Optimization Under Uncertainty

ORI397-18928

Lecture Notes
Fall 2016

Grani A. Hanasusanto

Graduate Program in Operations Research and Industrial Engineering
The University of Texas at Austin

Chapter 1

Background Material

In this chapter, we review some of the relevant concepts that will be used throughout the course.

Notation: We denote the set of extended real numbers as $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. We use lower-case bold face letters to denote vectors and upper-case bold face letters to denote matrices. For any $I \in \mathbb{N}$, we define $[I]$ as the index set $\{1, \ldots, I\}$. We denote by $\mathbb{I}$ the identity matrix and by $\mathbf{e}$ the vector of all ones. Their dimensions will be clear from the context. All random variables are designated by tilde signs (*e.g.*, $\tilde{\boldsymbol{\xi}}$), while their realizations are denoted without tildes (*e.g.*, $\boldsymbol{\xi}$). The characteristic function of a set $\mathcal{S}$ is defined as $\chi_{\mathcal{S}}(\boldsymbol{\xi}) = 0$ if $\boldsymbol{\xi} \in \mathcal{S}$; $= \infty$ otherwise.

1.1 Convex Optimization

A large class of decision making problems can be formulated as a formal mathematical optimization model of the form

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^N \\ & f_i(\boldsymbol{x}) \leq 0 \quad \forall i \in [I]. \end{array} \tag{$\mathbf{P}$}$$

Here $f_0 : \mathbb{R}^N \to \overline{\mathbb{R}}$ is the objective function (cost, negative profit, etc.) that we seek to minimize, and $f_i : \mathbb{R}^N \to \overline{\mathbb{R}}$, $i \in [I]$, are constraint functions (budget, capacity, etc.) that define the feasible region of the decision variable $\boldsymbol{x}$. Note the constraint system in ($\mathbf{P}$) is equivalent to a single constraint given by

$$\max_{i \in [I]} f_i(\boldsymbol{x}) \leq 0.$$

We will find this representation useful when we delve further into optimization under uncertainty.

Definition 1 (Domain)**.** *The domain of a function f is defined as*

$$\mathrm{dom}(f) = \left\{ \boldsymbol{x} \in \mathbb{R}^N : f(\boldsymbol{x}) < +\infty \right\}.$$

For a cleaner presentation, we henceforth encapsulate all constraint functions into the set $\mathcal{X} \subseteq \mathbb{R}^N$ defined as

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : f_i(\boldsymbol{x}) \leq 0 \; \forall i \in [I] \right\}.$$

Definition 2 (Infimum)**.** *The infimum of a minimization problem* $(\mathbf{P})$ *is the largest number $z^\star$ such that $f_0(\boldsymbol{x}) \geq z^\star \; \forall \boldsymbol{x} \in \mathcal{X}$. We denote the infimum of* $(\mathbf{P})$ *as* $\inf(\mathbf{P}) \in \overline{\mathbb{R}}$.

Definition 3 (Global Minima)**.** *A point $\boldsymbol{x}^\star \in \mathcal{X}$ is called a global minima for* $(\mathbf{P})$ *if $f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^\star) \; \forall \boldsymbol{x} \in \mathcal{X}$. We further call $f_0(\boldsymbol{x}^\star)$ a global minimum of* $(\mathbf{P})$.

Definition 4 (Local Minima)**.** *A point $\boldsymbol{x}^\star \in \mathcal{X}$ is called a local minima for* $(\mathbf{P})$ *if there exists $\delta > 0$ such that $f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^\star) \; \forall \boldsymbol{x} \in \mathcal{X}$ with $\|\boldsymbol{x} - \boldsymbol{x}^\star\| \leq \delta$. We further call $f_0(\boldsymbol{x}^\star)$ a local minimum of* $(\mathbf{P})$.

Definition 5 (Feasibility)**.** *The problem* $(\mathbf{P})$ *is called feasible if $\mathcal{X} \neq \emptyset$, in which case* $\inf(\mathbf{P}) < +\infty$. *Otherwise, it is called infeasible and* $\inf(\mathbf{P}) = +\infty$.

Definition 6 (Unbounded Problem)**.** *The problem* $(\mathbf{P})$ *is called unbounded if* $\inf(\mathbf{P}) = -\infty$.

In this course, we mostly concern ourselves with *convex optimization*, in which the functions f_0 and f_i, $i \in [I]$, have convex domains and satisfy the convexity property

$$f_i(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f_i(\boldsymbol{x}) + (1-\lambda) f_i(\boldsymbol{y}) \qquad \forall \lambda \in [0,1], \; \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f_i),$$

for all $i \in \{0\} \cup [I]$.

Definition 7 (Proper Convex Function)**.** *A convex function f is called proper if* $\mathrm{dom}(f)$ *is non-empty and $f(\boldsymbol{x}) > -\infty$ for every $\boldsymbol{x} \in \mathbb{R}^N$.*

Operations that preserve convexity are:

1. **Composition with affine functions:** If f is convex then $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is also convex.

2. **Non-negative weighted sum:** if $f_1, \ldots, f_K$ are convex functions and $w_1, \ldots, w_K$ are non-negative numbers, then the combination $w_1 f_1 + \cdots + w_K f_K$ is convex. We can generalize this result to the integral $F(\boldsymbol{x}) = \int_{\mathcal{Y}} w(\boldsymbol{y}) f(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y}$.

3. **Pointwise supremum:** If $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$ for every fixed $\boldsymbol{y} \in \mathcal{Y}$, then the pointwise supremum $\sup_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y})$ yields a convex function.

4. **Partial minimization:** If $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $(\boldsymbol{x}, \boldsymbol{y})$ and $\mathcal{C}$ is a convex set then the function $F(\boldsymbol{x}) = \inf\{f(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in \mathcal{C}\}$ is convex in $\boldsymbol{x}$.

If f is differentiable then f is convex if and only if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f),$$

where the gradient ∇f is defined as

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \cdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_N} \end{bmatrix}.$$

For every instance of $(\mathbf{P})$ we associate with it a *dual* problem defined as

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\theta}) \\ \text{subject to} & \boldsymbol{\theta} \in \mathbb{R}^I_+, \end{array} \tag{$\mathbf{D}$}$$

where $g(\boldsymbol{\theta}) = \inf_{\boldsymbol{x} \in \mathbb{R}^N} f_0(\boldsymbol{x}) + \sum_{i \in [I]} \theta_i f_i(\boldsymbol{x})$. We can similarly define the optimal value of the maximization problem $(\mathbf{D})$ as $\sup(\mathbf{D})$.

Proposition 1. *We have* $\inf(\mathbf{P}) \geq \sup(\mathbf{D})$.

Proof. For every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\theta} \in \mathbb{R}^I_+$ we have

$$f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}) + \sum_{i \in [I]} \theta_i f_i(\boldsymbol{x}),$$

since $f_i(\boldsymbol{x}) \leq 0$, $i \in [I]$. Taking infimum over $\mathcal{X}$ on both sides, we find

$$\begin{aligned} \inf(\mathbf{P}) = \inf\{f_0(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\} \quad & \geq \inf_{\boldsymbol{x} \in \mathcal{X}} f_0(\boldsymbol{x}) + \sum_{i \in [I]} \theta_i f_i(\boldsymbol{x}) \\ & \geq \inf_{\boldsymbol{x} \in \mathbb{R}^N} f_0(\boldsymbol{x}) + \sum_{i \in [I]} \theta_i f_i(\boldsymbol{x}) = g(\boldsymbol{\theta}), \end{aligned}$$

where the second inequality holds because we have enlarged the feasible set of $\boldsymbol{x}$ from $\mathcal{X}$ to the full space $\mathbb{R}^N$. As the arising inequality holds for any $\boldsymbol{\theta} \in \mathbb{R}^I_+$, the desired relation is thus obtained by taking the supremum of the right-hand side expression. Thus the claim follows. □

For convex optimization problems, we can often have *strong duality* where $\inf(\mathbf{P}) = \sup(\mathbf{D})$. A sufficient condition is described in the following theorem.

Theorem 1 (Slater's Constraint Qualification)**.** *Let $\mathcal{I} \subseteq [I]$ be the set of indices for which the functions f_i, $i \in \mathcal{I}$, are non-affine in $\boldsymbol{x}$. If there exists $\boldsymbol{x}$ such that $f_i(\boldsymbol{x}) < 0$, $i \in \mathcal{I}$, and $f_i(\boldsymbol{x}) \leq 0$, $i \in [I] \setminus \mathcal{I}$, then* $\inf(\mathbf{P}) = \sup(\mathbf{D})$.

A subclass of convex optimization problems that will be of particular interest to us is *linear optimization* or *linear programming* (LP) problems, in which the functions f_0 and f_i, $i \in [I]$, are affine in $\boldsymbol{x}$. In this case, we can without loss of generality define $f_0(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$ and $f_i(\boldsymbol{x}) = \boldsymbol{a}_i^\top \boldsymbol{x} - b_i$, $i \in [I]$, for some problem specific vectors $\boldsymbol{c} \in \mathbb{R}^N$, $\boldsymbol{a}_i \in \mathbb{R}^N$, $i \in [I]$, and scalars $b_i \in \mathbb{R}$, $i \in [I]$. This gives rise to the optimization problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^N \\ & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array} \tag{$\mathbf{P}$-LP}$$

To formulate the dual of this problem, we derive the explicit expression for $g(\boldsymbol{\theta})$ in $(\mathbf{D})$

$$g(\boldsymbol{\theta}) = \inf_{\boldsymbol{x} \in \mathbb{R}^N} \left[\boldsymbol{c}^\top \boldsymbol{x} + \sum_{i \in [I]} \theta_i \boldsymbol{a}_i^\top \boldsymbol{x} - \sum_{i \in [I]} \theta_i b_i \right] = -\boldsymbol{b}^\top \boldsymbol{\theta} + \inf_{\boldsymbol{x} \in \mathbb{R}^N} \left[\boldsymbol{c}^\top \boldsymbol{x} + \sum_{i \in [I]} \theta_i \boldsymbol{a}_i^\top \boldsymbol{x} \right].$$

The last minimization problem evaluates to $-\infty$ if $\boldsymbol{A}^\top \boldsymbol{\theta} \neq -\boldsymbol{c}$. Thus, for the dual problem $(\mathbf{D})$ to be feasible, necessarily we must have $\boldsymbol{A}^\top \boldsymbol{\theta} = -\boldsymbol{c}$. This yields the explicit dual linear optimization problem

$$\begin{array}{ll} \text{maximize} & -\boldsymbol{b}^\top \boldsymbol{\theta} \\ \text{subject to} & \boldsymbol{\theta} \in \mathbb{R}^I_+ \\ & \boldsymbol{A}^\top \boldsymbol{\theta} = -\boldsymbol{c}. \end{array} \tag{$\mathbf{D}$-LP}$$

For linear optimization problems, feasibility of either the primal problem ($\mathbf{P}$-LP) or the dual problem ($\mathbf{D}$-LP) is sufficient to guarantee strong duality.

Theorem 2 (LP Strong Duality)**.** *If there exists $\boldsymbol{x} \in \mathbb{R}^N$ such that $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ or if there exists $\boldsymbol{\theta} \in \mathbb{R}^I_+$ such that $\boldsymbol{A}^\top \boldsymbol{\theta} = -\boldsymbol{c}$ then* $\inf(\mathbf{P}\text{-LP}) = \sup(\mathbf{D}\text{-LP})$.

1.2 Probability Theory

In a random experiment, the sample space Ω is the set containing all posible outcomes $\boldsymbol{\xi} \in \Omega$. Typically, we set $\Omega = \mathbb{R}^K$. All subsets $\mathcal{A} \subseteq \Omega$ are called events. A probability measure $\mathbb{P}$ assigns every event $\mathcal{A} \subseteq \Omega$ a probability $\mathbb{P}(\tilde{\boldsymbol{\xi}} \in \mathcal{A}) \in [0, 1]$. Without loss of generality, we may henceforth use the shorthand notation $\mathbb{P}(\mathcal{A}) = \mathbb{P}(\tilde{\boldsymbol{\xi}} \in \mathcal{A})$. The probability measure $\mathbb{P}$ satisfies the following properties:

- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.
- For any disjoint sets $\mathcal{A}, \mathcal{B} \subseteq \Omega$, we have $\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B})$.
- Let $\mathcal{A}^c$ be the complement of $\mathcal{A} \subseteq \Omega$ (*i.e.*, $\mathcal{A} \cap \mathcal{A}^c = \emptyset$ and $\mathcal{A} \cup \mathcal{A}^c = \Omega$). Then, $\mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{A}^c) = \mathbb{P}(\mathcal{A} \cup \mathcal{A}^c) = \mathbb{P}(\Omega) = 1$.
- If $\mathcal{A} \subseteq \mathcal{B} \subseteq \Omega$, then $\mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\mathcal{B})$ and $\mathbb{P}(\mathcal{B} \setminus \mathcal{A}) = \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A})$.

- If $\mathcal{A}, \mathcal{B} \subseteq \Omega$ arbitrary, then $\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B})$.

Definition 8 (Support). *The smallest closed set $\Xi \subseteq \Omega$ such that $\mathbb{P}(\Xi) = 1$ is called the support of $\tilde{\boldsymbol{\xi}}$.*

Definition 9 (Conditional Probability). *If $\mathcal{A}, \mathcal{B} \subseteq \Omega$ are two events, then*

$$\mathbb{P}(\mathcal{A}|\mathcal{B}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{B})}$$

is the conditional probability of $\mathcal{A}$ given $\mathcal{B}$, which is well defined if $\mathbb{P}(\mathcal{B}) > 0$.

Definition 10 (Independence). *If $\mathcal{A}, \mathcal{B} \subseteq \Omega$ are two events and $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$, then $\mathcal{A}$ and $\mathcal{B}$ are called independent.*

If $\mathcal{A}$ and $\mathcal{B}$ are independent then

$$\mathbb{P}(\mathcal{A}|\mathcal{B}) = \frac{\mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})}{\mathbb{P}(\mathcal{B})} = \mathbb{P}(\mathcal{A}).$$

Theorem 3 (Law of Total Probability). *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_I$ be mutually exclusive and collectively exhaustive events such that:*

- *$\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for all $i \neq j$.*
- *$\cup_{i \in [I]} \mathcal{A}_i = \Omega$.*

Then for any event $\mathcal{B}$, we have

$$\begin{aligned} \mathbb{P}(\mathcal{B}) &= \mathbb{P}(\mathcal{B} \cap \mathcal{A}_1) + \cdots + \mathbb{P}(\mathcal{B} \cap \mathcal{A}_I) \\ &= \mathbb{P}(\mathcal{B}|\mathcal{A}_1)\mathbb{P}(\mathcal{A}_1) + \cdots + \mathbb{P}(\mathcal{B}|\mathcal{A}_I)\mathbb{P}(\mathcal{A}_I). \end{aligned}$$

A *discrete random variable* $\tilde{\boldsymbol{\xi}}$ is described by a finite number of scenarios $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_S$ with occurence probabilities $p_1, \ldots, p_S$, that satisfies $\sum_{s \in [S]} p_s = 1$ and

$$p_s = \mathbb{P}(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}_s) \qquad \forall s \in [S].$$

A *continuous random variable* $\tilde{\boldsymbol{\xi}}$ is described by a probability density function $p(\boldsymbol{\xi})$ satisfying

$$\mathbb{P}(\mathcal{A}) = \int_{\mathcal{A}} p(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi},$$

for any events $\mathcal{A} \subseteq \Omega$. Two discrete univariate random variables $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are called independent if the probability of any outcome factors into the form

$$p_s(\boldsymbol{\xi}) = p_x(\xi_{1,s}) p_y(\xi_{2,s}).$$

Two continuous univariate random variables $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are called independent if the joint density function factors into the form

$$p(\boldsymbol{\xi}) = p_x(\xi_1) p_y(\xi_2).$$

Definition 11 (Expectation)**.** *Expectation of a univariate random variable* $\tilde{\xi}$ *is defined as*

$$\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \int_{\Omega} \xi \ \mathbb{P}(\mathrm{d}\xi). \tag{1.1}$$

For a discrete random variable the expectation (1.1) reduces to $\sum_{s \in [S]} p_s \xi_s$ while for a continuous random variable it reduces to $\int_{\Omega} \xi \ p(\xi) \mathrm{d}\xi$.

Definition 12 (Generalized Expectation)**.** *For a function* $f : \mathbb{R}^K \to \mathbb{R}$ *of a random vector* $\tilde{\boldsymbol{\xi}}$*, its expectation is given by*

$$\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})] = \int_{\Omega} f(\boldsymbol{\xi}) \ \mathbb{P}(\mathrm{d}\boldsymbol{\xi}).$$

Definition 13 (Variance)**.** *Variance of a random variable* $\tilde{\xi}$ *is defined as*

$$\mathrm{Var}(\tilde{\xi}) = \mathbb{E}_{\mathbb{P}} \left[\left(\tilde{\xi} - \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] \right)^2 \right],$$

while its standard deviation is defined as $\sigma(\tilde{\xi}) = \sqrt{\mathrm{Var}(\tilde{\xi})}$.

Definition 14 (Covariance)**.** *Covariance of two random variables* $\tilde{\xi}_1$ *and* $\tilde{\xi}_2$ *is defined as*

$$\mathrm{Cov}(\tilde{\xi}_1, \tilde{\xi}_2) = \mathbb{E}_{\mathbb{P}} \left[\left(\tilde{\xi}_1 - \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1] \right) \left(\tilde{\xi}_2 - \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_2] \right) \right],$$

while their correlation is defined as

$$\rho(\tilde{\xi}_1, \tilde{\xi}_2) = \frac{\mathrm{Cov}(\tilde{\xi}_1, \tilde{\xi}_2)}{\sigma(\tilde{\xi}_1)\sigma(\tilde{\xi}_2)}.$$

If $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are independent, then they are uncorrelated.

A continuous univariate random variable $\tilde{\xi}$ is said to be normal (or has a normal distribution) if its probability density function is of the form

$$p(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\xi - \mu)^2}.$$

We have $\mathbb{E}[\tilde{\xi}] = \mu$ and $\mathrm{Var}(\tilde{\xi}) = \sigma^2$. A standard normal random variable is a random variable that has a normal distribution with $\mu = 0$ and $\sigma = 1$. To express that $\tilde{\xi}$ is a normal random variable with mean μ and variance σ^2, we use the shorthand notation:

$$\tilde{\xi} \sim \mathcal{N}(\mu, \sigma^2).$$

Let $\tilde{\xi}_1$, $\tilde{\xi}_2$, $\tilde{\xi}_3$, … be an infinite sequence of independent and identically distributed (i.i.d.) random variables, each with an expected value μ. The *strong law of large numbers* (SLLN) asserts that

$$\mathbb{P} \left(\lim_{I \to \infty} \frac{1}{I} \sum_{i \in [I]} \tilde{\xi}_i = \mu \right) = 1.$$

Let $\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \ldots$ be an infinite sequence of independent and identically distributed (i.i.d.) random variables, each with expected value μ and variance σ^2. Define $\tilde{\chi}_I = \sum_{i\in[I]} \tilde{\xi}_i$. Note that $\mathbb{E}[\tilde{\chi}_I] = I \times \mu$ and $\mathrm{Var}(\tilde{\chi}_I) = I \times \sigma^2$. The *Central Limit Theorem* (CLT) states that for large I the random variable $(\tilde{\chi}_I - I\mu)/(\sigma\sqrt{I})$ is approximately standard normally distributed. More precisely, letting $\tilde{\xi} \sim \mathcal{N}(0,1)$, we have

$$\mathbb{P}\left(\frac{\tilde{\chi}_I - I\mu}{\sigma\sqrt{I})} \le r\right) \to \mathbb{P}(\tilde{\xi} \le r) \qquad \text{as } I \to \infty \ (\forall r \in \mathbb{R}).$$

1.3 Using YALMIP and MOSEK

YALMIP is a modeling language for convex optimization problems. Using YALMIP, one can interface MATLAB with various off-the-shelf solvers (CPLEX, GUROBI, MOSEK, etc.). YALMIP can be downloaded from `http://users.isy.liu.se/johanl/yalmip/pmwiki.php?n=Main.Download`. The installation manual can be found at `http://users.isy.liu.se/johanl/yalmip/pmwiki.php?n=Tutorials.Installation`.

We also need an optimization solver. In this course, we shall utilize MOSEK (`https://mosek.com/`) which is excellent for solving generic conic programs. MOSEK has free academic license which can be requested online from `https://license.mosek.com/academic/`.

To this end, let us try to use the YALMIP and MOSEK combination to solve a simple mean-variance portfolio optimization given by:

$$\begin{aligned} \text{minimize} \quad & \lambda \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} - (1-\lambda)\boldsymbol{\mu}^\top \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w} \in \mathbb{R}^N_+ \\ & \mathbf{e}^\top \boldsymbol{w} = 1. \end{aligned}$$

An example of implementation of the mean-variance portfolio optimization problem in YALMIP is given as follows.

Listing 1.1. Mean-Variance Optimization

```
clear all
yalmip clear
options = sdpsettings('verbose', 0, 'dualize', 0, 'solver', 'mosek');

N = 3; % number of assets
mu = [10; 20; 30]; % mean returns
sigma = 0.3*mu; % std deviations
corr_mat = gallery('randcorr',N); % correlation matrix
Sigma = diag(sigma)*corr_mat*diag(sigma); % covariance matrix
```

```
lambdas = [0:0.1:1];
variances = zeros(length(lambdas),1);
means = zeros(length(lambdas),1);

for i=1:length(lambdas)
lambda = lambdas(i);

w = sdpvar(N,1); % decision variable
obj = lambda*w'*Sigma*w - (1-lambda)*mu'*w; % objective value

% generate the constraints
constraints = {};
constraints{end+1} = w >= 0;
constraints{end+1} = sum(w) == 1;

% solve the problem
optimize([constraints{:}], obj);

% collect the outputs
variances(i) = double(w'*Sigma*w);
means(i) = double(mu'*w);
end

% plot the efficient frontier
plot(variances,means);
xlabel('Variance');
ylabel('Expected Return');
```

Optimization under uncertainty

HW. 30%
Project 30% due Friday. Nov. 18. 5:00 p.m.
Final 40% Wed. Dec. 14. 2:00 – 5:00 p.m.

Lecture 1

— Example: portfolio optimization.
Allocate wealth W into a number of assets.

$$\max \; x^T \xi$$
$$\text{s.t.} \; x \geq c$$
$$c^T w = W$$

Parameter $\xi = (\xi_G, \xi_A, \xi_E)$ is uncertainty

— Example: newsvendor problem.
demand is uncertainty.

— Example: machine learning.
given observation points with labels.
seek to classify a new data point into one of several categories.
the data points come from unknown distribution.

— Example: PILOT4 from NETLIB
coefficients are 0.1% accurate, → the optimal solution can violate 450%.

$$\min \; c^T x$$
$$\text{s.t.} \; Ax \leq b.$$

c, A, b all can be uncertainty.

<u>Definition</u>
Extended reals: $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$
$[I] = \{1, \cdots, I\}$, $I \in \mathbb{N}$

<u>Definition of general optimization problem:</u>

$$\inf \; f_0(x)$$
$$\text{s.t.} \; x \in \mathbb{R}^N \qquad (P)$$
$$f_i(x) \leq 0. \quad \forall i \in [I]$$

Feasible set:
$\mathcal{X} = \{x \in \mathbb{R}^N : f_i(x) \leq 0, \forall x \in [I]\}$

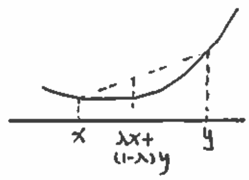
<u>Def</u>: $\mathrm{dom}(f) = \{x \in \mathbb{R}^N : f(x) < +\infty\}$?

<u>Def</u>: (Infimum). $\inf(P)$: largest number z^* such that $f(x) \geq z^*$, $\forall x \in X$

Def. (Feasibility)
Problem (P) is feasible, if $\exists x : f_i(x) < 0, \forall i \in [I]$.

Def (Unbounded)
Problem (P) is unbounded, if $\inf (P) = -\infty$

Def. (Convex.).
$f(\lambda x + (1-\lambda)x) \le \lambda f(x) + (1-\lambda) f(x) \quad \forall \lambda \in [0,1] \quad \forall x, y \in \mathrm{dom}(f)$



Convexity preserving operations.

1) Composition with affine function:
$f(x)$ is convex in x, then $g(x) = f(Ax+b)$ is also convex in x.

2) Nonnegative weighted sum:
If $f_1, \ldots, f_K$ are convex in x, then $g(x) = w_1 f(x) + \cdots + w_K f(x)$ is also convex in x. for $w_1, \ldots, w_K \ge 0$

3) Pointwise supremum:
If $f(x,y)$ is convex in x for every fixed $y \in Y$, then $g(x) = \sup_{y \in Y} f(x,y)$ is also convex in x.

4) Partial minimization:
If $f(x,y)$ jointly convex in x & y. & C is a convex set, then $g(x) = \inf \{ f(x,y) : y \in C \}$ is convex.

Duality theory : Largest possible lower bound

$$\sup_{\theta} \inf_{x \in R^N} f_0(x) + \sum_i \theta_i f_i(x)$$

$$\text{s.t.} \quad \theta \in R^I, \quad \theta \ge 0 \qquad (D)$$

(Weak Duality): $\inf(P) \ge \sup(D)$
(Strong Duality): if convex + mild conditions (Slater's condition)
then $\inf(P) = \sup(D)$

Linear Program (LP): $f_0(x) = c^T x$, $f_i(x) = a_i^T x - b_i$. $\forall i \in [I]$

$$\inf \; c^T x$$
$$\text{s.t.} \quad x \in \mathbb{R}^N, \quad Ax \le b$$

$$\sup_{\substack{\theta \in R^I \\ \theta \geq 0}} \inf_{x \in \mathbb{R}^N} \left[c^T x + \sum_{i \in [I]} \theta_i a_i^T x - \sum_{i \in [I]} \theta_i b_i \right]$$

$$= \sup_{\substack{\theta \in R^I \\ \theta \geq 0}} -b^T \theta + \inf_{x \in \mathbb{R}^N} \left[c^T x + \sum_{i \in [I]} \theta_i a_i^T x \right] \quad , \text{ if } -c \neq \sum_{i \in [I]} a_i^T x$$

$$= \sup \; -b^T \theta$$
$$\text{s.t.} \quad \theta \in R^I$$
$$A^T \theta = -c$$
$$\theta \geq 0$$

(LP Strong Duality): If there exists $x \in R^N$, $Ax \leq b$.
or $\theta \in R^I$, $A^T\theta = -c$, $\theta \geq 0$
then $\inf (P-LP) = \sup (D-LP)$

Set Ω contains all possible outcomes.

example: Dice $\Omega = \{1, 2, 3, 4, 5, 6\}$

All subsets $A \subseteq \Omega$ are called events.

$A = \{\text{outcome} \geq 4\} = \{4, 5, 6\}$

A probability measure $\mathbb{P}$ assigns any event $A \subseteq \Omega$ a probability $\mathbb{P}(A) \in [0, 1]$

$\mathbb{P}$ satisfies:

- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$
- A, B disjoint, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
- A^c complement of A, $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$
- $A \subseteq B$: $\mathbb{P}(A) \leq \mathbb{P}(B)$ and $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
- A, B arbitrary: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Def. (Support)

Support $\mathbb{P}$ is defined as the smallest closet set Ξ such that $\mathbb{P}(\Xi) = 1$

Example: unfair dice: $\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = \frac{1}{4}$, $\mathbb{P}(5) = \mathbb{P}(6) = 0$

then $\Xi = \{1, 2, 3, 4\}$

Def. (Conditional Probability)

$A \cdot B$ are two events, then the conditional probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Example:

$$\mathbb{P}(\text{car is behind door } 2 \mid \text{host open } 3)$$

$$= \frac{\mathbb{P}(\{\text{car behind } 2\} \cap \{\text{host open } 3\})}{\mathbb{P}(\{\text{host open } 3\})}.$$

$$= \frac{\mathbb{P}(\text{host open } 3 \mid \text{car behind } 2) \cdot \mathbb{P}(\text{car behind } 2)}{\mathbb{P}(\text{host open } 3)}$$

$$= \frac{1 \cdot 1/3}{\frac{1}{6} + \frac{1}{3}} = \frac{2}{3}$$

Def. (Independence).

$A \cdot B$ are two events and $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, then A & B are independent.

$\mathbb{P}(A \mid B) = \mathbb{P}(A) \cdot \mathbb{P}(B) / \mathbb{P}(B) = \mathbb{P}(A)$

<u>Random Variable</u>. $\tilde{\xi}: \Omega \to \mathbb{R}^k$

We typically set $\Omega = \mathbb{R}^k$, w.l.o.g. $\tilde{\xi}(\omega) = \omega$. $\forall \omega \in \Omega$

<u>Discrete random variable</u>: $\tilde{\xi}$ is supported on finitely many scenarios $\xi_1, \dots, \xi_s$ with probabilities $p_1, \cdots, p_s$, $\sum_{s \in [s]} p_s = 1$

<u>Continuous random variable</u>: described by a density function $p: \mathbb{R}^k \to \mathbb{R}_+$

$\mathbb{P}(A) = \int_A p(\xi) d\xi$. $\forall A \subseteq \Omega$

Normal: $p(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\xi-\mu)^2}$

Def (Expectation).

Expectation of $\tilde{\xi}$ is $E[\tilde{\xi}] = \int_\Omega \xi \, \mathbb{P}(d\xi)$

If $\tilde{\xi}$ is discrete, $E[\tilde{\xi}] = \sum_{s\in[S]} p_s \xi_s$

If $\tilde{\xi}$ is continuous, $E[\tilde{\xi}] = \int_\Omega \xi \, p(\xi) \, d\xi$

Def. (Generalized expectation). $f: \mathbb{R}^k \to \mathbb{R}$

$E[f(\tilde{\xi})] = \int_\Omega f(\xi) \, \mathbb{P}(d\xi)$

Def. (Variance)

$Var(\tilde{\xi}) = E[(\tilde{\xi} - E[\tilde{\xi}])^2]$

std dev $\sigma(\tilde{\xi}) = \sqrt{Var(\tilde{\xi})}$

Def. (Conditional Expectation).

$E[\tilde{\xi} \mid A] = \dfrac{E[\tilde{\xi} \, \mathbb{I}_A(\tilde{\xi})]}{\mathbb{P}(A)}$, where $\mathbb{I}_A(\xi) = \begin{cases} 1, & \text{if } \xi \in A \\ 0, & \text{if } \xi \notin A \end{cases}$

Example: $A = \{\xi \geq 4\}$

$E[\tilde{\xi} \mid A] = \frac{1}{6}(1\times 0 + 2\times 0 + 3\times 0 + 4\times 1 + 5\times 1 + 6\times 1) / \tfrac{1}{2} = 5$

Law of large numbers:

Let $\tilde{\xi}_1, \tilde{\xi}_2, \cdots$ be an infinite sequence of independent & identically distributed random variables with mean μ, then.

$$\mathbb{P}\left(\lim_{I\to\infty} \frac{1}{I} \sum_{i\in[I]} \tilde{\xi}_i = \mu \right) = 1$$

Central Limit Theorem: $\cdots$ with mean μ & variance σ^2

Let $\bar{X}_I = \sum_{i\in I} \tilde{\xi}_i$ & $\tilde{p}$ a std normal $\mu = 0$, $\sigma^2 = 1$.

then $\mathbb{P}\left(\dfrac{\bar{X}_I - I\cdot\mu}{\sigma\cdot\sqrt{I}} \leq r \right) \to \mathbb{P}(\tilde{p} \leq r)$ as $N \to \infty$, $\forall r \in \mathbb{R}$

$\inf \ f_0(x)$ Deterministic, for ξ fixed.

s.t. $x \in R^N$ $\quad \underbrace{[I]}_{\{1,2,\dots,I\}}$

$f_i(x) \le 0$, $\forall i \in [I]$

uncertainty $\to \xi \in \mathbb{R}^k$

obj. $f_0(x, \xi)$.

constraints: $f_i(x, \xi)$

example: (LP): $\xi = (c, A, b)$

$$\begin{aligned} &\inf \ c^T x \\ &\text{s.t. } x \in \mathbb{R}^N \\ &\qquad Ax \le b \end{aligned}$$

example (portfolio): ξ = return vector

$$\begin{aligned} &\inf \quad -x^T \xi \\ &\text{s.t.} \quad x \in \mathbb{R}^N \\ &\qquad\quad x \ge 0 \\ &\qquad\quad e^T x = 1 \end{aligned}$$

<u>Stochastic Optimization.</u>

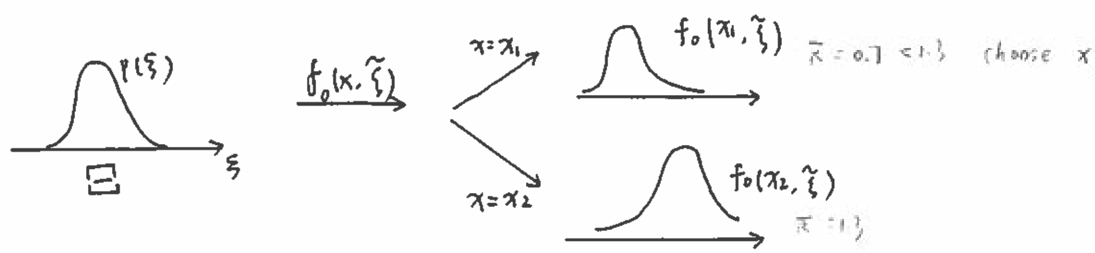
$\tilde{\xi} \sim \mathbb{P}$. tilde sign means random

any generic realization: ξ

$\tilde{\xi}$ discrete, the realizations ξ^s, $s \in [S]$

$\tilde{\xi} \sim \mathbb{P} \to$ $\inf \ f_0(x, \tilde{\xi})$ Stochastic.

$$\begin{aligned} &\text{s.t.} \quad x \in R^N \\ &\qquad f_i(x, \tilde{\xi}) \le 0, \ \forall i \in [I] \end{aligned}$$



Def. (Risk measure). $\mathcal{R} : \mathcal{L} \to \overline{\mathbb{R}}$

$\uparrow$ all univariate r.v.

$$\begin{aligned} &\inf \quad \mathcal{R}[f_0(x, \tilde{\xi})] \qquad \text{Deterministic.} \\ &\text{s.t.} \quad x \in R^N \\ &\qquad \mathcal{R}[f_i(x_i, \tilde{\xi})] \le 0, \quad \forall i \in [I] \end{aligned}$$

Constraint :

$$\sup_{\mathbb{P}\in\mathcal{P}} R_{\mathbb{P}}[f_i(x,\tilde{\xi})] \le 0 \iff R_{\mathbb{P}}[f_i(x,\tilde{\xi})] \le 0,\ \forall \mathbb{P}\in\mathcal{P}$$

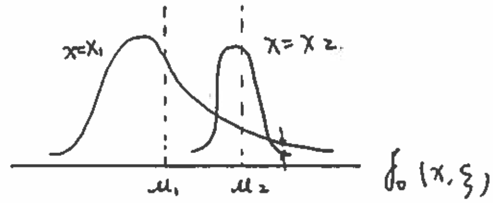
Example of $\mathcal{P}$

- Support only : $\mathcal{P} = \{\mathbb{P} \in \underbrace{\mathcal{P}_0(\Xi)}_{\text{set of all distributions supported on } \Xi}\}$ All possible prob. dist. of r.v. $\tilde{\xi}$
- mean + support : $\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : E[\tilde{\xi}] = \mu\}$
- mean + covariance + support : $\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : E_{\mathbb{P}}[\tilde{\xi}] = \mu,\ E_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] = \Sigma + \mu\mu^T$
- mean + mean-absolute deviation + support :

  $\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : E_{\mathbb{P}}[\tilde{\xi}] = \mu,\ E_{\mathbb{P}}[|\tilde{\xi} - \mu|] \le 6$
- symmetry : distribution is symmetric around μ
- unimodality

Example. (Risk neutral) : R = expected value.

$$\inf \ \mathbb{E}[f_0(x, \tilde{\xi})]$$



$\mu_1 < \mu_2$ choose x_1.
But x_1 has high chance to get higher obj, x_2 low chance get higher obj.

Example. (Risk averse : mean variance).

$$\inf \ \lambda \, \mathrm{Var}[f_0(x, \tilde{\xi})] + (1-\lambda) \, E[f_0(x, \tilde{\xi})]$$
$$\lambda \in [0, 1]$$

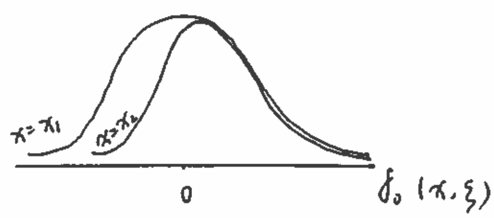
famous economic finder

Portfolio : $\inf \ \lambda \cdot \mathrm{Var}[-x^T \tilde{\xi}] + (1-\lambda) E[-x^T \tilde{\xi}] \ \rightarrow \inf \ \lambda \cdot x^T \Sigma x - (1-\lambda) x^T \mu$

s.t. $x \in R^N$

$x \geq 0, \ e^T x = 1$

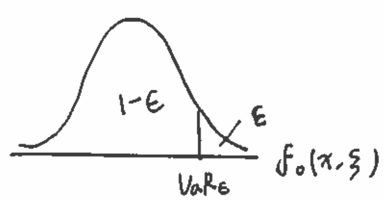
$$
\begin{aligned}
E[-x^T \tilde{\xi}] &= -x^T E[\tilde{\xi}] = -x^T \cdot \mu \\
\mathrm{Var}[+x^T \tilde{\xi}] &= E[(x^T \tilde{\xi} - E[x^T \tilde{\xi}])^2] \\
&= E[(x^T(\tilde{\xi} - \mu))^2] \\
&= E[x^T(\tilde{\xi} - \mu)(\tilde{\xi} - \mu)^T x] \\
&= x^T E[(\tilde{\xi} - \mu)(\tilde{\xi} - \mu)^T] x \\
&= x^T \cdot \Sigma \cdot x
\end{aligned}
$$



We only prefer left side of 0
but with variance, we [illegible] both
side [illegible]

Example (Risk averse: Value of risk)

$1-\epsilon$ quantile.

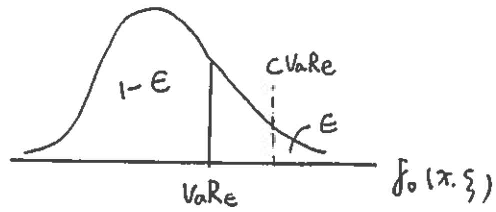


$$
\begin{aligned}
& VaR_\epsilon[f_0(x, \tilde{\xi})] \\
&= \inf \ t \\
& \quad \text{s.t. } \mathbb{P}(f_0(x, \tilde{\xi}) \leq t) \geq 1-\epsilon
\end{aligned}
$$

chance const

$$
\begin{aligned}
& \inf \ t \\
& \text{s.t. } x \in R^N, \ t \in R. \\
& \quad \mathbb{P}(f_0(x, \tilde{\xi}) \leq t) \geq 1-\epsilon
\end{aligned}
$$

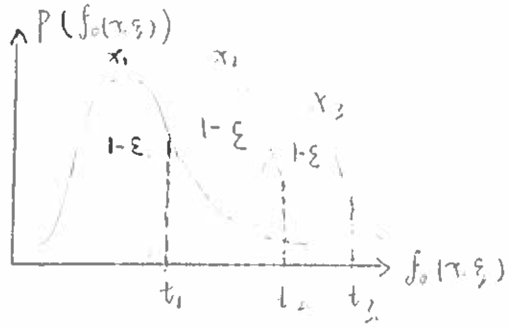
Example. (Risk averse : conditional value - at - risk)



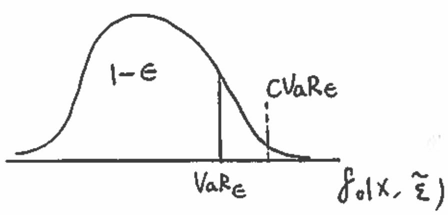
$$g(x) = \mathrm{CVaR}_\epsilon[f_0(x, \tilde{\xi})] = \inf_\beta \; \beta + \frac{1}{\epsilon} E[\max\{f_0(x, \tilde{\xi}) - \beta, 0\}]$$

convex

if this convex



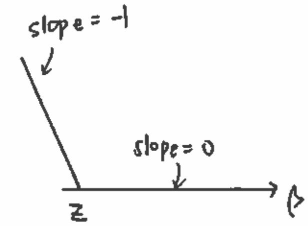
VaR_ϵ & $CVaR_\epsilon$



$$\tilde{z} = f_0(x, \tilde{\xi})$$

$$CVaR_\epsilon[\tilde{z}] = \inf_{\beta \in \mathbb{R}} \beta + \frac{1}{\epsilon} E[\max\{\tilde{z} - \beta, 0\}]$$

$$= E[\tilde{z} \mid \tilde{z} \geq VaR_\epsilon[\tilde{z}]]$$



$$\frac{\partial}{\partial \beta}\left(\beta + \frac{1}{\epsilon} E[\max\{\tilde{z} - \beta, 0\}]\right) = 0$$

$$1 + \frac{1}{\epsilon} E[-\mathbb{I}_{(\tilde{z} \geq \beta^*)}] = 0$$

$$1 - \frac{1}{\epsilon} P[\tilde{z} \geq \beta^*] = 0$$

$$\mathbb{P}[\tilde{z} \geq \beta^*] = \epsilon \Rightarrow \beta^* = VaR_\epsilon[\tilde{z}]$$

when $z = \beta$, it doesn't matter, since we assume continuous distribution.

since continuous, "$\geq$" can be substitude by "$>$"

$$CVaR_\epsilon[\tilde{z}] = \beta^* + \frac{1}{\epsilon} E[\max\{\tilde{z} - \beta^*, 0\}]$$

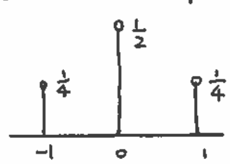
$$= \beta^* + \frac{1}{\epsilon} E[(\tilde{z} - \beta^*)\mathbb{I}\{\tilde{z} \geq \beta^*\}]$$

$$= \beta^* + \frac{1}{P(\tilde{z} \geq \beta^*)} E[\tilde{z} \cdot \mathbb{I}_{[\tilde{z} \geq \beta^*]}] - \frac{1}{P(\tilde{z} \geq \beta^*)} E[\beta^* \cdot \mathbb{I}_{[\tilde{z} \geq \beta^*]}]$$

$$= \beta^* + E[\tilde{z} \mid \tilde{z} \geq \beta^*] - \beta^*$$

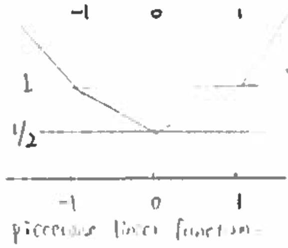
$$= E[\tilde{z} \mid \tilde{z} \geq VaR_\epsilon[\tilde{z}]]$$

Discrete example: $P(\tilde{z} = 1) = \frac{1}{4}$, $P(\tilde{z} = 0) = \frac{1}{2}$, $P(\tilde{z} = -1) = \frac{1}{4}$



$$VaR_{\frac{1}{2}}[\tilde{z}] = \inf_{t \in \mathbb{R}} t \quad \text{s.t. } \mathbb{P}(\tilde{z} \leq t) \geq 1 - \frac{1}{2}$$

$$= 0 = \beta^*$$



$$CVaR_{\frac{1}{2}}[\tilde{z}] = \inf_{\beta \in \mathbb{R}} \beta + \frac{1}{1/2}\left(\frac{1}{4}\max\{1-\beta, 0\} + \frac{1}{2}\max\{-\beta, 0\} + \frac{1}{4}\max\{-1-\beta, 0\}\right)$$

$$= \frac{1}{2}$$

$$E[\tilde{z} \mid \tilde{z} \geq VaR_\epsilon[\tilde{z}]] = \frac{\frac{1}{2} \times 0 + \frac{1}{4} \times 1}{\frac{1}{2} + \frac{1}{4}} = \frac{1}{3}$$

$$E[\tilde{z} \mid \tilde{z} > VaR_\epsilon[\tilde{z}]] = \frac{\frac{1}{4} \times 1}{\frac{1}{4}} = 1$$

$$E[\tilde{z} \mid \tilde{z} \geq VaR_\epsilon[\tilde{z}]] = \frac{\frac{1}{4} \times 0 + \frac{1}{4} \times 1}{\frac{1}{2}} = \frac{1}{2}$$

(A) & (B) are different ways to handle uncertainty in the constraints. Sometimes you may want to apply R to each constraint, in which case you get (A). Other times you may want to employ the equivalence of (C) & (D) and then apply R to (D). Usually, (A) and (B) lead to different feasible regions for x.

$$\min R[f_0(x,\tilde{\xi})]$$
$$\text{s.t.}\quad x \in X$$
$$\underbrace{R[f_i(x,\tilde{\xi})] \le 0,\ \forall i \in [I]}_{(A)} \quad \text{vs} \quad \underbrace{R[\max_{i\in[I]} f_i(x,\tilde{\xi})] \le 0}_{(B)}$$

$$\underbrace{f_i(x,\tilde{\xi}) \le 0\ .\ \forall i \in [I]}_{(C)} \iff \underbrace{\max_{i\in[I]} f_i(x,\tilde{\xi}) \le 0}_{(D)}$$

VaR_ϵ constraint $\iff$ chance constraint — even though f is convex, chance constraint maybe non-convex and leads to intractability

$$\text{VaR}_\epsilon[f_0(x,\tilde{\xi})] \le 0 \quad \mathbb{P}(f(x,\tilde{\xi}) \le 0) \ge 1-\epsilon$$
$$\iff \inf_{t\in\mathbb{R}} \{t : \mathbb{P}(f_i(x,\tilde{\xi}) \le t) \ge 1-\epsilon\} \le 0$$
$$\iff \exists t \le 0 : \mathbb{P}(f_i(x,\tilde{\xi}) \le t) \ge 1-\epsilon$$
$$\iff \mathbb{P}(f_i(x,\tilde{\xi}) \le 0) \ge 1-\epsilon$$

$$\text{CVaR}_\epsilon[f_i(x,\tilde{\xi})] \le 0$$
$$\Rightarrow \mathbb{P}(f_i(x,\tilde{\xi}) \le 0) \ge 1-\epsilon$$

VaR_ϵ CVaR_ϵ

VaR_ϵ is always smaller than CVaR_ϵ. Once $\text{CVaR}_\epsilon[\ ] \le 0$, we have $\text{VaR}_\epsilon[\ ] \le 0$ then chance constraint is satisfied. Since if $f(\ )$ is convex, CVaR_ϵ must be convex, then we find a convex approximation of chance constraint

$\tilde{\xi} \sim P$

<u>Distributionally robust model</u> distribution only partially known

ambiguity set $\mathcal{P}$, $\mathbb{P} \in \mathcal{P}$.

$\mathcal{P}$ contains all distributions consistent with prior information:

decision maker vs nature

$x \in X$ $\quad$ $\mathbb{P} \in \mathcal{P}$

$$\inf_{x\in X} \underbrace{\sup_{\mathbb{P}\in\mathcal{P}} R_{\mathbb{P}}[f_0(x,\tilde{\xi})]}_{R[f_0(x,\tilde{\xi})]}$$

Example (Ambiguity averse risk neutral: worst-case expectation)

$$\inf_{x\in X} \sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}[f_0(x,\tilde{\xi})]$$

Example (Ambiguity averse risk averse: worst-case VaR)

$$\inf_{x\in X} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{VaR}_\epsilon[f_0(x,\tilde{\xi})]$$

Piecewise Linear model

$$f(x,\varepsilon) = \max_{j \in [J]} a_j(x)^T \cdot \varepsilon + b_j(x)$$

any fix x, affine in ε
any fix ε, affine in x

where : $a_j(x) = A_j x + \hat{a}_j$ vector defined affinely by x.
$b_j(x) = b_j^T x + \hat{b}_j$ scalar defined by affine function in x

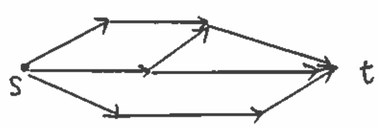
For any fixed ε, $f(x,\varepsilon)$ is convex in X.
For any fixed x, $f(x,\varepsilon)$ is convex in ε.

- Portfolio : $\inf_{x \in X} R[-\varepsilon^T x]$ only one piece, $|J| = 1$

- Newsvendor : $\inf_{x \in X} R[\underbrace{cx - d\min\{x, \tilde{\varepsilon}\}}]$

$$= cx + d\max\{-x, -\tilde{\varepsilon}\}$$
$$= cx + \max\{-dx, -d\tilde{\varepsilon}\}$$
$$= \max\{cx - dx,\ cx - d\tilde{\varepsilon}\}$$ two pieces.

- Shortest path:



$$\inf_{x \in X \cap \{0,1\}^{|E|}} R[\sum_{ij} \tilde{w}_{ij} x_{ij}]$$

uncertainty

flow conservation.

- Regression : $Fx = g$.

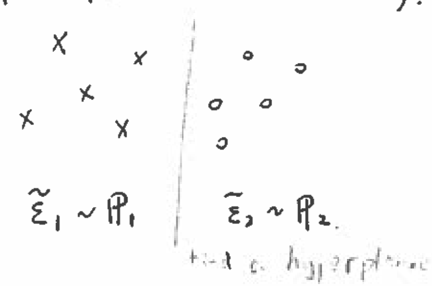
$$\inf_{x \in \mathbb{R}^N} R(\|\tilde{F}x - \tilde{g}\|_1)$$ 2 norm also works

$$\Leftrightarrow \inf_{x \in \mathbb{R}^N} \|\hat{F}x - \hat{g}\|_1 + \lambda \|x\|_r$$ equivalent to the deterministic form.

- Classification (Chance Constraints).

r ?

$$\begin{aligned} \sup\ & \alpha. \\ \text{s.t. }\ & (x, x_0) \in X \\ & \mathbb{P}_1(\tilde{\varepsilon}_1^T x \geq x_0) \geq \alpha \\ & \mathbb{P}_2(\tilde{\varepsilon}_2^T x \leq x_0) \geq \alpha. \end{aligned}$$



$\mathbb{P}$ is discrete with S scenerios

— Expectation as risk measure :

$$\inf_{x\in X} E_{\mathbb{P}}\left[\max_{j\in[J]} a_j(x)^T \tilde{\xi} + b_j(x)\right]$$

$$= \inf_{x\in X} \sum_{s\in[S]} p_s \max_{j\in[J]} a_j(x)^T \xi^s + b_j(x).$$

p_s – probability of scenario ξ^s

$$= \inf_{\substack{x\in X \\ r\in\mathbb{R}^S}} \sum_{s\in[S]} p_s r_s$$

$$\text{s.t.} \quad a_j(x)^T \xi^s + b_j(x) \le r_s \quad \forall s\in[S],\ \forall j\in[J]$$

$$\Leftrightarrow \max_{j\in[J]} a_j(x)^T\cdot \xi^s + b_j(x) \le r_s, \quad \forall s\in[S]$$

— $CVaR_\epsilon$:

when $\epsilon = 1$.

$$CVaR_\epsilon[\tilde{\xi}] = E[\tilde{\xi} \mid \tilde{\xi} \geq VaR_\epsilon[\tilde{\xi}]]$$

$$= E[\tilde{\xi}]$$

$$\inf_{x\in X} CVaR_\epsilon\left[\max_{j\in[J]} a_j(x)^T \tilde{\xi} + b_j(x)\right]$$

$$= \inf_{x\in X} \inf_{\beta\in\mathbb{R}} \beta + \frac{1}{\epsilon} E\left[\max\left\{\max_{j\in[J]} a_j(x)^T \xi + b_j(x) - \beta,\ 0\right\}\right]$$

$$= \inf_{\substack{x\in X \\ \beta\in\mathbb{R}.}} \beta + \frac{1}{\epsilon} \sum_{s\in[S]} p_s \cdot \max\left\{\max_{j\in[J]} a_j(x)^T \xi^s + b_j(x) - \beta,\ 0\right\} \le r_s \qquad r_s \in \mathbb{R}^S_+$$

— VaR_ϵ :

$$\inf_{x\in X} VaR_\epsilon\left[\max_{j\in[J]} a_j(x)^T \tilde{\xi} + b_j(x)\right]$$

$$= \inf_{t\in\mathbb{R},\ x\in X} t.$$

$$\text{s.t.} \quad \mathbb{P}\left(\max_{j\in[J]} a_j(x)^T\cdot\tilde{\xi} + b_j(x) \le t\right) \geq 1-\epsilon$$

$$= \inf_{\substack{t\in\mathbb{R} \\ z\in\{0,1\}^S \\ x\in X}} t. \qquad \text{s.t.} \quad \left(t - a_j(x)^T\cdot\xi^s - b_j(x)\right)\cdot z_s \overset{+(1-z_s)\cdot M}{} \geq 0 \qquad \forall j\in[J],\ \forall s\in S.$$

$$\sum_{s\in[S]} p_s z_s \geq 1-\epsilon$$

$\mathbb{P}$ is continuous

$E[\max\{a^T\xi + b, 0\}]$

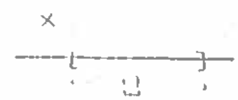
- Monte Carlo methods: sample S scenarios from $\mathbb{P}$
- bounding methods: obtain upper and lower bounds.

Distributionally Robust Models

- Support only $\mathcal{P} = \{ \mathbb{P} \in \underbrace{\mathcal{P}_0(\Xi)}_{\text{all dist. supported on } \Xi} \}$

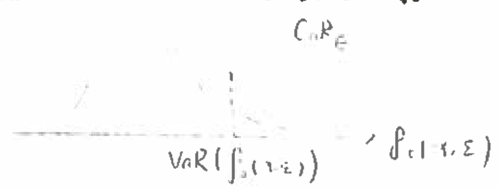
  the dist $\mathbb{P} \in \mathcal{P}$ can only assign positive mass to events in Ξ

✓



$\Leftrightarrow$ Robust optimization: the worst-case dist. is Dirac distribution with single scenario.

dist. on $\tilde{\xi}$ induces a dist. on $f_0(x, \tilde{\xi})$



Nature will place a scenario $\xi^* \in \Xi$ that maximizes $f_0(x,\xi)$

$\Rightarrow$ worst case dist $\mathbb{P}^* = \delta_{\xi^*}$

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-}VaR_\epsilon[f_0(x,\tilde{\xi})] = \mathbb{P}^*\text{-}VaR_\epsilon[f_0(x,\tilde{\xi})]$$
$$= \mathbb{P}^*\text{-}CVaR_\epsilon[f_0(x,\tilde{\xi})]$$
$$= E_{\mathbb{P}^*}[f_0(x,\tilde{\xi})] = f_0(x,\xi^*)$$

Assume Ξ is a polytope $\Xi = \{\xi \in \mathbb{R}^k : S\xi \le t\}$

$$\inf_{x\in X} \sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}[f_0(x,\tilde{\xi})]$$

find a scenario $\xi \in \Xi$ that maximizes $f(x,\xi)$ following the above argument

$$= \inf_{x\in X} \sup_{\xi\in\Xi} f_0(x,\xi)$$

$$= \inf_{\substack{x\in X \\ r\in\mathbb{R}}} r$$

$$\text{s.t.} \quad \sup_{\xi\in\Xi} f_0(x,\xi) \le r$$

Linear Programming

$$\sup_{\xi\in\Xi} \max_{j\in[J]} a_j(x)^T\xi + b_j(x) \le r$$

$$\Leftrightarrow \max_{j\in[J]} \sup_{\xi\in\Xi} a_j(x)^T\xi + b_j(x) \le r$$

$$\Leftrightarrow \sup_{\xi\in\Xi} a_j(x)^T\xi + b_j(x) \le r, \quad \forall j\in[J]$$

$$\sup_{\varepsilon:\, S\varepsilon \le t} a_j(x)^T \varepsilon = \sup_{\varepsilon \in \mathbb{R}^k} \inf_{\theta_j \in \mathbb{R}_+^m} a_j(x)^T \varepsilon + \theta_j^T t - \theta_j^T S \varepsilon$$

$$= \inf_{\theta_j \in \mathbb{R}_+^m} \sup_{\varepsilon \in \mathbb{R}^k} a_j(x)^T \varepsilon + \theta_j^T t - \theta_j^T S \varepsilon$$

Dual

$$= \inf_{\theta_j \in \mathbb{R}_+^m} t^T \theta_j$$

$$\text{s.t.} \quad S^T \theta_j = a_j(x)$$

o.w. the problem will be unbounded

$$\Leftrightarrow \left.\begin{aligned} &\inf_{\theta_j \in \mathbb{R}_+^m} t^T \theta_j + b_j(x) \\ &\text{s.t. } S^T \theta_j = a_j(x) \end{aligned}\right\} \le r \quad \forall j \in [J]$$

$$\Leftrightarrow \exists \theta_j \in \mathbb{R}_+^m : \ t^T \theta_j + b_j(x) \le r, \quad S^T \theta_j = a_j(x), \ \forall j \in [J]$$

$$\Leftrightarrow \inf_{\substack{x \in X \\ r \in \mathbb{R} \\ \theta_j \in \mathbb{R}_+^m}} r \quad \text{s.t. } \begin{aligned} &t^T \theta_j + b_j(x) \le r \quad \forall j \in [J] \\ &S^T \theta_j = a_j(x) \end{aligned}$$

— mean & covariance matrix

$$P = \{ \mathbb{P} \in P(\mathbb{R}^k) : \mathbb{E}_P[\tilde{\varepsilon}] = \mu, \ \mathbb{E}[(\tilde{\varepsilon} - \mu)(\tilde{\varepsilon} - \mu)^T] \le \Sigma \}$$

$\Leftrightarrow E[\tilde{\varepsilon}\tilde{\varepsilon}^T] \le \Sigma + \mu\mu^T$

- Delage & Ye (2010): generic framework
- H Scarf (1958): single-item newsvendor problem.

$$M \succeq 0 \Leftrightarrow \varepsilon^T M \varepsilon \ge 0, \ \forall \varepsilon \in \mathbb{R}^k$$

$$\Sigma - E[(\tilde{\varepsilon} - \mu)(\tilde{\varepsilon} - \mu)^T] \succeq 0$$

If we let $\Sigma = E[(\tilde{\varepsilon} - \mu)(\tilde{\varepsilon} - \mu)^T]$

- if $\Xi = \mathbb{R}^k$: (=) equivalent to ($\le$)
- if Ξ is a polytope : (=) leads to NP hard problem.

$$E[(\tilde{\varepsilon} - \mu)(\tilde{\varepsilon} - \mu)^T] = E[\tilde{\varepsilon}\tilde{\varepsilon}^T] - 2\mu\mu^T + \mu\mu^T = E[\tilde{\varepsilon}\tilde{\varepsilon}^T] - \mu\mu^T$$

<u>Lemma:</u> $\begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \succeq 0 \Leftrightarrow c + b^T \varepsilon + \varepsilon^T A \varepsilon \ge 0 \quad \forall \varepsilon \in \mathbb{R}^k$

Proof:

$$(\Rightarrow) \begin{bmatrix} \varepsilon \\ \rho \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \varepsilon \\ \rho \end{bmatrix} \ge 0 \quad \forall (\varepsilon, \rho) \in \mathbb{R}^k \times \mathbb{R}$$

$$\text{set } \rho = 1 \Rightarrow \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix} \ge 0 \quad \forall \varepsilon \in \mathbb{R}^k$$

$$\Leftrightarrow c + b^T \varepsilon + \varepsilon^T A \varepsilon \ge 0, \quad \forall \varepsilon \in \mathbb{R}^k$$

$$(\Leftarrow) \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix} \ge 0 \quad \forall \varepsilon \in \mathbb{R}^k$$

$$\Rightarrow \rho^2 \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix} \geq 0, \ \forall \varepsilon \in \mathbb{R}^k, \ \forall \rho \in \mathbb{R}, \ \rho \neq 0$$

$$\Rightarrow \begin{bmatrix} \varepsilon\rho \\ \rho \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \varepsilon\rho \\ \rho \end{bmatrix} \geq 0, \ \forall \varepsilon \in \mathbb{R}^k, \ \forall \rho \in \mathbb{R}, \ \rho \neq 0$$

$$\overset{\theta = \varepsilon\rho}{\Rightarrow} \begin{bmatrix} \theta \\ \rho \end{bmatrix}^T \begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & c \end{bmatrix} \begin{bmatrix} \theta \\ \rho \end{bmatrix} \geq 0, \ \forall \theta \in \mathbb{R}^k, \ \forall \rho \in \mathbb{R}, \ \rho \neq 0$$

... θ ... ε, but

for any fixed $\theta \in \mathbb{R}^k$, the LHS term is quadratic function in ρ and therefore is continuous. $\Rightarrow \rho = 0$

thus means the LHS term is also ≥ 0 for $\rho = 0$

— worse-case expectation: $\sup\limits_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [f_0(x, \tilde{\xi})]$

$$= \sup \int_{\mathbb{R}^k} f_0(x, \tilde{\varepsilon}) \, \mathbb{P}(d\varepsilon)$$

s.t. $\mathbb{P}$ is a probability measure.

$$\int_{\mathbb{R}^k} \varepsilon \, \mathbb{P}(d\varepsilon) = \mu$$

$$\int_{\mathbb{R}^k} \varepsilon\varepsilon^T \, \mathbb{P}(d\varepsilon) \leq \Sigma + \mu\mu^T$$

$$= \sup \int_{\mathbb{R}^k} f_0(x, \varepsilon) \, \nu(d\varepsilon)$$

s.t. ν is a nonnegative measure

$$\int_{\mathbb{R}^k} \nu \, d(\varepsilon) = 1$$

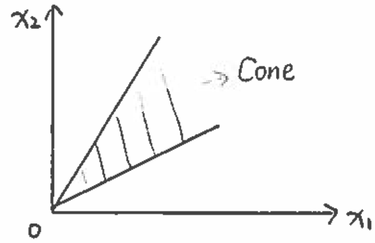
$$\int_{\mathbb{R}^k} \varepsilon \, \nu(d\varepsilon) = \mu$$

$$\int_{\mathbb{R}^k} \varepsilon\varepsilon^T \, \nu(d\varepsilon) \leq \Sigma + \mu\mu^T$$

LP with ∞ many variables

Conic Programming

Linear Optimization Problems over a convex cone.



$\mathcal{C}$ is a <u>cone</u> if $\forall x \in \mathcal{C}$. we have $\lambda x \in \mathcal{C}$ $\forall \lambda \geq 0$

$\mathcal{C}$ is a <u>convex cone</u> if it is cone and is convex.

The shadow part in graph is a convex cone.

Conic Program: $\inf \quad c^T x$

s.t. $x \in R^N$

$Ax \preceq_{\mathcal{C}} b \quad \Leftrightarrow \quad b - Ax \in \mathcal{C}$

$\uparrow$ cone $\mathcal{C}$

Example:

(1) $\mathcal{C} = R_+^k \Rightarrow b - Ax \in R_+^k \Rightarrow Ax \leq b$ (LP)

(2) $\mathcal{C} = S_+^k \Rightarrow$ Semidefinite programming (SDP)

S^k : set of symmetric matrices in $R^{k \times k}$ $(A^T = A)$

S_+^k : set of positive semidefinite matrices in $R^{k \times k}$

$M \in S_+^k \overset{\Leftrightarrow}{\text{iif}} \Sigma^T M \Sigma \geq 0 \quad . \forall \Sigma \in R^k$

SDP, $\inf \quad c^T x$

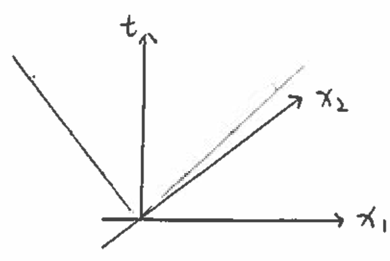
s.t. $x \in R^N$

$A_1 x_1 + \cdots + A_N x_N \preceq_{S_+^k} B$

where $A_1, A_2, \cdots, A_N, B \in S^k$

$x_1, x_2, \cdots x_n$ are univariate variables.

(3) $\mathcal{C} = \{ (\underset{\text{vector}}{x}, \underset{\text{scalar}}{t}) \in R^N \times R : \|x\|_2 \leq t \}$ — second-order cone (SOC)



Dual cone: $\mathcal{C}^* : \{ y \in R^k : \langle y, x \rangle \geq 0 \ . \ \forall x \in \mathcal{C} \}$

wher. $\langle y, x \rangle$ is inner product

Self Dual Cones ($\mathcal{C} = \mathcal{C}^*$) :

- $\mathcal{C} = R_+^k$. $\mathcal{C}^* = R_+^k$
- $\mathcal{C} = S_+^k$
- $\mathcal{C} = \text{soc}$ (second - order cone)

Conic Program :
$$\inf \; c^T x \quad \text{s.t.} \; x \in R^N, \quad Ax \preccurlyeq_{\mathcal{C}} b \iff b - Ax \in \mathcal{C}$$

Conic Duality :
$$\inf_{x \in R^N} \sup_{\theta \in \mathcal{C}^*} c^T x - \theta^T (b - Ax)$$
$$= \sup_{\theta \in \mathcal{C}^*} \inf_{x \in R^N} -\theta^T b + c^T x + \theta^T A x$$
$$= \sup \; -\theta^T b \quad \text{s.t.} \; \theta \in \mathcal{C}^*, \quad A^T \theta = -c$$

Lemma :
$$\begin{bmatrix} A & \frac{1}{2} b \\ \frac{1}{2} b^T & c \end{bmatrix} \succcurlyeq 0 \iff c + b^T \varepsilon + \varepsilon^T A \varepsilon \geq 0, \; \forall \varepsilon \in R^k$$

— Mean & Covariance : $\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}(R^k) : E[\tilde{\varepsilon}] = \mu, \; E[\tilde{\varepsilon} \cdot \tilde{\varepsilon}^T] \preccurlyeq \Sigma + \mu\mu^T \}$

upper bound of covariance is in terms of SDP equality

— worst — case Expectation:

loss function

fix x : $\sup_{\mathbb{P} \in \mathcal{P}} E_{\mathbb{P}} [f_0 (x, \tilde{\varepsilon})]$

$$= \sup \int_{R^k} f_0 (x, \varepsilon) \, v(d\varepsilon)$$

s.t. $v(\cdot)$ is a nonnegative probability measure. (Decision Variable)

$$\int_{R^k} v(d\varepsilon) = 1, \quad \int_{R^k} \varepsilon \cdot v(d\varepsilon) = \mu, \quad \int_{R^k} \varepsilon \varepsilon^T v(d\varepsilon) \preccurlyeq \Sigma + \mu\mu^T$$

dual variables: $\alpha \in R$ $\quad \beta \in R^k$ $\quad \gamma \in S_+^k$

$(v : R^k \to R_+)$

$$= \sup_{v(\cdot) \geq 0} \inf_{\alpha \in R, \beta \in R^k, \gamma \in S_+^k} \int_{R^k} f_0 (x, \varepsilon) v(d\varepsilon) + \alpha - \int_{R^k} \alpha \, v(d\varepsilon)$$
$$+ \beta^T \mu - \int_{R^k} \beta^T \varepsilon \, v(d\varepsilon) + \langle \gamma, \Sigma + \mu\mu^T \rangle$$
$$- \int \langle \gamma, \varepsilon\varepsilon^T \rangle \, v(d\varepsilon)$$

where $\langle \gamma, \varepsilon\varepsilon^T \rangle = \operatorname{tr}(\gamma \varepsilon \varepsilon^T) = \operatorname{tr}(\varepsilon^T \gamma \varepsilon) = \varepsilon^T \gamma \varepsilon$.

Assume strong duality holds, exchange

$$= \inf_{\alpha, \beta, \gamma} \sup_{v(\cdot) \geq 0} \alpha + \beta^T \mu + \langle \gamma, \Sigma + \mu\mu^T \rangle + \int_{R^k} [f_0 (x, \varepsilon) - \alpha - \beta^T \mu - \varepsilon^T \gamma \varepsilon^T] \, v(d\varepsilon)$$

nonnegative

$$= \inf \; \alpha + \beta^T \mu + \langle \gamma, \Sigma + \mu\mu^T \rangle$$
s.t. $\alpha \in R$, $\beta \in R^k$, $\gamma \in S_+^k$

$f_0(x,\varepsilon) \le \alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon, \quad \forall \varepsilon \in R^k \quad (*)$

where $(*)$

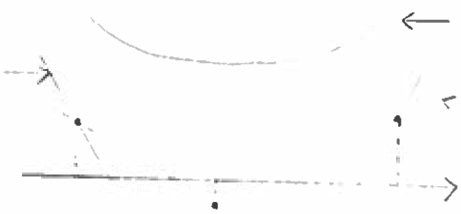
$$\overset{\Rightarrow}{\Longleftrightarrow} \; a_j(x)^T\varepsilon + b_j(x) \le \alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon, \; \forall \varepsilon \in R^k, \; \forall j \in [J]$$

$$\Longleftrightarrow \; 0 \le \alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon - a_j(x)^T\varepsilon - b_j(x), \; \forall \varepsilon \in R^k, \; \forall j \in [J]$$

$$\overset{\text{Lemma}}{\Longleftrightarrow} \begin{bmatrix} \gamma & \frac{1}{2}(\beta - a_j(x)) \\ \frac{1}{2}(\beta - a_j(x))^T & \alpha - b_j(x) \end{bmatrix} \succeq 0, \quad \forall j \in [J]$$

Informal Iterpretation:

$a_1(x)^T\varepsilon + b_1(x) \to$



$\leftarrow \alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon$

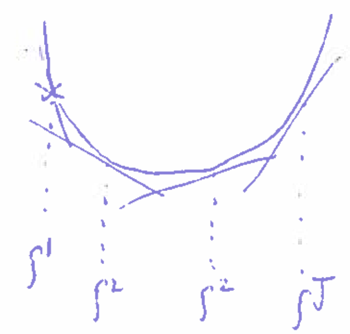
$\leftarrow a_j(x)^T\varepsilon + b_j(x)$

$$E_P[\alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon] = \alpha + \beta^T\mu + \langle \gamma, \Sigma + \mu\mu^T \rangle$$

Complementary slackness:

$$\Big(\alpha + \beta^T\varepsilon + \varepsilon^T\gamma\varepsilon - \max_{j\in[J]} \{a_j(x)^T\varepsilon + b_j(x)\}\Big) \cdot v(\varepsilon) = 0, \quad \forall \varepsilon \in R^k$$

where $\max_{j\in[J]} \{a_j(x)^T\varepsilon + b_j(x)\}$ is loss function.



By complementary slackness, whenever $\alpha + \beta^T\xi + \xi^T\Gamma\xi > \max_{j\in[J]} a_j(x)^T\xi + b_j(x)$ we must have $v(\xi) = 0$. This means that the measure $v(\cdot)$ can only assign positive mass to the set

$$\{\xi \in \mathbb{R}^k : \alpha + \beta^T\xi + \xi^T\Gamma\xi = \max_{j\in[J]} a_j(x)^T\xi + b_j(x)\},$$

which is a finite set.

- Worst-case CVaR:

$$\inf_{x\in X} \inf_{\tau\in\mathbb{R}} \tau + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max\{f_0(x,\tilde{\xi})-\tau, 0\}\right]$$

$$= \inf \tau + \frac{1}{\epsilon}\left(\alpha + \beta^T\mu + \langle \Gamma, \Sigma + \mu\mu^T\rangle\right)$$

Dual [illegible] α, β, Γ

$$\text{s.t. } x\in X,\ \tau\in\mathbb{R},\ \alpha\in\mathbb{R},\ \beta\in\mathbb{R}^k,\ \Gamma\in S_+^k$$

$$f_0(x,\xi) - \tau \le \alpha + \beta^T\xi + \xi^T\Gamma\xi,\quad \forall \xi\in\mathbb{R}^k$$

$$\Leftrightarrow \max_{j\in[J]} a_j(x)^T\xi + b_j(x) - \tau \le \alpha + \beta^T\xi + \xi^T\Gamma\xi,\quad \forall\xi\in\mathbb{R}^k$$

$$\Leftrightarrow a_j(x)^T\xi + b_j(x) - \tau \le \alpha + \beta^T\xi + \xi^T\Gamma\xi,\quad \forall\xi\in\mathbb{R}^k,\ \forall j\in[J],$$

$$0 \le \alpha + \beta^T\xi + \xi^T\Gamma\xi,\quad \forall\xi\in\mathbb{R}^k$$

$$\begin{bmatrix} \Gamma & \frac{1}{2}(\beta - a_j(x)) \\ \frac{1}{2}(\beta - a_j(x))^T & \alpha - b_j(x) + \tau \end{bmatrix} \succeq 0 \quad \forall j\in[J] \qquad \begin{bmatrix} \Gamma & \frac{1}{2}\beta \\ \frac{1}{2}\beta^T & \alpha \end{bmatrix} \succeq 0$$

- Worst-case VaR
  - NP-hard for $J > 1$
  - If $J = 1$, second order cone programming (SOCP). due to El-Ghaoui, Oks, Oustry (2003)

    SOCP formulated as LP, easy to solve

    SDP can [illegible] in polynomial time, but [illegible] performance

    - $$\inf_{x\in X, t\in\mathbb{R}} t.$$

      $$\text{s.t. } \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(a(x)^T\xi + b(x) \le t\right) \ge 1-\epsilon$$

      ($\text{assume } \epsilon < 1$, not redundant)

    - Fix $x\in X$, $t\in\mathbb{R}$: ([illegible])

      $$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(a^T\xi + b \le t) = \inf \int_{\mathbb{R}^k} \mathbb{I}[a^T\xi + b \le t]\, \nu(d\xi)$$

      $$\text{s.t. } \nu(\cdot) \ge 0$$

      $$\int_{\mathbb{R}^k} \nu(d\xi) = 1$$

      $$\int_{\mathbb{R}^k} \xi\, \nu(d\xi) = \mu \qquad \text{[illegible] definition of}$$

      $$\int_{\mathbb{R}^k} \xi\cdot\xi^T\, \nu(d\xi) \preceq \Sigma + \mu\mu^T \qquad \text{ambiguity [illegible]}$$

Dualize:

$$\inf_{\nu(\cdot)\ge 0} \inf_{\substack{\alpha\in\mathbb{R} \\ \beta\in\mathbb{R}^k \\ \gamma\in S_+^k}} \int \mathbb{I}[\cdot]\,\nu(d\xi) + \alpha - \int \alpha\cdot\nu(d\xi) + \beta^T\mu - \int \beta^T\xi\,\nu(d\xi) - \langle\Gamma, \Sigma + \mu\mu^T\rangle + \int \xi^T\Gamma\xi\,\nu(d\xi)$$

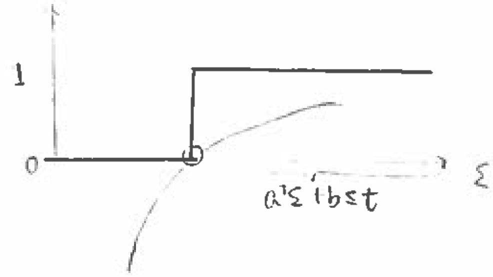
exchange: then solve ν analytically, only dual left

$$= \sup \alpha + \beta^T\mu - \langle\Gamma, \Sigma + \mu\mu^T\rangle$$

$$\text{s.t. } \alpha\in\mathbb{R},\ \beta\in\mathbb{R}^k,\ \Gamma\in S_+^k$$

$$\mathbb{I}[a^T\xi + b \le t] \ge \alpha + \beta^T\xi - \xi^T\Gamma\xi,\quad \forall\xi\in\mathbb{R}^k$$

[illegible] function by indicator function



$-\alpha+\beta^T\xi-\xi^T\Gamma\xi$

$$\Leftrightarrow \begin{cases} 1 \geq \alpha+\beta^T\xi-\xi^T\Gamma\xi, & \forall \xi\in\mathbb{R}^k \\ 0 \geq \alpha+\beta^T\xi-\xi^T\Gamma\xi, & \forall \xi: a^T\xi+b\geq t \quad (**) \end{cases}$$

change this from ">" to "≥" because the quadratic function is continuous.

$$(**) \Leftrightarrow \quad 0 \geq \begin{cases} \sup & \alpha+\beta^T\xi-\xi^T\Gamma\xi \\ \text{s.t.} & \xi\in\mathbb{R}^k \\ & a^T\xi+b\geq t \end{cases}$$

$$\overset{\text{dualize}}{\Leftrightarrow} \quad 0 \geq \sup_{\xi\in\mathbb{R}^k} \inf_{\theta\geq 0} \alpha+\beta^T\xi-\xi^T\Gamma\xi+\theta a^T\xi+\theta b-\theta t$$

$$0 \geq \inf_{\theta\geq 0} \sup_{\xi\in\mathbb{R}^k} \cdots$$

$$\Leftrightarrow \quad \exists\, \theta\geq 0 : \ 0 \geq \sup_{\xi\in\mathbb{R}^k} \alpha+\beta^T\xi-\xi^T\Gamma\xi+\theta a^T\xi+\theta b-\theta t$$

$$\Leftrightarrow \exists\, \alpha,\beta,\Gamma,\theta: \ \alpha+\beta^T\mu-\langle\Gamma,\Sigma+\mu\mu^T\rangle \geq 1-\epsilon$$

$$1\geq \alpha+\beta^T\xi-\xi^T\Gamma\xi, \ \forall\xi\in\mathbb{R}^k$$

$$0\geq \alpha+\beta^T\xi-\xi^T\Gamma\xi+\theta\cdot a(x)^T\xi+\theta\, b(x)-\theta t, \ \forall\xi\in\mathbb{R}^k$$

production of two decision variables! nonconvex

- Claim: $\theta>0$ for $\overset{\text{chance constraint}}{(CC)}$ to be satisfied.

suppose $\theta=0$: $\ 0\geq\alpha+\beta^T\xi-\xi^T\Gamma\xi \ \ \forall\xi\in\mathbb{R}^k$

take expectation of both sides:

$$0\geq \mathbb{E}_P[\alpha+\beta^T\xi-\xi^T\Gamma\xi] \quad =\langle\Gamma,\xi\xi^T\rangle$$

$$=\alpha+\beta^T\mu-\langle\Gamma,\mathbb{E}[\xi\xi^T]\rangle$$

$$\geq \alpha+\beta^T\mu-\langle\Gamma,\Sigma+\mu\mu^T\rangle. \quad (\text{since } \mathbb{E}_P[\xi\xi^T]\preceq\Sigma+\mu\mu^T)$$

$$\geq 1-\epsilon>0 \quad (\text{since } \epsilon<1)$$

$\Rightarrow$ replace $\theta\in\mathbb{R}_+$ with $\theta\in\mathbb{R}_{++}$

- Devide all the constraints by θ:

$$\Leftrightarrow \exists\, \alpha\in\mathbb{R},\ \beta\in\mathbb{R}^k,\ \Gamma\in S_+^k,\ \theta\in\mathbb{R}_{++}: \quad \frac{\alpha}{\theta}+\frac{\beta^T}{\theta}\mu-\langle\frac{\Gamma}{\theta},\Sigma+\mu\mu^T\rangle \geq \frac{1-\epsilon}{\theta}$$

$$\frac{1}{\theta}\geq\frac{\alpha}{\theta}+\frac{\beta^T}{\theta}\xi-\xi^T\frac{\Gamma}{\theta}\xi, \quad \forall\xi\in\mathbb{R}^k$$

$$0\geq\frac{\alpha}{\theta}+\frac{\beta^T}{\theta}\xi-\xi^T\frac{\Gamma}{\theta}\xi+a(x)^T\xi+b(x)-t, \quad \forall\xi\in\mathbb{R}^k$$

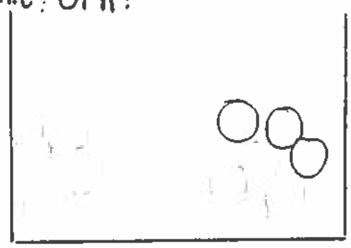
New decision variables:

$$\alpha'\leftarrow\frac{\alpha}{\theta}, \quad \beta'\leftarrow\frac{\beta}{\theta}, \quad \Gamma'\leftarrow\frac{\Gamma}{\theta}, \quad \theta'\leftarrow\frac{1}{\theta}$$

$$\Leftrightarrow \exists \alpha' \in \mathbb{R},\ \beta' \in \mathbb{R}^k,\ \Gamma' \in S_+^k,\ \theta' \in \mathbb{R}_{++} : \quad \begin{aligned} &\alpha' + \beta'^T \mu - \langle \Gamma', \Sigma + \mu\mu^T \rangle \geq (1-\epsilon)\theta' \\ &\theta' \geq \alpha' + \beta'^T \xi - \xi^T \Gamma' \xi, \quad \forall \xi \in \mathbb{R}^k \\ &0 \geq \alpha' + \beta'^T \xi - \xi^T \Gamma' \xi + a(x)^T \xi + b(x) - t, \quad \forall \xi \in \mathbb{R}^k \end{aligned}$$

$$\Leftrightarrow \exists \alpha' \in \mathbb{R},\ \beta' \in \mathbb{R}^k,\ \Gamma' \in S_+^k,\ \theta' \in \mathbb{R}_{++} : \quad \begin{aligned} &\alpha' + \beta'^T \mu - \langle \Gamma', \Sigma + \mu\mu^T \rangle \geq (1-\epsilon)\theta' \\ &\begin{bmatrix} \Gamma' & -\frac{1}{2}\beta' \\ (-\frac{1}{2}\beta')^T & \theta' - \alpha' \end{bmatrix} \succeq 0 \\ &\begin{bmatrix} \Gamma' & -\frac{1}{2}(\beta' + a(x)) \\ -\frac{1}{2}(\beta' + a(x))^T & t - b(x) - \alpha' \end{bmatrix} \succeq 0 \end{aligned}$$

Experiment : Urn :



Gamble A:
If you draw a red ball : \$100

Gamble B:
If you draw a blue ball : \$100

Gamble C:
If draw red or yellow : \$100

Gamble D:
If draw blue or yellow : \$100.

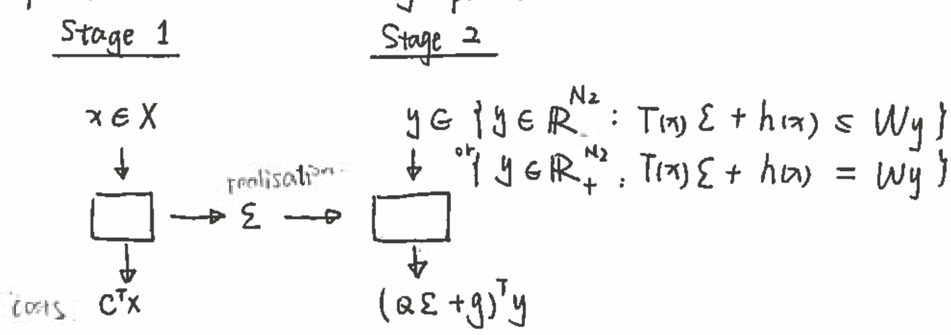
If $A \succ B$: $P(\text{draw} = \text{red}) > P(\text{draw} = \text{blue})$.
If $C \prec D$: $P(\text{draw} = \text{red}) < P(\text{draw} = \text{blue})$

$P_{red} = \frac{1}{3}$, $P_{blue} \in [0, \frac{2}{3}]$. ambiguity averse person : $A \succ B$.
$P_{red} + P_{yellow} \in [\frac{1}{3}, 1]$, $P_{blue} + P_{yellow} = \frac{2}{3}$. ambiguity averse person : $C \prec D$.

Lecture 9 Sept. 22

Two-stage models.

Sequential decision-making problems



$y(\Sigma)$ is a recourse / corrective action.

— $\inf\limits_{x \in X \subseteq \mathbb{R}^{N_1}} \; c^T x + \overset{\text{Arbitrary}}{\mathcal{R}}\left[Z(x, \tilde{\Sigma}) \right]$

where $\overset{\text{recourse function.}}{Z(x, \tilde{\Sigma})} = \inf \; (Q\Sigma + g)^T y$

$\quad$ s.t. $y \in \mathbb{R}_+^{N_2}$

$\quad\quad T(x)\Sigma + h(x) = W y$

$Q \in \mathbb{R}^{N_2 \times k}$

$W \in \mathbb{R}^{J \times N_2}$

$g \in \mathbb{R}^k$

$T(x) = \begin{bmatrix} T_1(x)^T \\ \vdots \\ T_J(x)^T \end{bmatrix}$

j-th row : $T_j(x) = T_j x + \hat{t}_j \qquad T_j \in \mathbb{R}^{K \times N_1}, \; \hat{t}_j \in \mathbb{R}^k$

$h(x) = Hx + \hat{h}, \quad H \in \mathbb{R}^{J \times N_1}, \; \hat{h} \in \mathbb{R}^J.$

— When $\mathcal{R}[\;] = \mathbb{E}[\;]$: $\inf\limits_{x \in X \subseteq \mathbb{R}^{N_1}} \; c^T x + \mathbb{E}\left[Z(x, \tilde{\Sigma}) \right]. \qquad$ (2S)

Example:

– Piecewise Linear Models.

$Z(x, \tilde{\Sigma}) = \inf \; y$

$\quad$ s.t. $y \in \mathbb{R}$

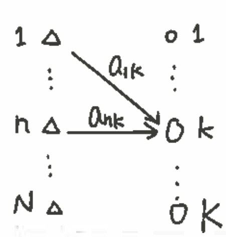
$\quad\quad a_j(x)^T \Sigma + b_j(x) \le y \qquad \forall j \in [J]$

$\quad \Leftrightarrow \max \; a_j(x)^T \Sigma + b_j(x) \le y.$

— Multi-product Assembly

K products.

N parts.



Data:

c_n per-unit cost of part $n \in [N]$

g_k per-unit selling price of product $k \in [K]$

a_{nk} number of units of part n required to produce 1 unit of product k.

Random Variable

$\tilde{\xi}_k$ demand of product $k \in [K]$

1st-stage decision

x_n number of units of part $n \in [N]$ to purchase

2nd-stage decision

$y_k(\xi)$ number of units of product $k \in [K]$ to produce.

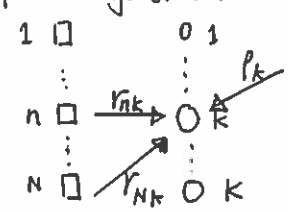
$$\min_{x \in \mathbb{R}_+^N} \; c^T x + E[z(x, \tilde{\xi})]$$

$$\text{where } z(x, \tilde{\xi}) = \min \; -g^T y$$
$$\text{s.t. } y \in \mathbb{R}_+^K$$
$$\sum_{k \in [K]} a_{nk} \cdot y_k \le x_n \quad , \forall n \in [N]$$
$$y_k \le \xi_k \quad , \forall k \in [K]$$

- Capacity expansion for electric power generator

N generators

K demand sites



Data

b bound on total generation capacity

c_n unit cost of installing capacity at generator n.

r_{nk} unit operating cost of sending energy from generator n to demand site k.

ρ_k unit subcontracting cost for demand site k

Random Variables

$\tilde{f}_n \in [0,1]$ available fraction of capacity for generator n.

$\tilde{d}_k$ demand site k

$\Rightarrow \tilde{\xi} = (\tilde{f}, \tilde{d})$

1st - stage decision

x_n capacity installed in generator n.

2nd - stage decisions

$y_{nk}(\xi)$ units of energy shipped from n to k.

$s_k(\xi)$ units of energy from subcontracting sent to demand site k.

$$\inf \; c^T x + \mathbb{E}[z(x,\tilde{\xi})]$$
$$\text{s.t.} \; x \in \mathbb{R}^N_+$$
$$e^T x \le b.$$
$$\text{where } z(x,\xi) = \inf \sum_{n\in[N]} \sum_{k\in[K]} r_{nk} y_{nk} + \sum_{k\in[K]} p_k s_k$$
$$\text{s.t. } y \in \mathbb{R}^{NK}_+ \;,\; s \in \mathbb{R}^K_+$$
$$\sum_{k\in[K]} y_{nk} \le f_n x_n \;,\; \forall n \in [N]$$
$$\sum_{n\in[N]} y_{nk} + s_k = d_k \;,\; \forall k \in [K]$$

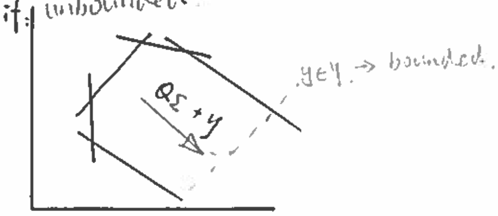
Properties of the recourse function.

$$z(x,\xi) = \inf \; (Q\xi + g)^T y$$
$$\text{s.t. } y \in \mathbb{R}^{N_2}_+ \quad y \in Y$$
$$T(x)\cdot\xi + h(x) = Wy.$$

$z(x,\xi) \in \mathbb{R} \cup \{-\infty, +\infty\}$

infeasible $z(x,\xi) = +\infty$, if $\{y \in \mathbb{R}^N_+ : T(x)\cdot\xi + h(x) = Wy\} = \emptyset$

$z(x,\xi) = -\infty$. if: unbounded



The recourse function induces a set of implicit constraints on x

$K = \{x \in \mathbb{R}^{N_1} : \mathbb{P}(\exists y \in \mathbb{R}^{N_2}_+ : T(x)\cdot\xi + h(x) = Wy) = 1\}$

K is the set of induced constraints.

- If $\mathbb{P}$ is discrete with S scenerios:

  $K = \{x \in \mathbb{R}^{N_1} : \forall s \in [S] \;.\; \exists y_s \in \mathbb{R}^{N_2}_+ : T(x)\cdot\xi^S + h(x) = W\cdot y_s\}$
- If $X \cap K = \emptyset$, then the problem is infeasible.

def (Relatively complete recourse): we say that (2S) has a relatively complete recourse, if $X \cap K = X$ or $X \subseteq K$

def (complete recourse): (25) has a complete recourse.
if $\{ Wy : y \in \mathbb{R}_+^{N_2} \} = \mathbb{R}^J$ span

Q: Suppose a problem doesn't have complete recourse, how to make it complete?
A:

$$Z(x, \xi) = \inf \ (Q\xi + g)^T y + \underbrace{\lambda e^T (z_+ + z_-)}_{\text{Penalty}}$$

λ: Large value

$$\text{s.t.} \quad y \in \mathbb{R}_+^{N_2}, \ z_+ \in \mathbb{R}_+^J, \ z_- \in \mathbb{R}_+^J$$
$$T(x) \cdot \xi + h(x) = Wy + z_+ - z_-$$

Lecture 10 Sept. 2

Recourse Problem: $Z(x, \xi) = \inf \ (Q\xi + g)^T y$ (RP)
$$\text{s.t.} \quad y \in \mathbb{R}_+^{N_2}$$
$$T(x) \xi + h(x) = Wy$$

Dual Problem: $Z_d(x, \xi) = \sup \ (T(x)\xi + h(x))^T \pi$ (RD)
$$\text{s.t.} \quad \pi \in \mathbb{R}^J$$
$$Q\xi + q \geq W^T \pi$$

Strong LP duality holds if either (RP) is feasible ($Z(x, \xi) < +\infty$)
or (RD) is feasible ($Z_d(x, \xi) > -\infty$)

- Lemma: $Z(x, \xi)$ is convex in x for any fixed $\xi \in \Xi$
  2 ways:
  1) partial infimum of (RP):
     $\{ (x, y) \in \mathbb{R}^{N_1} \times \mathbb{R}_+^{N_2} : T(x)\xi + h(x) = Wy \}$ is convex
     & $(Q\xi + g)^T y$ is jointly convex in (x, y)

  2) pointwise supremum of (RD)
     $(T(x)\xi + h(x))^T \pi$ is convex in x for any fixed π.

$\Rightarrow$ (25) is convex optimization problem.

- Lemma: if $Q = 0$, then $Z(x, \xi)$ is convex in ξ for any fixed $x \in X$.
  no uncertainty in obj.

$$Z(x, \xi) = \inf \ g^T y$$
$$\text{s.t.} \quad y \in \mathbb{R}^{N_2}$$
$$T(x)\xi + h(x) = Wy$$

$\{ (\xi, y) \in \mathbb{R}^k \times \mathbb{R}_+^{N_2} : T(x)\xi + h(x) = Wy \}$ is convex in ξ & y
& $g^T y$ is jointly convex in ξ & y.

$\Rightarrow$ <u>Expected value lower bound</u>

if $Q=0$, then $\mathbb{E}[\,Z(x,\tilde{\xi})\,] \geq Z(x, \mathbb{E}[\tilde{\xi}])$ for any fixed $x \in X$

by Jensen's inequality

$$x^*_{EV} \in \arg\inf \; c^T x + Z(x, \mathbb{E}[\tilde{\xi}])$$

$$EV = \inf_{x} \underbrace{\overset{\text{easy}}{c^T x + Z(x, \mathbb{E}[\tilde{\xi}])}}_{\text{Lower Bound}} \leq \inf_{x\in X} c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,]$$

$$\leq \underbrace{c^T x^*_{EV} + \mathbb{E}[\,Z(x^*_{EV}, \tilde{\xi})\,]}_{\text{Upper Bound}}$$

sometimes difficult

since x_{EV} is not optimal for (2S)

if infeasible, then $+\infty$ also upper bound

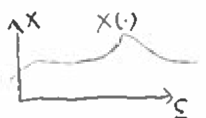
<u>Wait-and-see lower bound</u>

$$WS = \mathbb{E}\left[\inf_{x\in X} c^T x + Z(x, \tilde{\xi})\right]$$

make decision after observing ξ

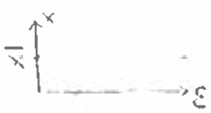
two-stage problem without first stage

$$= \mathbb{E}\left[\inf_{x(\tilde{\xi})\in X} c^T x(\tilde{\xi}) + Z(x(\tilde{\xi}), \tilde{\xi})\right]$$



$$= \inf_{x(\cdot),\, \Xi \to X} \mathbb{E}\left[c^T x(\tilde{\xi}) + Z(x(\tilde{\xi}), \tilde{\xi})\right]$$

try to find a function $x(\cdot)$



$$\leq \inf_{\substack{x(\cdot):\, \Xi \to X \\ x(\xi)=\bar{x},\ \forall \xi \in \Xi}} \mathbb{E}\left[c^T x(\tilde{\xi}) + Z(x(\tilde{\xi}), \tilde{\xi})\right]$$

$$= \inf_{\bar{x}\in X} c^T \bar{x} + \mathbb{E}[\,Z(\bar{x}, \tilde{\xi})\,] \quad (2S)$$

<u>How much is it worth to solve the real problem?</u> show in Report *

Value of stochastic solution

$$VSS(x) = \frac{c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,] - \left(\inf_{x\in X} c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,]\right)}{\inf_{x\in X} c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,]} \geq 0$$

normalize to percentage

$VSS(x^*_{EV})$

<u>What are you willing to pay to know the future?</u>

$$x^* \in \arg\inf_{x\in X} c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,] \quad (2S)$$

value of perfect information for a particular realization ξ:

$$VPI(\xi) = c^T x^* + Z(x^*, \xi) - \left(\inf_{x\in X} c^T x + Z(x,\xi)\right) \geq 0$$

Expected value of perfect information

$$EVPI = \mathbb{E}[\,VPI(\tilde{\xi})\,] = \inf_{x\in X} c^T x + \mathbb{E}[\,Z(x,\tilde{\xi})\,] - WS$$

How to solve (2S):

- $\mathbb{P}$ is discrete distribution with S scenarios

$$\inf \; c^T x + \mathbb{E}[Z(x,\tilde{\xi})]$$
$$= \inf_{x\in X} c^T x + \sum_{s\in[S]} p^s \, Z(x, \xi^s)$$

$$= \inf_{x\in X} c^T x + \sum_{s\in[S]} p^s \cdot \inf \{ (Q\xi^s + g)^T y^s : y^s \in \mathbb{R}_+^{N_2},\ T(x)\xi^s + h(x) = W y^s \}$$

$$= \inf \; c^T x + \sum_{s\in[S]} p^s (Q\xi^s + g^T y^s)$$ It is LP if Ξ is polytope

s.t. $x \in X$, $y^s \in \mathbb{R}_+^{N_2}$, $\forall s \in [S]$

$T(x)\xi^s + h(x) = W y^s$, $\forall s \in [S]$ computational complexity depends on S

- Later: Bender's decomposition algorithm.

- $\mathbb{P}$ is continuous distribution.
  1) Monte-Carlo: sample S scenarios from $\mathbb{P}$
  2) upper & lower bounds on (2S)

Naive bounds:

- Assume $Q = 0 \Rightarrow Z(x,\xi)$ is convex in ξ for any fixed x.
- Assume $\mathbb{E}[\tilde{\xi}] = \mu$

Ambiguity set $\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu\}$

$$\inf_{x\in X} c^T x + \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(x,\tilde{\xi})] \qquad \text{(NLB)}$$

$$\mathbb{E}_{\mathbb{P}}[Z(x,\tilde{\xi})] \geq Z(x, \mathbb{E}_{\mathbb{P}}[\tilde{\xi}]), \quad \forall \mathbb{P}\in\mathcal{P}$$

$$\Leftrightarrow \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(x,\tilde{\xi})] \underset{=}{\geq} Z(x,\mu)$$

Dirac distribution $\mathbb{P}^*$ at μ, s.t. $\mathbb{P}^*(\tilde{\xi} = \mu) = 1$ is feasible in $\mathcal{P}$

Lecture 11. Sept. 29

$$\inf_{x \in X} c^T x + \mathbb{E}[z(x,\tilde{\xi})]$$

where. $$z(x,\xi) = \inf \ (Q\xi + q)^T y \quad \text{s.t. } y \in \mathbb{R}^{N_2}, \ T(x)\xi + h(x) \le Wy$$

Naive bounds.

- Assume $Q = 0 \Rightarrow z(x,\xi)$ is convex in ξ for any fixed x
- Assume $\mathbb{E}[\xi] = \mu$

Lower bound (JENSEN'S BOUND)

$$\mathbb{E}[z(x,\tilde{\xi})] \ge \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[z(x,\tilde{\xi})] \qquad \mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu\}$$

$$\mathbb{E}_{\mathbb{P}}[z(x,\tilde{\xi})] \ge z(x, \mathbb{E}_{\mathbb{P}}[\tilde{\xi}]), \ \forall \mathbb{P} \in \mathcal{P}$$

$$\Leftrightarrow \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[z(x,\tilde{\xi})] \ge z(x,\mu)$$

Dirac distribution $\mathbb{P}^*$ at μ : $\mathbb{P}^*(\tilde{\xi} = \mu) = 1$ is feasible. $\mathbb{P}^* \in \mathcal{P}$

$$\Rightarrow \mathbb{E}_{\mathbb{P}^*}[z(x,\tilde{\xi})] = z(x,\mu)$$

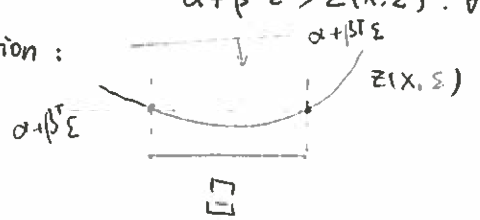
$$\Rightarrow \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[z(x,\tilde{\xi})] = z(x,\mu)$$

$$\inf_{x \in X} c^T x + \mathbb{E}[z(x,\tilde{\xi})] \ge \inf_{x \in X} c^T x + z(x,\mu) = EV$$

Upper bound. (EDMUNSON-MADANSKY'S BOUND)

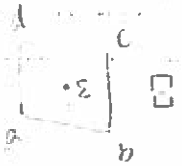
$$\begin{aligned}
\mathbb{E}[z(x,\tilde{\xi})] &\le \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[z(x,\tilde{\xi})] \\
&= \sup \int z(x,\xi)\, v(d\xi) \\
&\quad \text{s.t. } v(\cdot) \ge 0 \\
&\quad\quad \int v(d\xi) = 1, \ \int \xi\, v(d\xi) = \mu \\
&= \inf \ \alpha + \beta^T \mu \\
&\quad \text{s.t. } \alpha \in \mathbb{R}, \ \beta \in \mathbb{R}^k \\
&\quad\quad \alpha + \beta^T \xi \ge z(x,\xi), \ \forall \xi \in \Xi
\end{aligned}$$

Iterpretation :



Let $\text{ext}(\Xi)$ be the set of extreme points of Ξ. Assume Ξ is bounded
for any $\xi \in \Xi$. we can find $p_e(\xi)$, $\forall e \in \text{ext}(\Xi)$, such that

$$1 = \sum_{e \in \text{ext}(\Xi)} p_e(\xi) \quad \text{and} \quad \xi = \sum_{e \in \text{ext}(\Xi)} p_e(\xi)\, e$$



Ξ $\quad \varepsilon = P_a(\varepsilon)\, a + P_b(\varepsilon)\, b + P_c(\varepsilon)\, c + P_d(\varepsilon)\, d$

__Theorem__. $\mathbb{E}[Z(x, \tilde{\varepsilon})] \leq \sum_{e \in \text{ext}(\Xi)} \mathbb{E}[P_e(\tilde{\varepsilon})]\, Z(x, e)$ prob. dist. supported on extreme poi

__Proof:__ for any $\varepsilon \in \Xi$, we have

$$Z\Big(x, \sum_{e \in \text{ext}(\Xi)} P_e(\varepsilon)\, e\Big) \leq \sum_{e \in \text{ext}(\Xi)} P_e(\varepsilon)\, Z(x, e)$$ (Jensen's ineq.)

Taking expectations on both sides,

$$\mathbb{E}[Z(x, \tilde{\varepsilon})] \leq \mathbb{E}\Big[\sum_{e \in \text{ext}(\Xi)} P_e(\tilde{\varepsilon})\, Z(x, e)\Big]$$

$$= \sum_{e \in \text{ext}(\Xi)} \mathbb{E}[P_e(\tilde{\varepsilon})]\, Z(x, e)$$

__Example 1__ say $\varepsilon \in [a, b] \in \mathbb{R}$ (one dimension) $\quad \text{ext}(\Xi) = \{a, b\}$

then $P_a(\varepsilon) = \dfrac{b - \varepsilon}{b - a}$, $P_b(\varepsilon) = \dfrac{\varepsilon - a}{b - a}$. $P_a(\varepsilon) + P_b(\varepsilon) = \dfrac{b-a}{b-a} = 1$

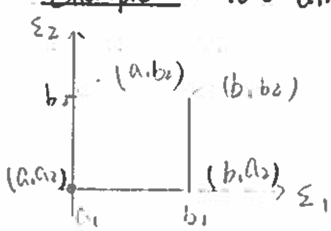
$$P_a(\varepsilon) \cdot a + P_b(\varepsilon) \cdot b = (ab - a\varepsilon + b\varepsilon - ab)/(b - a) = \varepsilon$$

$$\mathbb{E}[Z(x, \tilde{\varepsilon})] \leq \mathbb{E}[P_a(\tilde{\varepsilon})]\, Z(x, a) + \mathbb{E}[P_b(\tilde{\varepsilon})]\, Z(x, b)$$

$$= \left(\frac{b - \mu}{b - a}\right) Z(x, a) + \left(\frac{\mu - a}{b - a}\right) Z(x, b)$$

$$\inf_{x \in X} c^T x + \mathbb{E}[Z(x, \tilde{\varepsilon})] \leq \inf_{x \in X} c^T x + \left(\frac{b - \mu}{b - a}\right) Z(x, a) + \left(\frac{\mu - a}{b - a}\right) Z(x, b)$$

__Example 2__ Two-dimension ε



$$P_{(a_1, a_2)}(\varepsilon) = \left(\frac{b_1 - \varepsilon_1}{b_1 - a_1}\right)\left(\frac{b_2 - \varepsilon_2}{b_2 - a_2}\right)$$

$$P_{(b_1, a_2)}(\varepsilon) = \left(\frac{\varepsilon_1 - a_1}{b_1 - a_1}\right)\left(\frac{b_2 - \varepsilon_2}{b_2 - a_2}\right)$$

$$P_{(a_1, b_2)}(\varepsilon) = \left(\frac{b_1 - \varepsilon_1}{b_1 - a_1}\right)\left(\frac{\varepsilon_2 - a_2}{b_2 - a_2}\right)$$

Difficult to compute [?] since high dimension

$$P_{(b_1, b_2)}(\varepsilon) = \left(\frac{\varepsilon_1 - a_1}{b_1 - a_1}\right)\left(\frac{\varepsilon_2 - a_2}{b_2 - a_2}\right)$$

$$\mathbb{E}[Z(x, \tilde{\varepsilon})] \leq \mathbb{E}[P_{(a_1, a_2)}(\tilde{\varepsilon})] \cdot Z\left(x, \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right)$$
$$+ \cdots$$
$$+ \mathbb{E}[P_{(b_1, b_2)}(\tilde{\varepsilon})] \cdot Z\left(x, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right)$$

If ε_1 & ε_2 are independent, then $\mathbb{E}[P_{(a_1, a_2)}(\tilde{\varepsilon})] = \left(\dfrac{b_1 - \mu_1}{b_1 - a_1}\right)\left(\dfrac{b_2 - \mu_2}{b_2 - a_2}\right)$

# of extreme points scale with dimension $n \Rightarrow 2^n$

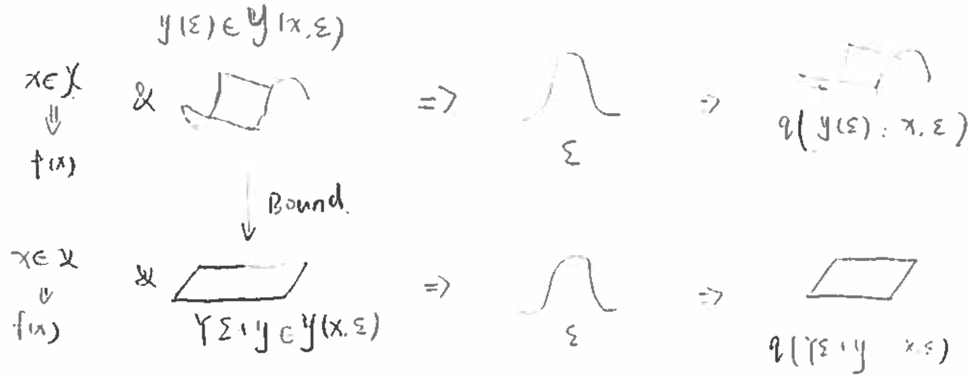
If $\Xi \subseteq \mathbb{R}^k$ is a convex hull of $K+1$ points.

$$\Xi = \text{conv}\{u_1, u_2, \cdots u_{K+1}\}$$

$$\sum_{k=1}^{K+1} P_k = 1 \quad , \quad \sum_{k=1}^{K+1} P_k u_k = \mu \in \mathbb{R}^K$$

$K+1$ constraints. unique $P_1, \cdots P_{K+1}$

Lecture 12. Oct 4



<u>Linear Decision Rules Bounds</u> (no assumption on $Q = 0$, but assume $\mathbb{E}[\tilde{\xi}] = \mu$, $\mathbb{E}[\tilde{\xi}\tilde{\xi}^T] = \Sigma + \mu\mu^T = \Omega$)

<u>Upper Bound</u>

$$(*) \quad \inf_{x \in X} \; c^T x + \mathbb{E}\left[\inf\{(Q\tilde{\xi} + g)^T y : y \in \mathbb{R}^{N_2}, \; T(x)\tilde{\xi} + h(x) \le W y\}\right]$$

after realization of ξ we decide y, so treat y as function of ξ

$$= \inf \; c^T x + \mathbb{E}\left[(Q\tilde{\xi} + g)^T \cdot y(\tilde{\xi})\right]$$

s.t. $x \in X$ $\quad y : \Xi \to \mathbb{R}^{N_2}$

Polytope

$y(\xi) \in \mathbb{R}^{N_2}$, $\forall \xi \in \Xi = \{\xi \in \mathbb{R}^k : S\xi \le t\}$

$T(x)\xi + h(x) \le W y(\xi)$

$y(\xi) = Y\xi + y$ it is a subset of all possible $y(\xi)$

$$\le \inf \; c^T x + \mathbb{E}\left[(Q\tilde{\xi} + g)^T (Y\tilde{\xi} + y)\right]$$

s.t. $x \in X$, $Y \in \mathbb{R}^{N_2 \times k}$, $y \in \mathbb{R}^{N_2}$

$T(x)\xi + h(x) \le W \cdot (Y\xi + y)$, $\forall \xi \in \Xi$

$$\begin{aligned}
\text{– Expectation: } & \mathbb{E}\left[\tilde{\xi}^T Q^T Y \tilde{\xi} + g^T Y \tilde{\xi} + \tilde{\xi}^T Q^T y + g^T y\right] \\
&= \mathbb{E}[\underbrace{\tilde{\xi}^T Q^T Y \tilde{\xi}}_{\text{scalar}}] + g^T Y \mu + \mu^T Q^T y + g^T y \\
&= \mathbb{E}[\mathrm{tr}(\tilde{\xi}^T Q^T Y \tilde{\xi})] + \cdots \\
&= \mathbb{E}[\mathrm{tr}(Q^T Y \tilde{\xi}\tilde{\xi}^T)] + \cdots \\
&= \mathrm{tr}(Q^T Y \,\mathbb{E}[\tilde{\xi}\tilde{\xi}^T]) + \cdots \\
&= \mathrm{tr}(Q^T Y \Omega) + g^T Y \mu + \mu^T Q^T y + g^T y
\end{aligned}$$

finite optimization

$$T(x) = \begin{bmatrix} T_1(x)^T \\ \vdots \\ T_J(x)^T \end{bmatrix} \qquad W = \begin{bmatrix} W_1^T \\ \vdots \\ W_J^T \end{bmatrix}$$

— Constraints: $T_j(x)^T \xi + h_j(x) \le W_j^T (Y\xi + y)$, $\forall j \in [J]$, $\forall \xi \in \Xi$

$\Leftrightarrow h_j(x) - W_j^T y \le (W_j^T Y - T_j(x)^T)\xi$, $\forall j \in [J]$, $\forall \xi \in \Xi$

$\Leftrightarrow h_j(x) - W_j^T y \le \inf_{S\xi \le t} (W_j^T Y - T_j(x)^T)\xi$ $\quad \forall j \in [J]$

$\Leftrightarrow h_j(x) - W_j^T y \le \sup_{\theta_j \ge 0} -\theta_j^T t$ s.t. $Y^T W_j + S^T \theta_j = T_j(x)$ $\quad \forall j \in [J]$

$\Leftrightarrow \exists \theta_j \in \mathbb{R}_+^m$: $h_j(x) - W_j^T y \le -\theta_j^T t$, $Y^T W_j + S^T \theta_j = T_j(x)$ $\quad \forall j \in [J]$

(*) $\Leftrightarrow$ $\le \inf\ c^T x + \mathrm{tr}(\Theta^T Y \Omega) + g^T Y \mu + \mu^T Q^T y + g^T y$

s.t. $x \in X$, $Y \in \mathbb{R}^{N_2 \times k}$, $y \in \mathbb{R}^{N_2}$, $\theta_j \in \mathbb{R}_+^m$, $\forall j \in [J]$

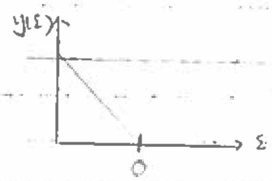
$Y^T W_j + S^T \theta_j = T_j(x)$, $\forall j \in [J]$

$h_j(x) - W_j^T y \le -\theta_j^T t$, $\forall j \in [J]$

Interpretation: $\inf\ [y(\xi)]$

where $y(\xi) \ge \xi$, $\forall \xi \in \Xi$

$y(\xi) \ge -\xi$, $\forall \xi \in \Xi$



Lower Bound (Assume $\dim(\Xi) = \mathbb{R}^k$)

Dual ※: $z_d(x, \xi) = \sup\ (T(x)\xi + h(x))^T \pi$

s.t. $\pi \in \mathbb{R}_+^J$

$Q\xi + g = W^T \pi$

For any fixed x: $\mathbb{E}[z(x, \tilde{\xi})] = \mathbb{E}[z_d(x, \xi)]$

$= \mathbb{E}[\sup\{(T(x)\tilde{\xi} + h(x))^T \pi : \pi \in \mathbb{R}_+^J, Q\tilde{\xi} + g = W^T \pi\}]$

$= \sup\ \mathbb{E}[(T(x)\tilde{\xi} + h(x))^T \pi(\tilde{\xi})]$

s.t. $\pi(\xi) \in \mathbb{R}_+^J$ $\quad \forall \xi \in \Xi$

$Q\xi + g = W^T \pi(\xi)$, $\forall \xi \in \Xi$

$\ge \sup\ \mathbb{E}[(T(x)\tilde{\xi} + h(x))^T (\Pi \tilde{\xi} + \rho)]$

s.t. $\Pi \in \mathbb{R}^{J \times k}$ $\quad \rho \in \mathbb{R}^J$

$Q\xi + g = W^T(\Pi\xi + \rho)$ $\quad \forall \xi \in \Xi$ $\quad$ (1)

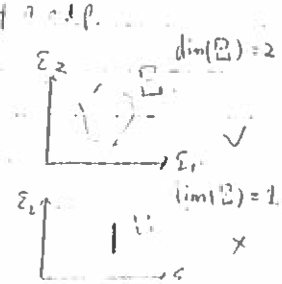
$\Pi\xi + \rho \ge 0$ $\quad \forall \xi \in \Xi$ $\quad$ (2)

Objective: $\mathbb{E}[\tilde{\xi}^T T(x)^T \Pi \tilde{\xi} + h(x)^T \Pi \tilde{\xi} + \tilde{\xi}^T T(x)^T \rho + h(x)^T \rho]$ linear function of Π and ρ.

$= \mathrm{tr}(T(x)^T \Pi (\Sigma + \mu\mu^T)) + h(x)^T \Pi \mu + \mu^T T(x)^T \rho + h(x)^T \rho$

(1) Constraints: $Q\xi + g = W^T \Pi \xi + W^T \rho$ $\quad \forall \xi \in \Xi$

assume $\dim(\Xi) = K$, meaning that Ξ has an interior



Dualization.

$$\begin{aligned}
&\min \quad c^T x \\
&\text{s.t.} \quad x \in \mathbb{R}^N \\
&\qquad A_1 x \le b_1 \qquad \theta_1 \ge 0 \\
&\qquad A_2 x = b_2 \qquad \theta_2
\end{aligned}$$

$$= \min_x \left\{ \max_{\substack{\theta_1 \ge 0 \\ \theta_2}} c^T x + \theta_1^T A_1 x - \theta_1^T b_1 + \theta_2^T b_2 - \theta_2^T A_2 x \right\}$$

$$= \max_{\substack{\theta_1 \ge 0 \\ \theta_2}} -\theta_1^T b_1 + \theta_2^T b_2 + \min_x c^T x + \theta_1^T A_1 x - \theta_2^T A_2 x$$

$$= \max_{\substack{\theta_1 \ge 0 \\ \theta_2}} -\theta_1^T b_1 + \theta_2^T b_2 \quad \text{s.t.} \quad c^T + \theta_1^T A = \theta_2^T A_2$$

Explain Transformation of (*) in Upper Bound.

$$\inf \; c^T x + \mathbb{E}\left[\inf\{(Q\hat{\xi}+g)^T y : y \in \mathbb{R}^{N_2},\ T(x)\hat{\xi} + h(x) \le Wy\}\right]$$

$$\begin{aligned}
= \inf \;& c^T x + \mathbb{E}\left[(Q\hat{\xi}+g)^T y(\hat{\xi})\right] \\
\text{s.t.}\;& x \in X,\ y : \Xi \to \mathbb{R}^{N_2} \\
& T(x)\xi + h(x) \le W y(\xi),\ \forall \xi \in \Xi
\end{aligned}$$

$\mathbb{P}$ is a discrete dist. with S scenarios, $\Xi = \{\xi^1, \cdots, \xi^S\}$

$$\inf \; c^T x + \sum_{s \in [S]} p_s \cdot \inf\{(Q\hat{\xi}^s + g)^T y^s : y^s \in \mathbb{R}^{N_2},\ T(x)\xi^s + h(x) \le W y^s\}$$

$$\begin{aligned}
= \inf \;& c^T x + \sum_{s\in[S]} p_s (Q\xi^s + g)^T y^s \\
\text{s.t.}\;& x \in X,\ y : \{\xi^1, \cdots \xi^S\} \to \mathbb{R}^{N_2} \\
& T(x)\xi^s + h(x) \le W y^s,\ \forall s \in [S]
\end{aligned}$$

Continue with Lower Bound.

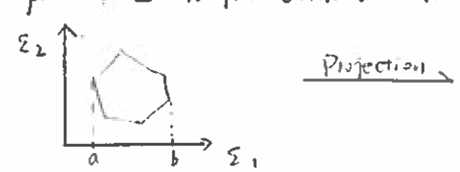
Constraint (1) $\Rightarrow$ $\underline{(Q - W^T\pi)}\,\xi = \underline{W^T\rho - g}$, $\forall \xi \in \Xi$

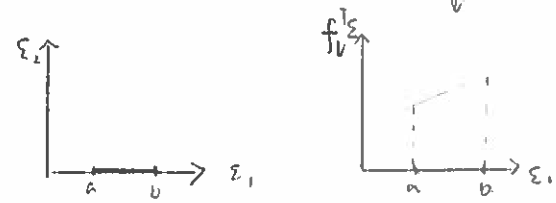
$$F \quad \xi = g,\ \forall \xi \in \Xi,\ F = \begin{bmatrix} f_1^T \\ \vdots \\ f_L^T \end{bmatrix}$$

each row: $f_l^T \xi = g_l$, $\forall \xi \in \Xi$ ☆

In this case.
☆ will not holds for $\forall \xi$ in interval $[a,b]$

Assumption: Ξ is full-dimensional.





Since Ξ is full-dimensional → the interval after projection must be an interval with positive length. → We want $f_l^T \xi = g_l$ to satisfy for $\forall \xi$ in the interval
→ Cannot hold in general case, only if $f_l^T = g_l = 0$
(If Ξ isn't full-dimensional → interval will be a point ! ...)

So the costraint (1) ⟺ $Q = W^T\pi$, $W^T\rho = g$

$\pi = \begin{bmatrix} \pi_1^T \\ \vdots \\ \pi_J^T \end{bmatrix}$ Constraint (3) : $\pi\varepsilon + \rho \geq 0, \ \forall \varepsilon \in \Xi$

$\iff \inf_{\varepsilon:\ S\varepsilon \leq t} \ \pi_j^T \varepsilon + \rho_j \geq 0, \ \forall j \in [J]$

$\iff \left\{ \sup_{\eta_j \geq 0} \ \eta_j^T t + \rho_j \ \ \text{s.t.} \ -\pi_j = S^T \eta_j \right\} \geq 0, \ \forall j \in [J]$

$\geq \sup \ tr\left(T(x)^T \pi \left(\Sigma + \mu\mu^T \right) \right) + h(x)^T \pi \mu + \mu^T T(x)^T \rho + h(x)^T \rho$

s.t. $\pi \in \mathbb{R}^{J \times K}$, $\rho \in \mathbb{R}^J$, $\eta_j \in \mathbb{R}_+^m$, $\forall j \in [J]$

$Q = W^T \pi$, $W^T \rho = g$

$\rho_j \geq \eta_j^T t$

$-\pi_j = S^T \eta_j$

Dualize and Combine with first stage problem. $= \inf \ tr(Q^T M) + g^T m + C^T x$

s.t. $M \in \mathbb{R}^{N_2 \times K}$, $m \in \mathbb{R}^{N_2}$, $r \in \mathbb{R}_+^J$, $R \in \mathbb{R}^{J \times K}$ (LLB)

$x \in X$

$T(x)\mu + h(x) + r = Wm$

$T(x)\left(\Sigma + \mu\mu^T \right) + h(x)\mu^T + R = WM$

$tr^T \geq SR^T$

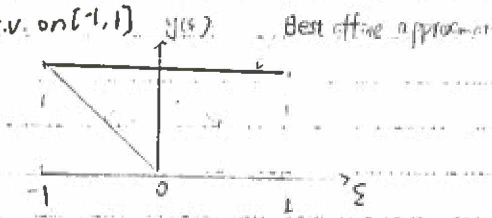
- Example : two stage problem without the first stage.

$\inf \ \mathbb{E}[y(\tilde{\xi})]$ — $\tilde{\xi}$ is uniform r.v. on $[-1,1]$

s.t. $y(\xi) \in \mathbb{R}$

$y(\xi) \geq \xi, \ \forall \xi \in \Xi = [-1, 1]$

$y(\xi) \geq -\xi, \ \forall \xi \in \Xi$



$\mathbb{E}[y(\tilde{\xi})] = 2\int_0^1 \xi \frac{1}{2} d\xi = \xi^2 \Big|_0^1 = \frac{1}{2}$

$\mathbb{E}[1] = 1$

$= \mathbb{E}[\max\{\tilde{\xi}, -\tilde{\xi}\}]$

$= \mathbb{E}[|\tilde{\xi}|]$

Lecture 14: Oct 11

$\Leftrightarrow$ $\inf \quad c^Tx + \mathbb{E}\left[(Q\tilde{\xi}+g)^T y(\tilde{\xi}) \right]$ add slack variable.

s.t. $\quad x \in X, \; y(\xi) \in \mathbb{R}^{N_2}, \; s(\xi) \in \mathbb{R}^J_+, \; \forall \xi \in \Xi$

$\quad T(x)\xi + h(x) + s(\xi) = W y(\xi), \; \forall \xi \in \Xi$

$\Leftrightarrow$ Introduce new variables and add redundant constraints:

$\inf \quad c^Tx + \underbrace{\mathbb{E}\left[(Q\tilde{\xi}+g)^T y(\tilde{\xi}) \right]}$

s.t. $\quad x \in X, \; y(\xi) \in \mathbb{R}^{N_2}, \; s(\xi) \in \mathbb{R}^J_+, \; \forall \xi \in \Xi$

$\quad m \in \mathbb{R}^{N_2}, \; M \in \mathbb{R}^{N_2 \times K}, \; r \in \mathbb{R}^J_+, \; R \in \mathbb{R}^{J \times K}$

$\quad T(x)\xi + h(x) + s(\xi) = W y(\xi), \; \forall \xi \in \Xi$

$\quad \mathbb{E}[y(\tilde{\xi})] = m, \; \mathbb{E}[y(\tilde{\xi})\tilde{\xi}^T] = M$

$\quad \mathbb{E}[s(\tilde{\xi})] = r, \; \mathbb{E}[s(\tilde{\xi})\tilde{\xi}^T] = R$

$\mathbb{E}\left[(Q\tilde{\xi}+g)^T y(\tilde{\xi}) \right]$ $\quad \xi^T Q^T y(\xi) = tr(Q^T y(\xi)\xi^T)$

$= \mathbb{E}\left[tr(Q^T y(\tilde{\xi})\tilde{\xi}^T) + g^T y(\tilde{\xi}) \right]$

$= tr(Q^T \mathbb{E}[y(\tilde{\xi})\tilde{\xi}^T]) + g^T \mathbb{E}[y(\tilde{\xi})]$

$= tr(Q^T M) + g^T m$

The objective becomes: $\inf \; tr(Q^T M) + g^T m + c^T x$, same as lower bound

$\Leftrightarrow$ $\inf \quad tr(Q^T M) + g^T m + c^T x$

s.t. $\quad x \in X, \; y(\xi) \in \mathbb{R}^{N_2}, \; s(\xi) \in \mathbb{R}^J_+, \; \forall \xi \in \Xi$

$\quad m \in \mathbb{R}^{N_2}, \; M \in \mathbb{R}^{N_2 \times K}, \; r \in \mathbb{R}^J_+, \; R \in \mathbb{R}^{J \times K}$

$\quad T(x)\xi + h(x) + s(\xi) = W y(\xi), \; \forall \xi \in \Xi \quad (\checkmark)$

$\quad T(x)\xi\xi^T + h(x)\xi^T + s(\xi)\xi^T = W y(\xi)\xi^T, \; \forall \xi \in \Xi \quad (\checkmark\checkmark)$ Because of "=" we can multiply ξ^T

$\quad s(\xi)(t - S\xi)^T \geq 0, \; \forall \xi \in \Xi \quad (\checkmark\checkmark\checkmark)$ — Because $s(\xi) \geq 0, \forall \xi \in \Xi = \{\xi \in \mathbb{R}^K, S\xi \leq t\}$

$\quad \mathbb{E}[y(\tilde{\xi})] = m, \; \mathbb{E}[y(\tilde{\xi})\tilde{\xi}^T] = M.$

$\quad \mathbb{E}[s(\tilde{\xi})] = r, \; \mathbb{E}[s(\tilde{\xi})\tilde{\xi}^T] = R.$

Take expectation of $(\checkmark)$ $\Leftrightarrow$ $T(x)\mu + h(x) + r = Wm$

Take expectation of $(\checkmark\checkmark)$ $\Leftrightarrow$ $T(x)(\Sigma + \mu\mu^T) + h(x)\mu^T + R = WM$

Take expectation of $(\checkmark\checkmark\checkmark)$ $\Leftrightarrow$ $tr^T \geq SR^T$

Other two become redundant, we get the lower bound again!

Monte Carlo Sampling Method

$$\inf_{x \in X} \mathbb{E}_{\mathbb{P}}[f_0(x, \tilde{\xi})]$$

f_0 → General random function.

For any fixed $x \in X$, computing $\mathbb{E}[f_0(x, \tilde{\xi})]$ is difficult due to multidimensional integration.

$$\mathbb{E}_{\mathbb{P}}[f_0(x, \tilde{\xi})] = \int_{\mathbb{R}^k} f_0(x, \xi)\, \mathbb{P}(d\xi)$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} f_0(x, \xi)\, p(\xi)\, d\xi_1 \cdots d\xi_k$$

$p(\xi)$ → density of ξ

How to evaluate the expectation?

- analytical integration (exact but rarely possible in practice)
- numerical integration (very accurate approximation, deterministic error bounds, limited to $K \leq 3$)
- Monto Carlo method (probabilistic error bounds, converges slowly $\propto 1/\sqrt{S}$ ← # of samples but rate is independent of K)

Def (almost surely): an event A happens almost surely (a.s. or $\mathbb{P}$-a.s.) if $\mathbb{P}(A) = 1$.

Assume: $\mathbb{E}[f_0(x, \tilde{\xi})^2] < +\infty$, $\forall x \in X$, $\tilde{\xi} \sim \mathbb{P}$

Fix $x \in X$: let $V(x) = \mathbb{E}[f_0(x, \tilde{\xi})]$ & $\sigma^2(x) = \mathrm{Var}(f_0(x, \tilde{\xi})) = \mathbb{E}[f_0(x, \tilde{\xi})^2] - \mathbb{E}[f_0(x, \tilde{\xi})]^2 < +\infty$

Let $\xi^1, \cdots, \xi^S$ be iid samples of $\mathbb{P}$.

We approximate $V(x)$ by $\overset{\text{r.v.}}{V_S(x)} = \frac{1}{S} \sum_{s \in [S]} f_0(x, \xi^s)$. randomness come from ξ^s

We have for any fix $x \in X$.

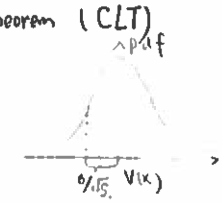
1) $\mathbb{E}[V_S(x)] = V(x)$. $V_S(x)$ is an unbiased estimator of $V(x)$.
2) $\lim_{S \to \infty} V_S(x) = V(x)$ as Law of Large Number (LLN)
3) $\sqrt{S}(V_S(x) - V(x)) \xrightarrow{d} N(0, \sigma^2(x))$ by Central Limit Theorem (CLT)

   d → convert in distribution

   $\Leftrightarrow V_S(x) - V(x) \xrightarrow{d} N(0, \sigma^2(x)/S)$

   Let $\tilde{\chi} \sim N(0,1)$

   then for large enough S, we have $V_S(x) \approx V(x) + \frac{\sigma}{\sqrt{S}} \tilde{\chi}$



rate of convergence is $\sqrt{S}$. (with 100x samples, you can improve the accuracy by 10x

$$\mathbb{P}\left(|V_S(x) - V(x)| \leq \frac{\sigma}{\sqrt{S}}\right) \geq 1 - \alpha$$

1) $\lim_{S\to\infty} \inf_{x\in X} V_S(x) \xrightarrow{?} \inf_{x\in X} V(x)$ a.s.

2) $\sqrt{S}\left(\left[\inf_{x\in X} V_S(x) \right] - \left[\inf_{x\in X} V(x) \right] \right) \xrightarrow{d?} \mathcal{N}(0, \sigma^2)$

3) $\lim_{S\to\infty} \arg\inf_{x\in X} V_S(x) \xrightarrow{?} \arg\inf_{x\in X} V(x)$ a.s.

Example:

$$\inf_{x\in[-1,1]} \mathbb{E}[f_0(x,\tilde{\xi})]$$

where $f_0(x,\tilde{\xi}) = \xi x$.

x is scalar, $\tilde{\xi} \sim N(0,1)$

$$= \inf_{x\in[-1,1]} \mathbb{E}[\tilde{\xi}]\, x$$
$$= 0$$

$$\inf_{x\in[-1,1]} \frac{1}{S} \sum_{s\in[S]} \xi^s x$$

$$= \inf_{x\in[-1,1]} \left(\frac{1}{S} \sum_{s\in[S]} \xi^s \right) x$$

$$= - \left| \frac{1}{S} \sum_{s\in[S]} \xi^s \right|$$

$$x_S^* = \begin{cases} -1, & \text{if } \frac{1}{S}\sum_{s\in[S]} \xi^s > 0 \\ [-1,1] & = 0 \\ 1, & \text{if } \frac{1}{S}\sum_{s\in[S]} \xi^s < 0 \end{cases}$$

$\xi^s \sim N(0,1)$

$\sum_{s\in[S]} \xi^s \sim N(0,S)$

$\frac{1}{S}\sum_{s\in[S]} \xi^s \sim N(0,\frac{1}{S})$

lecture 15. Oct. 13

Conclude from previous example:

1) $\mathbb{E}\left[\inf_{x\in X} V_S(x)\right] \leq \inf_{x\in X} V(x)$ (negative bias)

2) $\lim_{S\to\infty} \inf_{x\in X} V_S(x) = \inf V(x)$ a.s. (Consistency of $\inf_{x\in X} V_S(x)$)

3) $\sqrt{S}\left(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)\right) \xrightarrow{d} -|N(0,1)|$ (non-normal errors)

Error will reduce as rate $\sqrt{S}$

4) $\lim_{S\to\infty} \operatorname{arginf}_{x\in X} V_S(x) \subseteq \operatorname{arginf}_{x\in X} V(x)$ (consistency of x_S^*)

<u>Theorem</u>: $\mathbb{E}\left[\inf_{x\in X} V_S(x)\right] \leq \inf_{x\in X} V(x)$.

Proof: $\mathbb{E}\left[\inf_{x\in X} V_S(x)\right] \leq \inf_{x\in X} \mathbb{E}\left[V_S(x)\right] = \inf_{x\in X} V(x)$

<u>Theorem</u>: $\mathbb{E}\left[\inf_{x\in X} V_{S+1}(x)\right] \geq \mathbb{E}\left[\inf_{x\in X} V_S(x)\right]$ increase S, converge to exact value

Proof: $V_{S+1}(x) = \frac{1}{S+1} \sum_{s\in[S+1]} f_0(x, \tilde{\xi}^s)$

$= \frac{1}{S+1} \sum_{s\in[S+1]} \left[\frac{1}{S} \sum_{\substack{t\neq s \\ t\in[S+1]}} f_0(x, \tilde{\xi}^t)\right]$

$a_1, \dots, a_{S+1}$
S
$a_1 \dots a_{S+1}$

$$\mathbb{E}\left[\inf_{x\in X} V_{S+1}(x)\right] = \mathbb{E}\left[\underline{\inf_{x\in X}} \frac{1}{S+1} \sum_{s\in[S+1]} \left[\frac{1}{S} \sum_{\substack{t\neq s \\ t\in[S+1]}} f_0(x, \tilde{\xi}^t)\right]\right]$$

$$\geq \mathbb{E}\left[\frac{1}{S+1} \sum_{s\in[S+1]} \left[\underline{\inf_{x\in X}} \frac{1}{S} \sum_{\substack{t\neq s \\ t\in[S+1]}} f_0(x^s, \tilde{\xi}^t)\right]\right]$$

$$= \frac{1}{S+1} \sum_{s\in[S+1]} \mathbb{E}\left[\inf_{x\in X} \frac{1}{S} \sum_{\substack{t\neq s \\ t\in[S+1]}} f_0(x^s, \tilde{\xi}^t)\right]$$

$$= \frac{1}{S+1} (S+1)\, \mathbb{E}\left[\inf_{x\in X} V_S(x)\right]$$

$$= \mathbb{E}\left[\inf_{x\in X} V_S(x)\right]$$

LLN / pointwise convergence: $\lim_{S\to\infty} |V_S(x) - V(x)| = 0$ a.s. $\forall x\in X$ $\not\Rightarrow$ $\lim_{S\to\infty} \inf_{x\in X} V_S(x) = \inf_{x\in X} V(x)$. a.s.

Explain: $\inf_{x\in\mathbb{R}} \mathbb{E}\left[f_0(x, \tilde{\xi})\right] = 0$ In previous example.

$\inf_{x\in\mathbb{R}} \left(\frac{1}{S} \sum \varepsilon\right) x = -\infty$

replace $\mathbb{E}[\cdot]$ with $\mathbb{R}$

A sufficient condition is <u>uniform convergence</u>:

$$\lim_{S\to\infty} \sup_{x\in X} |V_S(x) - V(x)| = 0 \quad \text{a.s.} \quad \text{(ULLN)}$$

<u>Theorem</u> i): If (ULLN) holds then $\lim_{S\to\infty} \inf_{x\in X} V_S(x) = \inf_{x\in X} V(x)$. a.s.

Proof: Let $x_s^* \in \underset{x\in X}{\arg\inf}\, V_s(x)$ & $x^* \in \underset{x\in X}{\arg\inf}\, V(x)$

$$\begin{aligned}
\left|\inf_{x\in X} V_s(x) - \inf_{x\in X} V(x)\right| &= \left|V_s(x_s^*) - V(x^*)\right| \\
&= \max\{V_s(x_s^*) - V(x^*),\ V(x^*) - V_s(x_s^*)\} \\
&\le \max\{V_s(x^*) - V(x^*),\ V(x_s^*) - V_s(x_s^*)\} \\
&\le \max\{|V_s(x^*) - V(x^*)|,\ |V(x_s^*) - V_s(x_s^*)|\} \\
&= \max_{x\in\{x^*, x_s^*\}} |V_s(x) - V(x)| \\
&\le \sup_{x\in X} |V_s(x) - V(x)|
\end{aligned}$$

$$\lim_{s\to\infty}\left|\inf_{x\in X} V_s(x) - \inf_{x\in X} V(x)\right| \le \lim_{s\to\infty}\sup_{x\in X}|V_s(x) - V(x)| = 0 \quad \text{(ULLN)} \qquad \rfloor$$

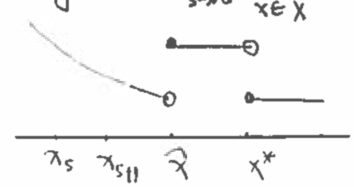
Facts for (ULLN) holds:
- If X is finite & (LLN) holds, then (ULLN) holds
- If X is convex & compact (close & bounded), $f_0(\cdot, \tilde{\xi})$ is convex & continuous a.s. on X & (LLN) holds, then (ULLN) holds.
- If X is compact, $f_0(\cdot, \tilde{\xi})$ is continuous a.s. on X. & there exists $g(\tilde{\xi})$ satisfying $\sup_{x\in X} |f(x, \tilde{\xi})| \le g(\tilde{\xi})$ a.s. and $\mathbb{E}[g(\tilde{\xi})] < \infty$, then (ULLN) holds.

<u>Theorem ii)</u> If (ULLN) holds then $\lim_{s\to\infty} V(x_s^*) = V(x^*)$ $\left(= \inf_{x\in X} V(x)\right)$

Proof:
$$\begin{aligned}
&|V(x_s^*) - V(x^*)| \\
&= V(x_s^*) - V(x^*) \\
&\le V(x_s^*) - V_s(x_s^*) + V_s(x^*) - V(x^*) \\
&\le |V(x_s^*) - V_s(x_s^*)| + |V_s(x^*) - V(x^*)| \\
&= \sup_{x\in\{x_s^*\}} |V_s(x) - V(x)| + \sup_{x\in\{x^*\}} |V_s(x) - V(x)| \\
&\le \sup_{x\in X} |V_s(x) - V(x)| + \sup_{x\in X} |V_s(x) - V(x)| \\
&= 2\sup_{x\in X} |V_s(x) - V(x)| \qquad \rfloor
\end{aligned}$$

→ This result doesn't guarantee $\lim_{s\to\infty} \underset{x\in X}{\arg\inf}\, V_s(x) \subseteq \underset{x\in X}{\arg\inf}\, V(x)$.

Example:



<u>Theorem iii)</u> Assume (ULLN) holds & X is closed, if X is finite or if $V(x)$ is continuous on X, then every limit point of $\{x_s^*\}$ solves $\inf_{x\in X} V(x)$.

$$\lim_{s\to\infty} \underset{x\in X}{\arg\inf}\, V_s(x) \subseteq \underset{x\in X}{\arg\inf}\, V(x)$$

Proof: Let $\hat{x}$ be a limit point of $\{x_s^*\}$ $\quad x_s^* \in \arg\inf_{x\in X} V_s(x)$

By the closedness of X we have $\hat{x} \in X$.

Since X is finite or $V(x)$ is continuous on X

we have: $\lim_{s\to\infty} V(x_s^*) = V(\hat{x})$

where $\lim_{s\to\infty} x_s^* = \hat{x}$ (pick a subsequence)

$$0 = \lim_{s\to\infty} |V(x_s^*) - V(x^*)|$$
$$= |V(\hat{x}) - V(x^*)| \qquad \text{(Theorem ii)} \qquad \rfloor$$

Remark: A more powerful type of convergence that yields the same results with more relaxed conditions is <u>epi-convergence</u>

Rate of convergence

$$\sqrt{S}\,(V_s(x) - V(x)) \xrightarrow{d} N(0, \sigma^2(x)) \qquad \text{(CLT)}$$

$$\sqrt{S}\,(\inf_{x\in X} V_s(x) - \inf_{x\in X} V(x)) \xrightarrow{d?} N(0, \sigma^2(x^*)) \qquad \text{where } x^* \in \arg\inf V(x),\ x\in X$$

If x^* is not unique then which one to use?

Let $U = \arg\inf_{x\in X} V(x)$ optimal solutions of true problem.

Fact: If X is compact & f_0 satisfies the Lipschitz condition:

$\exists g: \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[g^2(\xi)] < +\infty$

and $|f_0(x,\xi) - f_0(y,\xi)| \le g(\xi)\,\|x-y\|,\ \forall x,y \in X$ a.s.

then by defining $\tilde{\rho}(\bar{x}) \sim N(0, \sigma^2(\bar{x}))$, we have:

$$\sqrt{S}\,(\inf_{x\in X} V_s(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \inf_{\bar{x}\in U} \tilde{\rho}(\bar{x})$$

If $x^* \in \arg\inf_{x\in X} V(x)$ is unique, then: $\sqrt{S}\,(\inf_{x\in X} V_s(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \rho(x^*)$

Taking expectation:

$$\sqrt{S}\,\{\mathbb{E}[\inf_{x\in X} V_s(x)] - \mathbb{E}[\inf_{x\in X} V(x)] \longrightarrow \overbrace{\mathbb{E}[\inf_{x\in X} \tilde{\rho}(\bar{x})]}^{\text{bias term}}$$

goes to 0 at a rate of $\frac{1}{\sqrt{S}}$

If x^* is unique:

$$\sqrt{S}\,(\mathbb{E}[\inf_{x\in X} V_s(x)] - \mathbb{E}[\inf_{x\in X} V(x)]) \to 0$$

bias term goes to 0 at a rate of $\frac{1}{S}$ (Shapiro)

Estimating solution quality

of problem $\inf_{x\in X} V(x)$, $V(x) = \mathbb{E}[f_0(x,\xi)]$

Candidate solution $\hat{x} \in X$, we would like to assess its quality.

Given $\hat{x} \in X$ and $\alpha \in (0,1)$, we want to find a confidence interval $[0, \tilde{\delta}]$, so that:

$$\mathbb{P}(V(\hat{x}) - \inf_{x\in X} V(x) \in [0, \tilde{\delta}]) \approx 1-\alpha$$

$\tilde{\delta}$: we want to construct this.

not "=" since we are going to use CLT which only holds for infinite number. So we use "≈" meaning it only holds approximately.

$$\Leftrightarrow \mathbb{P}(V(\hat{x}) - \inf_{x\in X} V(x) \le \tilde{\delta}) \approx 1-\alpha$$

Lecture 17. Oct. 20

Observation: we have $\mathbb{E}[\inf_{x\in X} V_S(x)] \leq \inf_{x\in X} V(x)$

$$\Rightarrow \underbrace{V(\hat{x}) - \inf_{x\in X} V(x)}_{\text{true optimality gap}} \leq V(\hat{x}) - \mathbb{E}[\inf_{x\in X} V_S(x)]$$
$$= \mathbb{E}[\underbrace{V_S(\hat{x}) - \inf_{x\in X} V_S(x)}_{\text{empirical optimality gap}}]$$

Maybe we can study the asymptotic property of empirical optimality gap to derive the CI.

However:

$$\left.\begin{array}{l} \sqrt{S}\,(V_S(\hat{x}) - V(x)) \xrightarrow{d} N(0, \sigma^2(x)) \\ \sqrt{S}\,(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \underbrace{\inf_{x\in X} N(0, \sigma^2(x))}_{\text{not normal}} \end{array}\right\} V_S(\hat{x}) - \inf_{x\in X} V_S(x) \not\xrightarrow{d} N$$

Multiple Replications Procedure (MRP) (Mak, Morton & Wood 1999)

We will have L batches of S samples and do the analysis using these L batches.

Let $\tilde{\rho} = (\tilde{\xi}^1, \tilde{\xi}^2, \dots \tilde{\xi}^S)^T$

$$G(\hat{x}, \tilde{\rho}) = V_S(\hat{x}) - \inf_{x\in X} V_S(x)$$
$$= \frac{1}{S}\sum_{s\in[S]} f_0(\hat{x}, \tilde{\xi}^s) - \left(\inf_{x\in X} \frac{1}{S}\sum_{s\in[S]} f_0(x, \tilde{\xi}^s)\right)$$

Sample estimate approximation

Take L iid samples of $\tilde{\rho}$ and consider the SAA of

$$\mathbb{E}[G(\hat{x}, \tilde{\rho})] = \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]$$
$$G_L(\hat{x}) = \frac{1}{L}\sum_{l\in[L]} G(\hat{x}, \rho^l)$$

We have: (unbiased estimator)

$$\mathbb{E}[G_L(\hat{x})] = \mathbb{E}[G(\hat{x}, \tilde{\rho})] = \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]$$

Morever by CLT:

$$\sqrt{L}\,(G_L(\hat{x}) - \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]) \xrightarrow{d} N(0, \sigma_g^2(\hat{x}))$$

where: $\sigma_g^2(\hat{x}) = \mathrm{Var}(G(\hat{x}, \tilde{\rho}))$

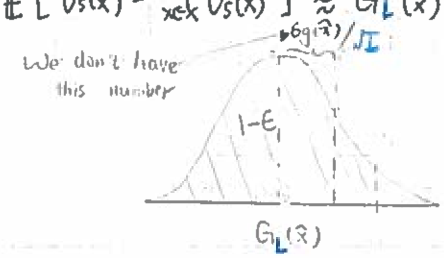
$L \geq 30$ For each k, we need to solve an optimization problem. We can paralize all these optimization problems. S should be hundreds or thousands.

Because of symmetry of $N(0, \sigma_g^2(\hat{x}))$.

$$\sqrt{L}\,(\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] - G_L(\hat{x})) \xrightarrow{d} N(0, \sigma_g^2(\hat{x}))$$

For large enough L:

$$\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] \approx G_L(\hat{x}) + (\sigma_g(\hat{x})/\sqrt{L})\cdot\tilde{w} \quad \text{where } \tilde{w} \sim N(0,1)$$



For large L:

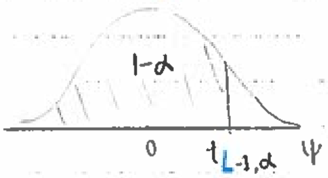
$$\frac{\mathbb{E}\left[V_S(\hat{x}) - \inf_{x\in X} V_S(x)\right] - G_L(\hat{x})}{?\ \sigma_g(\hat{x})/\sqrt{L}} \overset{\text{Approx}}{\sim} N(0,1)$$

We use the sample variance as a surrogate of $\sigma_g^2(\hat{x})$

$$SV_L^2(\hat{x}) = \frac{1}{L-1}\sum_{\ell\in[L]}\left(G(\hat{x},\rho^\ell) - G_L(\hat{x})\right)^2$$

then for large k:

$$\tilde{\Psi} = \frac{\mathbb{E}\left[V_S(\hat{x}) - \inf_{x\in X} V_S(x)\right] - G_L(\hat{x})}{SV_L(\hat{x})/\sqrt{L}} \overset{\text{Approx}}{\sim} \text{t-distribution with } L-1 \text{ degree of freedom.}$$



$$\mathbb{P}(\tilde{\Psi} \le t_{L-1,\alpha}) \approx 1-\alpha$$

$t_{L-1,\alpha}$: $(1-\alpha)$ quantile of a t-dist with $L-1$ degrees of freedom.

$$\mathbb{P}\left(\frac{\mathbb{E}\left[V_S(\hat{x}) - \inf_{x\in X} V_S(x)\right] - G_L(\hat{x})}{SV_L(\hat{x})/\sqrt{L}}\right) \approx 1-\alpha$$

$$\Leftrightarrow \mathbb{P}\left(\mathbb{E}\left[V_S(\hat{x}) - \inf_{x\in X} V_S(x)\right] \le G_L(\hat{x}) + t_{L-1,\alpha}\frac{SV_L(\hat{x})}{\sqrt{L}}\right) \approx 1-\alpha$$

since $V(\hat{x}) - \inf_{x\in X} V_S(x) \le \mathbb{E}\left[V_S(\hat{x}) - \inf_{x\in X} V_S(x)\right]$

$$\Rightarrow \mathbb{P}\left(V(\hat{x}) - \inf_{x\in X} V_S(x) \le \underbrace{G(\hat{x}) + \frac{t_{L-1,\alpha}\cdot SV_L(\hat{x})}{\sqrt{L}}}_{\delta}\right) \geq 1-\alpha$$

<u>Procedure for estimating solution quality</u>

Input:
- a candidate solution $\hat{x}\in X$
- tolerance α
- batch size S
- sample size L

Output: approximate $(1-\alpha)$-level confidence interval $[0,\delta]$ on the optimality gap $V(\hat{x}) - \inf_{x\in X} V(x)$

Steps:

1) Generate L batches of size S

$\xi^{11} \cdots \xi^{1S} - \rho^1$

$\xi^{21} \cdots \xi^{2S} - \rho^2$

$\vdots$

$\xi^{L1} \cdots \xi^{LS} - \rho^L$

2) Compute:

$$G(\hat{x},\rho^\ell) = \frac{1}{S}\sum_{s\in[S]} f_0(\hat{x},\xi^{\ell s}) - \inf_{x\in X}\frac{1}{S}\sum_{s\in[S]} f_0(x,\xi^{\ell s}),\ \forall \ell\in[L]$$

3) Let:

$$G_L(\hat{x}) = \frac{1}{L}\sum_{\ell\in[L]} G(\hat{x},\rho^\ell)$$

$$SV_L^2(\hat{x}) = \frac{1}{L-1}\sum_{\ell\in[L]}\left(G(\hat{x},\rho^\ell) - G_L(\hat{x})\right)^2$$

4) $\delta = G_L(\hat{x}) + t_{L-1,\alpha}\, S_{V_L}(\hat{x})/\sqrt{L}$

$[0, \delta]$ is the $(1-\alpha)$ level confident interval on $V(x) - \inf_{x \in X} V(x)$

CI is wide if:

1) $\hat{x}$ is a poor solution: $V(\hat{x}) - \inf_{x \in X} V(x)$ is large

2) $\inf_{x \in X} V(x) - \inf_{x \in X} V_S(x)$ is large, meaning that we have large negative bias $E[\inf_{x \in X} V_S(x)] \ll \inf_{x \in X} V(x)$. (CI also wide)

3) large sample error $t_{L-1,\alpha}\, S_{V_L}(\hat{x})/\sqrt{L}$ (rare to be the primary contribution)

Remedy:

1) Improve the quality of approx scheme for obtaining $\hat{x}$
2) Increase S.
3) Increase S or L.

* Remark: Increasing S is more expensive because problem becomes harder.
Recommendation in practice is to fix $L = 20 - 30$ to induce CLT

Proof: Let $\hat{x}$ be a limit point of $\{x_s^*\}$, $x_s^* \in \arg\inf_{x\in X} V_S(x)$.

By the closedness of X, we have $\hat{x} \in X$.

Since X is finite or $V(x)$ is continuous on X

we have: $\lim_{s\to\infty} V(x_s^*) = V(\hat{x})$

where $\lim_{s\to\infty} x_s^* = \hat{x}$ (pick a subsequence)

$0 = \lim_{s\to\infty} |V(x_s^*) - V(x^*)|$

$= |V(\hat{x}) - V(x^*)|$ (Theorem ii) $\lrcorner$

Remark: A more powerful type of convergence that yields the same results with more relaxed conditions is epi-convergence

Rate of convergence

$\sqrt{S}(V_S(x) - V(x)) \xrightarrow{d} N(0, \sigma^2(x))$ (CLT)

$\sqrt{S}(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)) \xrightarrow{d?} N(0, \sigma^2(x^*))$ where $x^* \in \arg\inf V(x)$, $x\in X$

If x^* is not unique then which one to use?

Let $U = \arg\inf_{x\in X} V(x)$ optimal solutions of true problem

Fact: If X is compact & f_0 satisfies the Lipschitz condition:

$\exists g: \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[g^2(\tilde{\xi})] < +\infty$

and $|f_0(x,\tilde{\xi}) - f_0(y,\tilde{\xi})| \le g(\tilde{\xi}) \|x-y\|$, $\forall x,y \in X$ a.s.

then by defining $\tilde{\rho}(\bar{x}) \sim N(0, \sigma^2(\bar{x}))$, we have:

$\sqrt{S}(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \inf_{\bar{x}\in U} \tilde{\rho}(\bar{x})$

If $x^* \in \arg\inf_{x\in X} V(x)$ is unique, then: $\sqrt{S}(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \rho(x^*)$

Taking expectation:

$\sqrt{S}(\mathbb{E}[\inf_{x\in X} V_S(x)] - \mathbb{E}[\inf_{x\in X} V(x)]) \to \overbrace{\mathbb{E}[\inf_{x\in X} \tilde{\rho}(\bar{x})]}^{\text{bias term}}$ goes to 0 at a rate of $\frac{1}{\sqrt{S}}$

If x^* is unique:

$\sqrt{S}(\mathbb{E}[\inf_{x\in X} V_S(x)] - \mathbb{E}[\inf_{x\in X} V(x)]) \to 0$ bias term goes to 0 at a rate of $\frac{1}{S}$ (Shapiro)

Estimating solution quality

of problem $\inf_{x\in X} V(x)$, $V(x) = \mathbb{E}[f_0(x,\tilde{\xi})]$

Candidate solution $\hat{x} \in X$, we would like to assess its quality.

Given $\hat{x} \in X$ and $\alpha \in (0,1)$, we want to find a confidence interval $[0, \inf_{x\in X} V(x) + \tilde{\delta}]$, so that:

$\mathbb{P}(V(\hat{x}) - \inf_{x\in X} V(x) \in [0, \tilde{\delta}]) \approx 1-\alpha$

we want to construct this.

not = since we are going to use CLT which only holds for infinite number. So we use ≈ meaning only holds approximately

$\Leftrightarrow \mathbb{P}(V(\hat{x}) - \inf_{x\in X} V(x) \le \tilde{\delta}) \approx 1-\alpha$

Lecture 17. Oct. 20

Observation: we have $\mathbb{E}[\inf_{x\in X} V_S(x)] \le \inf_{x\in X} V(x)$

$$\Rightarrow \underbrace{V(\hat{x}) - \inf_{x\in X} V(x)}_{\text{true optimality gap}} \le V(\hat{x}) - \mathbb{E}[\inf_{x\in X} V_S(x)] = \underbrace{\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]}_{\text{empirical optimality gap}}$$

Maybe we can study the asymptotic property of empirical optimality gap to derive the CI.

However:

$$\left.\begin{array}{l} \sqrt{S}\,(V_S(\hat{x}) - V(x)) \xrightarrow{d} N(0, \sigma^2(x)) \\ \sqrt{S}\,(\inf_{x\in X} V_S(x) - \inf_{x\in X} V(x)) \xrightarrow{d} \inf_{x\in X} N(0, \sigma^2(x)) \end{array}\right\} V_S(\hat{x}) - \inf_{x\in X} V_S(x) \overset{d}{\not\to} N \ldots$$

Multiple Replications Procedure (MRP) (Mak, Mortos & Wood 1999)

We will have K batches of S samples and do the analysis using these K batches.

let $\tilde{\rho} = (\tilde{\xi}^1, \tilde{\xi}^2, \ldots \tilde{\xi}^S)^T$

$$G(\hat{x}, \tilde{\rho}) = V_S(\hat{x}) - \inf_{x\in X} V_S(x)$$
$$= \frac{1}{S}\sum_{s\in[S]} f_0(\hat{x}, \tilde{\xi}^s) - \left(\inf_{x\in X} \frac{1}{S}\sum_{s\in[S]} f_0(x, \tilde{\xi}^s)\right)$$

Sample estimate approximation

Take K iid samples of $\tilde{\rho}$ and consider the SAA of

$\mathbb{E}[G(\hat{x}, \tilde{\rho})] = \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]$

$$G_K(\hat{x}) = \frac{1}{K}\sum_{k\in[K]} G(\hat{x}, \rho^k)$$

We have: (unbiased estimator)

$$\mathbb{E}[G_K(\hat{x})] = \mathbb{E}[G(\hat{x}, \tilde{\rho})] = \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]$$

Morever by CLT:

$$\sqrt{K}\,(G_K(\hat{x}) - \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]) \xrightarrow{d} N(0, \sigma_g^2(\hat{x}))$$

where: $\sigma_g(\hat{x}) = \mathrm{Var}(G(\hat{x}, \tilde{\rho}))$

$K \geq 30$ For each k we need to solve an optimization problem. We can paralize all these optimization problems. S should be hundreds or thousands.

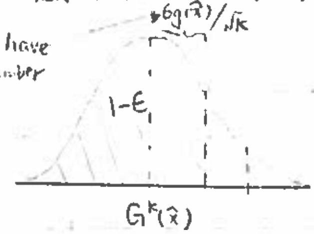
Because of symmetry of $N(0, \sigma_g^2(\hat{x}))$.

$$\sqrt{K}\,(\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] - G_K(\hat{x})) \xrightarrow{d} N(0, \sigma_g^2(\hat{x}))$$

For large enough K:

$$\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] \approx G_K(\hat{x}) + (\sigma_g(\hat{x})/\sqrt{K})\cdot\tilde{w} \quad \text{where } \tilde{w} \sim N(0,1)$$

We don't have this number



For large K:

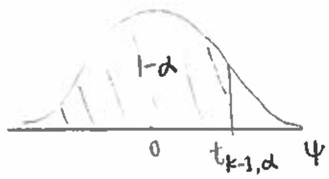
$$\frac{\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] - G_K(\hat{x})}{? \to \sigma_g(\hat{x})/\sqrt{K}} \overset{\text{Approx}}{\sim} \mathcal{N}(0,1)$$

We use the sample variance as a surrogate.

$$SV_K^2(\hat{x}) = \frac{1}{K+1} \sum_{k\in[K]} \left(G(\hat{x}, \rho^k) - G^K(\hat{x})\right)^2$$

then for large K:

$$\tilde{\Psi} = \frac{\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] - G_K(\hat{x})}{SV_K(\hat{x})/\sqrt{K}} \overset{\text{Approx}}{\sim} \text{t-distribution with } K-1 \text{ degree of freedom.}$$



$$\mathbb{P}(\tilde{\Psi} \le t_{K-1,\alpha}) \approx 1-\alpha$$

$(1-\alpha)$ quantile of a t-dist with K-1 degrees of freedom.

$$\mathbb{P}\left(\frac{\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] - G_K(\hat{x})}{SV_K(\hat{x})/\sqrt{K}} \le t_{K-1,\alpha}\right) \approx 1-\alpha$$

$$\Leftrightarrow \mathbb{P}\left(\mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)] \le G_K(\hat{x}) + t_{K-1,\alpha}\frac{SV_K(\hat{x})}{\sqrt{K}}\right) \approx 1-\alpha$$

since $V(\hat{x}) - \inf_{x\in X} V_S(x) \le \mathbb{E}[V_S(\hat{x}) - \inf_{x\in X} V_S(x)]$

$$\Rightarrow \mathbb{P}\left(V(\hat{x}) - \inf_{x\in X} V_S(x) \le \underbrace{G_K(\hat{x}) + \frac{t_{K-1,\alpha}\cdot SV_K(\hat{x})}{\sqrt{K}}}_{\delta}\right) \geq 1-\alpha$$

<u>Procedure for estimating solution quality</u>

Input: - a candidate solution $\hat{x} \in X$
- tolerance α
- batch size S
- sample size K

Output: approximate $(1-\alpha)$-level confidence interval $[0, \delta]$ on the optimality gap $V(\hat{x}) - \inf_{x\in X} V(x)$

Steps:

1) Generate K batches of size S.

$$\begin{array}{llll} \xi^{11} \cdots & \xi^{1S} & - \rho^1 & \text{Batch 1} \\ \xi^{21} \cdots & \xi^{2S} & - \rho^2 & \\ & \vdots & & \\ \xi^{K1} \cdots & \xi^{KS} & - \rho^K & \text{Batch K} \end{array}$$

2) Compute:

$$G(\hat{x}, \rho^k) = \frac{1}{S}\sum_{s\in[S]} f_0(\hat{x}, \xi^{ks}) - \inf_{x\in X} \frac{1}{S}\sum_{s\in[S]} f_0(x, \xi^{ks}), \ \forall k \in [K]$$

3) Let:

$$G_K(\hat{x}) = \frac{1}{K}\sum_{k\in[K]} G(\hat{x}, \rho^k)$$

$$SV_K^2(\hat{x}) = \frac{1}{K+1}\sum_{k\in[K]} \left(G(\hat{x}, \rho^k) - G_K(\hat{x})\right)^2$$

4) $\delta = G_k(\hat{x}) + t_{k-1,\alpha}\ SV_k(\hat{x})/\sqrt{k}$

$[0, \delta]$ is the $(1-\alpha)$ level confident interval on $V(x) - \inf_{x\in X} V(x)$

CI is wide if:

1) $\hat{x}$ is a poor solution: $V(\hat{x}) - \inf_{x\in X} V(x)$ is large

2) $\inf_{x\in X} V(x) - \inf_{x\in X} V_S(x)$ is large, meaning that we have large negative bias
$E[\inf_{x\in X} V_S(x)] << \inf_{x\in X} V(x)$ (CI also wide)

3) large sample error $t_{k-1,\alpha}\ SV_k(\hat{x})/\sqrt{k}$ (rare to be the primary contribution)

Remedy:

1) Improve the quality of approx scheme for obtaining $\hat{x}$

2) Increase S.

3) Increase S or K.

* Remark: Increasing S is more expensive because problem becomes harder.
Recommendation in practice is to fix $K = 20-30$ to induce CLT

Lecture 18. Oct. 25

Bender's decomposition for 2-stage problems with discrete distributions.

$$\inf_{x \in X} \; c^T x + \mathbb{E}[z(x, \xi)] \qquad (25)$$

where $z(x,\xi) = \inf \; (Q\xi + g)^T y$ $z_d(x,\xi) = \sup \; (T(x)\xi + h(x))^T \pi$

s.t. $y \in \mathbb{R}_+^{N_2}$ s.t. $\pi \in \mathbb{R}^J$

$T(x)\,\xi + h(x) = Wy$ $Q\xi + g \geq W^T \pi$

$$\begin{aligned} \inf \;\; & c^T x + \sum_{s \in [S]} p^s (Q\xi^s + g)^T y^s \\ \text{s.t.} \;\; & x \in X, \; y^1, y^2, \dots y^S \in \mathbb{R}_+^{N_2} \\ & T(x)\,\xi^s + h(x) = W y^s, \quad \forall s \in [S] \end{aligned}$$

$S = 3$. $\xi^1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ $\xi^2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ $\xi^3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$p^1 = 1/3$ $p^2 = 1/3$ $p^3 = 1/3$

$$\left.\begin{aligned} \inf \;\; & c^T x + p^1 (Q\xi^1 + g)^T y^1 + p^2 (Q\xi^2 + g)^T y^2 + \cdots + p^S (Q\xi^S + g)^T y^S \\ \text{s.t.} \;\; & x \in X, \; y^1, y^2, \dots y^S \in \mathbb{R}_+^{N_2} \\ & T(x)\,\xi^1 + h(x) = W y^1 \\ & T(x)\,\xi^2 + h(x) = W y^2 \\ & \quad \dots \\ & T(x)\,\xi^S + h(x) = W y^S \end{aligned}\right\} \text{large number of variables and constraints } N_2 \times S$$

For a fixed $x \in X$, the problem decomposes into S smaller subproblems
$\Rightarrow$ algorithm that exploits this.

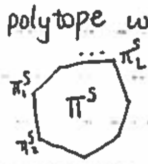
Idea: we are going to reduce the number of variables but increase considerably the number of constraints.
$\Rightarrow$ delayed constraint generation.

How to reduce the # of variables?

Let $\Pi^s = \{ \pi \in \mathbb{R}^J : Q\xi^s + g \geq W^T \pi \}$. $\forall s \in [S]$

Observations:

- for a fixed ξ^s, the feasible set is a polytope with finitely (but can be exponentially) many extreme points.
- it doesn't depend on x.



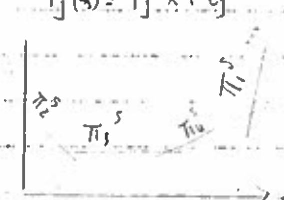
- Π^s is bounded with extreme points $\pi_1^s, \pi_2^s, \dots \pi_L^s$

(where L depends on S L can be exponentially in $\underset{\dim(y)}{N_2}$ & $\underset{\dim(\pi)}{J}$)

$$\begin{aligned} z_d(x, \xi^s) = \; & \sup \; (T(x)\,\xi^s + h(x))^T \pi \\ & \text{s.t. } \pi \in \mathbb{R}^J \\ & \quad\;\; Q\xi^s + g \geq W^T \pi \\ = \; & \max_{l \in [L]} \; (T(x)\,\xi^s + h(x))^T \pi_l^s \end{aligned}$$

$h(x) = H x + \hat{h}$

$T_j(x) = T_j x + \hat{t}_j$



piecewise linear convex

$$(2S) = \inf_{x \in X} \; c^T x + \sum_{s \in [S]} p^s \, Z(x, \varepsilon^s)$$

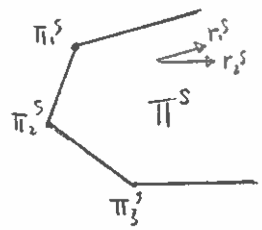
$$= \inf_{x \in X} \; c^T x + \sum_{s \in [S]} p^s \underbrace{\max_{l \in [L]} \left(T(x)\varepsilon^s + h(x)\right)^T \pi_l^s}$$

$$\begin{aligned}
= \inf \;& c^T x + \sum_{s \in [S]} p^s y^s \\
\text{s.t. } & x \in X, \; y \in \mathbb{R}^S \\
& \left(T(x)\varepsilon^s + h(x)\right)^T \pi_l^s \le y^s, \; \forall s \in [S], \forall l \in [L] \\
(*) \;& \left(T(x)\varepsilon^s + h(x)\right)^T r_m^s \le 0, \; \forall s \in [S], \forall m \in [M]
\end{aligned}$$

} LP with fewer # of variables but larger # of constraints
$S \times L$ (exponential in N_2.)]

- Π^s is unbounded

Any point in a polytope can be described by a convex combination of extreme points + a nonnegative linear combination of extreme rays



$$\Pi^s = \left\{ \pi \in \mathbb{R}^J : \pi = \sum_{l \in [L]} \lambda_l \pi_l^s + \sum_{m \in [M]} \mu_m r_m^s \text{ for some } \lambda \in \mathbb{R}_+^L, \lambda^T e = 1 \;\&\; \mu \in \mathbb{R}_+^M \right\}$$

For a fixed ε^s:

$$\begin{aligned}
Z_d(x, \varepsilon^s) = \sup \;& \left(T(x)\varepsilon^s + h(x)\right)^T \pi \\
\text{s.t. } & \pi \in \mathbb{R}^J \\
& Q\varepsilon^s + g \ge W\pi
\end{aligned}$$

$$Z_d(x, \varepsilon^s) < +\infty \iff \left(T(x)\varepsilon^s + h(x)\right)^T r_m^s \le 0, \; \forall m \in [M]$$

$$Z_d(x, \varepsilon^s) < +\infty \Rightarrow Z_d(x, \varepsilon^s) = \max_{l \in [L]} \left(T(x)\varepsilon^s + h(x)\right)^T \pi_l^s$$

Add constraint $(*)$ to previous reformulation

<u>Bender's decomposition algorithm for (2S)</u>

It will not use all the extreme points and rays. It will produce them iteratively.

Inputs: parameters of (2S)

Outputs: optimal solution x^* to (2S)

- Step 0: Let $U^s = \phi$, $V^s = \phi$, $\forall s \in [S]$ (empty set)
  - U^s: current set of extreme points of π^s
  - V^s: current set of extreme rays of Π^s

- Step 1: Solve the master problem.

At first, U^s and V^s are empty. The optimal value is $-\infty$

$$\begin{aligned}
\inf \;& c^T x + \sum_{s \in [S]} p^s y^s \\
\text{s.t. } & x \in X, \; y \in \mathbb{R}^S \\
& \left.\begin{aligned} &\left(T(x)\varepsilon^s + h(x)\right)^T \theta \le y^s, \; \forall \theta \in U^s \\ &\left(T(x)\varepsilon^s + h(x)\right)^T \eta \le 0, \; \forall \eta \in V^s \end{aligned}\right\} \forall s \in [S]
\end{aligned}$$

Give the optimal solution $\{\hat{x}, \{\hat{y}^s\}_{s \in [S]}\}$

• Step 2: For all $s \in [S]$ solve:

$$Z_d(\hat{x}, \varepsilon^s) = \sup \ (T(\hat{x})\varepsilon^s + h(\hat{x}))^T \cdot \pi$$
$$\text{s.t. } \pi \in \mathbb{R}^J$$
$$Q\varepsilon^s + g \geq W^T \pi$$

If $Z_d(\hat{x}, \varepsilon^s) < \hat{y}^s$, $\forall s \in [S]$

terminate $x^* = \hat{x}$ is the optimal solution

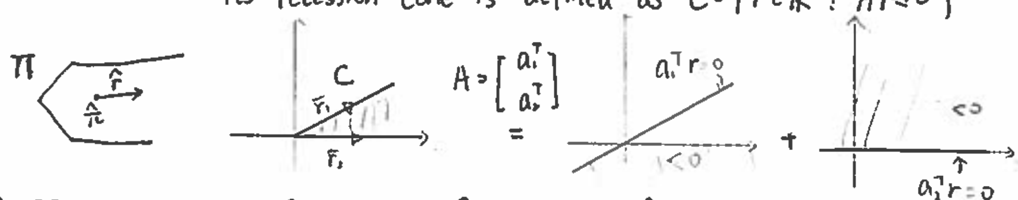
Otherwise go to step 3.

lecture 19. Oct 27.

$$\inf \ c^T x + \sum_{s \in [S]} p^s Z(x, \varepsilon^s)$$
$$= \inf \ c^T x + \sum_{s \in [S]} p^s y^s$$
$$\text{s.t. } x \in X, \ y \in \mathbb{R}^S$$
$$(T(x)\varepsilon^s + h(x))^T \underbrace{\pi^l}_{\text{extreme point of dual recourse}} \leq y^s, \ \forall s \in [S], \ \forall l \in [L]$$
$$(T(x)\varepsilon^s + h(x))^T \underbrace{r^m}_{\text{extreme ray of dual recourse.}} \leq 0, \ \forall s \in [S], \ \forall m \in [M]$$

$\Pi = \{\pi \in \mathbb{R}^J, A\pi \leq b\}$

Def (<u>Recession Cone</u>): For any polyhedral set $\Pi = \{\pi \in \mathbb{R}^J, A\pi \leq b\}$, its recession cone is defined as $C = \{r \in \mathbb{R}^J : Ar \leq 0\}$



If $\hat{\pi} \in \Pi$, then for any $\hat{r} \in C$, $A(\hat{\pi} + \hat{r}) \leq b$, $\underbrace{A\hat{\pi}}_{\leq b} + \underbrace{A\hat{r}}_{\leq 0} \leq b$

Def (<u>extreme ray</u>): $\bar{r} \in C$ is an extreme ray of Π if there are $J-1$ linearly independent constraints from $A\bar{r} \leq 0$ that are active

$$\sup \ c^T \pi$$
$$\text{s.t. } A\pi \leq b$$

- The LP problem is unbounded $\Leftrightarrow \exists \hat{r} \in C, \ c^T \hat{r} > 0$
  $\overset{(*)}{\Leftrightarrow} \exists \underbrace{\bar{r}}_{\text{extreme rays}} \in \text{ext}(C), \ c^T \bar{r} > 0$

Proof (*): $\Leftarrow$, if $c^T \bar{r} > 0$, $\bar{r} \in \text{ext}(C)$, since $\text{ext}(C) \subseteq C$, $\bar{r} \in C$

$\Rightarrow$, $\hat{r} = \sum_{m \in [M]} \mu_m \bar{r}_m$, where $\bar{r}_m \in \text{ext}(C)$, $\forall m \in [M]$

& $\mu_m \geq 0$, $\forall m \in [M]$

$\rightarrow c^T \hat{r} = \sum_{m \in [M]} \mu_m c^T \bar{r}_m > 0$

$\rightarrow c^T \bar{r}_m > 0$ for some $m \in [M]$

- Bounded $\Leftrightarrow \forall \bar{r} \in \text{ext}(C), \ c^T \bar{r} \leq 0$

Continue Bender's decomposition algorithm for (2S)

- Step 3: If for a $s \in [S]$ the dual recourse problem is unbounded.

  $Z_d(\hat{x}, \xi) = +\infty \Rightarrow \hat{x}$ is an infeasible solution

  Solver will give a direction of extreme ray $\bar{r}$.

  such that $(T(\hat{x})\xi + h(\hat{x}))^T \bar{r} > 0$

  Add $V^s = V^s \cup \{\bar{r}\}$.

  If $Z_d\{\hat{x}, \xi^s\} > \hat{\gamma}^s$ then

  let $\hat{\pi} \in \operatorname{argsup} (T(\hat{x})\xi^s + h(\hat{x}))^T \pi$

  s.t. $\pi \in \mathbb{R}^J$

  $Q\xi^s + q \geq W^T \pi$

  Add $U^s = U^s \cup \{\hat{\pi}\}$

  Go to step 1.

Assigment 3.

3. 1) $\inf \ \lambda + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} E_{\mathbb{P}}[\max\{(c-v)^T x + [\sum_{n \in [N]} \max\{v_n x_n - v_n \tilde{\xi}_n, 0\}] - \lambda, 0\}]$

s.t. $x \in \mathbb{R}_+^N, \lambda \in \mathbb{R}$

$= \inf \ \lambda + \frac{1}{\epsilon}(\alpha + \beta^T \mu + \langle \Gamma, \Omega \rangle)$

s.t. $x \in X, \alpha \in \mathbb{R}, \beta \in \mathbb{R}^N, \Gamma \in S_+^N$

$\alpha + \beta^T \xi + \xi^T \Gamma \xi \geq (c-v)^T x + \sum_{n \in [N]} \max\{v_n x_n - v_n \xi_n, 0\} - \lambda, \ \forall \xi \in \mathbb{R}^N$ (*)

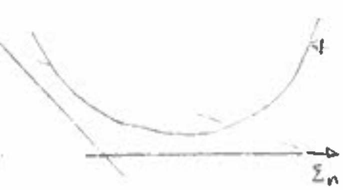
$\alpha + \beta^T \xi + \xi^T \Gamma \xi \geq 0, \ \forall \xi \in \mathbb{R}^N \Leftrightarrow \begin{bmatrix} \Gamma & \frac{1}{2}\beta \\ \frac{1}{2}\beta^T & \alpha \end{bmatrix} \succeq 0$

(*) $\Leftrightarrow \alpha + \beta^T \xi + \xi^T \Gamma \xi \geq (c-v)^T x + l^T (v \circ x - v \circ \xi) - \lambda, \ \forall l \in \{0,1\}^N$

$v \circ x = [v_1 x_1, v_2 x_2, \dots]^T$

$= (c-v)^T x + (l \circ v)^T x - (l \circ v)^T \xi - \lambda$

$\Leftrightarrow \begin{bmatrix} \Gamma & \frac{1}{2}(\beta + l \circ v) \\ \frac{1}{2}(\beta + l \circ v)^T & \alpha - (c-v)^T x - (l \circ v)^T x + \lambda \end{bmatrix} \succeq 0, \ \forall l \in \{0,1\}^N$



Support of the problem is unbounded, we can never find a linear upperbound. Any line will exceeds the support.

2) $\inf \ \lambda + \frac{1}{\epsilon} \sup E_{\mathbb{P}}[\max\{(c-v)^T x + [\sum_{n \in [N]} \xi^T Q_n \xi + r_n^T \xi + s_n] - \lambda, 0\}]$

s.t. $x \in \mathbb{R}_+^N, \lambda \in \mathbb{R}, Q^n \in S_+^N, r^n \in \mathbb{R}^N, s^n \in \mathbb{R}, \ \forall n \in [N]$

$\xi^T Q^n \xi + r^{nT} \xi + s^n \geq v_n x_n - \underbrace{v_n \xi_n}_{= (e_n \circ v)^T \xi}$

$\xi^T Q^n \xi + r^{nT} \xi + s^n \geq 0$

$= \inf \ \lambda + \frac{1}{\epsilon}(\alpha + \beta^T \mu + \langle \Gamma, \Omega \rangle)$

s.t. $\alpha + \beta^T \xi + \xi^T \Gamma \xi \geq (c-v)^T x + \xi^T (\sum_{n \in [N]} Q^n) \xi + (\sum_{n \in [N]} r^n)^T \xi + (\sum_{n \in [N]} s^n) - \lambda \quad \forall \xi \in \mathbb{R}^N$

$\alpha + \beta^T \xi + \xi^T \Gamma \xi \geq 0, \ \forall \xi \in \mathbb{R}^N$

$$(*) \Leftrightarrow \begin{bmatrix} \Gamma - \sum_{n\in[N]} Q^n & \frac{1}{2}(\beta - \sum_{n\in[N]} r^n) \\ \frac{1}{2}(\beta - \sum_{n\in[N]} r^n & \alpha - (C-V)^T x - \sum_{n\in[N]} S^n + \lambda \end{bmatrix} \succeq 0$$

$*$:

Since ξ is demand, $\forall \xi \in \{A\xi \le b\}$, $\inf_{A\xi\le b} \alpha + \beta^T \xi + \xi^T \Gamma \xi \ge 0$ ≥ 0 dual

Lecture 20. Nov. 1

$$\inf_{x\in X} c^T x + \sup_{\xi \in \Xi} Z(x,\xi) \qquad (2R)$$ worse case recourse function

where $Z(x,\xi) = \inf \; (Q\xi + g)^T y$
s.t. $y \in \mathbb{R}^{N_2}$
$T(x)\xi + h(x) \le Wy$

and $\Xi = \{\xi \in \mathbb{R}^k : S\xi \le t\}$

<u>Summary</u>

1) Piecewise-linear model (subset of 2-stage)
   - Stochastic : discrete dist. with S scenarios $\Rightarrow$ tractable $O(S)$
   - Stochastic in general is NP-hard
   - DRO tractable under reasonable assumptions
     Distributed robust model

2) 2-Stage mode
   - Stochastic discrete dist. with S scenarios $\Rightarrow$ tractable $O(S)$
   - Stochastic in general NP-hard — upper & lower bounds
     Monte Carlo
   - DRO in general NP-hard — upper & lower bounds
     $\hookrightarrow$ robust opt. : support of dist.

Why proving NP-hardness is useful:
- better understanding of the complexity of the problem.
- avoid future embarassment
  - want to solve a problem you perceive to be hard
  - develop approx. scheme, but didn't prove NP-hardness
  - some time later.
    - Someone proves problem NP-hard ✓
    - someone provides a polynomial time algorithm. ×

<u>Thm</u>: (2R) is NP-hard even if $Q=0$ & there is no first stage decision x.

Strategy: ① Pick an NP-hard problem (Q).
② Generate a polynomial-time reduction from any instance of (Q) to an instance of (2R).

Proof: ① 0/1 INTEGER PROGRAMMING FEASIBILITY.

INSTANCE: Given $S \in \mathbb{R}^{m\times k}$ & $t \in \mathbb{R}^{m}$

QUESTION: Is there a binary vector $\xi \in \{0,1\}^k$ such that $S\xi \le t$ ...

② Instance: $c=0$, $Q=0$, $g=e\in\mathbb{R}^k$

$$T(x)=\begin{bmatrix} I\\ -I\end{bmatrix} \in \mathbb{R}^{2k\times k}, \quad h(x)=\begin{bmatrix}0\\ e\end{bmatrix}\in\mathbb{R}^{2k\times 1}$$

$$W=\begin{bmatrix} I\\ I\end{bmatrix}\in\mathbb{R}^{2k\times k} \qquad \Xi=\{[0,1]^k : S\xi\le t\}$$

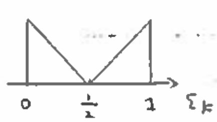
polynomial in k

$$z(x,\xi)=\inf\ e^Ty \quad \text{s.t.}\ \xi\le y,\ -\xi+e\le y \quad = \sum_{k\in[K]}\max\{\xi_k,1-\xi_k\}$$

IP feasibility is satisfied $\iff \sup_{\xi\in\Xi}\sum_{k\in[k]}\max\{\xi_k,1-\xi_k\}\ge K$

Proof: "$\Rightarrow$": $\exists\,\hat{\xi}\in\{0,1\}^k : S\xi\le t \Rightarrow \hat{\xi}\in\Xi$ & $\sum_{k\in[k]}\underbrace{\max\{\xi_k,1-\xi_k\}}_{1}=K$

"$\Leftarrow$"



$\max\{\xi_k,1-\xi_k\}\iff\xi_k\in\{0,1\}$

$\{\xi\in[0,1]^k : \sum_{k\in[k]}\max\{\xi_k,1-\xi_k\}\ge K\}=\{0,1\}^k$

Right hand Side: $\exists\,\xi\in\Xi : \sum_{k\in[k]}\max\{\xi_k,1-\xi_k\}\ge k$

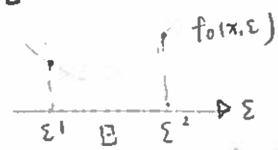
$\iff \exists\,\xi\in\Xi\cap\{0,1\}^k$

$\Rightarrow \exists\,\xi\in\{0,1\}^k,\ S\xi\le t$

Tractability results for $Q=0$:

① $\Xi=\mathrm{conv}\{\xi^1,\dots,\xi^L\}$ where L is polynomial in K & N_2. extreme points

if $f_0(x,\xi)$ is convex in ξ then: $\sup_{\xi\in\Xi} f_0(x,\xi)=\max_{l\in[L]} f_0(x,\xi^l)$



We have previously shown that $Z(x,\xi)$ is convex in ξ if $Q=0$

$$\inf_{x\in X} c^Tx+\sup_{\xi\in\Xi} Z(x,\xi)$$

$$=\inf_{x\in X} c^Tx+\max_{l\in[L]} Z(x,\xi^l)$$

$$=\inf_{x\in X} c^Tx+\tau \quad \text{s.t. } x\in X,\ \tau\in\mathbb{R},\ Z(x,\xi^l)\le\tau\ \ \forall l\in[L]$$

$$=\inf_{x\in X} c^Tx+\tau \quad \text{s.t. } x\in X,\ \tau\in\mathbb{R},\ \inf\{g^Ty : y\in\mathbb{R}^{N_2},\ T(x)\xi^l+h(x)\le Wy\}\le\tau\ \ \forall l\in[L]$$

$$
\begin{aligned}
= \inf_{x \in X} \quad & C^T x + \tau \\
\text{s.t.} \quad & x \in X, \ \tau \in \mathbb{R}, \ y^l \in \mathbb{R}^{N_2}, \ \forall l \in [L] \\
& \left. \begin{aligned} & g^T y^l \le \tau \\ & T(x)\,\xi^l + h(x) \le W y^l \end{aligned} \right\} \ \forall l \in [L]
\end{aligned}
$$

②

$$
\begin{aligned}
Z_d = \ \sup \quad & (T(x)\xi + h(x))^T \pi \\
\text{s.t.} \quad & \pi \in \mathbb{R}^J_+ \\
& g = W^T \pi
\end{aligned}
$$

Let $\Pi = \{ \pi \in \mathbb{R}^J_+ : g = W^T \pi \}$

Π is bounded & has extreme points $\pi^1, \dots, \pi^R$ where R is polynomial in K & N_2.

$$Z(x,\xi) = Z_d(x,\xi) = \max_{r \in [R]} (T(x)\xi + h(x))^T \pi^r$$

We also assume $Q = 0$

o.w. $\Pi(\xi) = \{ \pi \in \mathbb{R}^J_+ : Q\xi + g = W^T \pi \}$

$$\inf \ c^T x + \sup_{\xi \in \Xi} Z(x,\xi)$$

$$
\begin{aligned}
= \inf \quad & c^T x + \tau \\
\text{s.t.} \quad & x \in X, \ \tau \in \mathbb{R} \\
& \sup_{\xi \in \Xi} \max_{r \in [R]} (T(x)\xi + h(x))^T \pi \le \tau
\end{aligned}
$$

$$
\begin{aligned}
& \Leftrightarrow \max_{r \in [R]} \sup_{\xi \in \Xi} (T(x)\xi + h(x))^T \pi^r \le \tau \\
& \Leftrightarrow \sup_{\xi \in \Xi} (T(x)\xi + h(x))^T \pi^r \le \tau, \ \forall r \in [R] \qquad S\xi \le t \\
& \Leftrightarrow \left\{ \begin{aligned} \sup_{\xi \in \mathbb{R}^K} \quad & (\pi^r)^T T(x)\, \xi \\ \text{s.t.} \quad & S\xi \le t \end{aligned} \right\} \le \tau - h(x)^T \pi^r \\
& \Leftrightarrow \exists\, \theta^r \in \mathbb{R}^M_+ : \left. \begin{aligned} & (\theta^r)^T t \le \tau - h(x)^T \pi^r \\ & S^T \theta^r = T(x)^T \pi^r \end{aligned} \right\} \forall r \in [R]
\end{aligned}
$$

$$
\begin{aligned}
= \inf \quad & C^T x + \tau \\
\text{s.t.} \quad & x \in X, \ \tau \in \mathbb{R}, \ \theta^r \in \mathbb{R}^M_+, \ \forall r \in [R] \\
& \left. \begin{aligned} & (\theta^r)^T t + h(x)^T \pi^r \le \tau \\ & S^T \theta^T = T(x)^T \pi^r \end{aligned} \right\} \ \forall r \in [R]
\end{aligned}
$$

<u>Bender's decomposition algorithm</u> ...

$$
\begin{aligned}
\inf \quad & C^T x + \tau \qquad \qquad (2R) \\
\text{s.t.} \quad & x \in X, \ \tau \in \mathbb{R} \\
& \sup_{\xi \in \Xi} \sup_{\pi \in \Pi} (T(x)\xi + h(x))^T \pi \le \tau
\end{aligned}
$$

Lecture 21 Nov. 3.

$$\inf_{x \in X} c^T x + \sup_{\xi \in \Xi} Z(x, \xi) \qquad (2R)$$

where $Z(x,\xi) = \inf g^T y$

s.t. $y \in \mathbb{R}^{N_2}$

$T(x)\xi + h(x) \le Wy$

$= Z_d(x,\xi) = \sup \ (T(x)\xi + h(x))^T \pi$

s.t. $\pi \in \mathbb{R}_+^J$

$g = W^T \pi$

$$\inf_{\substack{x \in X \\ \tau \in \mathbb{R}}} c^T x + \tau \quad \text{s.t.} \sup_{\xi \in \Xi} Z(x,\xi) \le \tau$$

Bender's Decomposition Algorithm

- assume complete recourse

  $Z(x,\xi) < +\infty, \quad \forall x \in X, \ \forall \xi \in \Xi$

  can be extended to the case where $Z(x,\xi) = +\infty$ for som x & ξ
- can be extended to other DRO models.

-Algorithm:

Input: Parameter of (2R)

Output: Optimal solution x^* to (2R)

- Step 0: Let $U = \phi$ (current set of extreme points of Ξ)
- Step 1: Solve the master problem

  $\inf \ c^T x + \tau$

  s.t. $x \in X$, $\tau \in \mathbb{R}$

  $Z(x,\xi) \le t, \ \forall \xi \in U$ $\iff$ $\exists y^\xi \in \mathbb{R}^{N_2}$: $\left.\begin{array}{l} g^T y^\xi \le \tau \\ T(x)\xi + h(x) \le W y^\xi \end{array}\right\} \forall \xi \in U$

  substitute recourse function with its definition

  Give us optimal solution $(\hat{x}, \hat{\tau})$
- Step 2: Solve $\sup_{\xi \in \Xi} Z(\hat{x}, \xi)$

  hard to perform

  If $\sup_{\xi \in \Xi} Z(\hat{x}, \xi) \le \hat{\tau}$, terminate. The solution $x^* = \hat{x}$ is optimal to (2R).

  Otherwise, go to step 3.
- Step 3: Let $\hat{\xi} = \operatorname{argsup}_{\xi \in \Xi} \underbrace{Z(\hat{x}, \xi)}_{\text{convex in } \xi}$. Add $U = U \cup \{\hat{\xi}\}$. Go to step 1.

Step 2 is hard but admits a MILP reformulation.

$$\sup_{\varepsilon \in \Xi} \ \inf \ g^T y \quad \text{s.t. } y \in \mathbb{R}^{N_2}, \quad T(\hat{x})\varepsilon + h(\hat{x}) \le Wy$$

— For any $\varepsilon \in \Xi$, from Karush - Kuhn - Tucker (KKT) conditions:

y is optimal $\iff$ $T(\hat{x})\varepsilon + h(\hat{x}) \le Wy$

$\exists \pi \in \mathbb{R}^J_+, \ g = W^T\pi$

$0 \le Wy - T(\hat{x})\varepsilon - h(\hat{x}) \quad \pi$

$$e_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j\text{th row}$$

(*) $\pi_j e_j^T (Wy - T(\hat{x})\varepsilon - h(\hat{x})) = 0$ complementary slackness.

Reformulate (*) as MILP:

Introduce a binary vector: $\exists z \in \{0,1\}^J$:

$$\left.\begin{aligned} e_j^T(Wy - T(\hat{x})\varepsilon - h(x)) &\le Mz_j \\ \pi_j &\le M(1-z_j) \end{aligned}\right\} \ \forall j \in [J]$$

Where M is a large constant.

— $\sup_{\varepsilon \in \Xi} Z(\hat{x}, \varepsilon) = \sup \ g^T y$

$$\begin{aligned} \text{s.t. } & \varepsilon \in \mathbb{R}^k, \ y \in \mathbb{R}^{N_2}, \ \pi \in \mathbb{R}^J_+, \ z \in \{0,1\}^J \\ & T(x)\varepsilon + h(x) \le Wy \\ & g = W^T\pi \\ & e_j^T(Wy - T(x)\varepsilon - h(x)) \le Mz_j \quad \forall j \in [J] \\ & \pi_j \le M(1-z_j) \end{aligned}$$

$$\inf_{x \in X} \ c^T x + \sup_{\varepsilon \in \Xi} Z(x, \varepsilon) \qquad (2R)$$

where: $Z(x, \varepsilon) = \inf \ Q\varepsilon + g$

$$\begin{aligned} \text{s.t. } & y \in \mathbb{R}^{N_2} \\ & T(x)\varepsilon + h(x) \le Wy \end{aligned}$$

$$= Z_d(\pi, \varepsilon) = \sup \ (T(x)\varepsilon + h(x))^T \pi \quad \text{s.t. } \pi \in \mathbb{R}^J_+, \ Q\varepsilon + g = W^T\pi$$

General Problem ($Q \ne 0$, & w.l.o.g. assume $\Xi \subseteq \mathbb{R}^k_+$)

$$\left.\begin{aligned} \inf \ & c^T x + \tau \\ \text{s.t. } & x \in X, \ \tau \in \mathbb{R} \\ & \sup_{\varepsilon \in \Xi} \ \sup \ (T(x)\varepsilon + h(x))^T \pi \le \tau \quad (*) \\ & \qquad \text{s.t. } \pi \in \mathbb{R}^J_+ \\ & \qquad Q\varepsilon + g = W^T\pi \end{aligned}\right\}$$

NP hard (quadratic

$\rightarrow$ copositive program

positive <u>Semidefinite cone</u> : $\mathbb{S}_+^k$

$$M \in \mathbb{S}_+^k \Leftrightarrow M \succeq 0 \Leftrightarrow \xi^T M \xi \geq 0, \ \forall \xi \in \mathbb{R}^k$$

<u>Copositive cone</u>: $\mathcal{C}$

$$M \in \mathcal{C} \Leftrightarrow M \succeq_{\mathcal{C}} 0 \Leftrightarrow \xi^T M \xi \geq 0, \ \forall \xi \in \mathbb{R}_+^k$$

$\mathbb{S}_+^k \subseteq \mathcal{C}$

Given $M \in \mathbb{S}^k$, checking $M \in \mathbb{S}_+^k$ is easy

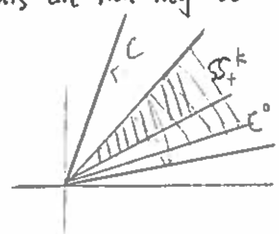
Given $M \in \mathbb{S}^k$, checking $M \in \mathcal{C}$ is NP-hard.

<u>Tractable inner approximation</u>

positive semidefinite, all elements are non-negative

if $M = P + N$, where $P \succeq 0$, $N \geq 0$

then $M \succeq_{\mathcal{C}} 0$



zero approximation of $\mathcal{C}$.

$\mathcal{C}^0 = \{M \in \mathbb{S}^k : M = P + N$

$\mathcal{C}^0 \subseteq \mathcal{C}^1 \subseteq \mathcal{C}^2 \cdots \subseteq \mathcal{C}$

<u>Lemma</u>:

$$\begin{bmatrix} A & \frac{1}{2}b \\ \frac{1}{2}b^T & 0 \end{bmatrix} \succeq_{\mathcal{C}} 0 \Leftrightarrow c + b^T\xi + \xi^T A \xi \geq 0, \ \forall \xi \in \mathbb{R}_+^k$$

<u>Notations</u>: $M_{j:}$: j-th row of matrix M.

$M_{:j}$: j-th column of M.

$$M = \begin{bmatrix} M_{1:}^T \\ \vdots \\ M_{k:}^T \end{bmatrix} = [M_{:1} \cdots M_{:L}]$$

$$(*) \quad \sup_{\substack{\xi \in \mathbb{R}_+^k \\ S\xi \leq t}} \ \sup \ (T(x)\xi + h(x))^T \pi \quad \leq \tau$$
$$\text{s.t.} \ \pi \in \mathbb{R}_+^J$$
$$Q\xi + g = W^T \pi$$

In order to combine two sup's, we should make them symmetric and the "$\leq$" should be "$=$" by slack variable

$$\Leftrightarrow \sup_{\substack{\xi \in \mathbb{R}_+^k \\ \rho \in \mathbb{R}_+^m \\ S\xi + \rho = t}} \ \sup \ (T(x)\xi + h(x))^T \pi \quad \leq \tau$$
$$\text{s.t.} \ \pi \in \mathbb{R}_+^J$$
$$Q\xi + g = W^T \pi$$

slack variable

$$\Leftrightarrow \sup \ (T(x)\xi + h(x))^T \pi$$
$$\text{s.t.} \ \xi \in \mathbb{R}_+^k, \ \rho \in \mathbb{R}_+^m, \ \pi \in \mathbb{R}_+^J$$
$$Q\xi + g = W^T \pi \quad \} \ N_2 \text{ rows}$$
$$S\xi - t = -\mathbb{I}\rho \quad \} \ m \text{ rows} \qquad \mathbb{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Define: $\bar{Q} = \begin{bmatrix} Q \\ S \end{bmatrix}$, $\bar{g} = \begin{bmatrix} g \\ -t \end{bmatrix}$, $\bar{W} = \begin{bmatrix} W & 0 \\ 0 & -\mathbb{I} \end{bmatrix}$ ← identity

$\bar{T}(x) = \begin{bmatrix} T(x) \\ 0 \end{bmatrix}$, $\bar{h}(x) = \begin{bmatrix} h(x) \\ 0 \end{bmatrix}$

$$\Leftrightarrow \sup \ (\bar{T}(x)\varepsilon + \bar{h}(x))^T \pi \quad \le \tau$$
$$\text{s.t.} \ \varepsilon \in \mathbb{R}_+^k, \ \pi \in \mathbb{R}_+^{M+1}$$
$$\bar{Q}\varepsilon + \bar{g} = \bar{W}^T \pi \ \} \ N_2 + M \text{ rows} \quad (v)$$

$$\begin{bmatrix} \bar{Q}_{1:}^T \\ \vdots \\ \bar{Q}_{N_2+M:}^T \end{bmatrix} \varepsilon + \begin{bmatrix} \bar{g}_1 \\ \vdots \\ \bar{g}_{N_2+M} \end{bmatrix} = [\bar{W}_{:1} \ \cdots \ \bar{W}_{:N_2+M}]^T \pi$$

Lecture 22. Nov. 8.

Slides Add $(\bar{Q}_{j:}^T \varepsilon - \bar{W}_{:j}^T \pi)^2 = \bar{g}^2$ redundant constraint

Lagrangian form:

$$\le \inf \ C^T x + \tau$$
$$\text{s.t.} \ \inf_{\psi, \varphi} \sup_{\varepsilon, \pi} \ (T(x)\varepsilon + h(x))^T \pi + \psi^T (Q\varepsilon + g - W\pi) + \varphi[(Q_{j:}^T \varepsilon - W_{:j}\pi)^2 - g^2]$$

...

Black board:

$$(v) \Leftrightarrow \bar{Q}_{l:}^T \varepsilon + \bar{g}_l = \bar{W}_{:l}\pi, \quad \forall l \in [N_2 + M]$$
$$\bar{g}_l^2 = (\bar{W}_{:l}^T \pi - \bar{Q}_{l:}^T \varepsilon)^2 = (\bar{Q}_{l:}^T \varepsilon)^2 - 2\bar{Q}_l \varepsilon \pi^T W_{:l} + (\bar{W}_{:l}^T \pi)^2$$
$$= \varepsilon^T \bar{Q}_{l:} \bar{Q}_{l:}^T \varepsilon - 2\varepsilon^T \bar{Q}_{l:} \bar{W}_{:l}^T \pi + \pi^T \bar{W}_{:l} \bar{W}_{:l}^T \pi, \quad \forall l \in [N_2 + M]$$

$$\Leftrightarrow \sup_{\varepsilon, \pi \ge 0} \inf_{\psi, \phi} \ (\bar{T}(x)\varepsilon + \bar{h}(x))^T \pi + \psi^T (\bar{Q}\varepsilon + \bar{g} - \bar{W}^T \pi)$$
$$+ \sum_{l \in [N_2+M]} \phi_l (\bar{g}_l^2 - \varepsilon^T \bar{Q}_{l:} \bar{Q}_{l:}^T \varepsilon + 2\varepsilon^T \bar{Q}_{l:} \bar{W}_{:l}^T \pi - \pi^T \bar{W}_{:l} \bar{W}_{:l}^T \pi) \le \tau$$

$$\Leftrightarrow \inf_{\psi, \phi} \sup_{\substack{\varepsilon \in \mathbb{R}_+^k \\ \pi \in \mathbb{R}_+^{M+1}}} \ \cdots \le \tau$$

$$\Leftrightarrow \exists \psi, \phi: \quad 0 \le \begin{bmatrix} \varepsilon \\ \pi \\ 1 \end{bmatrix}^T \begin{bmatrix} \sum_{l \in [N_2+M]} \phi_l \bar{Q}_{l:} \bar{Q}_{l:}^T & -\frac{1}{2}\bar{T}(x)^T - \sum_{l \in [N_2+M]} \phi_l \bar{Q}_{l:} \bar{W}_{:l}^T & -\frac{1}{2} \bar{Q}_{l:}^T \psi \\ * & \sum_{l \in [N_2+M]} \phi_l \bar{W}_{:l} W_{:l}^T & \frac{1}{2}(\bar{W}\psi - \bar{h}(x)) \\ * & * & \tau \end{bmatrix} \begin{bmatrix} \varepsilon \\ \pi \\ 1 \end{bmatrix}$$

↑ symmetric matrix

$\forall \varepsilon, \pi \ge 0$

$$(2R) \Leftrightarrow \inf \ C^T x + \tau$$
$$\text{s.t.} \ x \in X, \ \tau \in \mathbb{R}, \ \psi, \phi \in \mathbb{R}^{N_2+1}$$
$$[\quad\quad] \succcurlyeq_c 0$$

Lecture 23

Wasserstein ambiguity set

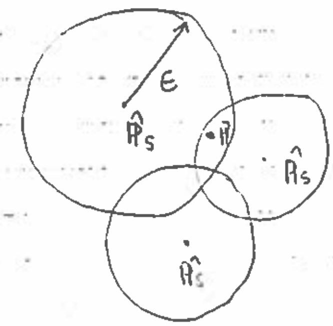
Use the data directly to construct the ambiguity set

given $\xi^1, \cdots, \xi^S$ independent samples from the true dist. $\mathbb{P}$

Empirical dist.: $\hat{\mathbb{P}}_S = \frac{1}{S} \sum_{s \in [S]} \delta_{\xi^s}$

matrix

Wasserstein ball: $B_\epsilon(\hat{\mathbb{P}}_S) = \{ \mathbb{P} \in \mathcal{P}_0(\Xi) : W(\mathbb{P}, \hat{\mathbb{P}}_S) \le \epsilon \}$



$\text{Prob}(\mathbb{P} \in B_\epsilon(\hat{\mathbb{P}}_S)) \ge 1 - \rho$

Location of the ball changes according to the samples

True dist. has high prob. contained in the ball

$\mathbb{P}^S = \underbrace{\mathbb{P} \times \cdots \times \mathbb{P}}_{S \text{ times}}$

$\epsilon = g(\rho)$ For any ρ, we can always find ϵ which satisfy the condition.

It's hard to derive. We will focus more on how to derive formulation using the ambiguity set.

If $\mathbb{P} \in B_\epsilon(\hat{\mathbb{P}}_S)$: $\mathbb{E}_{\mathbb{P}}[f_0(x, \tilde{\xi})] \le \sup_{\mathbb{P} \in B_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}[f_0(x, \tilde{\xi})] \quad \forall x$

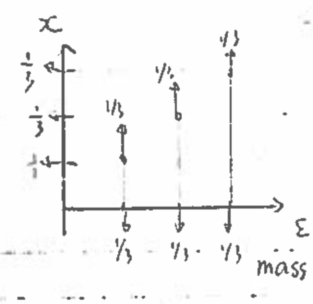
$\inf_{x \in X} \sup_{\mathbb{P} \in B_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}[f_0(x, \tilde{\xi})] = \hat{J}_S \rightarrow$ certificate

If $\mathbb{P} \in B_\epsilon(\hat{\mathbb{P}}_S)$: $\mathbb{E}_{\mathbb{P}}[f_0(\hat{x}_S, \tilde{\xi})] \le \hat{J}_S$

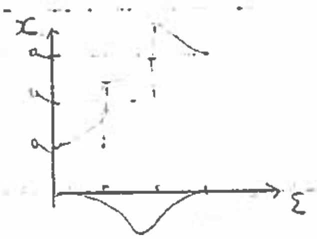
$\text{Prob}(\mathbb{E}_{\mathbb{P}}[f_0(\hat{x}_S, \tilde{\xi})] \le \hat{J}_S) \ge 1 - \rho$

$W = (\mathbb{P}, \mathbb{P}') = \inf \int_{\Xi \times \Xi} \|\xi - x\| \, \Pi(d\xi, dx)$

s.t. Π is a joint distribution of $\tilde{\xi}$ & $\tilde{x}$ with marginal dist. $\mathbb{P}$ & $\mathbb{P}'$, respectively.



Assume x and ξ have the same discrete dist.



Assume x has discrete dist.

ξ have continuous ξ.

fix x : $\sup\limits_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}[f_0(x,\tilde{\xi})]$ (*)

$= \sup\limits_{\pi,\mathbb{P}} \mathbb{E}_\pi[f_0(x,\tilde{\xi})]$

$W(\mathbb{P},\hat{\mathbb{P}}_S) \le \epsilon$

π has marginals $\mathbb{P}$ & $\hat{\mathbb{P}}_S$

$= \sup \mathbb{E}_\pi[f_0(x,\tilde{\xi})]$

s.t. $\mathbb{E}_\pi[\|\tilde{\xi}-\tilde{\chi}\|] \le \epsilon$

π has marginals $\mathbb{P}$ & $\hat{\mathbb{P}}_S$

$\tilde{\xi}\sim\mathbb{P}$, $\tilde{\chi}\sim\hat{\mathbb{P}}_S$

$$\pi(\tilde{\xi}\in A,\ \tilde{\chi}=\xi^s) = \pi(\tilde{\xi}\in A \mid \tilde{\chi}=\xi^s)\cdot\pi(\tilde{\chi}=\xi^s)$$

Bayes' Theorem

$$= \underbrace{\pi(\tilde{\xi}\in A \mid \tilde{\chi}=\xi^s)}_{Q_s(\tilde{\xi}\in A)}\cdot\frac{1}{S}$$

Law of total expectation

$$\mathbb{E}_\pi[f_0(x,\tilde{\xi})] = \sum_{s\in[S]} \underbrace{\mathbb{E}_\pi[f_0(x,\tilde{\xi}) \mid \hat{\chi}=\xi^s]}_{\mathbb{E}_{Q_s}[f_0(x,\xi)]}\ \underbrace{\pi(\hat{\chi}=\xi^s)}_{\frac{1}{S}}$$

$$= \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{Q_s}[f_0(x,\xi)]$$

$$(*) = \sup_{Q_s,\,s\in[S]} \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{Q_s}[f_0(x,\tilde{\xi})]$$

$$\text{s.t. } \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{Q_s}[\|\tilde{\xi}-\xi^s\|] \le \epsilon$$

$$= \sup_{Q_s,\,s\in[S]} \frac{1}{S}\sum_{s\in[S]} \int_\Xi f_0(x,\xi)\,\mu_s(d\xi)$$

s.t. $\mu_s(\cdot)\ge 0$, $\forall s\in[S]$

λ $\quad \frac{1}{S}\sum_{s\in[S]} \int \|\xi-\xi'\|\,\mu_s(d\xi) \le \epsilon$

α_s $\quad \int_\Xi \mu_s(ds) = 1$, $\forall s\in[S]$

$$= \sup_{\mu_s(\cdot)\ge 0}\ \inf_{\substack{\lambda\ge 0\\ \alpha_s}} \frac{1}{S}\sum_{s\in[S]}\int_\Xi f_0(x,\xi)\,\mu_s(d\xi) + \lambda\epsilon - \frac{\lambda}{S}\sum_{s\in[S]}\int_\Xi \|\xi-\xi'\|\,\mu_s(d\xi) + \frac{1}{S}\sum_{s\in[S]}\alpha_s - \frac{1}{S}\sum_{s\in[S]}\int_\Xi \alpha_s\,\mu_s(d\xi)$$

$$= \inf_{\substack{\lambda\ge 0\\ \alpha_s}}\ \sup_{\mu_s(\cdot)\ge 0} \cdots$$

$$= \inf\ \lambda\epsilon + \frac{1}{S}\sum_{s\in[S]}\alpha_s$$

s.t. $\lambda\in\mathbb{R}_+$, $\alpha\in\mathbb{R}^S$

$$f_0(x,\xi)\ \underbrace{-\lambda\|\xi-\xi^s\|}_{\text{Concave in }\xi} \le \alpha\ ,\quad \forall s\in[S],\ \forall \xi\in\Xi \qquad (v)$$

If concave in ξ or max of concave in ξ.

concave quatric function

$$f_0(x,\xi) = \max_{j\in[J]} a_j(x)^T\xi + b_j(x) - \xi^T T_j \xi \qquad \Xi = \{\xi \in \mathbb{R}^k : \xi S \le t\}$$

Exam

$$(V) \Leftrightarrow \sup_{\xi\in\Xi} a_j(x)^T\xi + b_j(x) - \xi^T \Gamma \xi - \lambda\|\xi - \xi^s\| \le \alpha_s, \quad \forall s\in[S], \ \forall j\in[J]$$

Assume $\|\ \|_1$: $\sup a_j(x)^T\xi + b_j(x) - \xi^T\Gamma\xi - \lambda\rho^T e \quad \le \alpha_j$

s.t. $S\xi \le t$, ρ

$\rho \geq \xi - \xi^s$

$\rho \geq \xi^s - \xi$

$$\sup_{\xi,\rho} \inf_{\substack{\gamma\ge0\\ \theta\ge0\\ \eta\ge0}} a_j(x)^T\xi + b_j(x) - \xi^T\Gamma\xi - \lambda\rho^T e + \gamma^T t - \gamma^T S\xi + \theta^T\rho - \theta^T\xi + \theta^T\xi^s + \eta^T\rho - \eta^T\xi^s + \eta^T\xi \le \alpha_j$$

$\exists \gamma\ge0, \theta\ge0, \eta\ge0$, $\cdots \le \alpha_j$, $\forall\xi, \forall\rho$

$$f_0(x,\xi) = \max_{j\in[J]} a_j(x)^T\xi + b_j(x)$$

$$(V) \Leftrightarrow \sup_{\xi\in\Xi} a_j(x)^T\xi + b_j(x) - \lambda\|\xi - \xi'\|_1 \le \alpha_s, \quad \forall s\in[S], \ \forall j\in[J]$$

$$\|c\| = \sup_{\|\xi\|_* \le 1} c^T\xi \qquad \Rightarrow \qquad \lambda\|c\| = \sup_{\|\xi\|_*\le\lambda} c^T\xi$$

dual norm

$$\Leftrightarrow \sup_{\xi\in\Xi} a_j(x)^T\xi + b_j(x) - \sup_{\theta_{js}: \|\theta_{js}\|_* \le \lambda} \theta_{js}^T(\xi - \xi') \le \alpha_s, \quad \forall s\in[S], \ \forall j\in[J]$$

Research Project.

- 2-stage problem solved using Wasserstein model
  - $\rightarrow$ copostive program $O(S)$ copositive constraints
- Benders decomposition.

Lecture 24.

$$\sup_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_s)} \mathbb{E}_{\mathbb{P}}[f_0(x,\tilde{\xi})] = \inf \lambda\epsilon + \frac{1}{S}\sum_{s\in[S]}\alpha_s$$

s.t. $\lambda\in\mathbb{R}_+$, $\alpha\in\mathbb{R}^S$

$f_0(x,\xi) - \lambda\|\xi - \xi^s\| \le \alpha_s$, $\forall s\in[S]$, $\forall \xi\in\Xi$ $(*)$

$$f_0(x,\xi) = \max_{j\in[J]} a_j(x)^T\xi + b_j(x), \quad \Xi = \{\xi\in\mathbb{R}^k : S\xi\le t\}$$

$(*) \Leftrightarrow a_j(x)^T\xi + b_j(x) - \lambda\|\xi - \xi^s\| \le \alpha_s$, $\forall s\in[S]$, $\forall j\in[J]$, $\forall s\in\Xi$

$\Leftrightarrow \sup_{\xi: S\xi\le t} a_j(x)^T\xi + b_j(x) - \lambda\|\xi - \xi^s\| \le \alpha_s$, $\forall s\in[S]$, $\forall j\in[J]$

<u>Dual Norm</u>

$$\|c\| = \sup_{\xi} c^T\xi \quad \text{s.t. } \|\xi\|_* \le 1 \qquad\qquad \lambda\|c\| = \sup_{\xi} c^T\xi \quad \text{s.t. } \|\xi\|_* \le \lambda$$

$\lambda \geq 0$

If $\lambda = 0$ ✓

norm is positive homogeneous

If $\lambda > 0$, $\lambda \|C\| = \|\lambda C\| = \sup \ C^T(\lambda\varepsilon)$ s.t. $\|\varepsilon\|_* \le 1$

Let $\varepsilon' = \lambda\varepsilon$, $= \sup \ C^T\varepsilon$ s.t. $\|\varepsilon'\|_* \le \lambda$

$$(*) \Leftrightarrow \sup_{\varepsilon: S\varepsilon \le t} a_j(x)^T\varepsilon + b_j(x) - \sup_{\theta_{js}: \|\theta_{js}\|_* \le \lambda} - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \sup_{\varepsilon: S\varepsilon \le t} a_j(x)^T\varepsilon + b_j(x) + \inf_{\theta_{js}: \|\theta_{js}\|_* \le \lambda} - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \sup_{\varepsilon: S\varepsilon \le t} \inf_{\theta_{js}: \|\theta_{js}\|_* \le \lambda} a_j(x)^T s + b_j(x) - \theta_{js}^T(\varepsilon - s^s) \le \alpha_s$$

change is allowed since sup problem is convex

$$\Leftrightarrow \exists \|\theta_{js}\|_* \le \lambda : \sup_{\varepsilon: S\varepsilon \le t} a_j(x)^T\varepsilon + b_j(x) - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \exists \|\theta_{js}\|_* \le \lambda : \inf_{\eta_{js} \ge 0} \eta_{js}^T t + b_j(x) + \theta_{js}^T \varepsilon^s \le \alpha_s$$
$$S^T\eta_{js} + \theta_{js} = a_j(x)$$

$$\Leftrightarrow \exists \eta_{js} \ge 0, \theta_{js} : \eta_{js}^T t + b_j(x) + \theta_{js}^T\varepsilon^s \le \alpha_s$$
$$\|\theta_{js}\|_* \le \lambda$$
$$S^T\eta_{js} + \theta_{js} = a_j(x)$$

$$\inf_x \sup_{\mathbb{P} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_s)} \mathbb{E}_\mathbb{P}\left[\max_{j \in [J]} a_j(x)^T\varepsilon + b_j(x)\right] = \inf \ \lambda\epsilon + \frac{1}{S}\sum_{s \in [S]} \alpha_s$$

s.t. $\lambda \in \mathbb{R}_+$, $\alpha \in \mathbb{R}^S$, $x \in X$

$\theta_{js} \in \mathbb{R}^k$, $\eta_{js} \in \mathbb{R}_+^m$, $\forall j \in [J]$, $\forall s \in [S]$

$\|\theta_{js}\|_* \le \lambda$

$S^T\eta_{js} + \theta_{js} = a_j(x)$

$\eta_{js}^T t + b_j(x) + \theta_{js}^T\varepsilon^s \le \alpha_s$

Fix x, dualize λ, α:

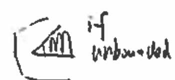
$$(*) \sup_{\mathbb{P} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_s)} \mathbb{E}_\mathbb{P}\left[\max_{j \in [J]} a_j(x)^T\hat{\varepsilon} + b_j(x)\right] = \sup \ \frac{1}{S}\sum_{s \in [S]}\sum_{j \in [J]} a_j(x)^T X_{js} + P_{js} b_j(x)$$

s.t. $P_{js} \in \mathbb{R}_+$, $X_{js} \in \mathbb{R}^k$, $\forall j \in [J]$, $\forall s \in [S]$

$\sum_{j \in [J]} P_{js} = 1$, $\forall s \in [S]$

(∨) $S X_{js} \le P_{js} t$, $\forall j \in [J]$, $\forall s \in [S]$.

$\frac{1}{S}\sum_{s \in [S]}\sum_{j \in [J]} \| X_{js} - P_{js}\varepsilon^s \| \le \epsilon$.



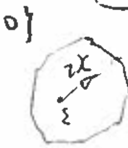
<u>Worst-case Distribution</u> (assume Ξ is bounded)

<u>Lemma</u>: If Ξ is bounded, then $\text{recc}(\Xi) = \{\varepsilon \in \mathbb{R}^k : S\varepsilon \le 0\} = \{0\}$

recession cone

Proof: If $\exists X \ne 0$, $SX \le 0$, then $S(\varepsilon + \tau X) \le t$, $\forall \tau \ge 0$

contradic with unbounded Ξ.



Assume Ξ is bounded & (X_{js}^*, P_{js}^*) is optimal.

If $P_{js}^* = 0$, then $S X_{js} \le 0 \rightarrow X_{js} = 0 \Rightarrow$ ignore all induces j & s for whic $P_{js}^* = 0$.

(∨)

Assume $P_{js}^* > 0$, $\forall j, s$.

Define discrete dist. $\mathbb{P}^*$ of $\hat{\varepsilon}$: $\mathbb{P}^*\left(\tilde{\varepsilon} = \frac{X_{js}^*}{P_{js}^*}\right) = \frac{P_{js}^*}{S}$, $\forall j \in [J]$, $\forall s \in [S]$

(iv) $\Rightarrow S\dfrac{x_{js}^*}{p_{js}^*} \le t \;\Rightarrow\; \dfrac{x_{js}^*}{p_{js}^*} \in \Xi = \{\xi \in \mathbb{R}^k : S\xi \le t\}$

$$\sum_{s\in[S]} \sum_{j\in[J]} \frac{p_{js}^*}{S} = \frac{1}{S}\sum_{s\in[S]}\sum_{j\in[J]} p_{js}^* = \frac{1}{S}\sum_{s\in[S]} 1 = 1$$ we indeed have a prob dist.

$$\sum_{s\in[S]}\sum_{j\in[J]} \frac{p_{js}^*}{S} \left\| \frac{x_{js}^*}{p_{js}^*} - \xi^s \right\| = \frac{1}{S}\sum_{s\in[S]}\sum_{j\in[J]} \| x_{js}^* - p_{js}^* \xi^s \| \le \epsilon$$

$$\sum_{s\in[S]}\sum_{j\in[J]} \frac{p_{js}^*}{S} \left(\max_{l\in[J]} a_l(x)^T \frac{x_{js}^*}{p_{js}^*} + b_l(x) \right)$$ ← under the dist. we just constructed $\mathbb{P}^{(*)}$ the obj ≥ than actual

$$\ge \sum_{s\in[S]}\sum_{j\in[J]} \frac{p_{js}^*}{S} \left(a_j(x)^T \frac{x_{js}^*}{p_{js}^*} + b_j(x) \right)$$

$$= \frac{1}{S}\sum_{s\in[S]}\sum_{j\in[J]} \left(a_j(x)^T x_{js}^* + b_j(x) \right) = \sup_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}\left[\max_{j\in[J]} a_j(x)^T \tilde{\xi} + b_j(x)\right]$$

Slides: "Optimizer's Curse"

SAA: sample average approximation

If $\lambda = 0$ ✓

If $\lambda > 0$, $\lambda \|c\| = \|\lambda c\|$ (norm is positive homogeneous) $= \sup \; c^T(\lambda \varepsilon)$ s.t. $\|\varepsilon\|_* \le 1$

Let $\varepsilon' = \lambda \varepsilon$, $= \sup \; c^T \varepsilon$ s.t. $\|\varepsilon'\|_* \le \lambda$

$$(*) \Leftrightarrow \sup_{\varepsilon : S\varepsilon \le t} a_j(x)^T \varepsilon + b_j(x) - \sup_{\theta_{js} : \|\theta_{js}\|_* \le \lambda} - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \sup_{\varepsilon : S\varepsilon \le t} a_j(x)^T \varepsilon + b_j(x) + \inf_{\theta_{js} : \|\theta_{js}\|_* \le \lambda} - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \sup_{\varepsilon : S\varepsilon \le t} \inf_{\theta_{js} : \|\theta_{js}\|_* \le \lambda} a_j(x)^T \varepsilon + b_j(x) - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

change is allowed since sup problem is convex

$$\Leftrightarrow \exists \|\theta_{js}\|_* \le \lambda : \sup_{\varepsilon : S\varepsilon \le t} a_j(x)^T \varepsilon + b_j(x) - \theta_{js}^T(\varepsilon - \varepsilon^s) \le \alpha_s$$

$$\Leftrightarrow \exists \|\theta_{js}\|_* \le \lambda : \inf_{\substack{\eta_{js} \ge 0 \\ S^T\eta_{js} + \theta_{js} = a_j(x)}} \eta_{js}^T t + b_j(x) + \theta_{js}^T \varepsilon^s \le \alpha_s$$

$$\Leftrightarrow \exists \eta_{js} \ge 0, \theta_{js} : \eta_{js}^T t + b_j(x) + \theta_{js}^T \varepsilon^s \le \alpha_s$$
$$\|\theta_{js}\|_* \le \lambda$$
$$S^T \eta_{js} + \theta_{js} = a_j(x)$$

$$\inf_x \sup_{\mathbb{P} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}\left[\max_{j \in [J]} a_j(x)^T \varepsilon + b_j(x)\right] = \inf \; \lambda \epsilon + \frac{1}{S} \sum_{s \in [S]} \alpha_s$$

s.t. $\lambda \in \mathbb{R}_+$, $\alpha \in \mathbb{R}^S$, $x \in \mathcal{X}$

$\theta_{js} \in \mathbb{R}^K$, $\eta_{js} \in \mathbb{R}_+^m$, $\forall j \in [J]$, $\forall s \in [S]$

$\|\theta_{js}\|_* \le \lambda$

$S^T \eta_{js} + \theta_{js} = a_j(x)$

$\eta_{js}^T t + b_j(x) + \theta_{js}^T \varepsilon^s \le \alpha_s$

Fix x, dualize λ, α:

$$(*) \sup_{\mathbb{P} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}\left[\max_{j \in [J]} a_j(x)^T \hat{\varepsilon} + b_j(x)\right] = \sup \; \frac{1}{S} \sum_{s \in [S]} \sum_{j \in [J]} a_j(x)^T \chi_{js} + p_{js} b_j(x)$$

s.t. $p_{js} \in \mathbb{R}_+$, $\chi_{js} \in \mathbb{R}^K$, $\forall j \in [J]$, $\forall s \in [S]$

$\sum_{j \in [J]} p_{js} = 1$, $\forall s \in [S]$

(V) $S \chi_{js} \le p_{js} t$, $\forall j \in [J]$, $\forall s \in [S]$.

$\frac{1}{S} \sum_{s \in [S]} \sum_{j \in [J]} \|\chi_{js} - p_{js} \varepsilon^s\| \le \epsilon$.

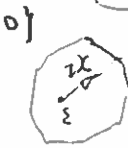
Worst-case Distribution (assume Ξ is bounded)

recession cone

Lemma: If Ξ is bounded, then $\text{recc}(\Xi) = \{\varepsilon \in \mathbb{R}^K : S\varepsilon \le 0\} = \{0\}$

Proof: If $\exists \chi \ne 0$, $S\chi \le 0$, then $S(\varepsilon + \tau \chi) \le t$, $\forall \tau \ge 0$

contradict with unbounded Ξ.



Assume Ξ is bounded & (χ_{js}^*, p_{js}^*) is optimal.

If $p_{js}^* = 0$, then $S\chi_{js} \le 0$ (V) $\to \chi_{js} = 0$ $\Rightarrow$ ignore all induces j & s, for which $p_{js}^* = 0$.

Assume $p_{js}^* > 0$, $\forall j, s$,

Define discrete dist. $\mathbb{P}^*$ of $\hat{\varepsilon}$: $\mathbb{P}^*\left(\tilde{\varepsilon} = \frac{\chi_{js}^*}{p_{js}^*}\right) = \frac{p_{js}^*}{S}$, $\forall j \in [J]$, $\forall s \in [S]$

$(v) \Rightarrow S \dfrac{x_{js}^*}{p_{js}^*} \le t \Rightarrow \dfrac{x_{js}^*}{p_{js}^*} \in \Xi = \{\xi \in \mathbb{R}^k : S\xi \le t\}$

$$\sum_{s\in[S]} \sum_{j\in[J]} \frac{p_{js}^*}{S} = \frac{1}{S} \sum_{s\in[S]} \sum_{j\in[J]} p_{js}^* = \frac{1}{S} \sum_{s\in[S]} 1 = 1$$ we indeed have a prob dist.

$$\sum_{s\in[S]} \sum_{j\in[J]} \frac{p_{js}^*}{S} \left\| \frac{x_{js}^*}{p_{js}^*} - \xi^s \right\| = \frac{1}{S} \sum_{s\in[S]} \sum_{j\in[J]} \| x_{js}^* - p_{js}^* \xi^s \| \le \epsilon$$

$$\sum_{s\in[S]} \sum_{j\in[J]} \frac{p_{js}^*}{S} \left(\max_{l\in[J]} a_l(x)^T \frac{x_{js}^*}{p_{js}^*} + b_l(x) \right)$$ ← under the dist. we just constructed $\mathbb{P}^*$ the obj ≥ than actual.

$$\ge \sum_{s\in[S]} \sum_{j\in[J]} \frac{p_{js}^*}{S} \left(a_j(x)^T \frac{x_{js}^*}{p_{js}^*} + b_j(x) \right)$$

$$= \frac{1}{S} \sum_{s\in[S]} \sum_{j\in[J]} \left(a_j(x)^T x_{js}^* + b_j(x) \right) = \sup_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_S)} \mathbb{E}_{\mathbb{P}}\left[\max_{j\in[J]} a_j(x)^T \tilde{\xi} + b_j(x)\right]$$

Slides: "Optimizer's Curse"

SAA. sample average approximation

Lecture 25. Nov. 22

<u>Multi-stage stochastic programs</u>

$$\underset{\substack{\downarrow \\ c_1^T x_1}}{\overset{\in X}{\boxed{x_1}}} \to \xi_2 \to \underset{\substack{\downarrow \\ c_2^T x_2}}{\overset{\in\{x_2 : T_2(x_1)\xi_2 + h_2(x_1) \le W_2 x_2\}}{\boxed{x_2}}} \to \xi_3 \to \underset{\in\{x_3,\ T_3(x_2)\xi_3 + h_3(x_2) \le W_3 x_3\}}{\boxed{x_3}} \to \cdots \to \xi_T \to \boxed{x_T}$$

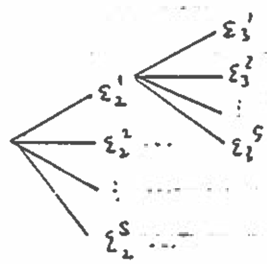
Assume $\xi_2, \dots, \xi_T$ are discrete random variables.

& they are stagewise independent.

$$\inf\ c^T x_1 + \mathbb{E}_{\xi_2}\left[\begin{array}{l} \inf\ c_2^T x_2 + \mathbb{E}_{\xi_3}\left[\begin{array}{l} \inf\ c_3^T x_3 + \mathbb{E}_{\xi_4}\left[\cdots \mathbb{E}_{\xi_T}\left[\begin{array}{l} \inf\ c_T^T x_T \\ \text{s.t. } T_T(x_{T-1})\tilde{\xi}_T + h_T(x_{T-1}) \\ \quad \le W_T x_T \end{array} \right. \right. \\ \text{s.t.}\ \cdots \end{array} \right. \\ \text{s.t. } T_2(x_1)\tilde{\xi}_2 + h_2(x_1) \le W_2 x_2 \end{array} \right.$$

s.t. $x_1 \in X$

How to solve:

— Scenario tree approximation 

$\Rightarrow$ Big LP $O(S^T)$ exponential !

number of stages (T); number of realization (S)

— decision rules approximation ⇒ tractable but cannot solve the problem exactly

— dynamic programming ⇒ stochastic dual dynamic programming ! (Perreira & Pinto 1991)
(general intractable) (SDDP)

Dynamic Programming

Start from $t=T$ and proceed backward until $t=1$.

Let $V_T(x_{T-1}, \xi_T) = \inf \; c_T^T x_T$

Decision from previous stage

realization for making this stage decision

$\quad$ s.t. $T_T(x_{T-1})\xi_T + h_T(x_{T-1}) \le W_T x_T$

solve for all $x_{T-1} \in \mathbb{R}^{N_{T-1}}$ & $\xi_T \in \Xi_T$

$$V_{T-1}(x_{T-2}, \xi_{T-1}) = \inf_{x_{T-1}} \; c_{T-1}^T x_{T-1} + \underline{\mathbb{E}_{\xi_T}[V_T(x_{T-1}, \tilde{\xi}_T)]}$$
$$\text{s.t. } T_{T-1}(x_{T-2})\xi_{T-1} + h_{T-1}(x_{T-2}) \le W_{T-1}x_{T-1}$$

solve for all $x_{T-2} \in \mathbb{R}^{N_{T-2}}$ & $\xi_{T-1} \in \Xi_{T-1}$

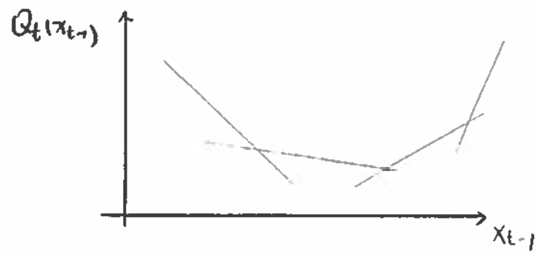
expectation of next step

Let $Q_t(x_{t-1}) = \mathbb{E}_{\xi_t}[V_t(x_{t-1}, \tilde{\xi}_t)] = \frac{1}{S}\sum_{s\in[S]} V_t(x_{t-1}, \xi_t^s)$

$$V_{T-1}(x_{T-2}, \xi_{T-1}) = \inf \; c_{T-1}^T x_{T-1} + Q_T(x_{T-1})$$
$$\text{s.t. } T_{T-1}(x_{T-2})\xi_{T-1} + h_{T-1}(x_{T-2}) \le W_{T-1}x_{T-1}$$
$$\vdots$$
$$V_t(x_{t-1}, \xi_t) = \inf \; c_t^T x_t + Q_{t+1}(x_t)$$
$$\text{s.t. } T_t(x_{t-1})\xi_t + h_t(x_{t-1}) \le W_t x_t$$
$$\vdots$$
$$V_2(x_1, \xi_2) = \cdots$$
$$\inf_{x_1\in X} c_1^T x_1 + \frac{1}{S}\sum_{s\in[S]} V_2(x_1, \xi^s)$$

$O(TS)$ problems. $\to O(|\mathbb{R}^{N_2}| \times \cdots \times |\mathbb{R}^{N_{T-1}}| \, ST)$ "curse of dimensionality"

In SDDP, we discretize $\mathbb{R}^{N_2} \times \cdots \times \mathbb{R}^{N_{T-1}}$ in a systematic way.
& approximate $Q_t(\cdot)$ with a piecewise affine function.



SDDP for two-stage problems

$$\inf_{x\in X} c_1^T x_1 + \frac{1}{S}\sum_{s\in[S]} \inf\{c_2^T x_2 : T_2(x_1)\xi_2^s + h_2(x_1) \le W_2 x_2\}$$

$$= \inf_{x\in X} c_1^T x_1 + Q_2(x_1)$$

$$\inf \; c_2^T x_2 \quad \text{s.t. } T_2(x_1)\xi_2^s + h_2(x_1) \le W_2 x_2 \quad = \quad \sup \; (T_2(x_1)\xi_2^s + h_2(x_1))^T \pi \quad \text{s.t. } \pi \ge 0, \; c_2 = W_2^T \pi$$

Algorithm

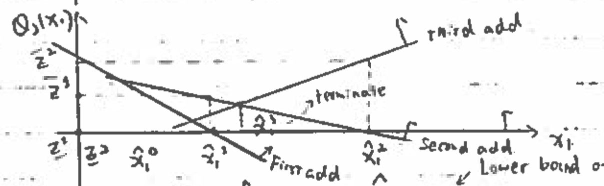
Input: A candidate solution $\hat{x}_1 \in X$ & tolerance ϵ

Output: ϵ - optimal solution $x_1^* \in X$

- Step 1: let $\hat{Q}_2(\cdot) = 0$
- Step 2: Solve for all $s \in [S]$

$$\pi^s \in \arg\sup \ (T_2(\hat{x}_1)\ \xi_2^s + h_2(\hat{x}_1))^T \pi$$
$$\text{s.t. } \pi^s \geq 0$$
$$c_2 = W_2^T \pi$$

Update $\hat{Q}_2(\,) = \max\{ \hat{Q}_2(), \frac{1}{S} \sum_{s \in [S]} (T_2(\cdot)\xi_2^s + h_2(\cdot))^T \pi^s \}$



Lower bound of true z since $\hat{Q}_2()$ is lower bound

- Step 3: Solve $\underline{z} = \inf_{x \in X} c_1^T x_1 + \hat{Q}_2(x_1)$

update $\hat{x}_1 \in \arg\inf_{x_1 \in X} c_1^T x_1 + \hat{Q}_2(x_1)$

- Step 4: we know that $\hat{x}_1$ is a suboptimal feasible solution

evaluate $\bar{z} = c_1^T \hat{x}_1 + Q_2(\hat{x}_1)$ ← the actual obj

If: $\bar{z} - \underline{z} \leq \epsilon$, $x_1^* = \hat{x}_1$ terminate.

else, go to Step 2

SDDP for multistage problems

points (not solution, solution should be a function in ξ)

Input: candidate solutions $\hat{x}_1, \cdots \hat{x}_{T-1}$ & tolerance ϵ

Output: ϵ - optimal solution $x_1^* \in X$

- Step 1: Set $\hat{Q}_t(\cdot) = 0$, for $t = 2, \cdots, T$
- Step 2: $t = T$ solve $\forall s \in [S]$ generate cuts

Backward step ↑

$$\pi^s \in \arg\sup \ (T_T(\hat{x}_{T-1})\ \xi_T^s + h_T(\hat{x}_{T-1}))^T \pi$$
$$\text{s.t. } \pi \geq 0$$
$$c_T = W_T^T \pi$$

Update: $\hat{Q}_T(\cdot) = \max\{ \hat{Q}_T(\cdot), \frac{1}{S} \sum_{s \in [S]} (T_T(\cdot)\ \xi_T^s + h_T(\cdot))^T \pi^s \}$

$t = T-1$ solve dual of

$$V_{T-1}(x_{T-2}, s_T) = \inf c_{T-1}^T x_{T-1} + Q_T(x_{T-1})$$
$$\text{s.t. } T_{T-1}(x_{T-2})\,\xi_{T-1}^s + h_{T-1}(x_{T-2}) \leq W_{T-1} x_{T-1}$$

But we don't have access to $Q_T(x_{T-1})$ so we solve approximation

$$\inf \ c_{T-1}^T x_{T-1} + \hat{Q}_T(x_{T-1})$$
$$\text{s.t. } T_{T-1}(x_{T-2})\,\xi_{T-1}^s + h_{T-1}(x_{T-2}) \leq W_{T-1} x_{T-1}$$

update: $\hat{Q}_{T-1}(\cdot) = \max\{ \hat{Q}_{T-1}(\cdot), \frac{1}{S} \sum_{s \in [S]} T_{T-1}(\cdot)\ \xi_{T-1}^s + h_{T-1}(\cdot))^T \pi^s \}$

repeat until $t = 2$

Lecture 26. Nov 29

- Step 3 : solve :

$$\underline{z} = \inf_{x \in X} c_1^T x_1 + \hat{Q}_2(x_1)$$

update . $\hat{x}_1 \in \arg\inf_{x \in X} c_1^T x_1 + \hat{Q}_2(x_1)$

- Step 4. Monte - Carlo

Forward step↑ Pick a realization l: $\hat{\xi}_2^l, \cdots, \hat{\xi}_T^l$ from $\mathbb{P}_2 \times \cdots \mathbb{P}_T$

Set $\hat{x}_1^l = \hat{x}_1$

solve for $t = 2, \cdots, T$

$$\hat{x}_t^l \in \arg\inf \quad c_t^T x_t + \hat{Q}_{t+1}(x_t)$$
$$\text{s.t.} \quad T_t(\hat{x}_{t-1}^l)\hat{\xi}_t^l + h_t(\hat{x}_{t-1}^l) \leq W_t x_t$$

$\hat{x}_1^l, \hat{x}_2^l, \cdots \hat{x}_T^l \Rightarrow c_1^T \hat{x}_1 + \sum_{t=2}^T c_t^T \hat{x}_t^l$ cost for realization l.

$$\bar{z} = c_1^T \hat{x}_1 + \mathbb{E}\left[\sum_{t=2}^T c_t^T \hat{x}_t\right]$$

$\approx c_1^T \hat{x}_1 + \frac{1}{L} \sum_{l \in [L]} c_t^T \hat{x}_t^l$ → use CLT to derive some confidence interval on $\bar{z}$

$\geq \inf_{x_1 \in X} c_1^T x_1 + Q_2(x_1)$ true objective.

$\geq \underline{z}$

If $\bar{z} - \underline{z} < \epsilon$

$x_1^* = \hat{x}_1$, terminate.

else go to step 2.

Lecture 26. Nov. 29.

- Step 3 : solve :

$$\underline{z} = \inf_{x \in X} c_1^T x_1 + \hat{Q}_2(x_1)$$

update . $\hat{x}_1 \in \arg\inf_{x \in X} c_1^T x_1 + \hat{Q}_2(x_1)$

- Step 4 : Monte - Carlo

Forward step↑ Pick a realization l: $\hat{\varepsilon}_2^l, \cdots, \hat{\varepsilon}_T^l$ from $\mathbb{P}_2 \times \cdots \mathbb{P}_T$

Set $\hat{x}_1^l = \hat{x}_1$

solve for $t = 2, \cdots, T$

$$\hat{x}_t^l \in \arg\inf \; c_t^T x_t + \hat{Q}_{t+1}(x_t)$$
$$\text{s.t.} \; T_t(\hat{x}_{t-1}^l)\hat{\varepsilon}_t^l + h_t(\hat{x}_{t-1}^l) \leq W_t x_t$$

$\hat{x}_1^l, \hat{x}_2^l, \cdots \hat{x}_T^l \Rightarrow c_1^T \hat{x}_1 + \sum_{t=2}^{T} c_t^T \hat{x}_t^l$ cost for realization l.

$$\bar{z} = c_1^T \hat{x}_1 + \mathbb{E}\left[\sum_{t=2}^{T} c_t^T \hat{x}_t\right]$$
$$\approx c_1^T \hat{x}_1 + \frac{1}{L}\sum_{l \in [L]} c_t^T \hat{x}_t^l$$ → use CLT to derive some confidence interval on $\bar{z}$
$$\geq \inf_{x_1 \in X} c_1^T x_1 + Q_2(x_1)$$ true objective.
$$\geq \underline{z}$$

If $\bar{z} - \underline{z} < \epsilon$

$x_1^* = \hat{x}_1$, terminate.

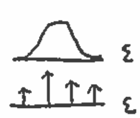
else. go to step 2.

<u>Review.</u>

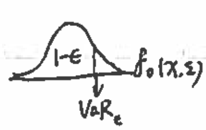
$\inf f_0(x)$ s.t. $f_i(x) \leq 0, \; \forall i \in [I]$ → $\inf f_0(x, \varepsilon)$ s.t. $f_i(x, \varepsilon) \leq 0. \; \forall i \in [I]$

Different paradigms :
- stochastic
- robust Ξ
- distributionally robust.

Stochastic : $\tilde{\varepsilon} \sim \mathbb{P}$ — continuous / discrete.



Decision maker can be :
- risk neutral . $\mathbb{E}[\cdot]$
- risk averse : $\mathrm{CVaR}_\epsilon[\cdot]$, $\mathrm{VaR}_\epsilon[\cdot]$



risk averse chance constraint, $\mathbb{P}(f_i(x,\tilde{\xi}) \le 0,\ \forall i \in [I]) \geq 1-\epsilon$

$\Leftrightarrow \mathrm{VaR}_\epsilon [\max_{i \in [I]} f_i(x, \tilde{\xi})] \le 0$

Distribution robust model:

$\mathbb{P}$ is ambiguous

$\mathbb{P} \in \mathcal{P} \to \mu$, $\mu \& \Sigma$, μ & MAD, Wasserstein.

(unimodality? symmetry?)

$f_0(x,\xi)$ is convex in x for any $\xi \in \Xi$.

- One-stage:
  - Stochastic: $\mathbb{P}$ is discrete - tractable

    $\mathbb{P}$ is continuous - intractable
  - dist. robust: tractability $\nearrow$ (streaming / online opt. $\xi^1, \cdots \xi^S$. $\xi \in \mathbb{P}$. $O(k \times S) \to O(k)$ S problems)

    $\searrow$ (treading: DR streaming)
  - chance constraint: $\mathbb{P}$ discrete - intractable.

    dist. robust - $\inf_{\mathbb{P} \in \mathcal{P}} (f(x,\xi) \le 0) \geq 1-\epsilon$

(→ Constraint sampling approach: $\inf\ c^T x$

$\text{s.t. } \mathbb{P}(f(x,\tilde{\xi}) \le 0) \geq 1-\epsilon$

$\inf\ c^T x$

$\text{s.t. } f(x, \xi^s) \le 0,\ \forall s \in [S]$

Solve this for each sample and solution is $\hat{x}^s$

To see how many S satisfies:

$\mathrm{Prob}\,(\mathbb{P}(f(\hat{x}^s, \tilde{\xi}) \le 0) \geq 1-\epsilon) \geq 1-\hat{\beta}$

$S \geq \frac{2}{\epsilon}(\ln\frac{1}{\beta} + N)$)

- Two-stage: $f_0(x,\xi) = c^T x + \inf\ (Q\xi + g)^T y$

$\text{s.t. } y \in \mathbb{R}^{N_2}$

$T(x)\xi + h(x) \le Wy$

$\mathbb{P}$ is discrete - tractable

$\mathbb{P}$ is continuous - intractable. - approx. schemes with guarantees.

dist. robust - intractable.

(x is discrete ✓

y is discrete . $y \in \mathbb{R}^{N_2} \cap \mathbb{Z}^n$

systematically portion the set and iterate to get updated feasible solution $\hat{y}^n$)

Decision rules:

- linear
- quadratic

(- polynomial, piecewise linear).

— Multi-stage:

$\mathbb{P}$ discrete — intractable.

— SDDP

(— scenario tree ×

— decision rules)

* (continue). chance constraint:

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\phi_i(x, \tilde{\xi}) \le 0, \ \forall i \in [I] \right) \ge 1 - \epsilon$$

In general, intractable.

Sometimes tractable $\mathcal{P}$ is μ & MAD

Exam

- Wed, Dec. 14. 2-5 P.M
- location: SZB 278.
- Open book.