

ロバスト最適化モデリング入門

後藤 順哉

中央大学 理工学部 ビジネスデータサイエンス学科

連続最適化および関連分野に関する夏季学校 2022

2022 年 8 月 8～10 日

$$\begin{array}{ll} \min_{\boldsymbol{x}} & \max_{\boldsymbol{y} \in U_0} f(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} & g_i(\boldsymbol{x}, \boldsymbol{y}) \leq 0, \forall \boldsymbol{y} \in U_i, i = 1, \dots, m, \\ & h_i(\boldsymbol{x}, \boldsymbol{y}) = 0, \forall \boldsymbol{y} \in U_i, i = m + 1, \dots, l \end{array}$$

「ロバスト最適化」「分布的ロバスト最適化」と呼ばれる不確実性下の最適化モデリングの枠組みにおいて現れる，基本的なモデリング技法およびそれらの動機について大雑把に理解する

ただし，

網羅的ではありません（むしろバイアス大）

必ずしも無条件にプロモートするものでもありません

個人的にはむしろ，ややシニカルな立場かも

① ロバスト最適化

- Soyster 1973, Ben-Tal, Nemirovski, 2000, Bertsimas, Sim, 2004
- 応用例
 - ロバストポートフォリオ選択
 - SVM

② 分布的ロバスト最適化 (DRO)

- リスク尺度と DRO
- DRO の含意

③ DRO の事後パフォーマンス

動機

現実問題における最適化問題のパラメータの多くはデータ

発電計画の例

$$\begin{array}{ll} \min_{x_1, x_2} & 46 x_1 + 135 x_2 \quad \leftarrow \text{コスト} \\ \text{s. t.} & x_1 + x_2 \geq 1000, \quad \leftarrow \text{需要量} \\ & 150 \leq x_1 \leq 600 \quad \leftarrow \text{石炭の発電上下限} \\ & 0 \leq x_2 \quad \leftarrow \text{石油の発電上下限} \end{array}$$

- 「最適化」は与えられたデータに対して「最適」（を目指す）
- データが（少しでも）変化した場合には変化後の最適性は保証できない
- 特に不確実性を含む場合には不確実性のモデルに対する理解（とその帰結に対する覚悟）が必要

cf. 感度分析

Ben-Tal, Nemirovski (2000) による動機

- そういったデータの多くは “�い (ugly)”

[例] NETLIB の LP のテスト問題集の中の問題例 PILOT4 (1000 変数, 410 制約) の 372 番目の制約:

$$\begin{aligned} a^\top x \equiv & -15.79081 x_{826} - 8.598819 x_{827} - 1.88789 x_{828} - 1.362417 x_{829} \\ & - 1.526049 x_{830} - 0.031883 x_{849} - 28.725555 x_{850} - 10.792065 x_{851} \\ & - 0.19004 x_{852} - 2.757176 x_{853} - 12.290832 x_{854} + 717.562256 x_{855} \\ & - 0.057865 x_{856} - 3.785417 x_{857} - 78.30661 x_{858} - 122.163055 x_{859} \\ & - 6.46609 x_{860} - 0.48371 x_{861} - 0.615264 x_{862} - 1.353783 x_{863} \\ & - 84.644257 x_{864} - 122.459045 x_{865} - 43.15593 x_{866} - 1.712592 x_{870} \\ & - 0.401597 x_{871} + x_{880} - 0.946049 x_{898} - 0.946049 x_{916} \\ \geq & b \equiv 23.387405 \end{aligned}$$

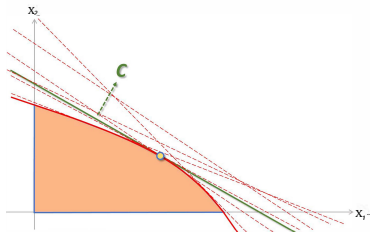
→ CPLEX による最適解の一部:

$$\begin{aligned} x_{826}^* &= 255.6112787181108, & x_{827}^* &= 6240.488912232100, \\ x_{828}^* &= 3624.613324098961, & x_{829}^* &= 18.20205065283259, \\ x_{849}^* &= 174397.0389573037, & x_{870}^* &= 14250.00176680900, \\ x_{871}^* &= 25910.00731692178, & x_{880}^* &= 104958.3199274139 \end{aligned}$$

→ もし制約式の係数がそれぞれ 0.1% ずつずれると制約式の違反は右辺の 450% !

ロバスト最適化 (Soyster,'73)

$$\begin{array}{ll} \max_{\mathbf{x} \geq 0} & \mathbf{c}^\top \mathbf{x} \\ \text{s. t.} & \sum_{j=1}^d \underbrace{a_j}_{\text{推定値}} x_j \leq b \end{array}$$



$a_1, \dots, a_d$ は正しくないかもしれない $\rightsquigarrow$ 最適解が現実には実行不可能かも

そこで $\delta|a_j|$ ぶれても実行可能なように条件を変更 ($\delta > 0$ はぶれの相対的な大きさ) :

$$a_j \rightsquigarrow [a_j - \underbrace{\delta|a_j|}_{\text{正定数}}, a_j + \delta|a_j|]$$

$$\sum_{j=1}^d \tilde{a}_j x_j \leq b, \quad \forall \tilde{a}_j \in [a_j - \delta|a_j|, a_j + \delta|a_j|]$$

$$\left| \begin{array}{ll} \max_{\mathbf{x} \in X} & \mathbf{c}^\top \mathbf{x} \\ \text{s. t.} & \sum_{j=1}^d \tilde{a}_j x_j \leq b, \quad \forall \tilde{a}_j \in [\underline{a}_j - \delta |a_j|, \underline{a}_j + \delta |a_j|] \end{array} \right|$$

$\Updownarrow$

$$\left| \begin{array}{ll} \max_{\mathbf{x} \in X} & \mathbf{c}^\top \mathbf{x} \\ \text{s. t.} & \sum_{j=1}^d a_j x_j + \sum_{j=1}^d \delta |a_j| |x_j| \leq b \end{array} \right|$$

$\Updownarrow$

$$\left| \begin{array}{ll} \max_{\mathbf{x} \in X, \mathbf{y}} & \mathbf{c}^\top \mathbf{x} \\ \text{s. t.} & \sum_{j=1}^d a_j x_j + \sum_{j=1}^d \delta |a_j| y_j \leq b \\ & y_j \geq x_j, y_j \geq -x_j \end{array} \right|$$

X が線形不等式・方程式で与えられていれば LP に帰着できる

Soyster, '73 流 RO の例

不確実なパラメータを持った 0 - 1IP :

$$\begin{array}{|l} \max_{x_1, x_2, x_3} \quad 5x_1 \quad \quad -x_2 \quad \quad +3x_3 \\ \text{s. t.} \quad \quad x_1 \quad + a_{1,2} x_2 \quad + a_{1,3} x_3 \leq 7, \quad \forall a_{1,2} \in [1, 3], a_{1,3} \in [1, 5], \\ \quad \quad a_{2,1} x_1 \quad - a_{2,2} x_2 \quad \quad + x_3 \leq 5, \quad \forall a_{2,1} \in [3, 5], a_{2,2} \in [4, 10], \\ \quad \quad \quad \quad \quad x_1, x_2, x_3 \in \{0, 1\} \end{array}$$

$\Updownarrow$

$$\begin{array}{|l} \max_{x_1, x_2, x_3} \quad 5x_1 \quad \quad -x_2 \quad \quad +3x_3 \\ \text{s. t.} \quad \quad x_1 \quad + [2 \pm 1]x_2 \quad + [3 \pm 2]x_3 \leq 7, \\ \quad \quad [4 \pm 1]x_1 \quad - [7 \pm 3]x_2 \quad \quad + x_3 \leq 5, \\ \quad \quad \quad \quad \quad x_1, x_2, x_3 \in \{0, 1\} \end{array}$$

$\Updownarrow$

$$\begin{array}{|l} \max_{x_1, x_2, x_3} \quad 5x_1 \quad -x_2 \quad +3x_3 \\ \text{s. t.} \quad \quad x_1 \quad +3x_2 \quad +5x_3 \leq 7, \\ \quad \quad 5x_1 \quad -4x_2 \quad +x_3 \leq 5, \\ \quad \quad \quad \quad \quad x_1, x_2, x_3 \in \{0, 1\} \end{array}$$

普通の 0 - 1IP

Soyster, '73 流 RO の例

不確実なパラメータを持った IP :

$$\begin{array}{|l}
 \max_{x_1, x_2, x_3} \quad 5x_1 \quad \quad \quad -x_2 \quad \quad \quad +3x_3 \\
 \text{s. t.} \quad \quad x_1 \quad + a_{1,2} x_2 \quad + a_{1,3} x_3 \leq 7, \quad \forall a_{1,2} \in [1, 3], a_{1,3} \in [1, 5], \\
 \quad \quad a_{2,1} x_1 \quad - a_{2,2} x_2 \quad \quad \quad + x_3 \leq 5, \quad \forall a_{2,1} \in [3, 5], a_{2,2} \in [4, 10], \\
 \quad \quad \quad -10 \leq x_1, x_2, x_3 \leq 10, \quad x_1, x_2, x_3 \in \mathbb{Z} \\
 \\
 \quad \quad \quad \Updownarrow \\
 \max_{x_1, x_2, x_3} \quad 5x_1 \quad \quad \quad -x_2 \quad \quad \quad +3x_3 \\
 \text{s. t.} \quad \quad x_1 \quad \quad \quad +2x_2 \quad \quad +|x_2| \quad +3x_3 \quad +2|x_3| \leq 7, \\
 \quad \quad 4x_1 \quad +|x_1| \quad -7x_2 \quad +3|x_2| \quad \quad +x_3 \quad \leq 5, \\
 \quad \quad \quad -10 \leq x_1, x_2, x_3 \leq 10, \quad x_1, x_2, x_3 \in \mathbb{Z} \\
 \\
 \quad \quad \quad \Updownarrow \\
 \max_{x_i, y_i} \quad 5x_1 \quad \quad \quad -x_2 \quad \quad \quad +3x_3 \\
 \text{s. t.} \quad \quad x_1 \quad \quad \quad +2x_2 \quad +y_2 \quad +3x_3 \quad +2y_3 \leq 7, \\
 \quad \quad 4x_1 \quad +y_1 \quad -7x_2 \quad +3y_2 \quad \quad +x_3 \quad \leq 5, \\
 \quad \quad \quad y_i \geq x_i, \quad y_i \geq -x_i, \quad i = 1, 2, 3, \\
 \quad \quad \quad -10 \leq x_1, x_2, x_3 \leq 10, \quad x_1, x_2, x_3 \in \mathbb{Z}
 \end{array}$$

(変数は増えるも) 普通の (M)IP

Soyster, '73 流 RO の問題点 (!?)

$$x_1 + a_{1,2}x_2 + a_{1,3}x_3 \leq 7, \quad \forall a_{1,2} \in [2 - 1, 2 + 1], a_{1,3} \in [3 - 2, 3 + 2]$$



$$x_1 + 2x_2 + |x_2| + 3x_3 + 2|x_3| \leq 7$$

- 同じ制約の中の異なる変数の不確実な係数がそれぞれが最悪なケースをとる場合の最適を指向 (いくらなんでもありえなさそう !?)

→ 過度に保守的な解になりがち

→ 少し保守性を抑えてみる

ξ_j を $[-1, 1]$ を分布の台とする対称な iid 確率変数とし,

$$\tilde{a}_j = (1 + \delta \xi_j) a_j, \quad j = 1, \dots, d$$

と表せることを仮定する

定数 $\Omega > 0$ とする. (x, y, z) が問題

$$\left| \begin{array}{ll} \max_{x \in X, y, z} & c^\top x \\ \text{s. t.} & \sum_{j=1}^d a_j x_j + \sum_{j=1}^d \delta |a_j| y_j + \delta \Omega \sqrt{\sum_{j=1}^d a_j^2 z_j^2} \leq b \\ & y_j \geq x_j - z_j, y_j \geq z_j - x_j \end{array} \right.$$

の実行可能解であるならば,

$$\mathbb{P}\left\{ \sum_{j=1}^d \tilde{a}_j x_j > b \right\} \leq \exp\left(-\frac{\Omega^2}{2}\right)$$

演習 (問題)

上記命題を示せ

上の問題は 2 次錐計画 (SOCP) と呼ばれる, 凸 2 次計画を一般化したクラス

楕円型

$$\sum_{j=1}^d \tilde{a}_j w_j \leq b, \quad \forall \tilde{a}_j \in [a_j - \hat{a}_j, a_j + \hat{a}_j] \quad : \text{Soyster 型}$$

$$\Leftrightarrow \sum_{j=1}^d a_j w_j + \sum_{j=1}^d \hat{a}_j |w_j| \leq b$$

〜 d 個の変数それぞれが同時に最悪の動きをする状況に対処しているため過度に保守的

$$\sum_{j=1}^d a_j w_j + \sqrt{\sum_{j=1}^d \hat{a}_j^2 w_j^2} \leq b$$

ℓ_1 vs. ℓ_2

ℓ_1 ノルムと ℓ_2 ノルム

- $\|(x_1, \dots, x_d)\|_1 := \sum_{i=1}^d |x_i|$: ℓ_1 ノルム
- $\|(x_1, \dots, x_d)\|_2 := \sqrt{\sum_{i=1}^d x_i^2}$: ℓ_2 ノルム (ユークリッドノルム)

$$\begin{aligned} & \sum_{j=1}^d a_j w_j + \sum_{j=1}^d \hat{a}_j |w_j| \leq b \\ \Leftrightarrow & \sum_{j=1}^d a_j w_j + \|(\hat{a}_1 w_1, \dots, \hat{a}_d w_d)\|_1 \leq b \end{aligned}$$

↪ d 個の変数それぞれが同時に最悪の動きをする状況に対処しているため過度に保守的

$$\begin{aligned} & \sum_{j=1}^d a_j w_j + \sqrt{\sum_{j=1}^d \hat{a}_j^2 w_j^2} \leq b \\ \Leftrightarrow & \sum_{j=1}^d a_j w_j + \|(\hat{a}_1 w_1, \dots, \hat{a}_d w_d)\|_2 \leq b \end{aligned}$$

ℓ_2 ノルム (ユークリッドノルム)

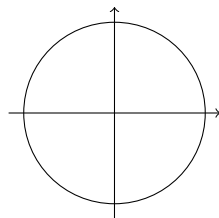
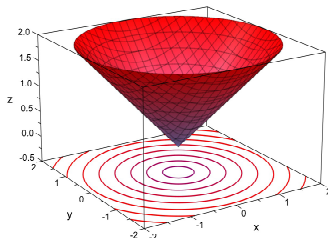
$$\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}^\top \mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

$\|\mathbf{x}\|_2$

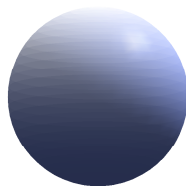
関数の概形

ノルム球 (レベル集合)

$$\begin{aligned} \|(x, y)\|_2 \\ = \sqrt{x^2 + y^2} \end{aligned}$$



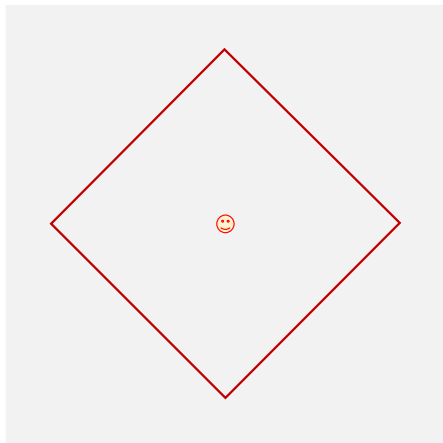
$$\begin{aligned} \|(x, y, z)\|_2 \\ = \sqrt{x^2 + y^2 + z^2} \end{aligned}$$



ℓ_1 ノルム (マンハッタンノルム)

$$\|\boldsymbol{x}\|_1 := |x_1| + \cdots + |x_n|$$

稠密なマンハッタンで人に会いに行くことを考える

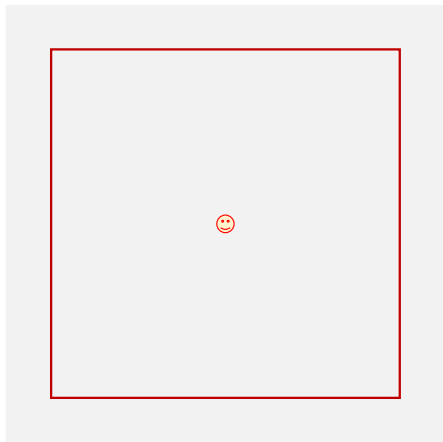


演習 (基礎問題)

- ℓ_1 ノルムが公理 1.~3. を満たすことを示せ

l_∞ ノルム (最大値ノルム)

$$\|\mathbf{x}\|_\infty := \max\{|x_1|, \dots, |x_n|\}$$



演習 (基礎問題)

- l_∞ ノルムが公理 1.~3. を満たすことを示せ

• $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto \|\mathbf{x}\|$

ノルムの公理

1. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\| \quad (a \in \mathbb{R})$

斉次性

2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

劣加法性

3. $\|\mathbf{x}\| = 0 \Rightarrow \mathbf{x} = \mathbf{0}$

$\leadsto \|\mathbf{x}\| \geq 0 \quad \text{かつ} \quad \langle \|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0} \rangle$

演習 (基礎問題)

公理 1.~3. を用いて $\forall \mathbf{x}, \|\mathbf{x}\| \geq 0$ を示せ

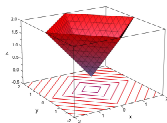
斉次性 + 劣加法性 $\Leftrightarrow$ 斉次性 + 凸性 :

$$\forall \lambda \in [0, 1], \forall \mathbf{x}, \mathbf{y}, \quad f((1 - \lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$$

ℓ_p ノルム

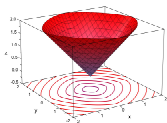
$$\|\mathbf{x}\|_p := \begin{cases} (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}}, & p \in [1, \infty) \\ \max\{|x_1|, \dots, |x_n|\} (= \lim_{p \rightarrow \infty} \|\mathbf{x}\|_p), & p = \infty \end{cases}$$

1

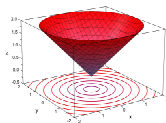


正八面体

1.5

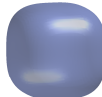
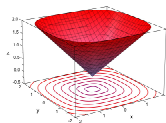


2

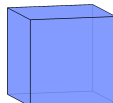
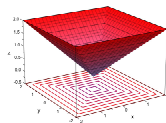


球

3



∞



立方体

ノルムを保存する演算の例

- 線形変換

$$\|\boldsymbol{x}\| = \|\boldsymbol{A}\boldsymbol{x}\|', \quad \boldsymbol{A} \in \mathbb{R}^{n \times n}: \text{正則}$$

- 非負結合

$$\|\boldsymbol{x}\| = \lambda_1 \|\boldsymbol{x}\|' + \lambda_2 \|\boldsymbol{x}\|'', \quad \lambda_1 > 0, \lambda_2 \geq 0$$

- 最大値演算

$$\|\boldsymbol{x}\| = \max\{\|\boldsymbol{x}\|', \|\boldsymbol{x}\|''\}$$

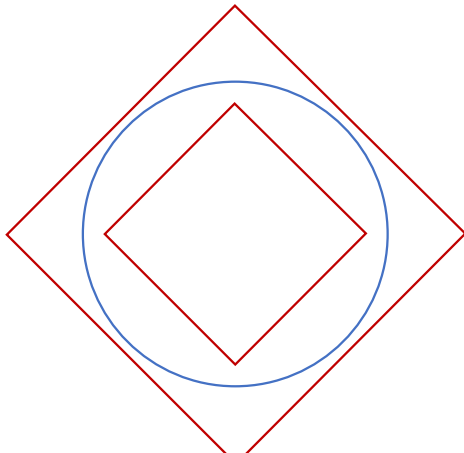
演習 (基礎問題)

- 上記の各ノルムが公理 1.~3. を満たすことを示せ

$\mathbb{R}^n$ 上のノルムの等価性

$$\exists L, U > 0, \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \quad L \leq \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|'} \leq U$$

$\mathbb{R}^n$ 上のノルムの等価性

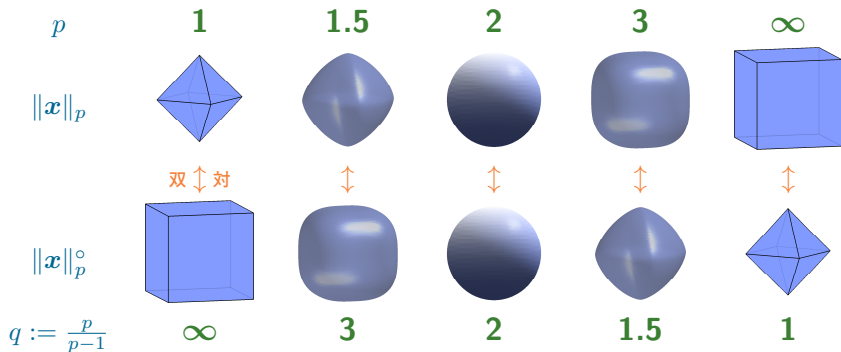


双対ノルム

$$\|x\|^\circ := \max_y \{x^\top y : \|y\| \leq 1\}$$

$\|\cdot\|$ の双対ノルム

例 $\|\cdot\|_p^\circ = \|\cdot\|_q \quad (\frac{1}{p} + \frac{1}{q} = 1)$



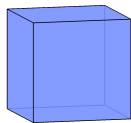
最大 k ノルム $\|\cdot\|_k$ (Watson,'92, Bertsimas, Pachamanova, Sim,'04)

- $x_{(1)}, \dots, x_{(n)}$: $(x_1, \dots, x_n)$ の要素の絶対値を降順で並べ替え
i.e., $|x_{(1)}| \geq |x_{(2)}| \geq \dots \geq |x_{(n)}|$
- $k \in \{1, \dots, n\}$: ユーザーが指定する自然数

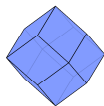
$$\|\mathbf{x}\|_k := |x_{(1)}| + |x_{(2)}| + \dots + |x_{(k)}|$$

最大 k ノルム ($k \in \{1, \dots, n\}$)

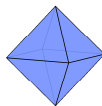
例 $\|(3, 0, 1, -5)\|_2 = |5| + |3| = 8$



$k = 1$ (ℓ_∞ ノルム)



$k = 2$



$k = 3$ (ℓ_1 ノルム)

- 異名色々... D ノルム (Bertsimas, Pachamanova, Sim, '04), ベクター k ノルム (Wu et al., '14), CVaR ノルム (Pavlikov, Uryasev, '13)

演習

$\|\mathbf{x}\|_k$ がノルムの公理を満たすことを示せ

Price of Robustness (Bertsimas, Sim,'04)

Ben-Tal, Nemirovski, 2000 のやり方では LP をロバスト化すると SOCP になる
保守性を抑えつつ LP のまま LP をロバスト化できるか？

$$\left| \begin{array}{ll} \min_{w \in W} & c^\top w \\ \text{s. t.} & \sum_{j=1}^n \tilde{a}_j w_j \leq b, \quad \forall \tilde{a}_j \in [a_j - \hat{a}_j, a_j + \hat{a}_j] \end{array} \right. \quad : \text{Soyster 型}$$
$$\Leftrightarrow \sum_{j=1}^n a_j w_j + \sum_{j=1}^n \hat{a}_j |w_j| \leq b$$
$$\Leftrightarrow \sum_{j=1}^n a_j w_j + \left\| (\hat{a}_1 w_1, \dots, \hat{a}_n w_n) \right\|_1 \leq b$$

↪ n 個の変数全部の最悪の動きを考えている

↪ 過度に保守的



高々 k 個の変数について最悪の動きを考える $k \leq n$ として

$$\sum_{j=1}^n a_j w_j + |\hat{a}_1 w_1|, \dots, |\hat{a}_n w_n| \text{ のうち高々 } k \text{ 個の和} \leq b$$

$\|\cdot\|_k$ の双対表現

$$\begin{aligned}\|x\|_k &= \min_{c \in \mathbb{R}} \left\{ kc + \sum_{i=1}^n \max\{|x_i| - c, 0\} \right\} \\&= \left| \begin{array}{ll} \min_{z, c} & kc + \mathbf{1}^\top z \\ \text{s. t.} & z \geq |x| - c\mathbf{1}, \\ & z \geq \mathbf{0} \end{array} \right. && \because \max \text{ を線形不等式で} \\&&& \text{分解} \\&= \left| \begin{array}{ll} \max_{\mathbf{q}} & |x_1|q_1 + \cdots + |x_n|q_n \\ \text{s. t.} & q_1 + \cdots + q_n = k, \\ & \mathbf{0} \leq \mathbf{q} \leq \mathbf{1} \end{array} \right. && \because \text{LP の双対定理} \\&= |x_{(1)}| + \cdots + |x_{(\lfloor k \rfloor)}| && \because \text{貪欲解法の要領で} \\&\quad + (k - \lfloor k \rfloor) |x_{(\lfloor k \rfloor + 1)}| && (k \text{ が実数の場合})\end{aligned}$$

$k \in \{1, \dots, n\}$ から $k \in [1, n] \curvearrowright$

$$\|x\|_k := \min_{c \in \mathbb{R}} \left\{ kc + \sum_{i=1}^n \max\{|x_i| - c, 0\} \right\}$$

$k \in [1, n]$ に拡張した最大 k ノルムの定義

k

3

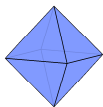
$\sqrt{2} + 1$

2

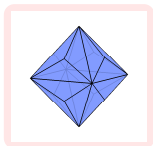
$\frac{3}{2}$

1

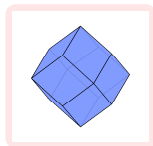
$\|x\|_k$



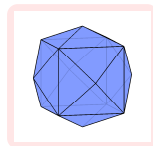
正八面体



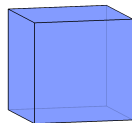
三方八面体



菱形十二面体



四方六面体



立方体

$k \leq n$ として

$$\sum_{j=1}^n a_j w_j + |\hat{a}_1 w_1|, \dots, |\hat{a}_n w_n| \text{ のうち高々 } k \text{ 個の和} \leq b$$

$\updownarrow$

$$\sum_{j=1}^n a_j w_j + \|(\hat{a}_1 w_1, \dots, \hat{a}_n w_n)\|_k \leq b \quad : w/ \text{ 最大 } k \text{ ノルム}$$

$\updownarrow$

$$\sum_{j=1}^n a_j w_j + \min_{c \in \mathbb{R}} \left\{ kc + \sum_{i=1}^n \max\{|x_i| - c, 0\} \right\} \leq b \quad : w/ \text{ 最適化}$$

Bertsimas, Sim, 2004: LP $\rightsquigarrow$ LP

$k \leq n$ として

$$\min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \sum_{j=1}^d a_j w_j + \underbrace{\left\| (\hat{a}_1 w_1, \dots, \hat{a}_d w_d) \right\|_k}_{|\hat{a}_1 w_1|, \dots, |\hat{a}_d w_d| \text{ のうち高々 } k \text{ 個の和}} \leq b$$

$$\Leftrightarrow \min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \sum_{j=1}^d a_j w_j + \min_{c \in \mathbb{R}} \left\{ kc + \sum_{i=1}^d \max\{|\hat{a}_i w_i| - c, 0\} \right\} \leq b$$

$$\Leftrightarrow \min_{\mathbf{w}, c} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \sum_{j=1}^d a_j w_j + kc + \sum_{i=1}^d \max\{|\hat{a}_i w_i| - c, 0\} \leq b$$

$$\Leftrightarrow \min_{\mathbf{w}, c, \mathbf{z}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \sum_{j=1}^d a_j w_j + kc + \sum_{i=1}^d z_i \leq b, \quad \begin{matrix} z_i \geq |\hat{a}_i w_i| - c, \\ z_i \geq 0, \quad i = 1, \dots, d \end{matrix}$$

$$\Leftrightarrow \min_{\mathbf{w}, c, \mathbf{z}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \sum_{j=1}^d a_j w_j + kc + \sum_{i=1}^d z_i \leq b, \quad \begin{matrix} z_i \geq \hat{a}_i y_i - c, \\ y_i \geq w_i, y_i \geq -w_i, \\ z_i \geq 0, \quad i = 1, \dots, d \end{matrix}$$

LP に帰着される

ロバスト LP $\rightsquigarrow$ LP

不確実性集合が非空な多面体 $U_i = \{\tilde{\mathbf{a}}_i \mid \mathbf{D}_i \tilde{\mathbf{a}}_i \leq \mathbf{d}_i\}$, $i = 1, \dots, m$ のケース：

$$\min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \tilde{\mathbf{a}}_i^\top \mathbf{w} \leq b_i, \quad \forall \tilde{\mathbf{a}}_i \in U_i, \quad i \in [m] := \{1, \dots, m\}$$

$$\Leftrightarrow \min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \max_{\tilde{\mathbf{a}}} \left\{ \tilde{\mathbf{a}}_i^\top \mathbf{w} \mid \mathbf{D}_i \tilde{\mathbf{a}}_i \leq \mathbf{d}_i \right\} \leq b_i, \quad i \in [m]$$

$\uparrow$ 双対変数 $\mathbf{z}_i \geq \mathbf{0}$

$$\Leftrightarrow \min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \min_{\mathbf{z}_i} \left\{ \mathbf{d}_i^\top \mathbf{z}_i \mid \mathbf{D}_i^\top \mathbf{z}_i = \mathbf{w}, \mathbf{z}_i \geq \mathbf{0} \right\} \leq b_i, \quad i \in [m] \quad \because \text{強双対性}$$

$$\Leftrightarrow \min_{\mathbf{w}, \mathbf{z}_i} \mathbf{c}^\top \mathbf{w} \quad \text{s.t.} \quad \mathbf{d}_i^\top \mathbf{z}_i \leq b_i, \quad \mathbf{D}_i^\top \mathbf{z}_i = \mathbf{w}, \quad \mathbf{z}_i \geq \mathbf{0}, \quad i \in [m]$$

LP に帰着される

ノルムで不確実性集合を表してみる

$$\begin{aligned} \sum_{j=1}^n \tilde{a}_j w_j &\leq b, \quad \forall \tilde{a}_j \in [a_j - \hat{a}_j, a_j + \hat{a}_j] \\ \Leftrightarrow \sum_{j=1}^n a_j w_j + \sum_{j=1}^n \hat{a}_j |w_j| &\leq b \end{aligned}$$

ノルムを用いて記述すると

$$\begin{aligned} \sum_{j=1}^n \tilde{a}_j w_j &\leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \left\| \left(\frac{\tilde{a}_1 - a_1}{\hat{a}_1}, \dots, \frac{\tilde{a}_n - a_n}{\hat{a}_n} \right) \right\|_{\infty} \leq 1 \\ \Leftrightarrow \sum_{j=1}^n a_j w_j + \left\| (\hat{a}_1 w_1, \dots, \hat{a}_n w_n) \right\|_1 &\leq b \end{aligned}$$

一般に任意のノルム $\|\cdot\|$ と定数 $\delta > 0$ を用いて

$$\begin{aligned} \tilde{\mathbf{a}}^{\top} \mathbf{w} &\leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq \delta && \text{無限本の線形不等式} \\ \Leftrightarrow \mathbf{a}^{\top} \mathbf{w} + \delta \|\mathbf{w}\|^{\circ} &\leq b && \text{1本の凸制約式} \end{aligned}$$

コスト (損失) が決定変数とパラメータの内積の場合

一般に任意のノルム $\|\cdot\|$ と定数 $\delta > 0$ を用いて

$$\tilde{\mathbf{a}}^\top \mathbf{w} \leq b, \quad \forall \tilde{\mathbf{a}} \in U = \{\tilde{\mathbf{a}} \mid \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq \delta\}$$

無限本の線形不等式

$$\Leftrightarrow \mathbf{a}^\top \mathbf{w} + \delta \|\mathbf{w}\|^\circ \leq b$$

1本の凸制約式

$$\therefore \quad \tilde{\mathbf{a}}^\top \mathbf{w} \leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq \delta$$

$$\Leftrightarrow \max_{\tilde{\mathbf{a}}} \{\tilde{\mathbf{a}}^\top \mathbf{w} \mid \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq \delta\} \leq b, \quad \rightsquigarrow \hat{\mathbf{a}} := \tilde{\mathbf{a}} - \mathbf{a} \text{ と置換え}$$

$$\Leftrightarrow \max_{\hat{\mathbf{a}}} \{(\mathbf{a} + \hat{\mathbf{a}})^\top \mathbf{w} \mid \|\hat{\mathbf{a}}\| \leq \delta\} \leq b,$$

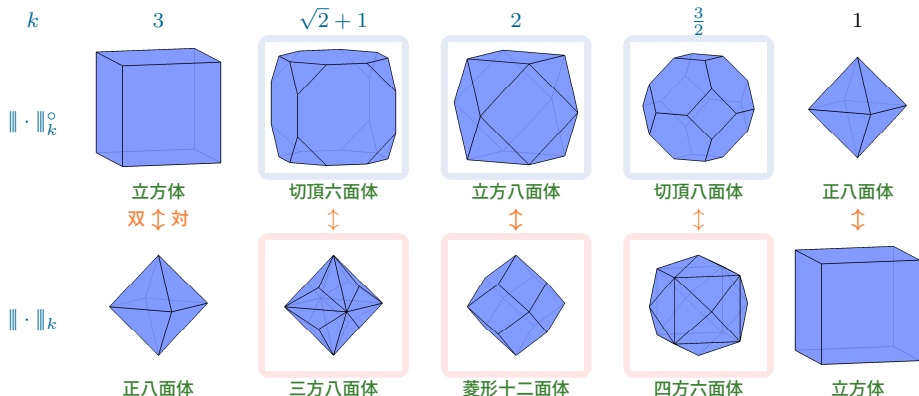
$$\Leftrightarrow \mathbf{a}^\top \mathbf{w} + \max_{\hat{\mathbf{a}}} \left\{ \hat{\mathbf{a}}^\top \mathbf{w} \mid \left\| \frac{\hat{\mathbf{a}}}{\delta} \right\| \leq 1 \right\} \leq b, \quad \rightsquigarrow \hat{\mathbf{a}}' := \hat{\mathbf{a}}/\delta \text{ と置換え}$$

$$\Leftrightarrow \mathbf{a}^\top \mathbf{w} + \delta \max_{\hat{\mathbf{a}}'} \{ \hat{\mathbf{a}}'^\top \mathbf{w} \mid \|\hat{\mathbf{a}}'\| \leq 1 \} \leq b,$$

$$\Leftrightarrow \mathbf{a}^\top \mathbf{w} + \delta \|\mathbf{w}\|^\circ \leq b$$

$\|\cdot\|_k$ の双対ノルム

$$\|x\|_k^\circ = \max\{\|x\|_1/k, \|x\|_\infty\}$$



- 上段はワイソフ切頂 (Wythoffian truncation) により得られる系列をなす
- 特に中 3 つは アルキメデスの立体 (半正多面体) , 下段の双対は カタランの立体

Bertsimas, Sim 流の不確実性集合

$$\mathbf{a}^\top \mathbf{w} + \|\mathbf{w}\|^\circ \leq b \quad \text{1本の凸制約式}$$

$$\Leftrightarrow \tilde{\mathbf{a}}^\top \mathbf{w} \leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq 1 \quad \text{無限本の線形不等式}$$

ノルムを用いて記述すると

$$\begin{aligned} & \sum_{j=1}^n \mathbf{a}_j w_j + \left\| (\hat{\mathbf{a}}_1 w_1, \dots, \hat{\mathbf{a}}_n w_n) \right\|_k \leq b \\ \Leftrightarrow & \sum_{j=1}^n \tilde{\mathbf{a}}_j w_j \leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \left\| \left(\frac{\tilde{\mathbf{a}}_1 - \mathbf{a}_1}{\hat{\mathbf{a}}_1}, \dots, \frac{\tilde{\mathbf{a}}_n - \mathbf{a}_n}{\hat{\mathbf{a}}_n} \right) \right\|_k^\circ \leq 1 \end{aligned}$$

ただし,

$$\|\mathbf{x}\|_k^\circ = \max\{\|\mathbf{x}\|_1/k, \|\mathbf{x}\|_\infty\}$$

ロバスト制約のセミパラメトリックな確率的保証

$$\begin{aligned} \min_{\mathbf{w} \in W} f(\mathbf{w}) \quad \text{s.t.} \quad & \tilde{\mathbf{a}}^\top \mathbf{w} \leq b, \quad \forall \tilde{\mathbf{a}} \text{ s.t. } \|\tilde{\mathbf{a}} - \mathbf{a}\| \leq \delta \\ \Leftrightarrow \min_{\mathbf{w} \in W} f(\mathbf{w}) \quad \text{s.t.} \quad & \mathbf{a}^\top \mathbf{w} + \delta \|\mathbf{w}\|^\circ \leq b \end{aligned} \quad \rightsquigarrow \mathbf{w}^*: \text{最適解}$$

命題 (Bertsimas, Pachamanova, Sim, '04)

$$\tilde{\mathbf{a}} \sim (\mathbf{a}, \Sigma) \text{ のとき, } \quad \mathbb{P}\{\tilde{\mathbf{a}}^\top \mathbf{w}^* \leq b\} \geq 1 - \frac{1}{1 + \left(\frac{\delta \|\Sigma^{1/2} \mathbf{w}^*\|^\circ}{\|\Sigma^{1/2} \mathbf{w}^*\|_2} \right)^2}$$

$$L \leq \|\mathbf{x}\|' / \|\mathbf{x}\|_2 \leq U \Leftrightarrow 1/U \leq \|\mathbf{x}\|'^\circ / \|\mathbf{x}\|_2 \leq 1/L$$

もし $\|\cdot\| = \|\cdot\|'$ ならば

$$\mathbb{P}\{\tilde{\mathbf{a}}^\top \mathbf{w}^* \leq b\} \geq 1 - \frac{1}{1 + (\delta/U)^2}$$

もし $\|\cdot\| = \|\cdot\|'^\circ$ ならば

$$\mathbb{P}\{\tilde{\mathbf{a}}^\top \mathbf{w}^* \leq b\} \geq 1 - \frac{1}{1 + (\delta L)^2}$$

最大 k ノルムと ℓ_p ノルムの近さ (Gotoh, Uryasev, '16)

$$\min\{1, kn^{-\frac{1}{p}}\} \leq \left\{ \frac{\|\mathbf{x}\|_k / \|\mathbf{x}\|_p}{\|\mathbf{x}\|_q / \|\mathbf{x}\|_k^\circ} \right\} \leq \left\{ \lfloor k \rfloor + (k - \lfloor k \rfloor)^{\frac{p}{p-1}} \right\}^{\frac{p-1}{p}}$$

タイトな下限 L

タイトな上限 U

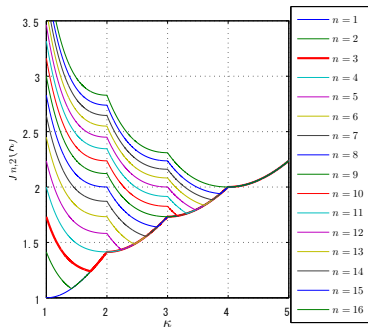
ただし, $\frac{1}{p} + \frac{1}{q} = 1$

$$f_{n,p}(k) := \frac{\left\{ \lfloor k \rfloor + (k - \lfloor k \rfloor)^{\frac{p}{p-1}} \right\}^{\frac{p-1}{p}}}{\min\{1, n^{-\frac{1}{p}} k\}}$$

→ $k = n^{1/p}$ で最小値

$$\left\{ \lfloor n^{\frac{1}{p}} \rfloor + (n^{\frac{1}{p}} - \lfloor n^{\frac{1}{p}} \rfloor)^{\frac{p}{p-1}} \right\}^{\frac{p-1}{p}}$$

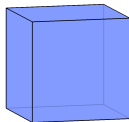
注: $\|\cdot\|_k$ と $\|\cdot\|_p$ の近さ = $\|\cdot\|_k^\circ$ と $\|\cdot\|_q$ の近さ



ノルムとノルムの近さ



← どれくらい似ているか? →



注意: $\mathbb{R}^n$ 上のノルムの等価性

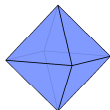
$$\exists L, U > 0 : \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \quad L \leq \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|'} \leq U, \quad (\text{or } \frac{1}{U} \leq \frac{\|\mathbf{x}\|'}{\|\mathbf{x}\|} \leq \frac{1}{L})$$

Definition ($\|\cdot\|$ と $\|\cdot\|'$ の近さ)

1. **タイトな** L と U を求める
2. U/L (または $\log U/L$) を “近さ” とする

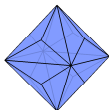
- $\log U/L \geq 0$ for any $\|\cdot\|, \|\cdot\|'$
- $\log U/L = 0$ for $\|\cdot\| = \|\cdot\|'$
- $U/L = \sqrt{n}$ for $(\|\cdot\|, \|\cdot\|') = (\|\cdot\|_1, \|\cdot\|_2)$ or $(\|\cdot\|_\infty, \|\cdot\|_2)$

$$\sqrt{x_1^2 + \cdots + x_n^2} \approx |x_{(1)}| + \cdots + |x_{(\sqrt{n})}|$$

 k
 3
 $\sqrt{2} + 1$
 2
 $\sqrt{3}$
 $\frac{3}{2}$
 1
 $\| \cdot \|_k$


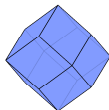
正八面体

双 $\updownarrow$ 对



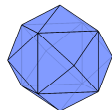
三方八面体

$\updownarrow$

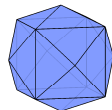


菱形十二面体

$\updownarrow$

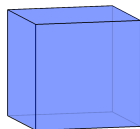


$\updownarrow$



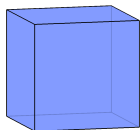
四方六面体

$\updownarrow$

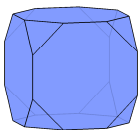


立方体

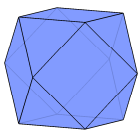
$\updownarrow$

 $\| \cdot \|_k^\circ$


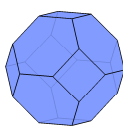
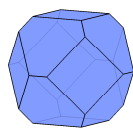
立方体



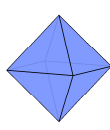
切頂六面体



立方八面体

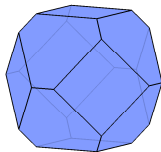


切頂八面体

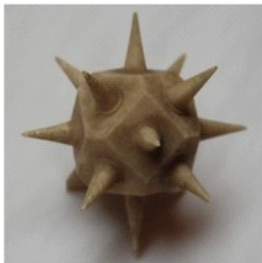


正八面体

一見少しイビツですが...



(a)



(b)



(c)

Lengyel,

カードモデル；ドイツの象牙の部品（17 世紀）；韓国 新羅時代（7-9 世紀）のサイコロ

Gáspár, Tarnai, '17 <http://www.mdpi.com/2073-8994/9/3/41/htm>

ここまでのまとめ

- RO は不確実なパラメータを含むような最適化問題に対して最悪な場合を考える.
- 最悪なパラメータが取りうる範囲を集合で記述した集合を与える
- ノルムによって不確実性集合が与えられていると, 双対ノルムを使って定式化を簡約できる
- 多面体のノルムを利用すると, 問題のクラスを変えないままロバスト化できる
- 定式化の簡約化において双対性が重要な役割を果たす
(最近の展開の例: Zhen, Kuhn, Wiesemann, 2021, arXiv)

平均・分散ポートフォリオ選択 (Markowitz, '52)

- $\mathbf{w} = (w_1, \dots, w_n)^\top \in \mathbb{R}^n$:

n 種の資産への投資配分比率 (ポートフォリオ)

- $\mathcal{R} = (\mathcal{R}_1, \dots, \mathcal{R}_n)^\top$:

n 種の資産の収益率 (確率変数)

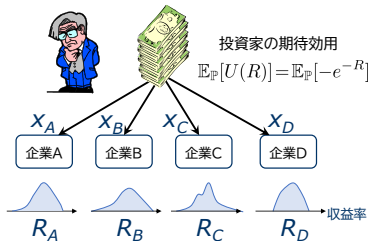
$$\leadsto \mathcal{R}^\top \mathbf{w} = R_1 w_1 + \dots + R_n w_n:$$

ポートフォリオ収益率

$$\text{MV}_\lambda := \left| \begin{array}{ll} \max_{\mathbf{w}} & \mathbb{E}[\mathcal{R}^\top \mathbf{w}] - \lambda \mathbb{V}[\mathcal{R}^\top \mathbf{w}] = \underbrace{\boldsymbol{\mu}^\top \mathbf{w}}_{\text{期待収益率ベクトル}} - \lambda \underbrace{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}_{\text{共分散行列}} \\ \text{s. t.} & \mathbf{1}^\top \mathbf{w} = 1, \mathbf{w} \geq 0 \quad \text{諸々の制約} \end{array} \right.$$

- 現実には $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ は推定値を用いる $\leadsto \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}$
- 色々な推定が考えられている (いずれも推定誤差があるはず)
- $\boldsymbol{\mu}$ の推定誤差の影響の方が $\boldsymbol{\Sigma}$ のそれより深刻 (e.g., Chopra, Ziemba, '93)

いくらずつ投資したら良いか?



μ の不確実性を考慮したロバスト最適化

$\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_n)^\top$ の推定誤差を考える

$$\text{MV}_\lambda(\delta) := \begin{cases} \max_w \min_{\mu \in U(\hat{\mu})} & \mu^\top w - \lambda w^\top \Sigma w \\ \text{s. t.} & \mathbf{1}^\top w = 1, w \geq 0 \end{cases}$$

不確実性集合 U の例：

- $U = U_\delta(\hat{\mu}) := \{\mu \in \mathbb{R}^d \mid |\mu_i - \hat{\mu}_i| \leq \delta_i, i = 1, \dots, n\}$
- $U = U_{\delta, \Sigma_\mu}(\hat{\mu}) := \{\mu \in \mathbb{R}^d \mid (\mu - \hat{\mu})^\top \Sigma_\mu^{-1} (\mu - \hat{\mu}) \leq \delta^2\}$

演習

- $U = U_\delta(\hat{\mu}), U_{\delta, \Sigma_\mu}(\hat{\mu})$ の場合の RO の目的関数がそれぞれ

$$\max_w \hat{\mu}^\top w - \delta^\top |w| - \lambda w^\top \hat{\Sigma} w$$

$$\max_w \hat{\mu}^\top w - \delta \sqrt{w^\top \Sigma_\mu w} - \lambda w^\top \hat{\Sigma} w$$

となることを確かめよ

分散共分散行列 $\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{1,1} & \cdots & \hat{\sigma}_{1,n} \\ \vdots & & \vdots \\ \hat{\sigma}_{n,1} & \cdots & \hat{\sigma}_{n,n} \end{pmatrix}$ の推定誤差を考える

$$\text{MV}_\lambda(\delta, \underline{\Sigma}, \bar{\Sigma}) := \left| \begin{array}{ll} \max_w \min_{(\mu, \Sigma) \in U_\mu(\hat{\mu}) \times U_\Sigma(\hat{\Sigma})} & \mu^\top w - \lambda w^\top \Sigma w \\ \equiv \max_w \min_{\mu \in U_\mu(\hat{\mu})} & \mu^\top w - \lambda \max_{\Sigma \in U_\Sigma(\hat{\Sigma})} w^\top \Sigma w \\ \text{s. t.} & \mathbf{1}^\top w = 1, \mathbf{A}w \leq \mathbf{b} \end{array} \right.$$

不確実性集合 U_Σ の例：

- $U_\Sigma = U_{\underline{\Sigma}, \bar{\Sigma}}(\hat{\Sigma}) := \{\Sigma \in \mathbb{S}^d \mid \underline{\Sigma} \leq \Sigma \leq \bar{\Sigma}, \Sigma \succeq O\}$

ただし

- $\mathbb{S}^d$ は $d \times d$ の実対称行列全体からなる集合
- $\Sigma \succeq O$ は Σ が半正定値行列であることを表す
- $\Sigma \succeq O \Leftrightarrow$ 対応する確率変数が存在する

$$\begin{aligned} \max_{\Sigma} \quad & w^\top \Sigma w \\ \text{s. t.} \quad & \underline{\Sigma} \leq \Sigma \leq \bar{\Sigma}, \Sigma \succeq O \end{aligned}$$

- 半正定値計画 (SDP) と呼ばれる問題のクラスの問題例
 $\leadsto$ (適当な制約想定の下) 双対問題との間に強双対定理が成り立つ:

$$\begin{aligned} \min_{\underline{\Lambda}, \bar{\Lambda}} \quad & \bar{\Sigma} \bullet \bar{\Lambda} - \underline{\Sigma} \bullet \underline{\Lambda} \\ \text{s. t.} \quad & \underbrace{\bar{\Lambda} - \underline{\Lambda} - ww^\top}_{\left(\begin{array}{cc} \bar{\Lambda} - \underline{\Lambda} & w \\ w^\top & 1 \end{array} \right) \succeq O} \succeq O, \quad \bar{\Lambda}, \underline{\Lambda} \geq O \end{aligned}$$

$MV_\lambda(\delta, \Sigma, \bar{\Sigma})$ は以下の線形 SDP に帰着:

$$\left| \begin{array}{l} \max_{w_+, w_-, \underline{\Lambda}, \bar{\Lambda}} \quad \hat{\mu}^\top (w_+ - w_-) - \delta^\top (w_+ + w_-) - \lambda (\bar{\Sigma} \bullet \bar{\Lambda} - \underline{\Sigma} \bullet \underline{\Lambda}) \\ \text{s. t.} \quad 1^\top (w_+ - w_-) = 1, w_+, w_- \geq 0, A(w_+ - w_-) \leq b \\ \left(\begin{array}{cc} \bar{\Lambda} - \underline{\Lambda} & (w_+ - w_-) \\ w_+^\top - w_-^\top & 1 \end{array} \right) \succeq O, \bar{\Lambda}, \underline{\Lambda} \geq O \end{array} \right.$$

- 実務において、収益率のヒストリカルデータのみを使ってパラメータ推定せずファクターモデルを介すことが多い
- ファクターモデルは各資産の収益率を経済変量による線形モデルを推定したもの
- 適当な仮定の下、ファクターモデルのパラメータに統計的な信頼領域を考えることができるので、適当な信頼水準の下で信頼領域を考え、それらを不確実性集合として与える
- 結果、ロバスト化した平均・分散モデルは SOCP になる

VaR, CVaR

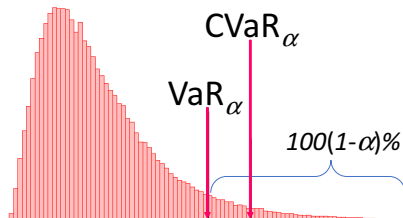
収益率 $\mathcal{R} = (\mathcal{R}_1, \dots, \mathcal{R}_n)^\top$ が多次元正規分布に従うならば平均分散モデルで十分
実際には、正規分布よりも裾が厚かったり、非対称であったりすることが知られている
裾の振る舞いに関する分かり易い尺度が VaR (損失の上側 $100\alpha\%$ 点)

$$\text{VaR}_\alpha[\mathcal{L}] := \min_{z \in \mathbb{R}} \{z \mid \mathbb{P}\{\mathcal{L} \leq z\} \geq \alpha\} \quad \approx \quad z \text{ s.t. } \mathbb{P}\{\mathcal{L} \leq z\} = \alpha$$

しかし、VaR は劣加法性を持たず (非凸!)、実務的にも好ましくない
VaR 以上の損失の条件付き期待値を意味する CVaR は重要な代替的尺度

$$\text{CVaR}_\alpha[\mathcal{L}] := \min_{c \in \mathbb{R}} \left\{ c + \frac{1}{1-\alpha} \mathbb{E}[\max\{\mathcal{L} - c, 0\}] \right\} \quad \approx \quad \mathbb{E}[\mathcal{L} \mid \mathcal{L} > \text{VaR}_\alpha[\mathcal{L}]]$$

確率



損失 L

CVaR 最小化ポートフォリオ選択

$\mathcal{L} = \mathcal{L}(\mathbf{x}) = -\mathcal{R}^\top \mathbf{x} = -(\mathcal{R}_1 x_1 + \cdots + \mathcal{R}_d x_d)$ とし,

$p_i := \mathbb{P}\{(\mathcal{R}_1, \dots, \mathcal{R}_d) = (R_{i,1}, \dots, R_{i,d})\} = \frac{1}{n}, \quad i = 1, \dots, n$

とすると, CVaR 最小化ポートフォリオ選択は LP に帰着される:

$$\begin{aligned} \min_{\mathbf{x}} \quad \text{CVaR}_\alpha[\mathcal{L}(\mathbf{x})] &= \min_{\mathbf{x}, c} \left\{ c + \frac{1}{1-\alpha} \mathbb{E}[\max\{\mathcal{L}(\mathbf{x}) - c, 0\}] \right\} \\ &= \min_{\mathbf{x}, c} \quad c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i \max\{-\mathbf{R}_i^\top \mathbf{x} - c, 0\} \\ &= \begin{cases} \min_{\mathbf{x}, c} & c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i z_i \\ \text{s. t.} & z_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, \quad z_i \geq 0, \quad i \in [n] \end{cases} \end{aligned}$$

- CVaR は分布の裾全体の $100(1 - \alpha)\%$ のシナリオにのみ依存

(通常 $\alpha = 0.95, 0.99, 0.999$ のように 1 に近い値で固定される)

→ シナリオの値 R_t のぶれに対するロバスト最適化を考えてみる

$$\mathbf{R}_i \rightsquigarrow \mathbf{R}_i - \Delta_i, \quad \forall \Delta_i \text{ s.t. } \|\Delta_i\| \leq \delta, \quad i \in [n]$$

ロバスト CVaR 最小化ポートフォリオ選択

$$\begin{aligned}
 & \min_{\mathbf{x}} \text{WCVaR}_{\alpha}[\mathcal{L}(\mathbf{x})] \\
 &= \left| \begin{array}{ll} \min_{\mathbf{x}, c, \mathbf{z}} & c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i z_i \\ \text{s. t.} & z_i \geq (-\mathbf{R}_i + \mathbf{\Delta}_i)^{\top} \mathbf{x} - c, \quad \forall \mathbf{\Delta}_i \text{ s.t. } \|\mathbf{\Delta}_i\| \leq \delta, \quad i \in [n] \\ & z_i \geq 0, \quad i \in [n] \end{array} \right. \\
 &= \left| \begin{array}{ll} \min_{\mathbf{x}, c, \mathbf{z}} & c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i z_i \\ \text{s. t.} & z_i \geq -\mathbf{R}_i^{\top} \mathbf{x} + \delta \|\mathbf{x}\|^{\circ} - c, \quad i \in [n] \quad \rightsquigarrow c = c' + \delta \|\mathbf{x}\|^{\circ} \text{とおく} \\ & z_i \geq 0, \quad i \in [n] \end{array} \right. \\
 &= \left| \begin{array}{ll} \min_{\mathbf{x}, c', \mathbf{z}} & c' + \frac{1}{1-\alpha} \sum_{i=1}^n p_i z_i + \delta \|\mathbf{x}\|^{\circ} \\ \text{s. t.} & z_i \geq -\mathbf{R}_i^{\top} \mathbf{x} - c', \quad i \in [n] \\ & z_i \geq 0, \quad i \in [n] \end{array} \right. \\
 &\equiv \min_{\mathbf{x}} \text{CVaR}_{\alpha}[\mathcal{L}(\mathbf{x})] + \delta \|\mathbf{x}\|^{\circ}
 \end{aligned}$$

- ロバスト化 = 正則化

① ロバスト最適化

- Soyster 1973, Ben-Tal, Nemirovski, 2000, Bertsimas, Sim, 2004
- 応用例
 - ロバストポートフォリオ選択
 - SVM

② 分布的ロバスト最適化 (DRO)

- リスク尺度と DRO
- DRO の含意

③ DRO の事後パフォーマンス

二値判別問題

- $\{(x_1, y_1), \dots, (x_n, y_n)\}$: 標本の集合 (所与)
 - ▶ $x_i \in \mathbb{R}^d$: 標本 i の属性
 - ▶ $y_i \in \{\pm 1\}$: 標本 i のラベル
- 既知の標本データに基づき未知の標本 x' のラベル y' を予測する決定関数 $d(\cdot)$ を構築したい

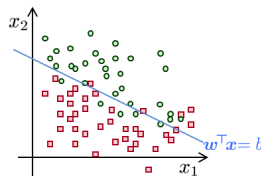
$$y' \leftarrow \boxed{d(x)} \leftarrow x'$$

応用例

- 信用スコアリング 入力 x = 財務指標; y = 信用あり vs なし
- 手書き文字認識 入力 x = 画素データ; y = 7 or 0,1,...,6,8,9?
- eメールフィルタリング 入力 x = 関連キーワードの含有; y = スパム vs 無害?
- 推薦システム 入力 x = 購買履歴; y = ビール vs ワイン?

二値判別の決定関数は通常以下の形：

$$d(x) = \begin{cases} +1, & \text{if } w^\top x \geq b \\ -1, & \text{if } w^\top x < b \end{cases}$$



線形分離可能性

- $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$: 標本集合 (所与)

- ▶ $\mathbf{x}_i \in \mathbb{R}^d$: 標本 i の属性
- ▶ $y_i \in \{\pm 1\}$: 標本 i のラベル

- $\mathbf{w}^\top \mathbf{x} = b$: 超平面, $\mathbf{w} \neq \mathbf{0}$

- 標本集合が線形分離可能 (linearly separable)

$$\stackrel{\text{def.}}{\iff} \exists(\mathbf{w}, b) \in \mathbb{R}^d \times \mathbb{R}, \quad \forall i = 1, \dots, n, \\ \left\{ \begin{array}{ll} y_i = +1 & \Rightarrow \mathbf{w}^\top \mathbf{x}_i > b \\ y_i = -1 & \Rightarrow \mathbf{w}^\top \mathbf{x}_i < b \end{array} \right\} \iff y_i(\mathbf{w}^\top \mathbf{x}_i - b) > 0$$

$$\iff y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1 \iff 1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b) \leq 0$$

- 線形分離不可能 $\iff \forall(\mathbf{w}, b), \exists i = 1, \dots, n, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b) > 0$

- $\min_{\mathbf{w}, b} \sum_{i \in [n]} \max\{1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b), 0\}$

$$\iff \left| \begin{array}{ll} \min_{\mathbf{w}, b} & \sum_{i \in [n]} z_i \\ \text{s. t.} & z_i \geq 1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b), \quad z_i \geq 0, \quad i \in [n] \end{array} \right.$$

$$\min_{\mathbf{w}, b} \sum_{i \in [n]} \max\{1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b), 0\}$$

データの摂動 $\mathbf{x}_i \rightsquigarrow \mathbf{x}_i + \Delta_i$ を考える

$$U = \{(\Delta_1, \dots, \Delta_n) \mid \sum_{i \in [n]} \|\Delta_i\| \leq \delta\}$$

$$\begin{aligned} & \min_{\mathbf{w}, b} \sum_{i \in [n]} \max\{1 - y_i[\mathbf{w}^\top (\mathbf{x}_i + \Delta_i) - b], 0\} \\ &= \min_{\mathbf{w}, b} \delta \|\mathbf{w}\|^\circ + \sum_{i \in [n]} \max\{1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b), 0\} \end{aligned}$$

ロバスト Robust LP $\rightsquigarrow$ C-SVM

また、回帰分析でのデータの摂動に対するロバスト化を考えると LASSO やリッジ回帰と関係づけられることが示されている (El Ghaoui, Le Bret, 1997)

この項のまとめ

- RO は双対性を用いることでシンプルな凸計画問題に帰着できる場合が結構ある
- 観測値や推定パラメータに対する不確実性を考えた RO は、判別/回帰などのモデルの推定やポートフォリオ選択などの問題で考えると、決定変数に対する正則化の役割と対応すると見なせる例が結構指摘されている

① ロバスト最適化

- Soyster 1973, Ben-Tal, Nemirovski, 2000, Bertsimas, Sim, 2004
- 応用例
 - ロバストポートフォリオ選択
 - SVM

② 分布的ロバスト最適化 (DRO)

- リスク尺度と DRO
- DRO の含意

③ DRO の事後パフォーマンス

確率計画

$$\begin{aligned} \min_{\boldsymbol{x} \in X} \quad & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{Y})] \\ \text{s. t.} \quad & \mathbb{E}_{\mathbb{P}}[g_i(\boldsymbol{x}, \boldsymbol{Y})] \leq b_i, \quad i = 1, \dots, m \end{aligned}$$

- f : コスト関数
- $\boldsymbol{x}$: 決定変数 (ベクトル)
- $\boldsymbol{Y}$: 確率変数 (ベクトル)
- $\mathbb{E}_{\mathbb{P}}[\cdot]$: 確率分布 $\mathbb{P}$ の下での期待値

【注】 $\min_{\boldsymbol{x}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{Y})] = -\max_{\boldsymbol{x}} \mathbb{E}_{\mathbb{P}}[-f(\boldsymbol{x}, \boldsymbol{Y})] \quad \rightsquigarrow -f : \text{報酬関数}$

分布的口バスト最適化 Distributionally Robust Optimization (DRO)

$$\begin{aligned} \min_{\boldsymbol{x} \in X} \quad & \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{Y})] \\ \text{s. t.} \quad & \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[g_i(\boldsymbol{x}, \boldsymbol{Y})] \leq b_i, \quad i = 1, \dots, m \end{aligned}$$

- f : コスト関数
- $\boldsymbol{x}$: 決定変数 (ベクトル)
- $\boldsymbol{Y}$: 確率変数 (ベクトル)
- $\mathbb{E}_{\mathbb{Q}}[\cdot]$: 確率分布 $\mathbb{Q}$ の下での期待値
- $\mathcal{Q}$: $\mathbb{Q}$ の不確実性集合 (ambiguity set とも言う)

$$\begin{aligned} \Leftrightarrow \max_{\boldsymbol{x} \in X} \quad & \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[-f(\boldsymbol{x}, \boldsymbol{Y})] \\ \text{s. t.} \quad & \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[-g_i(\boldsymbol{x}, \boldsymbol{Y})] \geq -b_i, \quad i = 1, \dots, m \end{aligned}$$

$$\min_{\boldsymbol{x} \in X} \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{Y})]$$

$$\Leftrightarrow \max_{\boldsymbol{x} \in X} \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[-f(\boldsymbol{x}, \boldsymbol{Y})]$$

「意思決定者が決めた $\boldsymbol{x}$ に対し，nature が意思決定者にとって最も都合悪い $\mathbb{Q}$ を $\mathcal{Q}$ から選ぶ」という枠組みを前提として，意思決定者が $\boldsymbol{x}$ を決める

【例】 新聞売り子問題

単価 c で仕入れ、売れば r だけ、売れなければ q だけ戻ってくる

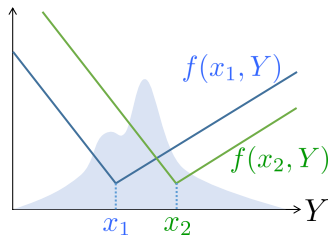
$$f(x, Y) = - \left[\underbrace{r \min\{x, Y\}}_{\text{売上}} + \underbrace{q \max\{x - Y, 0\}}_{\text{残余価値}} - \underbrace{s \max\{Y - x, 0\}}_{\text{機会費用}} - \underbrace{cx}_{\text{仕入費用}} \right]$$

負の利益
売上
残余価値
機会費用
仕入費用

- x : 仕入れ数 ($x \geq 0$) [決定変数]
- Y : 需要 [確率変数] ($Y \geq 0$)

$E := c - r$, $U = q + s - c$, $V = q - c$ とおくと

$$f(x, Y) = E[x - Y]^+ + U[Y - x]^+ - VY$$



期待利益最大化：

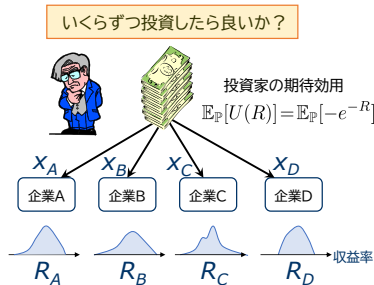
$$\min_x \mathbb{E}_{\mathbb{P}}[f(x, Y)] = \int_0^{\infty} f(x, Y) dF_Y$$

- $0 \leq q < c < r$, $s \geq 0$ を仮定すると凸最小化
- Y の分布関数 F_Y が逆関数を持つならば、最適解は $x^* = F^{-1}\left(\frac{U}{E+U}\right)$

[例] ポートフォリオ選択問題

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{Y})] \\ \text{s.t.} \quad & \mathbf{1}^{\top} \mathbf{x} = 1, \dots \end{aligned}$$

- $\mathbf{x} \in \mathbb{R}^d$: ポートフォリオ [決定変数]
- $\mathbf{R} \in \mathbb{R}^d$: 収益率 [確率変数]



$f(\mathbf{x}, \mathbf{R}) := -\mathbf{R}^{\top} \mathbf{x} \rightsquigarrow$ 期待収益率最大化

$f(\mathbf{x}, \mathbf{R}) := \exp(-\mathbf{R}^{\top} \mathbf{x}) \rightsquigarrow$ 期待（指数）効用最大化

$f(\mathbf{x}, \mathbf{R}) := \min_c \left\{ c + \frac{1}{1 - \alpha} [-\mathbf{R}^{\top} \mathbf{x} - c]^+ \right\} \rightsquigarrow$ CVaR 最小化

$f(\mathbf{x}, \mathbf{R}) := -\mathbf{R}^{\top} \mathbf{x} + \lambda \min_c \left\{ (\mathbf{R}^{\top} \mathbf{x} - c)^2 \right\} \rightsquigarrow$ 平均分散モデル

標本平均近似 (SAA)

$$\min_x \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{Y})]$$

しばしば $\mathbb{P} =$ 経験分布 $\mathbb{P}_n$ で代用

確率	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$
取りうる値	$\mathbf{Y}_1$	$\mathbf{Y}_2$	$\dots$	$\mathbf{Y}_n$

標本平均近似 (Sample Average Approximation: SAA)

$$\min_x \left\{ \mathbb{E}_{\mathbb{P}_n}[f(\mathbf{x}, \mathbf{Y})] := \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}, \mathbf{Y}_i) \right\}$$

- しかし SAA の解の事後性能 ($\mathbb{P}$ の下でのパフォーマンス) の劣化が著しい (特に n が少ない場合や標本に偏りがある場合など)

$\therefore$ 経験分布 $\mathbb{P}_n \neq$ 真の確率モデル $\mathbb{P}$

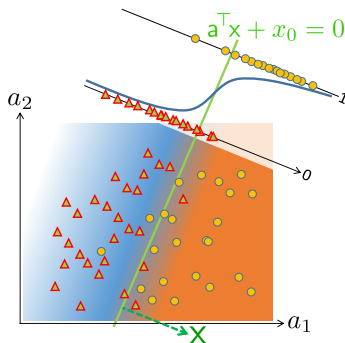
[例] ロジスティック回帰

$$f((\mathbf{x}, x_0), (b, \mathbf{a})) = \ln(1 + \exp(-b(\mathbf{x}^\top \mathbf{a} + x_0)))$$

- f : 負の対数尤度の 1 標本分

$$\leadsto \mathbb{E}_{\mathbb{P}_n}[f((\mathbf{x}, x_0), (b, \mathbf{a}))] = \frac{1}{n} \sum_{i=1}^n f((\mathbf{x}, x_0), (b_i, \mathbf{a}_i))$$

- $(\mathbf{x}, x_0) \in \mathbb{R}^{d+1}$: 線形モデルの係数ベクトルと切片項
- $\mathbf{a} \in \mathbb{R}^d$, $b \in \{\pm 1\}$: 線形モデルの説明変数とバイナリクラスラベル



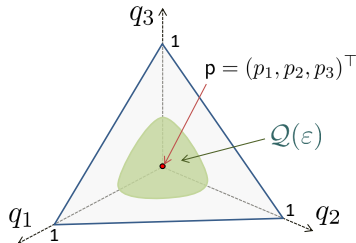
SAA から DRO へ

$$\min_{\mathbf{x}} \left\{ \mathbb{E}_{\mathbb{P}_n} [f(\mathbf{x}, \mathbf{Y})] := \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}, \mathbf{Y}_i) \right\} \quad \text{標本平均近似 (SAA)}$$

$$\min_{\mathbf{x}} \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [f(\mathbf{x}, \mathbf{Y})] \quad \text{分布的ロバスト最適化 (DRO)}$$

- $\mathcal{Q}$: 不確実性集合 (凸集合に限定して OK)

$\leadsto \mathcal{Q} = \mathcal{Q}(\varepsilon)$: 大きさのパラメータ $\varepsilon \geq 0$ に依存 ; $\mathcal{Q}(0) = \{\mathbb{P}_n\}$



期待収益率最大化ポートフォリオの DRO

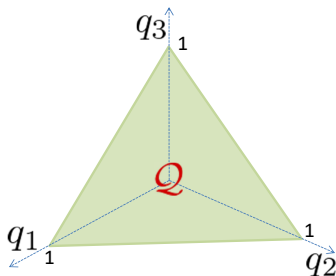
試しに $f(x, Y) = -R^\top x$ とする

$$\min_x \max_{Q \in \mathcal{Q}} \left\{ \mathbb{E}_Q[-R^\top x] = -(q_1 R_1^\top x + \cdots + q_n R_n^\top x) \right\}$$

- $\mathcal{Q} = \Pi := \{q = (q_1, \dots, q_n)^\top \mid \mathbf{1}^\top q = 1, q \geq 0\}$ のとき

$$\begin{aligned} \max_{Q \in \mathcal{Q}} \mathbb{E}_Q[-R^\top x] &= \max_q \left[\begin{array}{l} q_1(-R_1^\top x) + \cdots + q_n(-R_n^\top x) \\ \text{s.t. } q_1 + \cdots + q_n = 1, q_1, \dots, q_n \geq 0 \end{array} \right] \\ &= \max\{-R_1^\top x, \dots, -R_n^\top x\} \end{aligned}$$

$$\leadsto \left| \begin{array}{ll} \min_{x, z} & z \\ \text{s.t.} & z \geq -R_i^\top x, \quad i = 1, \dots, n, \\ & \mathbf{1}^\top x = 1, \dots \end{array} \right.$$



期待収益率最大化ポートフォリオの DRO

試しに $f(x, Y) = -R^\top x$ とする

$$\min_x \max_{Q \in \mathcal{Q}} \left\{ \mathbb{E}_Q[-R^\top x] = \sum_{i=1}^n -(q_1 R_1^\top x + \cdots + q_n R_n^\top x) \right\}$$

- $\mathcal{Q}(\varepsilon) = \left\{ q := (q_1, \dots, q_n) \mid \sum_{i=1}^n q_i \ln \frac{q_i}{p_i} \leq \varepsilon, \mathbf{1}^\top q = 1, q \geq 0 \right\}$ のとき

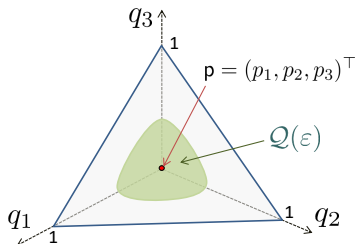
$$\max_{Q \in \mathcal{Q}} \mathbb{E}_Q[-R^\top x] = \begin{cases} \max_q & q_1(-R_1^\top x) + \cdots + q_n(-R_n^\top x) \\ \text{s.t.} & q_1 + \cdots + q_n = 1, q_1, \dots, q_n \geq 0, \sum_{i=1}^n q_i \ln \frac{q_i}{p_i} \leq \varepsilon \end{cases}$$

$$D_{\text{KL}}(Q|\mathbb{P}) = \sum_{i=1}^n q_i \ln \frac{q_i}{p_i}$$

: Kullback-Leibler ダイバージェンス

一般に

- $D(Q|\mathbb{P}) \geq 0$ for 任意の確率分布 $Q, \mathbb{P}$
- $D(\mathbb{P}|\mathbb{P}) = 0$



リスク尺度としての最悪ケースの期待値

f をコスト（損失）を表す確率変数とし，DRO の目的関数の部分を考える：

$$\rho[f] := \max_{Q \in \mathcal{Q}} \mathbb{E}_Q[f]$$

これはコスト f のリスクを返す（汎）関数「リスク尺度」である
上の ρ が以下を満たすことは簡単に確認できる

- 単調（非減少）性

$$f_i \leq g_i \Rightarrow \rho[f] \leq \rho[g]$$

- 正 1 次同次性

$$t > 0 \Rightarrow \rho[tf] = t\rho[f]$$

- 平行移動に関する不変性

$$a \in \mathbb{R} \Rightarrow \rho[f + a] = \rho[f] + a$$

- 劣加法性

$$\rho[f + g] \leq \rho[f] + \rho[g]$$

リスク尺度としての最悪ケースの期待値

f をコスト（損失）を表す確率変数とし，DRO の目的関数の部分を考える：

$$\rho[f] := \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f] \quad (*)$$

これはコスト f のリスクを返す（汎）関数「リスク尺度」である
上の ρ が以下を満たすことは簡単に確認できる

- ① 単調性 $f_i \leq g_i \Rightarrow \rho[f] \leq \rho[g]$
- ② 正 1 次同次性 $t \geq 0 \Rightarrow \rho[tf] = t\rho[f]$
- ③ 平行移動に関する不変性 $a \in \mathbb{R} \Rightarrow \rho[f + a] = \rho[f] + a$
- ④ 劣加法性 $\rho[f + g] \leq \rho[f] + \rho[g]$

この 4 つの性質（公理）を満たすリスク尺度を**コヒレント**という

逆に，任意のコヒレントなリスク尺度 ρ に対して確率分布の集合 $\mathcal{Q}$ が存在して
(*) のように書けることが知られている (Huber, 1981)

つまり，DRO はコヒレントリスクの最小化とも見れる

コヒレントリスク尺度の双対表現 Huber 1981, Artzner et al.1999

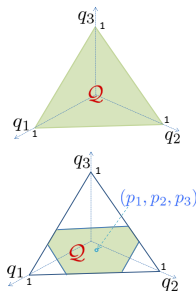
ρ がコヒレントリスク尺度 $\Leftrightarrow \exists \mathcal{Q}$: 確率分布の (閉凸な) 集合

$$\rho[\mathcal{L}] = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[\mathcal{L}]$$

たとえば $\mathbb{P}\{\mathcal{L} = L_t\} = p_i, i = 1, \dots, n$ である離散型確率変数 $\mathcal{L}$ に対して

$$\rho[\mathcal{L}] = \max\{L_1, \dots, L_n\} \stackrel{\text{強双対性}}{=} \begin{cases} \max_q q^\top L \\ \text{s.t.} \quad \mathbf{1}^\top q = 1, q \geq 0 \end{cases}$$

$$\rho[\mathcal{L}] = \begin{cases} \min_{c, z} \alpha + \frac{1}{1-\beta} p^\top z \\ \text{s.t.} \quad z \geq L - c\mathbf{1}, \\ \quad \quad z \geq 0 \end{cases} \stackrel{\text{強双対性}}{=} \begin{cases} \max_q q^\top L \\ \text{s.t.} \quad \mathbf{1}^\top q = 1, q \geq 0, \\ \quad \quad q \leq \frac{1}{1-\beta} p \end{cases}$$



CVaR 再訪

$$\text{CVaR}_\alpha[\mathcal{L}] = \min_c \left\{ c + \frac{1}{1-\alpha} \mathbb{E}[\max\{\mathcal{L} - c, 0\}] \right\}$$

演習

CVaR が 4 つの公理を満たすことを確認せよ

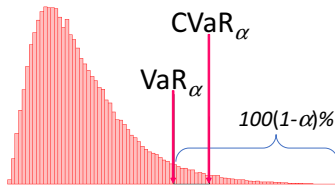
一方, VaR は劣加法性を満たさない:

$$\text{VaR}_\alpha[\mathcal{L}] := \min_{z \in \mathbb{R}} \{ z \mid \mathbb{P}\{\mathcal{L} \leq z\} \geq \alpha \}$$

演習

VaR が劣加法性を満たさない分布の組を 1 つ例示せよ

確率



損失 L

CVaR の双対表現 (離散分布の場合)

$$\begin{aligned}
 \text{CVaR}_\alpha[\mathcal{L}] &= \min_c \left\{ c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i \max\{L_i - c, 0\} \right\} \\
 &= \left| \begin{array}{ll} \min_{c, \mathbf{z}} & c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i z_i \\ \text{s.t.} & z_i \geq L_i - c, \quad i = 1, \dots, n \quad \leftarrow q_i \geq 0 \\ & z_i \geq 0, \quad i = 1, \dots, n \end{array} \right. \\
 &= \left| \begin{array}{ll} \max_{\mathbf{q}} & \sum_{i=1}^n q_i L_i \\ \text{s.t.} & \sum_{i=1}^n q_i = 1, \quad \leftarrow c \\ & 0 \leq q_i \leq \frac{1}{1-\alpha} p_i, \quad i = 1, \dots, n \quad \leftarrow z_i \geq 0 \end{array} \right. \quad \because \text{強双対性}
 \end{aligned}$$

したがって,

$$\mathcal{Q} = \left\{ \mathbf{q} \in \Pi \mid \mathbf{q} \leq \frac{1}{1-\alpha} \mathbf{p} \right\} \subset \Pi := \{ \mathbf{q} \mid \mathbf{1}^\top \mathbf{q} = 1, \mathbf{q} \geq \mathbf{0} \}$$

とすれば

$$\text{CVaR}_\alpha[\mathcal{L}] = \max_{\mathbf{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbf{Q}}[L]$$

Zhu, Fukushima, 2009—CVaR 最小化ポートフォリオの DRO

$$\begin{aligned} \min_{\mathbf{x} \in X} \max_{\mathbf{Q} \in \mathcal{Q}} \left\{ \text{CVaR}_{\mathbf{Q}, \alpha}[\mathcal{L}(\mathbf{x})] = \min_c \left\{ c + \frac{1}{1-\alpha} \sum_{i=1}^n q_i \max \{ L_i(\mathbf{x}) - c, 0 \} \right\} \right\} \\ = \min_{\mathbf{x} \in X} \min_c \max_{\mathbf{Q} \in \mathcal{Q}} \left\{ c + \frac{1}{1-\alpha} \sum_{i=1}^n q_i \max \{ L_i(\mathbf{x}) - c, 0 \} \right\} \quad \because \mathcal{Q} \text{ が有界\&閉} \end{aligned}$$

たとえば $\mathcal{Q} = \{ \mathbf{q} | \mathbf{q} = \hat{\mathbf{p}} + \mathbf{S}\mathbf{u} \geq \mathbf{0}, \mathbf{1}^\top \mathbf{S}\mathbf{u} = 0, \|\mathbf{u}\| \leq 1 \}$ のとき

$$\begin{aligned} \max_{\mathbf{Q} \in \mathcal{Q}} \left\{ c + \frac{1}{1-\alpha} \sum_{i=1}^n q_i \underbrace{\max \{ L_i(\mathbf{x}) - c, 0 \}}_{v_i} \right\} \\ = c + \frac{1}{1-\alpha} \left[\hat{\mathbf{p}}^\top \mathbf{v} + \max_{\mathbf{u}} \{ \mathbf{v}^\top \mathbf{S}\mathbf{u} | \mathbf{1}^\top \mathbf{S}\mathbf{u} = 0, \hat{\mathbf{p}} + \mathbf{S}\mathbf{u} \geq \mathbf{0}, \|\mathbf{u}\| \leq 1 \} \right] \\ = c + \frac{1}{1-\alpha} \left[\hat{\mathbf{p}}^\top \mathbf{v} + \min_{\zeta, \boldsymbol{\omega}, \boldsymbol{\xi}, \boldsymbol{\theta}} \{ \zeta + \hat{\mathbf{p}}^\top \boldsymbol{\omega} - \mathbf{S}^\top (\boldsymbol{\omega} - \mathbf{1}\boldsymbol{\theta} + \mathbf{v}) = \boldsymbol{\xi}, \|\boldsymbol{\xi}\| \leq \zeta, \boldsymbol{\omega} \geq \mathbf{0} \} \right] \end{aligned}$$

これより錐計画問題に帰着できる

演習

$L_i(\mathbf{x}) = -\mathbf{R}_i^\top \mathbf{x}, \|\cdot\| = \|\cdot\|_2, \mathbf{X} = \{ \mathbf{x} | \mathbf{A}\mathbf{x} \leq \mathbf{b} \}$ として, 2次錐計画として表せ.

コヒレントリスク尺度の DRO

$$\max_{\theta \in \Theta(\hat{p})} \max_{q \in \mathcal{Q}(\theta)} \mathbb{E}_q[\mathcal{L}]$$

は不確実性集合を拡大したコヒレント尺度．たとえば CVaR の場合で $\hat{p}$ に楕円型を考えると

$$\begin{aligned}\mathcal{Q} &= \{q \in \mathbb{R}^n \mid q \in \mathcal{Q}_{\text{CVaR}}(\theta), \theta \in \Theta(\hat{p})\} \\ &= \left\{ q \in \mathbb{R}^n \mid \begin{array}{l} \mathbf{1}^\top q = 1, \mathbf{0} \leq q \leq \frac{1}{1-c} \theta, \\ \mathbf{1}^\top \theta = 1, \theta \geq \mathbf{0}, \|S(\theta - \hat{p})\|_2 \leq \delta \end{array} \right\}\end{aligned}$$

→ より大きな不確実性集合に基づくコヒレントリスク尺度

コヒレントリスク尺度最小化に基づくポートフォリオ選択の RO

有界&閉かつ $0 \in \text{int}\mathcal{V}$ なる集合 $\mathcal{V} \subset \mathbb{R}^n$ に対し

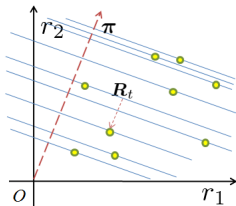
$\mathcal{Z} = \{(\Delta_1, \dots, \Delta_n) | \Delta_i \in \mathcal{V}_i, i = 1, \dots, n\}$ とし, 以下を定義する

$$\rho^e(x) := \max_{q \in \mathcal{Q}} \sum_{i=1}^n q_i [-R_i^\top x]$$

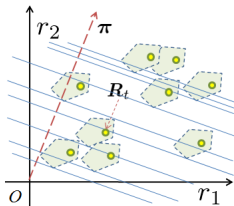
$$\rho^w(x) := \max_{(\Delta_i) \in \mathcal{Z}} \max_{q \in \mathcal{Q}} \sum_{i=1}^n q_i [-(R_i - \Delta_i)^\top x]$$

このとき次が成り立つ:

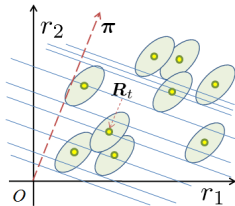
$$\rho^w(x) = \rho^e(x) + \max_s \{s^\top x : s \in \mathcal{V}\}$$



経験 CVaR



有限シナリオ型摂動
 $d = 2$ 資産の場合の含意



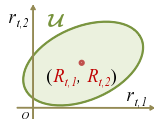
楕円型摂動

[補足] $\max_s \{s^\top x : s \in \mathcal{V}\}$ による正則化 (CVaR 最小化ポートフォリオ選択)

$$\min_{x \in X} \max_{(\Delta R_i) \in \mathcal{Z}} \min_{c \in \mathbb{R}} \left\{ c + \frac{1}{(1-\alpha)n} \sum_{i=1}^n \max\{-(R_i - \Delta R_i)^\top x - c, 0\} \right\}$$

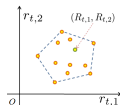
$\mathcal{V} = \mathcal{V}_N := \{v : \|v\| \leq \delta\}$ の場合 $\rightsquigarrow$ 凸最適化 ($\|\cdot\|^\circ$ に依存) :

$$\begin{cases} \min & c + \frac{1}{(1-\alpha)n} \mathbf{1}^\top \mathbf{y} + \delta \|\mathbf{x}\|^\circ \\ \text{s. t.} & y_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, \quad i = 1, \dots, n; \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{x} \in X. \end{cases}$$



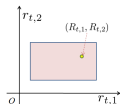
$\mathcal{V} = \mathcal{V}_I := \{v_1, \dots, v_k\}, (\mathbf{0} \in \text{co}\mathcal{V}_I)$ の場合 $\rightsquigarrow$ LP

$$\begin{cases} \min & c + \frac{1}{(1-\alpha)n} \mathbf{1}^\top \mathbf{y} + \zeta \\ \text{s. t.} & y_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, \quad i = 1, \dots, n; \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{x} \in X, \\ & \zeta \geq \mathbf{v}_h^\top \mathbf{x}, \quad h = 1, \dots, k. \end{cases}$$



$\mathcal{V} = \mathcal{V}_B := \{v : v_L \leq v \leq v_U\}, (v_{Lj} < 0 < v_{Uj})$ の場合 $\rightsquigarrow$ LP:

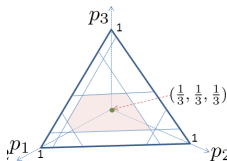
$$\begin{cases} \min & c + \frac{1}{(1-\alpha)n} \mathbf{1}^\top \mathbf{y} + (\mathbf{v}_U - \mathbf{v}_L)^\top \mathbf{w} + \mathbf{v}_L^\top \mathbf{x} \\ \text{s. t.} & y_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, \quad i = 1, \dots, n; \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{x} \in X, \\ & \mathbf{w} \geq \mathbf{x}, \quad \mathbf{w} \geq \mathbf{0}. \end{cases}$$



二重に最悪ケースを考えた CVaR ポートフォリオ選択：RO+DRO

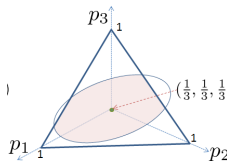
$$\min_{\mathbf{x} \in X} \max_{(\Delta \mathbf{R}_i) \in \mathcal{Z}} \max_{\mathbf{q} \in \mathcal{Q}} \min_{c \in \mathbb{R}} \left\{ c + \frac{1}{1-\alpha} \sum_{i=1}^n q_i \max\{-(\mathbf{R}_i - \Delta_i)^\top \mathbf{x} - c, 0\} \right\}$$

例 $\mathcal{Q} = \{\mathbf{q} \in \mathbb{R}^n : \mathbf{q} = \hat{\mathbf{p}} + \boldsymbol{\eta}, \mathbf{1}^\top \boldsymbol{\eta} = 0, \boldsymbol{\eta}_L \leq \boldsymbol{\eta} \leq \boldsymbol{\eta}_U\}, \boldsymbol{\eta}_L, \boldsymbol{\eta}_U \text{ 所与}$



$$\begin{aligned} \min \quad & \theta + \max_{\mathbf{s} \in \mathcal{V}} \{\mathbf{s}^\top \boldsymbol{\pi} : \mathbf{s} \in \mathcal{V}\} \\ \text{s.t.} \quad & c + \frac{1}{1-\alpha} (\hat{\mathbf{p}}^\top \mathbf{z} + \boldsymbol{\eta}_L^\top \boldsymbol{\xi} + \boldsymbol{\eta}_U^\top \boldsymbol{\omega}) \leq \theta \\ & \mathbf{1}^\top \mathbf{u} + \boldsymbol{\xi} + \boldsymbol{\omega} = \mathbf{z}, \boldsymbol{\xi} \geq \mathbf{0}, \boldsymbol{\omega} \leq \mathbf{0} \\ & \mathbf{x} \in X, z_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, z_i \geq 0, i = 1, \dots, n. \end{aligned}$$

例 $\mathcal{Q} = \{\mathbf{q} \in \mathbb{R}^n : \mathbf{q} = \hat{\mathbf{p}} + \mathbf{S}\boldsymbol{\eta} \geq \mathbf{0}, \mathbf{1}^\top \mathbf{S}\boldsymbol{\eta} = 0, \sqrt{\boldsymbol{\eta}^\top \boldsymbol{\eta}} \leq 1\}, \mathbf{S} : \text{正則}$



$$\begin{aligned} \min \quad & \theta + \max_{\mathbf{s} \in \mathcal{V}} \{\mathbf{s}^\top \boldsymbol{\pi} : \mathbf{s} \in \mathcal{V}\} \\ \text{s.t.} \quad & c + \frac{1}{1-\alpha} (\hat{\mathbf{p}}^\top \mathbf{z} + \zeta + \hat{\mathbf{p}}^\top \boldsymbol{\omega}) \leq \theta \\ & \boldsymbol{\xi} = \mathbf{S}^\top (\mathbf{1} - \boldsymbol{\omega} - \mathbf{z}), \sqrt{\boldsymbol{\xi}^\top \boldsymbol{\xi}} \leq \zeta, \boldsymbol{\omega} \geq \mathbf{0} \\ & \mathbf{x} \in X, z_i \geq -\mathbf{R}_i^\top \mathbf{x} - c, z_i \geq 0, i = 1, \dots, n. \end{aligned}$$

KL ダイバージェンスに基づく不確実性集合

Kullback-Leibler (KL) ダイバージェンスに基づく不確実性集合

$$\mathcal{Q}(\varepsilon) = \left\{ \mathbf{q} \in \mathbb{R}^n : \mathbf{1}^\top \mathbf{q} = 1, \mathbf{q} \geq \mathbf{0}, \sum_{i=1}^n q_i \ln \frac{q_i}{p_i} \leq \varepsilon \right\}$$

このとき

$$\max_{\mathbf{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbf{Q}}[f] = \left| \begin{array}{ll} \max_{\mathbf{q}} & \mathbf{f}^\top \mathbf{q} \\ \text{sub.to} & \sum_{i=1}^n q_i \ln \frac{q_i}{p_i} \leq \varepsilon, \quad \leftarrow \lambda \geq 0 \\ & \mathbf{1}^\top \mathbf{q} = 1, \quad \leftarrow \eta \\ & (q_i > 0, \quad i = 1, \dots, n) \end{array} \right.$$

双対性を使うと

$$\max_{\mathbf{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbf{Q}}[f] = \min_{\lambda > 0, \eta} \left\{ \varepsilon \lambda + \eta + \lambda \sum_{i=1}^n p_i \exp \left(\frac{f_i - \eta}{\lambda} - 1 \right) \right\}$$

KL ダイバージェンスに基づく DRO

$$\min_{\mathbf{x}} \max_{\mathbb{Q} \in \mathcal{Q}(\varepsilon)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x}, \mathbf{Y})] = \min_{x, \lambda > 0, \eta} \left\{ \varepsilon \lambda + \eta + \lambda \sum_{i=1}^n p_i \exp \left(\frac{f(\mathbf{x}, \mathbf{Y}_i) - \eta}{\lambda} - 1 \right) \right\}$$

新聞売り子問題の DRO

$$\begin{aligned} \min_{\lambda > 0, \eta, x, \mathbf{f}} \quad & \varepsilon \lambda + \eta + \lambda \sum_{i=1}^n p_i \exp \left(\frac{f_i - \eta}{\lambda} - 1 \right) \\ \text{sub.to} \quad & f_i \geq cx - rY_i, \quad f_i \geq (c - r)x, \quad i = 1, \dots, n, \\ & x \geq 0, \quad (\lambda > 0). \end{aligned}$$

ポートフォリオ選択問題の DRO

$$\begin{aligned} \min_{\lambda, \eta, \mathbf{x}} \quad & \varepsilon \lambda + \eta + \lambda \sum_{i=1}^n p_i \exp \left(\frac{\exp(-\mathbf{R}^\top \mathbf{x}) - \eta}{\lambda} - 1 \right) \\ \text{sub.to} \quad & \mathbf{1}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq \mathbf{0}, \quad (\lambda > 0). \end{aligned}$$

ロジスティック回帰の DRO

$$\begin{aligned} \min_{\lambda, \eta, \mathbf{x}, x_0} \quad & \varepsilon \lambda + \eta + \lambda \exp \left(-\frac{\eta}{\lambda} - 1 \right) \sum_{i=1}^n p_i \left(1 + \exp \left(-b_i (\mathbf{a}_i^\top \mathbf{x} + x_0) \right) \right) \\ \text{sub.to} \quad & (\lambda > 0). \end{aligned}$$

[トピック] 最適輸送距離に基づく DRO

輸送問題

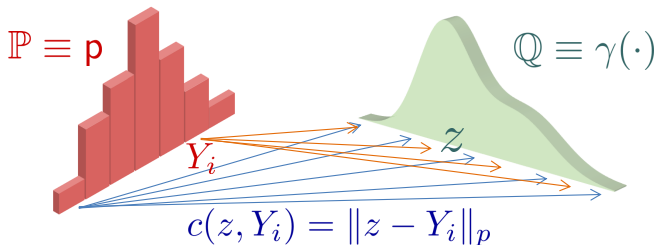
n 地点 $S_1, \dots, S_n$ から n 地点 $T_1, \dots, T_n$ へ品物を運ぶとき、輸送費が最も安くなるようにしたい。 S_i での供給量を p_i 単位、 T_j での消費量を q_j 単位とし、 $\sum_{i=1}^n p_i = \sum_{j=1}^n q_j$ が成り立っているとする。 S_i から T_j のルートで品物 1 単位を運ぶのにかかる輸送費が c_{ij} 円するとき、輸送費用の合計を最小にする輸送計画を求めよ。

$$d(\mathbb{Q}|\mathbb{P}) = \left| \begin{array}{ll} \min_{(\pi_{ij})} & \sum_{i=1}^n \sum_{j=1}^n c_{ij} \pi_{ij} \\ \text{s. t.} & \sum_{j=1}^n \pi_{ij} = q_i, \quad i = 1, \dots, n, \\ & \sum_{i=1}^n \pi_{ij} = p_j, \quad j = 1, \dots, n, \\ & \pi_{ij} \geq 0, \quad i, j = 1, \dots, n \end{array} \right|$$

[トピック] Wasserstein 距離に基づく DRO

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

- $\mathbb{Q}$ 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



【例】 ロジスティック回帰に対する Wasserstein DRO (Abadeh et al., 2015)

輸送費用を $c((\mathbf{x}_i, y_i), (\mathbf{x}, y)) = \|\mathbf{x}_i - \mathbf{x}\|_2 + \kappa|y_i - y|$ とする (ただし $\kappa \geq 0$) と, 凸計画問題に帰着できる:

$$\begin{aligned} \underset{\lambda \geq 0, \mathbf{s}, \mathbf{w}}{\text{minimize}} \quad & \varepsilon \lambda + \frac{1}{n} \sum_{i=1}^n s_i \\ \text{subject to} \quad & s_i \geq \ln \left(1 + \exp \left(-y_i (\mathbf{x}_i^\top \mathbf{w}) \right) \right), \quad i = 1, \dots, n, \\ & s_i \geq \ln \left(1 + \exp \left(y_i (\mathbf{x}_i^\top \mathbf{w}) \right) \right) - \kappa \lambda, \quad i = 1, \dots, n, \\ & \lambda \geq \|\mathbf{w}\|_2 \end{aligned}$$

特に $\kappa = \infty$ のとき

$$\underset{\lambda \geq 0, \mathbf{s}, \mathbf{w}}{\text{minimize}} \quad \varepsilon \|\mathbf{w}\|_2 + \frac{1}{n} \sum_{i=1}^n \ln \left(1 + \exp \left(-y_i (\mathbf{x}_i^\top \mathbf{w}) \right) \right)$$

→ Wasserstein DRO = 正則化付きロジスティック回帰推定

この項のまとめ

- 分布の不確実性を考えた RO が DRO
- コヒレント尺度の双対表現における分布の集合が不確実性集合
- ダイバージェンスにより異なる偏差尺度と関連付けられる
- 経験分布に対する ϕ ダイバージェンスを用いた DRO では分布の台は固定，Wasserstein DRO では分布の台も動きうる
- Wasserstein DRO では RO のときと同じような正則化付経験リスク最小化との等価性が示される例が知られている

① ロバスト最適化

- Soyster 1973, Ben-Tal, Nemirovski, 2000, Bertsimas, Sim, 2004
- 応用例
 - ロバストポートフォリオ選択
 - SVM

② 分布的ロバスト最適化 (DRO)

- リスク尺度と DRO
- DRO の含意

③ DRO の事後パフォーマンス

結局、ロバスト最適化とは何なのか？

- ロバスト最適化の枠組みの自由度は大きい
 - ▶ どの不確実性を考えるか：RO vs. DRO
 - ▶ どういう不確実性集合を考えるか
 - ▶ 不確実性集合の大きさはどう決めるべきなのか

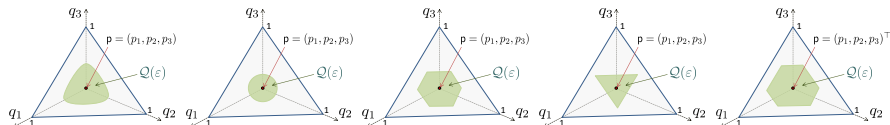
適当に決めると正則化付き経験リスク最小化が得られそうだが、それは本質なのだろうか？

- そもそもロバスト最適化の含意は何なのか？
(最悪ケースを考える意味は何なのか？)

案外このあたりの議論は未成熟

不確実性集合 Q はどのように与えられるべきか？

- どのような形？ (どのダイバージェンスを用いるべきか？)



- どれが良いかの議論は多くない
- Sutter, Van Parys, Kuhn, 2020, arXiv ... 統計的な議論

ペナルティ型 DRO

※この項は f を報酬とした期待報酬最大化で記述しています

$$\max_x \hat{\psi}_n(x) := \frac{1}{n} \sum_{i=1}^n f(x, Y_i) = \mathbb{E}_{\mathbb{P}_n}[f(x, Y)] \quad : \text{経験的最適化}$$

$$\max_x \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}}[f(x, Y)] + \frac{1}{\delta} d(\mathbb{Q} | \mathbb{P}_n) \right\}, \quad : \text{ロバスト最適化}$$

ただし $\delta > 0$ は定数, $\mathbb{Q} \equiv (q_i)$ は $\mathbb{P}_n \equiv (\hat{p}_i)$ と同じ台を持つ離散確率分布,

$$d(\mathbb{Q} | \mathbb{P}_n) = \begin{cases} \sum_{i: \hat{p}_i > 0} q_i \ln \left(\frac{q_i}{\hat{p}_i} \right), & \text{if } \sum_{i: \hat{p}_i > 0} q_i = 1, q_i \geq 0, \\ +\infty, & \text{その他,} \end{cases}$$

は $\mathbb{Q}$ の $\mathbb{P}_n$ に対する Kullback-Leibler ダイバージェンス

- $d(\mathbb{Q} | \mathbb{P}_n)$ は $\mathbb{Q}$ について凸
- $\forall \mathbb{Q}, d(\mathbb{Q} | \mathbb{P}_n) \geq 0; \quad d(\mathbb{Q} | \mathbb{P}_n) = 0 \Leftrightarrow \mathbb{Q} = \mathbb{P}_n$
- $\delta > 0$ は曖昧さパラメータ $\rightsquigarrow$ 意思決定者の $\mathbb{P}_n$ に対する確信度

事実1：KL ダイバージェンスを用いたロバスト最適化 = 平均-偏差最適化

$$g_{\delta}(x) := \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(x, Y)] + \frac{1}{\delta} d(\mathbb{Q} | \mathbb{P}_n) \right\} = \hat{\psi}_n(x) - \mathcal{D}_{\delta}(f(x, Y) | \hat{\psi}_n(x)),$$

$$\text{ただし } \mathcal{D}_{\delta}(f(x, Y) | \hat{\psi}_n(x)) = \frac{1}{\delta} \ln \mathbb{E}_{\mathbb{P}_n} \left[\exp \left(-\delta (f(x, Y) - \hat{\psi}_n(x)) \right) \right].$$

∴

$$\begin{aligned} g_{\delta}(x) &= -\frac{1}{\delta} \ln \mathbb{E}_{\mathbb{P}_n} \left[\exp \left(-\delta f(x, Y) \right) \right] && (\because \text{双対関係}) \\ &= \underbrace{\mathbb{E}_{\mathbb{P}_n} [f(x, Y)]}_{\hat{\psi}_n(x)} - \frac{1}{\delta} \ln \mathbb{E}_{\mathbb{P}_n} \left[\exp \left(-\delta (f(x, Y) - \mathbb{E}_{\mathbb{P}_n} [f(x, Y)]) \right) \right]. \quad \square \end{aligned}$$

注意： $\mathcal{D}_{\delta}$ は次の意味で期待値周りの偏差尺度：

- ① 任意の定数 $c \in \mathbb{R}$ に対して $\mathcal{D}_{\delta}(c | \mathbb{E}[c]) = 0$
- ② 一定でない任意の確率変数 Z に対して $\mathcal{D}_{\delta}(Z | \mathbb{E}[Z]) > 0$

事実2：KL ダイバージェンスから得られる偏差 $\approx$ 分散

$$\begin{aligned}\mathcal{D}_\delta(f(x, Y) \mid \hat{\psi}_n(x)) &= \frac{1}{\delta} \ln \mathbb{E}_{\mathbb{P}_n} \left[\exp \left(-\delta (f(x, Y) - \hat{\psi}_n(x)) \right) \right] \\&= \frac{1}{\delta} \ln \mathbb{E}_{\mathbb{P}_n} \left[1 - \delta (f(x, Y) - \hat{\psi}_n(x)) + \frac{\delta^2}{2} (f(x, Y) - \hat{\psi}_n(x))^2 + \cdots \right] \\&= \frac{1}{\delta} \ln \left\{ 1 + \frac{\delta^2}{2} \mathbb{E}_{\mathbb{P}_n} \left[(f(x, Y) - \hat{\psi}_n(x))^2 \right] + \cdots \right\} \\&= \frac{\delta}{2} \left\{ \frac{1}{n} \sum_{i=1}^n (f(x, Y_i) - \hat{\psi}_n(x))^2 \right\} + o(\delta) \\&= \frac{\delta}{2} \mathbb{V}_{\mathbb{P}_n}(f(x, Y)) + o(\delta)\end{aligned}$$

KL ダイバージェンスに基づく DRO $\approx$ 平均-分散最適化 (MVO)

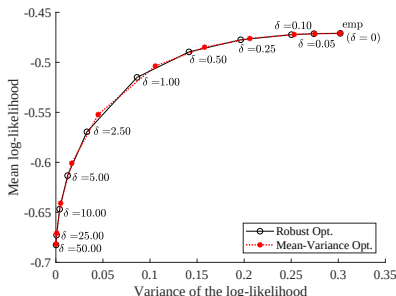
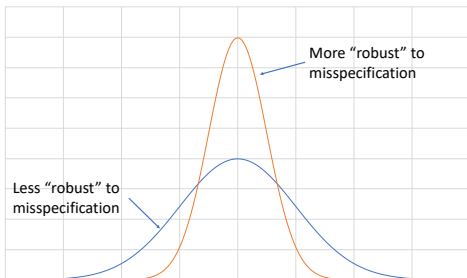
$$\min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}}[f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\}$$

DRO の目的関数

$$= \underbrace{\mathbb{E}_{\mathbb{P}_n}[f(x, Y)]}_{\text{経験平均}} - \frac{\delta}{2\phi''(1)} \underbrace{\mathbb{V}_{\mathbb{P}_n}[f(x, Y)]}_{\text{経験分散}} + o(\delta)$$

cf. Lam 2016, Namkoong, Duchi, 2017

$$\min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}}[f(x, Y)] \mid d(\mathbb{Q} | \mathbb{P}_n) \leq \eta^2 \right\} = \mathbb{E}_{\hat{\mathbb{P}}_n}[f(x, Y)] - \sqrt{2\mathbb{V}_{\mathbb{P}_n}[f(x, Y)]}\eta + o(\eta),$$



DRO と MVO の in-sample 平均分散フロンティア (G., Kim, Lim, 2020)

Csiszár の ϕ -ダイバージェンスを用いた DRO

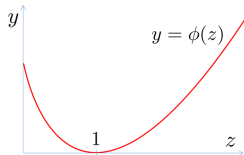
$\forall z, \phi(z) \geq \phi(1) = 0$ を満たす閉真凸関数 $\phi: \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ に対し $\mathbb{Q}$ の $\mathbb{P}_n$ に対する ϕ -ダイバージェンスを以下で定義する:

$$\mathcal{H}_\phi(\mathbb{Q} | \mathbb{P}_n) := \begin{cases} \sum_{i: \hat{p}_i > 0} \hat{p}_i \phi\left(\frac{q_i}{\hat{p}_i}\right), & \sum_{i: \hat{p}_i > 0} q_i = 1, q_i \geq 0, \\ +\infty, & \text{その他.} \end{cases}$$

例

- KL ダイバージェンス if $\phi(z) = z \ln z - z + 1$
- χ^2 -ダイバージェンス if $\phi(z) = \frac{1}{z}(z-1)^2$
- 修正 χ^2 -ダイバージェンス if $\phi(z) = (z-1)^2$
- Hellinger 距離 if $\phi(z) = (\sqrt{z} - 1)^2$

他の例については Ben-Tal et al., 2013 など



ϕ -ダイバージェンスを用いた罰則型 DRO の目的関数:

$$g_{\delta, \phi}(x) := \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(x, Y)] + \frac{1}{\delta} \mathcal{H}_\phi(\mathbb{Q} | \mathbb{P}_n) \right\},$$

ϕ -ダイバージェンスを用いた DRO = 平均-偏差 (= リスク最小化)

仮定 (★) 各 x に対し, 最悪ケースの確率測度

$$\mathbb{Q}^* \in \arg \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\}$$

は $\hat{\mathbb{P}}_n$ と同値 (つまり $q_i^* > 0$ iff $\hat{p}_i > 0$)

(★) が満たされるとする. このとき, 次の 2 つの関数は等しい:

1. $g_{\delta, \phi}(x) = \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\}$ (ロバストな目的関数)

2. $\hat{\psi}_n(x) - \mathcal{D}_{\delta, \phi, \mathbb{P}_n}(f(x, Y)) | \hat{\psi}_n(x)$. (平均-偏差)

注意: 1. と 2. の最大化はそれぞれあるリスク尺度 $\mathcal{R}_{\delta, \phi, \hat{\mathbb{P}}}(x)$ の最小化と等価.

Regular な偏差尺度

$$\mathcal{D}_{\delta, \phi, \mathbb{P}}(Z | \mathbb{E}_{\mathbb{P}}[Z]) := \inf_c \left\{ c + \frac{1}{\delta} \mathbb{E}_{\mathbb{P}}[\phi^*(\delta(\mathbb{E}_{\mathbb{P}}[Z] - Z - c))] \right\}$$

ただし ϕ^* は ϕ の共役, i.e., $\phi^*(\zeta) = \sup_z \{z\zeta - \phi(z)\}$.

regular な偏差尺度

任意の確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ に対し, 2 乗可積分な確率変数の空間を $\mathcal{L}^2(\Omega)$ と記す. i.e., $\mathbb{E}[X^2] < \infty$. ある汎関数 $\mathcal{D} : \mathcal{L}^2(\Omega) \rightarrow [0, \infty]$ が **regular な偏差尺度** であるとは, それが閉凸で以下を満たすときをいう

- 1 任意の定数 $c \in \mathbb{R}$ に対して $\mathcal{D}(c) = 0$
- 2 任意の一定でない確率変数 $Z \in \mathcal{L}^2(\Omega)$ に対して $\mathcal{D}(Z) > 0$

ϕ を $\phi(1) = 0$ and $\phi'(1) = 0$ なる閉真凸関数とする. このとき任意の確率変数 $Z \in \mathcal{L}^2(\Omega)$ に対して $\mathcal{D}(Z) = \mathcal{D}_{\delta, \phi, \mathbb{P}}(Z | \mathbb{E}_{\mathbb{P}}[Z])$ は regular な偏差尺度である.

ϕ -ダイバージェンスを用いた $\text{DRO} \approx \text{MVO}$

ϕ が凸で、2 回連続微分可能、かつ $\phi(1) = 0$, $\phi'(1) = 0$, $\phi''(1) > 0$ を満たすとする。このとき偏差尺度 $\mathcal{D}_{\delta, \phi, \mathbb{P}_n}(f(x, Y) | \hat{\psi}_n(x))$ は次を満たす：

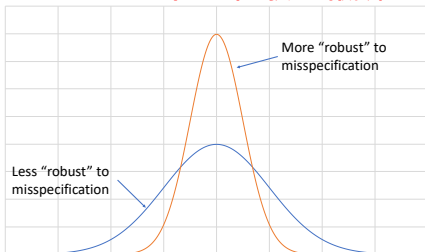
$$\mathcal{D}_{\delta, \phi, \mathbb{P}_n}(f(x, Y) | \hat{\psi}_n(x)) = \frac{\delta}{2\phi''(1)} \hat{\sigma}_n^2(x) + o(\delta),$$

ただし、 $\hat{\sigma}_n^2(x) := \mathbb{V}_{\mathbb{P}_n}(f(x, Y))$ 。

まとめると、

$$\min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}}[f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\} = \hat{\psi}_n(x) - \frac{\delta}{2\phi''(1)} \hat{\sigma}_n^2(x) + o(\delta)$$

〜 ロバスト性は経験的報酬のばらつきを抑えることで達成される



注意： 報酬関数 $f(x, Y)$ が x について凹であっても平均-分散最適化は凸最適化にはならないかもしれない。一方、ロバスト最適は $f(x, Y)$ が x について凹であれば凸最適化である。これは DRO のメリット。

⟨Extension⟩ Robust risk minimization

Consider to robustify the min. of a risk measure of a *loss* $\ell \equiv -f$

$$\mathcal{R}_{v,\mathbb{P}}[\ell] := \inf_{\alpha} \{\alpha + \mathbb{E}_{\mathbb{P}}[v(\ell - \alpha)]\},$$

where v is a convex func. on $\mathbb{R}$ (Ben-Tal, Teboulle,'07; Rockafellar, Uryasev,'13)

- Examples:
 - (i) CVaR ($v(z) = [z]^+ / (1 - \beta)$ with $\beta \in [0, 1)$)
 - (ii) mean-variance ($v(z) = x + \lambda z^2$ with $\lambda > 0$)
 - (iii) log-sum-exp ($v(z) = \exp(z) - 1$), etc.

Under some technical assumptions, we have

$$\sup_{\mathbb{Q}} \left\{ \mathcal{R}_{v,\mathbb{Q}}[\ell] - \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\} = \mathcal{R}_{v,\mathbb{P}_n}[\ell] + \frac{\delta}{2\phi''(1)} \mathbb{V}_{\mathbb{P}}[v(\ell - a_0)] + o(\delta)$$

robust risk measure

where $a_0 \in \arg \min_{\alpha} \{\alpha + \mathbb{E}_{\mathbb{P}}[v(\ell - \alpha)]\}$

[Example] Robust CVaR minimization

$$\begin{aligned}\text{CVaR}_{\beta, \mathbb{P}_n}[\ell] &:= \min_{\alpha} \left\{ \alpha + \frac{1}{1-\beta} \mathbb{E}_{\mathbb{P}_n} \left[[\ell - \alpha]^+ \right] \right\} \\ &= a_0 + \frac{1}{1-\beta} \mathbb{E}_{\mathbb{P}_n} \left[[\ell - a_0]^+ \right]\end{aligned}$$

where $[x]^+ = \max\{x, 0\}$ and a_0 is a minimizer, which approximately equals $\text{VaR}_{\beta, \mathbb{P}_n}[\ell] := \min \{x : \mathbb{P}_n\{\ell \leq x\} \geq \beta\}$.

We have

$$\max_{\mathbb{Q}} \left\{ \text{CVaR}_{\beta, \mathbb{Q}}[\ell] - \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\}$$

robust CVaR

$$= \text{CVaR}_{\beta, \mathbb{P}_n}[\ell] + \frac{\delta}{2(1-\beta)\phi''(1)} \mathbb{V}_{\mathbb{P}_n} \left[[\ell - a_0]^+ \right] + o(\delta)$$

不確実性集合 $\mathcal{Q}(\varepsilon) = \{\mathbb{Q} | d(\mathbb{Q} | \mathbb{P}) \leq \varepsilon\}$ の例

ϕ -ダイバージェンス (Csiszár, 1963)

$$d(\mathbb{Q} | \mathbb{P}) = \mathcal{H}_\phi(\mathbb{Q} | \mathbb{P}) := \begin{cases} \sum_i p_i \phi\left(\frac{q_i}{p_i}\right), & \mathbf{1}^\top \mathbf{q} = 1, \mathbf{q} \geq \mathbf{0}, \\ +\infty, & \text{otherwise.} \end{cases}$$

ただし $\phi : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$: 閉凸 & $\forall z \in \mathbb{R}, \phi(z) \geq \phi(1) = 0$

[例] 滑らかなケース $\phi''(1) > 0$

[例] 滑らかでないケース (i) $\phi(t) = |t - 1|$

[例] 滑らかでないケース (ii) $\phi(t) = \delta_I(t) = \begin{cases} 0, & t \in I \\ \infty, & t \notin I \end{cases}$

- $I = [0, 1 + \varepsilon] \xrightarrow{\varepsilon = \frac{\alpha}{1-\alpha}} [0, \frac{1}{1-\alpha}] \rightsquigarrow \mathcal{Q}_{\text{CVaR}}(\alpha)$

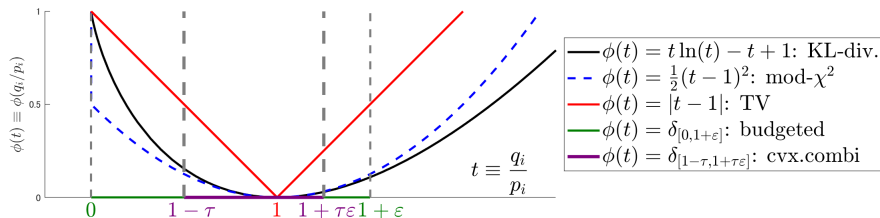
- $I = [1 - \tau, 1 + \tau\varepsilon] \rightsquigarrow \mathcal{Q} = \mathcal{Q}_c(\tau) := (1 - \tau)\{\mathbf{p}\} + \tau\mathcal{Q}_{\text{CVaR}}(\alpha)$

不確実性集合 $\mathcal{Q}(\varepsilon) = \{\mathbb{Q} | d(\mathbb{Q} | \mathbb{P}) \leq \varepsilon\}$ の例

ϕ -ダイバージェンス (Csiszár, 1963)

$$d(\mathbb{Q} | \mathbb{P}) = \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}) := \begin{cases} \sum_i p_i \phi\left(\frac{q_i}{p_i}\right), & \mathbf{1}^{\top} \mathbf{q} = 1, \mathbf{q} \geq \mathbf{0}, \\ +\infty, & \text{otherwise.} \end{cases}$$

ただし $\phi: \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$: 閉凸 & $\forall z \in \mathbb{R}, \phi(z) \geq \phi(1) = 0$



DRO の目的関数 $\approx$ SAA + 感度

$$V(\varepsilon, x) := \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(x, Y)]$$

以下で x に対する依存性が重要でないとき $V(\varepsilon)$ と記します
一般に、凹で単調増加 & $g(0) = 0$ であるような g に対し、
 $V(\varepsilon) - V(0) \sim O(g(\varepsilon))$ のとき、WCS

$$\mathcal{S}_{\mathbb{P}}[f] = \lim_{\varepsilon \downarrow 0} \frac{V(\varepsilon) - \sum_{i=1}^n p_i f(Y_i)}{g(\varepsilon)}$$

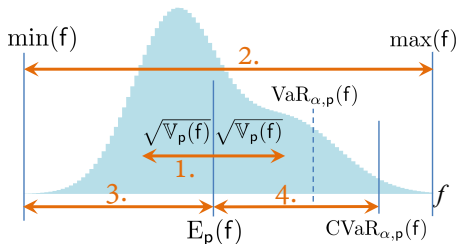
を定義しておく。

このとき $\varepsilon > 0$ に対して

$$V(\varepsilon, x) = \sum_{i=1}^n p_i f(x, Y_i) + g(\varepsilon) \mathcal{S}_{\mathbb{P}}[f(x, \cdot)] + o(g(\varepsilon))$$

$$V(\varepsilon, x) = \sum_{i=1}^n p_i f(x, Y_i) + g(\varepsilon) \mathcal{S}_{\mathbb{P}}[f(x, \cdot)] + o(g(\varepsilon)) \text{ の } \mathcal{S}_{\mathbb{P}}[f(x, \cdot)]$$

ダイバージェンス／目的関数	$\mathcal{S}_{\mathbb{P}}[f]$
滑らかな ϕ	$\sqrt{\frac{2\mathbb{V}_{\mathbf{p}}(\mathbf{f})}{\phi''(1)}}$
全変動 ($\phi(z) = z - 1 $)	$\frac{1}{2}(\max(\mathbf{f}) - \min(\mathbf{f}))$
予算型 (CVaR 型) ($\phi = \delta_{[0, 1+\varepsilon]}$)	$\mathbb{E}_{\mathbf{p}}(\mathbf{f}) - \min(\mathbf{f})$
$(1 - \varepsilon)$ “期待コスト” + ε “最大コスト”	$\max(\mathbf{f}) - \mathbb{E}_{\mathbf{p}}(\mathbf{f})$
$\phi = \delta_{[\frac{1}{1+\varepsilon}, 1+\varepsilon]}$	$\text{CVaR}_{\mathbf{p}, \frac{1}{2}}(\mathbf{f}) - \mathbb{E}_{\mathbf{p}}(\mathbf{f})$
$(1 - \varepsilon)$ “期待コスト” + ε “ α -CVaR”	$\text{CVaR}_{\mathbf{p}, \alpha}(\mathbf{f}) - \mathbb{E}_{\mathbf{p}}(\mathbf{f})$
Wasserstein 距離 w/ $f(z_i) - f(Y_i) \leq L\ z_i - Y_i\ _p$	$\max_{i=1, \dots, n} \max_{z_i} \frac{f(z_i) - f(Y_i)}{\ z_i - Y_i\ _p}$



1. は標準偏差,
2. はレンジ,
3. は下半レンジ (期待値と最小値の差),
4. は CVaR 偏差

偏差尺度

$$\bullet S_p(f) = V'_\phi(0^+) = \sqrt{\frac{2\mathbb{V}_p(f)}{\phi''(1)}}$$

滑らかな ϕ

$$\bullet S_p(f) = V'_{TV}(0^+) = \frac{\max(f) - \min(f)}{2}$$

全変動型 ϕ

→ 共通点は何か？

偏差尺度 (Rockafellar, Uryasev, Zabrankin, 2006)

確率変数 f に対して以下を満たす関数 $\mathcal{H}[f]$ を f の偏差尺度 (Generalized Measure of Deviation) と呼ぶ：

- ① $\mathcal{H}[f] \geq 0$ で、等号成立と f が定数であることが等価
- ② 任意の非負定数 $\beta \geq 0$ に対して $\mathcal{H}[\beta f] = \beta \mathcal{H}[f]$
- ③ 任意の定数 $\alpha \in \mathbb{R}$ に対して $\mathcal{H}[\alpha + f] = \mathcal{H}[f]$

新聞売り子問題に対する Wasserstein DRO

単価 c で仕入れ, 売れば r だけ, 売れなければ q だけ戻ってくる

$$f(x, Y) = cx - \underbrace{r \min\{x, Y\}}_{\text{売上}} - \underbrace{q \max(x - Y, 0)}_{\text{残余価値}} + \underbrace{s \max(Y - x, 0)}_{\text{機会費用}}$$

ただし $0 \leq q < c < r, \quad s \geq 0$

$x \in (\min_i Y_i, \max_i Y_i)$ とする. $p = 1$ の Wasserstein 距離を用いた場合の WCS は

$$\mathcal{S}_{\mathbb{P}}[f(x, \cdot)] = \max\{r - q, s\}$$

→ Wasserstein DRO 解 = SAA 解

ここまでのまとめ

- $\text{DRO} \approx$ 平均と感度の多目的最適化
- 特に不確実性集合の大きさ ε が小さいとき,
DRO の目的関数 $\approx$ 平均 + WCS
 $\leadsto$ WCS が正則化項 !?
- 代表的な不確実性集合に対する明示的な WCS

① ロバスト最適化

- Soyster 1973, Ben-Tal, Nemirovski, 2000, Bertsimas, Sim, 2004
- 応用例
 - ロバストポートフォリオ選択
 - SVM

② 分布的ロバスト最適化 (DRO)

- リスク尺度と DRO
- DRO の含意

③ DRO の事後パフォーマンス

パラメータ ε のチューニング

$$(\text{DRO}(\varepsilon)) \quad \min_{\mathbf{x}, \lambda > 0, \eta} \left\{ \varepsilon \lambda + \eta + \lambda \sum_{i=1}^n p_i \exp \left(\frac{f(\mathbf{x}, \mathbf{Y}_i) - \eta}{\lambda} - 1 \right) \right\}$$

- $\varepsilon > 0$ を固定して $(\text{DRO}(\varepsilon))$ をソルバーに投げればその最適解 $\mathbf{x}_n(\varepsilon)$ を得る
- ではどの ε (の最適解) を用いるべきか？
- 実現するコスト $\mathbb{E}_{\mathbb{P}}[f(\mathbf{x}_n(\varepsilon), \mathbf{Y})]$ が小さくなるように？
- そもそも真の分布 $\mathbb{P}$ がわからない (手元にあるのは $\mathbb{P}_n$ のみ)

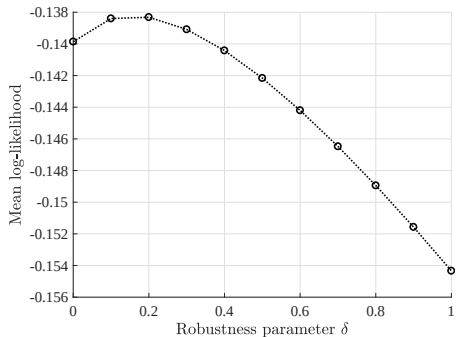
→ 再標本法で $\mathbb{P}$ (or $\mathbb{P}_n$ に対する感度) をシミュレートする

ブートストラップ法

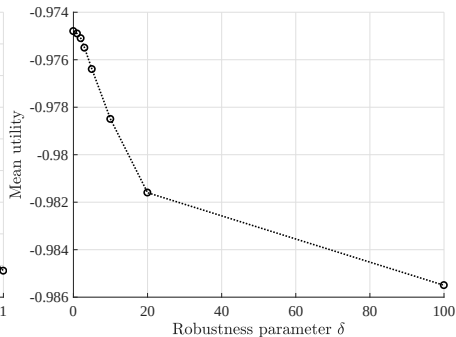
確率	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\leadsto \left\{ \begin{array}{l} \begin{array}{ c c c c } \hline \mathbf{Y}_4 & \mathbf{Y}_2 & \mathbf{Y}_1 & \mathbf{Y}_2 \\ \hline \mathbf{Y}_1 & \mathbf{Y}_4 & \mathbf{Y}_4 & \mathbf{Y}_2 \\ \hline \mathbf{Y}_2 & \mathbf{Y}_3 & \mathbf{Y}_1 & \mathbf{Y}_1 \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \mathbf{Y}_2 & \mathbf{Y}_3 & \mathbf{Y}_2 & \mathbf{Y}_2 \\ \hline \end{array} \begin{array}{l} \leadsto \mathbf{x}_4^{(1)}(\varepsilon) \\ \leadsto \mathbf{x}_4^{(2)}(\varepsilon) \\ \leadsto \mathbf{x}_4^{(3)}(\varepsilon) \\ \\ \leadsto \mathbf{x}_4^{(B)}(\varepsilon) \end{array} \end{array} \right.$
取りうる値	$\mathbf{Y}_1$	$\mathbf{Y}_2$	$\mathbf{Y}_3$	$\mathbf{Y}_4$	

これを ε の K 個の候補 $\varepsilon_1, \dots, \varepsilon_K$ について繰り返す $\leadsto K \times B$ 回の最適化計算

ブートストラップの平均に基づく δ の選択



(i) ロジスティック回帰の DRO の例



(ii) ポートフォリオ選択の例

注意：縦軸が $-\mathbb{E}[f]$, つまり「大きいほど良い」になっています

DRO の事後パフォーマンスの分析 (Gotoh, Kim, Lim, 2021+)

$$\mathbf{x}_n(\delta) := \arg \min_{\mathbf{x}} \underbrace{\max_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(\mathbf{x}, \mathbf{Y})] - \frac{1}{\delta} \text{KL}(\mathbb{Q}|\mathbb{P}_n) \right\}}_{\approx \mathbb{E}_{\mathbb{P}_n} [f(\mathbf{x}, \mathbf{Y})] + \delta \mathbb{V}_{\mathbb{P}_n} [f(\mathbf{x}, \mathbf{Y})] + o(\delta)}$$

$$\begin{aligned} \text{cf. } \min_{\mathbf{x}} \max_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(\mathbf{x}, \mathbf{Y})] \mid \text{KL}(\mathbb{Q}|\mathbb{P}_n) \leq \varepsilon \right\} \\ \approx \min_{\mathbf{x}} \mathbb{E}_{\mathbb{P}_n} [f(\mathbf{x}, \mathbf{Y})] + \sqrt{2\varepsilon} \sqrt{\mathbb{V}_{\mathbb{P}_n} [f(\mathbf{x}, \mathbf{Y})]} + o(\sqrt{\varepsilon}) \end{aligned}$$

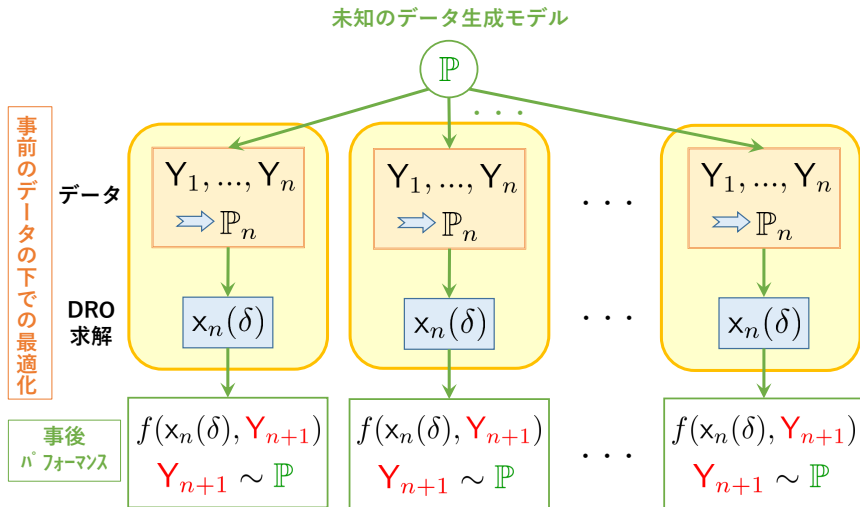
事前の問題 (*In-sample problem*):

- $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim \mathbb{P}$: 独立に同一な分布 $\mathbb{P}$ から得られた n 標本
 $\rightsquigarrow \begin{cases} \mathbf{x}_n(0) \equiv \text{SAA} \\ \mathbf{x}_n(\delta) \equiv \text{DRO} \quad \text{for } \delta > 0 \end{cases}$

事後パフォーマンス (*Out-of-sample performance*):

- $\mathbf{Y}_{n+1} \sim \mathbb{P}$: $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ とは独立に $\mathbb{P}$ の新しい標本
 $\rightsquigarrow \begin{cases} f(\mathbf{x}_n(0), \mathbf{Y}_{n+1}) : \text{SAA の事後コスト} \\ f(\mathbf{x}_n(\delta), \mathbf{Y}_{n+1}) : \text{DRO の事後コスト} \quad \text{for } \delta > 0 \end{cases}$

事後パフォーマンス分析の視点



DRO solution as an M-estimator

$\phi^*(\zeta) := \sup_z \{z\zeta - \phi(z)\}$ (凸共役) とすると,

$$\min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}} [f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}_n) \right\} = - \min_c \left\{ c + \frac{1}{\delta} \mathbb{E}_{\mathbb{P}_n} \left[\phi^*(\delta(-f(x, Y) - c)) \right] \right\}$$

であるから, DRO は以下のような最適化問題:

$$(x_n(\delta), c_n(\delta)) = \arg \min_{x, c} \left\{ c + \frac{1}{\delta} \mathbb{E}_{\mathbb{P}_n} \left[\phi^*(\delta(-f(x, Y) - c)) \right] \right\}$$

$f(x, Y)$ が x について滑らかな狭義凹ならば 1 次の条件が最適解を特徴づける:

$$\begin{bmatrix} \mathbb{E}_{\mathbb{P}_n} \left[(\phi^*)' \left(-\delta(f(x_n(\delta), Y) + c_n(\delta)) \right) \nabla_x f(x_n(\delta), Y) \right] \\ \mathbb{E}_{\mathbb{P}_n} \left[-\frac{\phi^{(2)}(1)}{\delta} \left((\phi^*)' \left(-\delta(f(x_n(\delta), Y) + c_n(\delta)) \right) - 1 \right) \right] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e., $\exists \psi : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^{d+1}$ s.t. $\mathbb{E}_{\mathbb{P}}[\psi(x_n(\delta), c_n(\delta))] = 0$

$\leadsto x_n(\delta)$ は M 推定量

Asymptotic normality of DRO solution

M 推定量の性質より，得られる DRO 解は真の分布の下での DRO 解を平均とする正規分布に漸近的に従う：

$$x_n(\delta) \sim N\left(x^*(\delta), \frac{\xi(\delta)}{n}\right)$$

where

- $x^*(\delta) := \arg \max_x \min_{\mathbb{Q}} \left\{ \mathbb{E}_{\mathbb{Q}}[f(x, Y)] + \frac{1}{\delta} \mathcal{H}_{\phi}(\mathbb{Q} | \mathbb{P}) \right\}$
- $\xi(\delta) := \xi(0) + O(\delta)$
- $\xi(0) := \mathbb{V}_{\mathbb{P}} \left[\left(\mathbb{E}_{\mathbb{P}} \left[\nabla_x^2 f(x^*(0), Y) \right] \right)^{-1} \nabla_x f(x^*(0), Y) \right]$

DRO adds bias to the solution

$$\begin{array}{ccc} x_n(0) & \overset{\text{asymptotic}}{\sim} & N(x^*(0), \xi(0)/n) \\ & & \updownarrow \text{bias} \\ x_n(\delta) & \overset{\text{asymptotic}}{\sim} & N(x^*(\delta), \xi(\delta)/n) \end{array}$$

$$x^*(\delta) = x^*(0) + \delta \pi + o(\delta) \quad (\text{解の平均})$$

$$\xi(\delta) = \xi(0) + O(\delta) \quad (\text{解の共分散})$$

where

$$\pi = \frac{1}{\phi''(1)} \left(\mathbb{E}_{\mathbb{P}} [\nabla_x^2 f(x^*(0), Y)] \right)^{-1} \text{Cov}_{\mathbb{P}} \left[f(x^*(0), Y), \nabla_x f(x^*(0), Y) \right]$$

→ DRO は期待報酬の減少と分散の減少のトレードオフを最適化するために、SAA 解にバイアス π を加えることで解のばらつきを変更する
(ここでのばらつきはサンプリングに伴うばらつき)

Out-of-sample reward of DRO

	empirical	true
mean	$\mu_n(\delta) := \mathbb{E}_{\mathbb{P}}[f(x_n(\delta), Y_{n+1})]$	$\mu^*(\delta) := \mathbb{E}_{\mathbb{P}}[f(x^*(\delta), Y_{n+1})]$
variance	$v_n(\delta) := \mathbb{V}_{\mathbb{P}}[f(x_n(\delta), Y_{n+1})]$	$v^*(\delta) := \mathbb{V}_{\mathbb{P}}[f(x^*(\delta), Y_{n+1})]$

報酬 $f(x, Y)$ が x について狭義凹のとき,

$$\begin{aligned} \mu_n(\delta) &= \mu^*(\delta) + \underbrace{\frac{1}{2n} \text{tr}\left(\xi(0) \textcolor{red}{H}(0)\right)}_{<0} + \frac{\rho\delta}{2n} + \overbrace{\frac{1}{n}O(\delta^2) + o\left(\frac{1}{n}\right) + o(\delta^2)}^{\text{higher order terms}}, \\ v_n(\delta) &= v^*(\delta) + \frac{1}{2n} \text{tr}\left(\xi(0) G(0)\right) + \frac{\theta\delta}{2n} + \frac{1}{n}O(\delta^2) + o\left(\frac{1}{n}\right) + O(\delta^2), \end{aligned}$$

where

$$\begin{aligned} \textcolor{red}{H}(0) &:= \mathbb{E}_{\mathbb{P}}[\nabla_x^2 \textcolor{red}{f}(x^*(0), Y_{n+1})], \quad G(0) := \nabla_x^2 \mathbb{V}_{\mathbb{P}}[f(x^*(0), Y_{n+1})], \\ \rho &:= \text{tr}\left(\xi'(0) \textcolor{red}{H}(0)\right) + \pi' \nabla_x \left[\text{tr}\left(\xi(0) \textcolor{red}{H}(0)\right) \right], \\ \theta &:= \text{tr}\left(\xi'(0) G(0)\right) + \pi' \nabla_x \left[\text{tr}\left(\xi(0) G(0)\right) \right], \\ \pi &:= \frac{1}{\phi''(1)} H(0)^{-1} \beta, \quad \beta := \text{Cov}_{\mathbb{P}}[\nabla_x f(x^*(0), Y_{n+1}), f(x^*(0), Y_{n+1})] \end{aligned}$$

Robust mean-variance frontier

$$\left\{ \left(\underbrace{\mathbb{E}_{\mathbb{P}}[f(x_n(\delta), Y_{n+1})]}_{\mu_n(\delta)}, \underbrace{\mathbb{V}_{\mathbb{P}}[f(x_n(\delta), Y_{n+1})]}_{v_n(\delta)} \right) : \delta \geq 0 \right\}$$

$$\underbrace{\mu_n(\delta)}_{\text{DRO}} = \underbrace{\mu_n(0)}_{\text{SAA}} + \frac{\rho}{2n}\delta + \frac{\delta^2}{2[\phi''(1)]^2} \underbrace{\beta' H(0)^{-1} \beta}_{<0} + \frac{1}{n}O(\delta^2) + o(\delta^2),$$

$$v_n(\delta) = v_n(0) + \left[\frac{2}{\phi''(1)} \underbrace{\beta' H(0)^{-1} \beta}_{<0} + \frac{\theta}{2n} \right] \delta + \frac{1}{n}O(\delta^2) + O(\delta^2).$$

where

$$\begin{aligned} H(0) &:= \mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(x^*(0), Y_{n+1})], \quad G(0) := \nabla_x^2 \mathbb{V}_{\mathbb{P}}[f(x^*(0), Y_{n+1})], \\ \rho &:= \text{tr}(\xi'(0)H(0)) + \pi' \nabla_x [\text{tr}(\xi(0)H(0))], \\ \theta &:= \text{tr}(\xi'(0)G(0)) + \pi' \nabla_x [\text{tr}(\xi(0)G(0))], \\ \pi &:= \frac{1}{\phi''(1)} H(0)^{-1} \beta, \quad \beta := \text{Cov}_{\mathbb{P}}[\nabla_x f(x^*(0), Y_{n+1}), f(x^*(0), Y_{n+1})] \end{aligned}$$

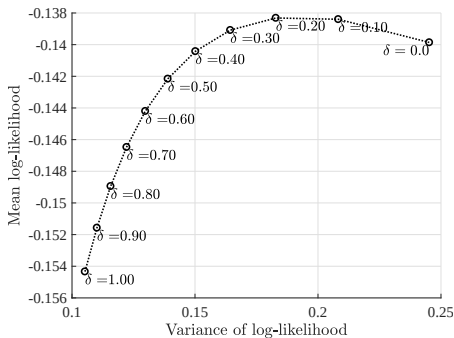
$\rightsquigarrow n$ が大きくなる際, $O(\delta)$ で残るのは分散の減少部分のみ

ロバスト平均・分散フロンティア

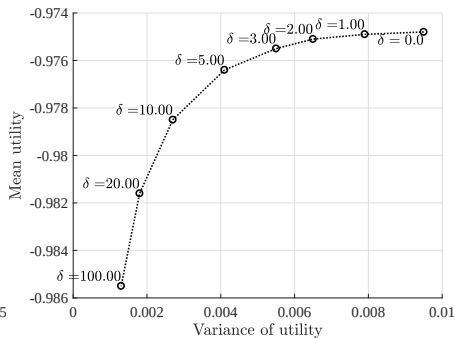
$$\left\{ (\mathbb{E}_{\mathbb{P}}[f(\mathbf{x}_n(\delta), \mathbf{Y}_{n+1})], \mathbb{V}_{\mathbb{P}}[f(\mathbf{x}_n(\delta), \mathbf{Y}_{n+1})]) : \delta \geq 0 \right\}$$

δ が小さい + f が x について狭義凸のとき

- 事後分散の減少効果が支配的: $O(\delta)$
- そのトレードオフとしての事後期待コストの増加は小さい: $O(\delta^2)$
- 事後期待コストの SAA からの改善: (あったとしても) $O(\delta/n)$



(i) ロジスティック回帰の DRO の例

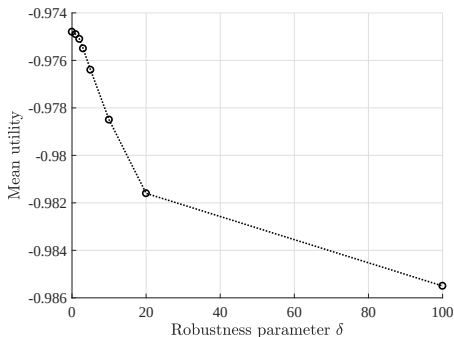


(ii) ポートフォリオ選択の例

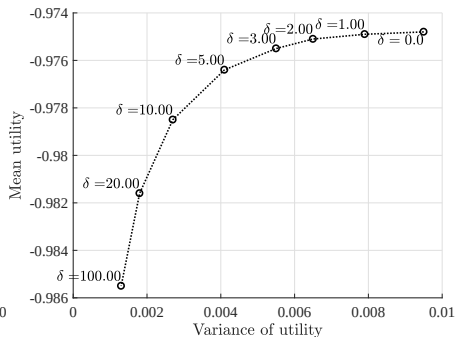
注意：縦軸が $-\mathbb{E}[f]$ になっています

DRO においてはばらつきを抑える視点が重要

ポートフォリオ選択の例



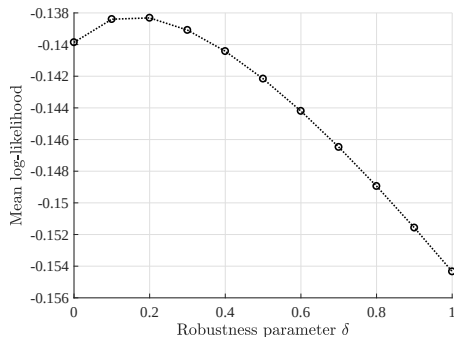
(i) 平均のみに着目した場合



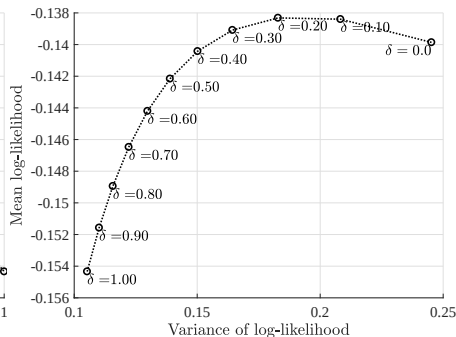
(ii) 平均・分散に着目した場合

DRO においてはばらつきを抑える視点が重要

ロジスティック回帰の例



(i) 平均のみに着目した場合



(ii) 平均・分散に着目した場合

この項のまとめ

- DRO において最悪の分布に基づくことの意味
 $\approx$ コスト f (or 報酬) 分布の広がりを抑える
- たとえば, 不確実性のサイズパラメータ ε/δ が小さいとき,
 KL ダイバージェンス $\leftrightarrow$ ばらつきは標準偏差/分散
 ※ 異なるダイバージェンスや不確実性集合に対しては異なるばらつきの尺度
- 事後的にも平均を改善させる効果よりは, ばらつきを減少させる効果の方が大きい (ペナルティー型 DRO の分析)
- 再標本法などと組み合わせたデータ駆動型の最適化