_{ik}\|_* \leq \lambda} \langle z_{ik}, \xi - \hat{\xi}_i \rangle \right) \leq s_i, \forall i \leq N, \forall k \leq K \\ \lambda \geq 0 \end{array} \right. \\ &\leq \left| \begin{array}{l} \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t.} \quad \min_{\|z_{ik}\|_* \leq \lambda} \sup_{\xi \in \Xi} (l_k(\xi) - \langle z_{ik}, \xi - \hat{\xi}_i \rangle) \leq s_i, \forall i \leq N, \forall k \leq K \\ \lambda \geq 0 \end{array} \right. \quad (12e) \\ &= \left| \begin{array}{l} \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \right. \end{aligned}$$

$$\begin{aligned}
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} \left(l_k(\xi) - \langle z_{ik}, \xi - \hat{\xi}_i \rangle \right) \leq s_i, \forall i \leq N, \forall k \leq K \\ \lambda \geq 0 \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} \left(l_k(\xi) - \langle z_{ik}, \xi - \hat{\xi}_i \rangle \right) \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} (l_k(\xi) - \langle z_{ik}, \xi \rangle) + \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} (l_k(\xi) + \langle z_{ik}, \xi \rangle) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} (\langle z_{ik}, \xi \rangle - (-l_k(\xi))) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad \sup_{\xi \in \Xi} (\langle z_{ik}, \xi \rangle - (-l_k(\xi) + \chi_{\Xi}(\xi))) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad [-l_k + \chi_{\Xi}]^*(z_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i, \forall i \leq N, \forall k \leq K \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \\
&= \begin{array}{|l} \text{s.t.} \quad [-l_k]^*(z_{ik} - v_{ik}) + \sigma_{\Xi}(v_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \\ \|z_{ik}\|_* \leq \lambda \\ \inf_{\lambda, s_i, z_{ik}, v_{ik}} \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \end{array} \quad (\because \text{補題4})
\end{aligned}$$

仮定4.1のもとで、式(12a)と式(12e)の不等号が等号になることを示す。仮定4.1のもとで、式(12a)は強双対性より等号が成立する。仮定4.1のもとで、minimax定理より式(12e)は等号が成立する。

よって、仮説4.1において、任意の $\epsilon \geq 0$ において、式(10)の最悪ケースの期待コスト問題は式(11)の凸最適化問題の最適値に等しい。

参考文献

-
- [1] P.M.Esfahani, D.Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Springer
 - [2] R. Chen and I. C. Paschalidis, Distributionally Robust Learning
 - [3] [ロバスト最適化モデリング入門, 閲覧日 2022-08-29, https://www.ism.ac.jp/~mirai/sscoke/2022/gotoh-lecture.pdf](https://www.ism.ac.jp/~mirai/sscoke/2022/gotoh-lecture.pdf)

