△ Stochastic Programming

remark:

difference between stochastic & robust

- stochastic programming:
  - we know the distribution
  - optimize the expected value
- robust programming:
  - we do not know the distr. information
  - optimize w/ worst-case !!!

△ The Farmer's Problem

- a farmer raises
  - x_1 wheat
  - x_2 corn
  - x_3 sugar beets
  - } ← 500 acres land

⇒ $x_1+x_2+x_3=500$

- 200 T wheat, 240 T corn } → cattle feed; can grow or buy
- 170 \$/T wheat, 150 \$/T corn } → selling price; if buy +40% higher
- 36 \$/T sugar beet } → selling price; 10 \$/T if surplus > 6000 T
- 2.5 T/acre wheat, 3 T/acre corn, 20 T/acre sugar beet
- 150 \$ wheat, 230 \$ corn, 260 \$ sugar beets } planting cost

⇒ max $-150x_1-230x_2-260x_3+170w_1+150w_2+36w_3+10w_4-170\cdot1.4\cdot y_1-150\cdot1.4\cdot y_2$

s.t. $x_1+x_2+x_3\le500$

$w_3\le6000$

$2.5x_1+y_1-w_1\ge200$

$3x_2+y_2-w_2\ge240$

$w_3+w_4\le20x_3$

$x_i\ge0$

$w_i\ge0$

$y_i\ge0$

solution: 118,600

△ Induction of stochastic programming

Normal year: 2.5 T/acre wheat, 3 T/acre corn, 20 T/acre sugar beet

Good year: 3 T/acre wheat, 3.6 T/acre corn, 24 T/acre sugar beet

Bad year: 2.083 T/acre wheat, 2.5 T/acre corn, 20 T/acre sugar beet

assume the occurrence of the respective normal, good, bad are 1/3, 1/3, 1/3

new formulation (maximizing the expected value)

min $-150x_1-230x_2-260x_3$
$+\frac{1}{3}(170w_{11}+150w_{21}+36w_{31}+10w_{41}-170\cdot1.4y_{11}-150\cdot1.4y_{21})$
$+\frac{1}{3}(170w_{12}+150w_{22}+36w_{32}+10w_{42}-170\cdot1.4y_{12}-150\cdot1.4y_{22})$
$+\frac{1}{3}(170w_{13}+150w_{23}+36w_{33}+10w_{43}-170\cdot1.4y_{13}-150\cdot1.4y_{23})$

s.t. $x_1+x_2+x_3\le500$

$w_{3i}\le6000$ i= 1,2,3

$2.5x_1+y_{11}-w_{11}\ge200$, $3x_2+y_{21}-w_{21}\ge240$, $w_{31}+w_{41}\le20x_3$ } scenario 1

$3x_1+y_{12}-w_{12}\ge200$, $3.6x_2+y_{22}-w_{22}\ge240$, $w_{32}+w_{42}\le24x_3$ } scenario 2

$2.083x_1+y_{13}-w_{13}\ge200$, $2.5x_2+y_{23}-w_{23}\ge240$, $w_{33}+w_{43}\le20x_3$ } scenario 3

sol: 108,390

much harder to solve, hard to satisfy each condition

△ Expected Value of Perfect Information (EVPI)

assume
- good year 167,667
- normal year 118,600
- bad year 59,950

if respectively 1/3 1/3 1/3 to each year

then expected value 115,406 "perfect information"

yet, from the above-listed stochastic problem, the solution is 108,390

perfect information 115,406, no information 108,390 } 7,016

"expected value of perfect information" ⇒ represents the loss of profit due to the presence of uncertainty

△ Value of stochastic solution

- perfect info: 115,406
- no info: 108,390 } 7,016
- expected yields: 107,240 } 1,150

expected value of perfect info 7016] EVPI

value of stochastic sol 1,150] VSS

remark:

EVPI measures the value of knowing the future w/ certainty

VSS measures the value of knowing the future w/ ending decision/distribution

△ General Model Formulation

- 1st stage decisions: x (w.o. full info on random events)
- 2nd-stage decisions: y, $y(s)$ (full info based on "realization of random vector ξ, $\xi(w)$") (corrective actions)

$Q(x,\xi)=\min\{q^Ty \mid Wy=h-Tx, y\ge0\}$

min $c^Tx+E_\xi Q(x,\xi)$, $Ax=b$, $x\ge0$

⇓

min $c^Tx+E_\xi\, Q(x,\xi)$

s.t. $Ax=b$

$x\ge0$

min $c^Tx+\mathcal{Q}(x)$, $Ax=b$, $x\ge0$

$Q(x,\xi)=\min\{q^Ty \mid Wy=h-Tx, y\ge0\}$

$\xi=[q,h,T]$ this is a random variable

e.g. rewrite the farmer problem

$Q(x,s)=\min\{238y_1-170w_1+210y_2-150w_2-36w_3-10w_4\}$

s.t. $t_1(s)x_1+y_1-w_1\ge200$

$t_2(s)x_2+y_2-w_2\ge240$

$w_3+w_4\le t_3(s)x_3$

$w_3\le6000$

$y,w\ge0$

$t_i(s)$ is the yield of the crop

△ Implicit Representation

condensed implicit representation

min $c^Tx+E_\xi Q(x,\xi)$

s.t. $Ax=b$

$x\ge0$

min $c^Tx+\mathcal{Q}(x)$

s.t. $Ax=b$

$x\ge0$

△ Revisit farmer's problem

- assume the yield for each crop i within interval $[l_i,u_i]$ → uniform random variable
- first stage decision: decisions on land allocation
- second stage decision: purchases, sales after the growing period

$\mathcal{Q}(x)=E_\xi Q(x,\xi)$ → is the value of second stage for a given realization of the random vector

$E_\xi Q(x,\xi)$

$=\sum_{i=1}^{3}E_\xi Q_i(x_i,\xi)$

$=\sum_{i=1}^{3}\mathcal{Q}_i(x_i)$ ↳ crop 1, 2, 3

e.g. sugar beet

- $Q_3(x_3,\xi)$

$=\min -36w_3(\xi)-10w_4(\xi)$

s.t. $w_3(\xi)+w_4(\xi)\le t_3(\xi)x_3$

$w_3(\xi)\le6000$

$w_3(\xi),w_4(\xi)\ge0$

optimal solution

$w_3(\xi)=\min[t_3(\xi)x_3,6000]$

$w_4(\xi)=\max[t_3(\xi)x_3-6000,0]$

∴ $Q_3(x_3,\xi)$

$=-36\min[t_3(\xi)x_3,6000]$
$-10\max[t_3(\xi)x_3-6000,0]$

- assume

$l_3x_3\le6000\le u_3x_3$

⇒ $l_3\le\frac{6000}{x_3}\le u_3$

l_3, u_3 → bounds on values of $t_3(\xi)$ → $t_3(\xi)$ is the yield

- Therefore, the expected value:

$\mathcal{Q}_3(x_3)=E_\xi Q_3(x_3,\xi_3)$

$=-\int_{l_3}^{\frac{6000}{x_3}}36t\,x_3 f(t)dt$

$-\int_{\frac{6000}{x_3}}^{u_3}[36\cdot6000+10\cdot t\cdot x_3-6000]f(t)dt$

if uniformly distributed

$\mathcal{Q}_3(x_3)=-18\frac{(u_3^2-l_3^2)x_3}{u_3-l_3}+\frac{13(u_3x_3-6000)^2}{x_3(u_3-l_3)}$

$=-36\bar{t}_3x_3+\frac{13(u_3x_3-6000)^2}{x_3(u_3-l_3)}$

now, we can apply the same thing for $\mathcal{Q}_1,\mathcal{Q}_2$, then:

❑ **Continuous Random Variables**

For wheat:

$$\mathcal{Q}_1(x_1)=\begin{cases}47600-595x_1 & \text{for } x_1\le200/3,\\ 119\frac{(200-2x_1)^2}{x_1}-85\frac{(200-3x_1)^2}{x_1} & \text{for } \frac{200}{3}\le x_1\le100,\\ 34000-425x_1 & \text{for } x_1\ge100,\end{cases}$$

For corn:

$$\mathcal{Q}_2(x_2)=\begin{cases}50400-630x_2 & \text{for } x_2\le200/3,\\ 87.5\frac{(240-2.4x_2)^2}{x_2}-62.5\frac{(240-3.6x_2)^2}{x_2} & \text{for } 200/3\le x_2\le100,\\ 36000-450x_2 & \text{for } x_2\ge100.\end{cases}$$

The global problem is therefore:

$$\begin{aligned}&\min 150x_1+230x_2+260x_3+\mathcal{Q}_1(x_1)+\mathcal{Q}_2(x_2)+\mathcal{Q}_3(x_3)\\&\text{s.t. } x_1+x_2+x_3\le500,\\&\quad x_1,x_2,x_3\ge0.\end{aligned}$$

△ News Vendor Problem

- News vendor buys x newpaper @ c price
- $c\le u$
- sell each newspaper @ q price
- return unsold paper @ r price
- $r<c$
- Q: $x=?$, w/ demands ξ (random)

y sales

w returning paper

⇓

min $cx+\mathcal{Q}(x)$

$0\le x\le u$

s.t. $\mathcal{Q}(x)=E_\xi Q(x,\xi)$ — ξ: demand variable; $Q(x,\xi)$ → profit based on ξ

$Q(x,\xi)=\min -q\,y(\xi)-r\,w(\xi)$

$y(\xi)\le\xi$

$y(\xi)+w(\xi)\le x$

$y(\xi),w(\xi)\ge0$

sold: $y^*(\xi)=\min(\xi,x)$

returned: $w^*(\xi)=\max(x-\xi,0)$

$\mathcal{Q}(x)=E_\xi[-q\min(\xi,x)-r\max(x-\xi,0)]$

solution omitted here

△ A Routing Example

solution omitted here.

△ Probability Space

- event outcome ω
- set of ω Ω
- subsets of Ω $\mathcal{A}$
- element of $\mathcal{A}$ $A\in\mathcal{A}$
- $A\longrightarrow P(A)$
- $(\Omega,\mathcal{A},P)$ probability space

△ cumulative distribution

$F_\xi(x)=P(\xi\le x)$

$F_\xi(x)=P(\{\omega \mid \xi\le x\})$

- distribution

$f(\xi^k)=P(\xi=\xi^k)$

s.t. $\sum_{k\in K}f(\xi^k)=1$

s.t. $P(a\le\xi\le b)=\int_a^b f(\xi)d\xi$

- expectation

$\mu=\sum_{k\in K}\xi^k f(\xi^k)$

$\mu=\int_{-\infty}^{\infty}\xi\,dF(\xi)$

- variance

$E[(\xi-\mu)^2]$

- expection of ξ^r is called r^{th} moment of ξ → $\bar{\xi}^{(r)}=E[\xi^r]$

central moment

- η is called α-quantile of ξ i.ff. $0<\alpha<1$

$\eta=\min\{x \mid F(x)\ge\alpha\}$

$\mathcal{Q}(x)=E_\xi[Q(x,\xi)]$

- $K_2=\{x \mid \mathcal{Q}(x)<\infty\}$
- $K_2^P=\bigcap_{\xi\in\Xi}K_2(\xi)$ $(K_2(\xi)=\{x \mid Q(x,\xi)<\infty\})$

e.g. $Q(x,\xi)=\min_y\{y \mid \xi y=1-x, y\ge0\}$

$\xi\in[0,1]$, $F(\xi\le u)=u^2$ $p(u)=2u$

△

$\xi\in(0,1]$ ⇒ y optimal: $\frac{1-x}{\xi}$, & $x\le1$ always ∴ $K_2(\xi)=\{x \mid x\le1\}$

⇒ $Q(x,\xi)=0$, $x\le1$ $<\infty$

$\xi=0$ ⇒ y optimal: 0, & $x=1$ always ∴ $K_2(0)=\{x \mid x=1\}$

⇒ $Q(x,0)=0$

△ Stochastic Linear Program

linear program w/ uncertain data

△ Recourse Program

decision/recourse action taken after uncertainty is disclosed

$\xi = \xi(\omega)$ known after experiment

1. 1st-stage decisions – prior to experiments
2. 2nd-stage decisions – after the experiments.

$x \longrightarrow \xi(\omega) \longrightarrow y(\omega, x)$

1st … uncertain of the random variable … known after experiment … 2nd

△ 2-stage stochastic linear program

$$\min z = c^T x + E_\xi[\min q(\omega)^T y(\omega)]$$

s.t. $Ax = b$ ($\mathbb{R}^{m_1}$), $T(\omega)x + Wy(\omega) = h(\omega)$ ($\mathbb{R}^{m_2}$), $x \ge 0$, $y(\omega) \ge 0$

fixed recourse $\mathbb{R}^{m_2 \times n_2}$; technology matrix; Recourse Matrix

notice that

$\omega \in \Omega$ ↓ @ 2nd stage, ω is known; $q(\omega)$ $\mathbb{R}^{n_2}$, $h(\omega)$ $\mathbb{R}^{m_2}$, $T(\omega)$ $\mathbb{R}^{m_2 \times n_1}$ is then known

random variable ξ

$\xi^T(\omega) = (q(\omega)^T, h(\omega)^T, \underbrace{T_1(\omega), \ldots, T_{m_2}(\omega)}_{\text{rows of } T})$

$\in \mathbb{R}^{n_2 + m_2 + n_1 \cdot m_2}$

△ support of ξ

$\Xi \in \mathbb{R}^N$, $N = n_2 + m_2 + n_1 \cdot m_2$

$P(\Xi) = 1$ (all outcomes of ξ)

★ Based on outcome $\omega \to \xi(\omega)$ decide y; for each ω, $y(\omega)$ is the solution to "a" linear program $q(\omega)^T y(\omega)$

↳ let

$$Q(x, \xi(\omega)) = \min_y \{q(\omega)^T y \mid Wy = h(\omega) - T(\omega)x,\ y \ge 0\}$$

$\Rightarrow \min z = c^T x + E_\xi\, Q(x, \xi(\omega))$ s.t. $Ax = b$, $x \ge 0$

⇕

$\Rightarrow \min z = c^T x + \mathcal{Q}(x)$ s.t. $Ax = b$, $x \ge 0$ — deterministic equivalent program (DEP)

△ Basic Properties

1. formulations of deterministic program equivalent to stochastic program
2. forms of feasible region, objective function
3. conditions for optimality

△ Recall

- $\min z = c^T x + E_\xi[\min q(\omega)^T y(\omega)]$

  s.t. $Ax = b$, $T(\omega)x + Wy(\omega) = h(\omega)$, $x \ge 0$, $y(\omega) \ge 0$ — recourse matrix (fixed); technology matrix

  we hv a lot of ω
- $\xi^T(\omega) = [q(\omega)^T, h(\omega)^T, T_1(\omega), \ldots, T_{m_2}(\omega)] \in \mathbb{R}^{n_2 + m_2 + m_2 \cdot n_1} = \mathbb{R}^N$
- $\exists\, \Xi \in \mathbb{R}^N$, support of ξ; $P(\Xi) = 1$
- DEP: $\min z = c^T x + \mathcal{Q}(x)$ s.t. $Ax = b$, $x \ge 0$

  $\mathcal{Q}(x) = E_\xi\, Q(x, \xi(\omega))$, $Q(x, \xi(\omega)) = \min_y \{q(\omega)^T y \mid Wy = h(\omega) - T(\omega)x,\ y \ge 0\}$

This representation clearly illustrates the sequence of events in the recourse problem. First-stage decisions x are taken in the presence of uncertainty about future realizations of ξ. In the second stage, the actual value of ξ becomes known and some corrective actions or recourse decisions y can be taken. First-stage decisions are, however, chosen by taking their future effects into account. These future effects are measured by the value function or recourse function, $\mathcal{Q}(x)$, which computes the expected value of taking decision x.

$Q(x, \xi) = \min\{q(\omega)y \mid Wy = h(\omega) - T(\omega)x,\ y \ge 0\}$

Q convex h, T
Q concave q
Q convex x

① Discrete Random Variable feasibility set

- ξ is discrete random variable
- $K_1 = \{x \mid Ax = b,\ x \ge 0\}$ set determined by the fixed constraints
- $K_2(\xi) = \{x \mid y \ge 0 \text{ exists s.t. } W(\omega)y = h(\omega) - T(\omega)x\}$

$K_2 = \bigcap_{\xi \in \Xi} K_2(\xi)$ — set determined by a given ξ (elementary feasibility set)

e.g. consider 2nd-stage program

$\min 2y_1 + y_2$
s.t. $y_1 + 2y_2 \ge \xi_1 - x_1$
$y_1 + y_2 \ge \xi_2 - x_1 - x_2$
$0 \le y_1 \le 1$, $0 \le y_2 \le 1$

w/ u.b. $\xi_1 - x_1 \le 3$, $\xi_2 - x_1 - x_2 \le 2$

$\therefore K_2(\xi) = \{x \mid x_1 \ge \xi_1 - 3,\ x_1 + x_2 \ge \xi_2 - 2\}$

$\therefore K_2 = \bigcap_{\xi \in \Xi} K_2(\xi)$

Theorem 1 (on feasibility set)

a. For a given ξ, the elementary feasibility set $K_2(\xi)$ is a convex polyhedron.
b. When ξ is a finite discrete random variable, K_2 is a convex polyhedron.

△ Discrete Random Variable 2nd-stage value function

- given x, ξ

$$Q(x, \xi) = \min_y \{q(\omega)^T y \mid W(\omega)y = h(\omega) - T(\omega)x,\ y \ge 0\}$$

- bound y — might not be solvable if • unbounded below • infeasible
- let $x \in K_2$ → solution

$\therefore x \in K_2$, $Q(x, \xi)$ finite

$\mathcal{Q}(x) = E_\xi\, Q(x, \xi) = \sum_{k=1}^{K} p_k Q(x, \xi_k)$

$\Rightarrow \min z(x) = c^T x + \mathcal{Q}(x)$, $x \in K_1 \cap K_2$

Theorem 2 (on 2nd-stage value function)

Theorem 2. *For a given ξ, the value function $Q(x, \xi)$ is*
(a) a piecewise linear convex function in (h, T);
(b) a piecewise linear concave function in q;
(c) a piecewise linear convex function in x for all $x \in K = K_1 \cap K_2$.
When ξ is a finite discrete random variable, $\mathcal{Q}(x)$ is piecewise linear and convex on K_2.

Properties of the Objective Function

- For a stochastic program with fixed recourse, $Q(x,\xi)$ is
  - a piecewise linear convex function in (h,T);
  - a piecewise linear concave function in q;
  - a piecewise linear convex function in x for all x in $K = K_1 \cap K_2$.
- Proof: Linear supports – use duality (we hv strong duality in LP)
- With finite second moments, $\mathcal{Q}(x)$ is: Lipschitzian, convex, finite on K_2, p.l. if Ξ finite, differentiable on K_2 if $F(\xi)$ abs. continuous

② general case feasibility set

$K_2(\xi) = \{x \mid Q(x, \xi) < \infty\}$
or
$K_2(\xi) = \{x \mid y \ge 0 \text{ exists s.t. } W(\omega)y = h(\omega) - T(\omega)x\}$

$\Rightarrow$ w/ $\mathcal{Q}(x) = E_\xi\, Q(x, \xi)$

$K_2 = \{x \mid \mathcal{Q}(x) < \infty\}$
or
$K_2^P = \bigcap_{\xi \in \Xi} K_2(\xi)$ (possibility interpretation of 2nd stage feasibility set)

★ K_2 & K_2^P can be sometimes different; $P(Q(x,\xi)) = 1$ yet $E_\xi[Q(x,\xi)] \to \infty$
↳ yet if W fixed, ξ is continuous, case never happens; won't happen

"This is why discrete / continuous discussed separately."

Proposition 3. *If ξ has finite second moments, then*
$$P(\omega \mid Q(x, \xi) < \infty) = 1 \text{ implies } \mathcal{Q}(x) < \infty.$$

This is also to make sure that K_2^P K_2 same

Theorem 4. *For a stochastic program with fixed recourse where ξ has finite second moments, the sets K_2 and K_2^P coincide.* sets are equal

Theorem 5. *When W is fixed and ξ has finite second moments:*
(a) K_2 is closed and convex.
(b) If T is fixed, K_2 is polyhedral.
(c) Let Ξ_T be the support of the distribution of T. If $h(\xi)$ and $T(\xi)$ are independent and Ξ_T is polyhedral, then K_2 is polyhedral.

② general case 2nd-stage value function

$Q(x, \xi)$ not $-\infty$
– has theorem 2

Theorem 6. *For a stochastic program with fixed recourse where ξ has finite second moments,*
(a) $\mathcal{Q}(x)$ is a Lipschitzian convex function and is finite on K_2.
(b) If $F(\xi)$ is an absolutely continuous distribution, $\mathcal{Q}(x)$ is differentiable on riK_2. ↳ relative interior

When the random variables are appropriately described by a finite distribution, the constraint set K_2 is best defined by the possibility interpretation and is easily seen to be polyhedral. The second-stage recourse function $\mathcal{Q}(x)$ is piecewise linear and convex on K_2. The decomposition techniques of Chapter 5 then apply. This is a category of programs for which computational methods can be made efficient, as we shall see.

When the random variables cannot be described by a finite distribution, they can usually be associated with some probability density. Many common probability densities are absolutely continuous and have finite second moments; so, the constraints set definitions K_2 and K_2^P coincide and the second-stage value function $\mathcal{Q}(x)$ is differentiable and convex. Classical nonlinear programming techniques could then be applied. A typical example was given in the farmer's problem in Chapter 1. There,

△ special case

- Complete recourse $K_2 = \Re^{n_1}$
- Relatively complete recourse $K_2 \supset K_1$
- Simple recourse $W = [I, -I]$, $q = (q^+, q^-)$ Holds for the news vendor problem

complete: $K_2 = \mathbb{R}^{n_1}$
relative complete: $K_1 \subset K_2$
simple: $W = [I, -I]$, $q = (q^+, q^-)$

recall $Q(x, \xi) = \min\{q(\xi)y \mid Wy = h(\xi) - T(\xi)x\}$

$\because K_2^P = \{x \mid x \ge 1\} \cap \{x \mid x \le 1\} = \{x \mid x = 1\}$

△ $\mathcal{Q}(x) = \int_0^1 \frac{1-x}{\xi} 2\xi\, d\xi = 2(1-x)$, $x \le 1$

$\therefore K_2 = \{x \mid x \le 1\}$

we want this s.t. constraints are finite

Feasibility

- Definitions
  - $K_1 = \{x \mid Ax = b,\ x \ge 0\}$
  - $K_2 = \{x \mid \mathcal{Q}(x) < \infty\}$, $K_2^P = \bigcap_{\xi \in \Xi}\{x \mid Q(x,\xi) < \infty\}$
- Results:
  - $K_2^P = K_2$ if Ξ finite or W is fixed and ξ has finite second moments, which also means:
  - K_2 is closed and convex.
  - If T is fixed, K_2 is polyhedral.
  - Let Ξ_T be the support of the distribution of T. If $h(\xi)$ and $T(\xi)$ are independent and Ξ_T is polyhedral, then K_2 is polyhedral.

bring $|K_2| < \infty$ doesn't mean that $|K_2^P| < \infty$

Results with Feasibility Regions

Example: $K_2 \neq K_2^P$:
- Setup: $q = 1$, $x + y(\omega) \ge \xi(\omega)$, $y(\omega) \ge 0$.
- Distribution: $\xi(\omega)$ has density $f(\xi) = \frac{1}{\xi^2}$ over $\Xi = [1, \infty)$
- Result: $E[Q(x)] = \infty$ for any x so $K_2 = \emptyset$ while $K_2^P = K_1$.

Showing polyhedral K_2 when T is fixed:

$$K_2 = \{x \mid \exists y \text{ s.t. } Wy = \xi - Tx,\ y \ge 0,\ \forall \xi \in \Xi\}$$
$$\iff \nexists \pi \text{ s.t. } \pi^T W \le 0,\ \pi^T(\xi - Tx) > 0$$
$$\iff \pi^T Tx \ge \pi^T \xi,\ \forall \pi^T W \le 0,\ \xi \in \Xi,$$

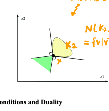
which is polyhedral since $\pi^T W \le 0$ has a finite representation.

△ optimality condition

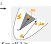
- Existence: suppose ξ has finite second moments and either K is bounded or $\mathcal{Q}$ becomes linear eventually, then optimum attained if it exists.
- Conditions:
  Suppose a finite optimal value. A solution $x^* \in K_1$, is optimal in DEP if and only if there exists some $\lambda^* \in \Re^{m_1}$, $\mu^* \in \Re^{n_1}_+$, $\mu^{*T}x^* = 0$, such that,
  $-c + A^T\lambda^* + \mu^* \in \partial\mathcal{Q}(x^*)$ subgradient
  $\partial\mathcal{Q}(x^*) = E[\partial Q(x^*, \xi)] + N(K_2, x^*)$ (recall: normal cone of K_2 @ x)

normal cone

In essence



$N(K_2, x) = \{v \mid v^T y \le 0\ \forall y \text{ s.t. } x + y \in K_2\}$



- **Optimality Conditions and Duality**
  - **Notations**
    - rc denotes the *recession cone*, $\{v \mid u + \lambda v \in S, \text{for all } \lambda \ge 0 \text{ and } u \in S\}$ when applied to a set, S
    - The *recession value*, $\sup_{x \in \text{dom} f}(f(x+v) - f(x))$ when applied to a proper convex function, f.

Theorem 8
Suppose that the random elements ξ have finite second moments and one of the following:
(a) the feasible region K is bounded; or
(b) the recourse function $\mathcal{Q}$ is eventually linear in all recession directions of K, i.e., $\mathcal{Q}(x + \bar\lambda v) = \mathcal{Q}(x+) + (\lambda - \bar\lambda)\text{rc}\mathcal{Q}(v)$ for some $\bar\lambda \ge 0$ (dependent on x), all $\lambda \ge \bar\lambda$, and some constant recession value, rc$\mathcal{Q}(v)$, for all v such that $x + \lambda v \in K$ for all $x \in K$ and $\lambda \ge 0$.
If (1.2) has a finite optimal value, it is attained for some $x \in \Re^n$.

Theorem 9
Suppose (1.2) has a finite optimal value. A solution $x^* \in K_1$, is optimal in (1.2) if and only if there exists some $\lambda^* \in \Re^{m_1}$, $\mu^* \in \Re^{n_1}_+$, $\mu^{*T}x^* = 0$, such that,
$$-c + A^T\lambda^* + \mu^* \in \partial\mathcal{Q}(x^*)$$
subgradient

max of inf $\mathcal{L}(\lambda, \mu)$, i.e., inf — might be un-continuous, the gradient, should be equal to subgradient @ optimal.

Corollary 10
Suppose (1.1) has simple recourse and a finite optimal value. Then $x^* \in K_1$ is optimal in (1.2) corresponding to this problem if and only if there exists some $\lambda^* \in \Re^{m_1}$, $\mu^* \in \Re^{n_1}_+$, $\mu^{*T}x^* = 0$, π_i^* such that:
$$-(q_i^+ - q_i F_i^-(T_{i\cdot}x^*)) \le \pi_i^* \le -(q_i^+ - q_i F_i^+(T_{i\cdot}x^*))$$
$$-c + A^T\lambda^* + \mu^* - (\pi^*)^T T = 0$$

$\min c^T x + \mathcal{Q}(x)$
s.t. $b - Ax = 0$
$-x \le 0$
$\Rightarrow c^T x + \mathcal{Q}(x) + \lambda(b - Ax) - \mu x$
$\Rightarrow c - A^T\lambda - \mu + \partial\mathcal{Q}(x) = 0$
$\Rightarrow -c + A^T\lambda + \mu \in \partial\mathcal{Q}(x)$

Theorem 11
If $x \in K$, then
$$\partial\mathcal{Q}(x) = \text{E}_\omega \partial Q(x, \xi(\omega)) + N(K_2, x)$$
where
$$N(K_2, x) = \{v \mid v^T y \le 0, \forall y \text{ such that } x + y \in K_2\}$$
the normal cone to K_2 at x.

Theorem 12
If (1.2) has relatively complete recourse, a solution x^* is optimal in (1.2) if and only if there exists some $\lambda^* \in \Re^{m_1}$, $\mu^* \in \Re^{n_1}_+$, $\mu^{*T}x^* = 0$, such that,
$$-c + A^T\lambda^* + \mu^* \in \text{E}_\omega \partial Q(x, \xi(\omega))$$

△ Chance constraint

- constraints are "Almost Surely"
- hold it w/ probabilistic level / reliability

$\Rightarrow P\{A^i(\omega)x \ge h^i(\omega)\} \ge \alpha^i$ ↳ $\in [0,1]$

$\Rightarrow E_\omega(f^i(\omega, x(\omega))) \ge \alpha^i$

where $f^i(\cdot) = \{\omega \mid A^i(\omega)x \ge h^i(\omega)\}$

△ Chance constraints w/ discrete random variables

- $\xi(\omega) \longrightarrow \xi_1, \xi_2, \ldots, \xi_K$

  $\sum_{k=1}^{K} p_k = 1$
- assume a constraint

  $g(x, y(\omega), \xi(\omega)) \le 0$ — it's reliability

  $\Rightarrow P(g(\cdot) \le 0) \ge \alpha$
- if $\eta(g(\cdot) \le 0) = 0$

  $\eta(g(\cdot) > 1) = 1$
- now

  $\sum_{k=1}^{K} p_k \cdot \eta(g_k(\cdot)) \le 1 - \alpha$

  & assume $g_k(\cdot) \le u_k$

$\iff \sum_{k=1}^{K} p_k \cdot w_k \le 1 - \alpha$
$g_k(\cdot) \le u_k w_k$
$w_k \in \{0, 1\}$

e.g. x_1 first-class, x_2 business-class

$P(x_1 \ge \xi_F) \ge 0.98$
$P(x_1 + x_2 \ge \xi_B) \ge 0.95$

△ $P(x_1 + x_2 \ge \xi_F + \xi_B) \ge 0.95$
$(x_1 + x_2 - 46)/8 \ge 1.645$

△ $P(x_1 \ge \xi_F) \ge 0.98$
$\Rightarrow F_F(x_1) \ge 0.98$ — 对应 z-score
$\Rightarrow (x_1 - 16)/4 \ge 2.054$ (z-score) — $\frac{x - \mu}{\sigma}$
$\Rightarrow x_1 \ge 24.216$

examples: location problem

- d_i, $i = 1, \ldots, m$ clients demand
- x_j, $j = 1, \ldots, n$ potential sites $\in \{0, 1\}$
- problem: $x_j = ?$ (open how many sites, maximize profit)

- c_j fixed cost of x_j
- v_j varied cost of x_j — decide by company: which provider should send → which customer
- y_{ij} fraction demand of i (customer) from j (provider)
- t_{ij} transport cost from j (provider) to i (customer)
- r_i unit price charge to i (customer)

- $g_{ij} = (r_i - v_j - t_{ij}) d_i$ — total revenue from i (custom) w.r.t. j (provider)

Uncapacitated Facility Location Problem (UFLP):

$$\max_{x, y} g(x, y) = -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij} y_{ij}$$

s.t. $\sum_{j=1}^{n} y_{ij} \le 1$, $i = 1, \ldots, m$
$0 \le y_{ij} \le x_j$, $i = 1, \ldots, m$; $j = 1, \ldots, n$
$x_j \in \{0, 1\}$, $j = 1, \ldots, n$

client i demand from each provider j: sum = 1. (y is fraction)
client only get from $x_j = 1$

based on the above uncertainties can then be added.

case A:

- uncertainties in
  - r_i unit price change
  - v_i varied cost
  - t_{ij} transportation cost
- 1st stage: location, distribution
- 2nd stage: measure of the cost

⇓

$$\max -\sum_{j=1}^{n} c_j x_j + E_\xi \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}(\omega) w_{ij}(\omega)$$

s.t. $g_{ij} = (r_i - v_j - t_{ij}) d_i$
$\sum_{j=1}^{n} y_{ij} \le 1$
$0 \le y_{ij} \le x_j$
$x_j \in \{0, 1\}$
$w_{ij}(\omega) = y_{ij}$ ⟶ fixed here

$\Rightarrow \max -\sum_{j=1}^{n} c_j x_j + E_\xi \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}(\omega) y_{ij}$
s.t. ....
expectation no need consider y
$\Rightarrow \max -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} E_\xi[g_{ij}(\omega)] y_{ij}$
s.t. ....

deterministic program

case B:

- uncertainties in
  - d_i demand
- r_i, v_i, t_{ij} ~~may~~ / not be uncertain
- $y_{ij} \triangleq$ quantity from j to i — fixed!

  (not fraction here, cause d is uncertain now)
- q_i^+ penalty: paid/unit of $\boxed{d_i \text{ cannot be satisfied}}$

  e.g. d_i not enough, need from other src
- q_i^- penalty: paid/unit of $\boxed{d_i \text{ over-satisfied}}$

  e.g. d_i too much, need to store in inventory.
- 2nd stage:

  $w_i^-(\omega)$ extra delivered
  $w_i^+(\omega)$ deficit delivered

⇓

$$\max -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n}(E_\xi[-v_j - t_{ij}]) y_{ij}$$
$$+ E_\xi\Big[-\sum_{i=1}^{m} q_i^+ w_i^+(\omega) - \sum_{i=1}^{m} q_i^- w_i^-(\omega)\Big]$$
$$+ E_\xi \sum_{i=1}^{m} r_i d_i(\omega)$$

here our demand is uncertain

s.t. $\sum_{i=1}^{m} y_{ij} \le M x_j$
不够 $w_i^+(\omega) - w_i^-(\omega) = d_i(\omega) - \sum_{j=1}^{n} y_{ij}$ 多的
$x_j \in \{0, 1\}$
$0 \le y_{ij}$
$w_i^-(\omega) \ge 0$
$w_i^+(\omega) \ge 0$

case C:

- uncertainties in
  - y_{ij} distribution pattern

    fraction of d_i (demand of client i) from distribution shop j
  - w_j new variable

    size of distribution shop j
  - g_j shop invest cost / unit

⇓

$$\max -\sum_{j=1}^{n} c_j x_j - \sum_{j=1}^{n} g_j w_j + E_\xi \max \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}(\omega) y_{ij}(\omega)$$

s.t. $x_j \in \{0, 1\}$
$w_j \ge 0$
$\sum_{j=1}^{n} y_{ij}(\omega) \le 1$
$\sum_{i=1}^{m} d_i(\omega) y_{ij}(\omega) \le w_j$ — cannot be over the capacity of the shop
$0 \le y_{ij}(\omega) \le x_j$
$g_{ij}(\omega) = (r_i - v_j - t_{ij}) d_i(\omega)$

△ Stochastic Integer Programs.

$$\min_{x \in X} z = c^T x + Q(x)$$

s.t. $Ax = b$

$$Q(x) = E_\xi \min\{ q(\omega)^T y(\omega) \mid W y(\omega) = h(\omega) - T(\omega)x,\ y(\omega) \in Y \}$$

$X \subset \mathbb{Z}$

$Y \subset \mathbb{Z}$

△ Recourse Problems

Proposition 20

The expected recourse function $Q(x)$ of an integer program is in general, lower semicontinuous, nonconvex and discontinuous.

Proposition 21

The expected recourse function $Q(x)$ of an integer program with an absolutely continuous random variable is continuous.

Proposition 22

The second-stage feasibility set $K_2(\xi)$ is in general nonconvex.

△ Simple Integer Recourse

$$\min z = c^T x + E_\xi\{ \min (q^+)^T y^+ + (q^-)^T y^- \mid y^+ \geq \xi - Tx,\ y^+ \in \mathbb{Z}_+^m;\ y^- \geq Tx - \xi,\ y^- \in \mathbb{Z}_+^m \}$$

s.t. $Ax = b$

$x \in X$ ⟶ non-negative continuous integer

---

recall

$W y(\omega) = h(\omega) - T(\omega) x$

In stochastic programming, the second-stage decision involves decisions made under uncertainty, where the future outcomes are not known with certainty. The constraints you've mentioned are part of the mathematical formulation used to represent this uncertainty.

The first form, $wy = h - Tx$, represents an equality constraint. This implies that the resources consumed or produced in the second stage, denoted by wy, exactly balance out against the predetermined future demands h minus the effect of first-stage decisions Tx. This form is more restrictive and may not be realistic in practice because it assumes perfect foresight and exact balancing of resources.

The second form, $wy \geq \xi - Tx$, where ξ represents some stochastic variable or random parameter, introduces inequality constraints. This allows for a more realistic representation of uncertainty by acknowledging that the second-stage decisions must be able to cover potential demand realizations ξ while considering the effects of the first-stage decisions. The term ξ typically represents scenarios or realizations of uncertain parameters, such as demand, prices, or resource availability.

In practice, the second form is more commonly used because it provides flexibility and allows for a broader range of possible outcomes. It enables decision-makers to account for uncertainties and manage risks effectively. By setting constraints as inequalities, the model allows for a buffer to accommodate variations in the uncertain parameters without violating feasibility.

So, the reason for preferring the inequality form in practice is to make the model more robust and adaptable to real-world uncertainties.

here, we are just setting $W = [I, -I, \dots]$ & determine I of $\pm$ based on the output of $h - Tx$

max z

min $z = 100x_1 + 150x_2$
$x_1 + x_2 \leq 120$
$x_1 \geq 40$
$x_2 \geq 20$

[ ]

$\min z = c^T x + \theta$
$Ax = b$
$D_\ell x \geq d_\ell$
$E_\ell x + \theta \geq e_\ell$
$x \geq 0$

$\min w = q_k^T y$

$Wy = h - Tx$
$y \geq 0$

$$E_\ell = \sum_{k=1}^{K} p_k (\pi_k^v)^T T_k$$

$$e_\ell = \sum_{k=1}^{K} p_k (\pi_k^v)^T h_k$$

$w_\ell = e_\ell - E_\ell$

△ L-method

$\min z = c^T x + Q(x)$
s.t. $Ax = b$
$x \geq 0$

⟺

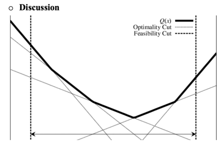
$$\min z = c^T x + \sum_{k=1}^{K} p_k q_k^T y_k$$ ↗ assume finite

s.t. $Ax = b$
$T_k x + W y_k = h_k$
$x \geq 0$
$y \geq 0$

$$\begin{array}{lllll} Ax & & & & = b \\ T_1 x + W y_1 & & & & = h_1 \\ T_2 x + & W y_2 & & & = h_2 \\ T_3 x + & & W y_3 & & = h_3 \\ \vdots & & & \ddots & \vdots \\ T_K x + & & & W y_K & = h_K \\ x \geq 0 & & & & \\ y_k \geq 0 & & & & \end{array}$$

● *L-Shaped Algorithm*

○ **Discussion**



---

e.g. Use example to explain L-shaped

$z = \min\ 100x_1 + 150x_2 + E_\xi(q_1 y_1 + q_2 y_2)$

s.t. $x_1 + x_2 \leq 120$
$6y_1 + 10y_2 \leq 60x_1$
$8y_1 + 5y_2 \leq 80x_2$
$y_1 \leq d_1$
$y_2 \leq d_2$
$x_1 \geq 40$
$x_2 \geq 20$
$y_1, y_2 \geq 0$

$\xi^T = (d_1, d_2, q_1, q_2)$ < 0.4: $(500, 100, -24, -28)$; 0.6: $(300, 300, -28, -32)$

□ Step 0.

$r = 0$ ~ size of constraints
$s = 0$
$v = 0$ ~ step

★ Iteration I.

□ Step 1

- $v = v + 1$
- $\min z = c^T x + \theta$
$Ax = b$
$D_\ell x \geq d_\ell \quad \ell = 1, \dots, r$
$E_\ell x + \theta \geq e_\ell \quad \ell = 1, \dots, s$
$x \geq 0$
if no $E_\ell x + \theta \geq e_\ell$, $\theta = -\infty$

△ ignore θ

△ master program

$\min z = 100x_1 + 150x_2$
$x_1 + x_2 \leq 120$
$x_1 \geq 40$
$x_2 \geq 20$

$\Big\}$ $x^1 = \begin{bmatrix} 40 \\ 20 \end{bmatrix}$, $\theta^1 = -\infty$

□ Step 2

$\begin{bmatrix} 40 \\ 20 \end{bmatrix} \in K_2$? < yes. go step 3; no. add 1 cut go step 1

△ yes.

□ Step 3

for $1, \dots, K$, solve
$\min w = q_k^T y$
s.t. $Wy = h_k - T_k x^v$
$y \geq 0$

$$E_{s+1} = \sum_{k=1}^{K} p_k \cdot (\pi_k^v)^T T_k$$

$$e_{s+1} = \sum_{k=1}^{K} p_k (\pi_k^v)^T h_k$$

$w^v = e_{s+1} - E_{s+1} x^v$

---

if $\theta^v \geq w^v$ stop, x^v is optimal

else $s = s+1$, return step 1

△ $k = 1$
$\xi_k = \xi_1$,

solve
$\min w = -24y_1 - 28y_2$
s.t. $6y_1 + 10y_2 \leq 2400$
$8y_1 + 5y_2 \leq 1600$
$0 \leq y_1 \leq 500$
$0 \leq y_2 \leq 100$

here
$k = 1 \Rightarrow w = -6100$
$y = \begin{bmatrix} 137.5 \\ 100 \end{bmatrix}$
$\pi = (0, -3, 0, -13)$

△ $k = 2$
$\xi_k = \xi_2$,

solve
$\min w = -28y_1 - 32y_2$
s.t. $6y_1 + 10y_2 \leq 2400$
$8y_1 + 5y_2 \leq 1600$
$0 \leq y_1 \leq 300$
$0 \leq y_2 \leq 300$

here
$k = 2 \Rightarrow w = -8384$
$y = \begin{bmatrix} 80 \\ 192 \end{bmatrix}$
$\pi = (-2.32, -1.76, 0, 0)$

△ @ $v = 1$

$x_{v=1} = (40, 20)^T$
$w_{v=1} = -520 - [83.52, 180.48] \begin{bmatrix} 40 \\ 20 \end{bmatrix} = -7470.4$

△ Add optimality cuts.

$w_{v1} < -\infty \Rightarrow$ false
$\Rightarrow$ add cut $E_\ell x + \theta \geq e_\ell$
$[83.52, 180.48] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \theta \geq -520$

now you got the idea

△ add feasibility cut,

so that solution outta step 1 $x \in K_2$.

$\min w' = e^T v^+ + e^T v^-$ ↖ (no. of variable) depends on the (no. of constraints)

s.t. $Wy + Iv^+ - Iv^- = h_k - T_k x^v$
$y \geq 0 \quad v^+ \geq 0$
$v^- \geq 0$
$e^T = [1, \dots, 1]$

v^+, v^- can be understood as upper/lower bound of the constraints concerning the 2nd-stage problem

---

solve the above until for some $k \longrightarrow w' > 0$

---

then σ^v (the associated simplex multipliers)

$\Rightarrow D_{r+1} = (\sigma^v)^T T_k$
$d_{r+1} = (\sigma^v)^T h_k$

$\Rightarrow$ return step 1.

---

if $\forall k$, $w' = 0$
go to step 3

e.g.

$\min 3x_1 + 2x_2 - E_\xi(15y_1 + 12y_2)$
s.t. $3y_1 + 2y_2 \leq x_1$
$2y_1 + 5y_2 \leq x_2$
$0.8\xi_1 \leq y_1 \leq \xi_1$
$0.8\xi_2 \leq y_2 \leq \xi_2$
$x, y \geq 0$

$\xi^T = (\xi_1, \xi_2)^T$
4, 6; 4, 8 ($\frac{1}{4}$ each)

---

realization $\xi = [6, 8]$

$x_{v=1} = [0, 0]^T$

$\min w' = v_1^+ + v_1^- + v_2^+ + v_2^- + v_3^+ + v_3^- + v_4^+ + v_4^- + v_5^+ + v_5^- + v_6^+ + v_6^-$

s.t.
$v_1^+ - v_1^- + 3y_1 + 2y_2 \leq 0$
$v_2^+ - v_2^- + 2y_1 + 5y_2 \leq 0$
$v_3^+ - v_3^- + y_1 \geq 4.8$
$v_4^+ - v_4^- + y_2 \geq 6.4$
$v_5^+ - v_5^- + y_1 \leq 6$
$v_6^+ - v_6^- + y_2 \leq 8$
$v^+, v^-, y \geq 0$

---

optimal $w' = 11.2$
non-zero variables $v_3^+ = 4.8$
$v_4^+ = 6.4$

dual variables $\sigma^{v=1} = \begin{bmatrix} -3/11 \\ -1/11 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$

$h = [0, 0, 4.8, 6.4, 6, 8]^T$

$T = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

$\begin{cases} D_{r+1 = 0+1} = (\sigma^{v=1})^T T_k \\ d_{r+1 = 0+1} = (\sigma^{v=1})^T h_k \end{cases}$

$\begin{cases} D_1 = [0.273, 0.091] \\ d_1 = 11.2 \end{cases}$

$\Rightarrow \frac{3}{11} x_1 + \frac{1}{11} x_2 \geq 11.2$

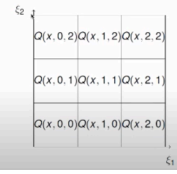
↳ plug this into step 1, repeat.

step 1 ⇄ step 2 → until $x \in K_2$ → step 3 → step 1; step 3 → until converge → x^*

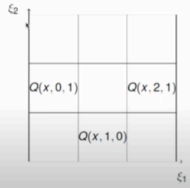
△ Monte-Carlo method.

Sample Average Approximation (SAA)

Expectation is an inescapable operator in stochastic programming: $V(x) = \sum_{\xi_1=0}^{2} \sum_{\xi_2=0}^{2} p(\xi_1, \xi_2) Q(x, \xi_1, \xi_2)$



Sample ξ according to $p(\xi)$,
$V(x) \simeq (Q(x, 0, 1) + Q(x, 2, 1) + Q(x, 1, 0))/3$



Pays off in higher dimensions

Idea: replace $V(x)$ by Monte Carlo estimate:

$$V^\nu(x) = \sum_{k=1}^{\nu} \frac{Q(x, \xi^k)}{\nu}$$

We are effectively solving the following extended form problem:

$$\min c^T x + \frac{1}{\nu} \sum_{k=1}^{K} q_k^T y_k$$

s.t. $Ax = b$
$T_k x + W y_k = h_k$
$x \geq 0, y_k \geq 0$

---

max cx s.t. $Ax \leq b$, $x \geq 0$ — min by, $A^T y \geq c$, $y \geq 0$

max cx s.t. $Ax \leq b$, $x \geq 0$ — min by, $A^T y \geq c$

max cx, $Ax \leq b$, $x \geq 0$ — min by, $A^T y \geq c$

△ Robust Optimization

$\min_x \; c^T x + d$

s.t. $Ax \le b$

$$D = \begin{bmatrix} c^T_{1\times n} & d_{1\times 1} \\ A_{m\times n} & b_{m\times 1} \end{bmatrix}$$

△ prediction errors

△ measurement errors

△ artificial data uncertainties

↳ implementation error

△ Definition

- $\left\{ \min_x \{ c^T x + d : Ax \le b \} \right\}_{D = \begin{bmatrix} c^T & d \\ A & b \end{bmatrix} \in \mathcal{U}}$
- $\mathcal{U} \subset \mathbb{R}^{(m+1)\times(n+1)}$
- $\mathcal{U} = \left\{ \begin{bmatrix} c^T & d \\ A & b \end{bmatrix} = \underbrace{\begin{bmatrix} c_0^T & d_0 \\ A_0 & b_0 \end{bmatrix}}_{\text{nominal data } D_0} + \sum_{l=1}^{L} \zeta_l \underbrace{\begin{bmatrix} c_l^T & d_l \\ A_l & b_l \end{bmatrix}}_{\text{basic shift } D_l} : \zeta \in \mathcal{Z} \subset \mathbb{R}^L \right\}$

e.g.

- $\mathcal{Z}$ is parallelotope

$\{ \zeta \in \mathbb{R}^k : -1 \le \zeta_j \le 1, \; j = 1 \sim k \}$

- $\mathcal{Z}$ is a ball

$\{ \zeta \in \mathbb{R}^k : \|x\|_2^2 = x^T x < 1 \}$

- A.1. All the decision variables in $(LO_{\mathcal{U}})$ represent "here and now" decisions; they should be assigned specific numerical values as a result of solving the problem *before* the actual data "reveals itself".
- A.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set $\mathcal{U}$ given by (1.2.1).
- A.3. The constraints in $(LO_{\mathcal{U}})$ are "hard" — we cannot tolerate violations of constraints, even small ones, when the data is in $\mathcal{U}$.

△ Robust feasible solution

$x \in \mathbb{R}^n$ is robust feasible solution

if $Ax \le b \quad \forall (c, d, A, b) \in \mathcal{U}$

△ Worst-Case-Oriented Assumptions

robust value

$$\hat{c}(x) = \sup_{(c,d,A,b) \in \mathcal{U}} [c^T x + d]$$

△ Robust Counterpart

$$\equiv \min_x \left\{ \hat{c}(x) = \sup_{(c,d,A,b)\in\mathcal{U}} [c^T x + d] : Ax \le b \right\}$$

x^* is the robust optimal solution

$\hat{c}(x^*)$ is the robust optimal value

△ observation

rewrite w/ epigraph

(A) RC

↕

$$\min_{x,t} \left\{ t : \begin{array}{l} c^T x + d \le t \\ Ax \le b \end{array} \; \Big| \; \forall (c,d,A,b) \in \mathcal{U} \right\}$$

upper bound

(B) decompose $\mathcal{U}$. $\mathcal{U}$ is direct product of $\mathcal{U}_i$'s w/ certain objective of $LO_{\mathcal{U}}$

$\min_x \{ c^T x + d : Ax \le b, \; \forall (A, b) \in \mathcal{U} \}$

$\Rightarrow \mathcal{U} = \mathcal{U}_1 \times \dots \times \mathcal{U}_m$

i.e., $a_i^T x \le b_i \quad \forall [a_i; b_i] \in \mathcal{U}_i$

(C) x is robust feasible solution

- $\bar{a}_i^T x = \sum_{j=1}^{J} \lambda_j [a_i^j]^T x \le \sum_j \lambda_j b_i^j = \bar{b}_i \qquad \sum_j \lambda_j = 1$

- $[\bar{a}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$

we loose nothing when assuming sets $\mathcal{U}_i$ are closed convex !

(D) • avoid adding slack variables

i.e., avoid converting ineq. → eq.

• eliminate state variables

- Sometimes a good for the RO methodology modeling requires eliminating "state variables" – those which are readily given by variables representing actual decisions – via the corresponding "state equations".

Example

Time dynamics of an inventory is given in the simplest case by the state equations

$$x_0 = c$$
$$x_{t+1} = x_t + q_t - d_t, \; t = 0, 1, \dots, T,$$

A wise approach to the RO processing would be eliminate the state variables x_t by setting:

$$x_t = c + \sum_{\tau=1}^{t-1} q_\tau, \; t = 0, 1, 2, \dots, T+1$$

△ NP hard

$\mathcal{U}$ is infinite.

- computationally intractable
- NP-hard

△ Tractability Analysis.

↓

23/4

△ chance constraint

- stochastic
- stochastic program
- robust programming

△ review

- stochastic
- robust

(before dinner)

△ Trackability Analysis

- as $\mathcal{U}$ has infinite elements

∴ has NP-hardness

∴ has computational intractability

- do tractability analysis

The RC of the uncertain LO problem with uncertainty set $\mathcal{U}$ is computationally tractable whenever the convex uncertainty set $\mathcal{U}$ itself is computationally tractable. assumed U is convex

The latter means that we know in advance the affine hull of $\mathcal{U}$, a point from the relative interior of $\mathcal{U}$, and we have access to an efficient membership oracle that, given on input a point u, reports whether $u \in \mathcal{U}$.

In computer science, an oracle is an abstract concept representing a hypothetical "black box" that can provide answers to certain questions or perform specific computations. Oracles are often used to analyze algorithms or problem-solving techniques without delving into the specifics of how those answers are obtained.

This is an oracle example.

e.g. consider

△ $\{ a^T x \le b \}_{[a;b] \in \mathcal{U}}$

s.t. $\mathcal{U} = \left\{ [a; b] = [a^0; b^0] + \sum_{l=1}^{L} \zeta_l [a^l; b^l] : \zeta \in \mathcal{Z} \right\}$

⇒ tractable representation

$a^T x \le b \quad \forall \left([a; b] = [a^0; b^0] + \sum_{l=1}^{L} \zeta_l [a^l; b^l] : \zeta \in \mathcal{Z} \right)$

△ let $\mathcal{Z} \equiv \{ \zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1 \}$

$\Rightarrow [a^0]^T x + \sum_{l=1}^{L} \zeta_l [a^l]^T x \le b^0 + \sum_{l=1}^{L} \zeta_l b^l$

$\Leftrightarrow \sum_{l=1}^{L} \zeta_l \left[[a^l]^T x - b^l \right] \le b^0 - [a^0]^T x$

$\Leftrightarrow \max_{-1 \le \zeta_l \le 1} \left[\sum_{l=1}^{L} \zeta_l \left[[a^l]^T x - b^l \right] \right] \le b^0 - [a^0]^T x$ $\quad \zeta_l$ 不見了!!!

$\Leftrightarrow \sum_{l=1}^{L} \left| [a^l]^T x - b^l \right| \le b^0 - [a^0]^T x$

tractable representation.

$\Leftrightarrow \begin{cases} -u_l \le [a^l]^T x - b^l \le u_l \\ \sum_{l=1}^{L} u_l \le b^0 - [a^0]^T x \end{cases}$

e.g.

$\mathcal{Z} = \{ \zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \Omega \}$

$\Rightarrow [a^0]^T x + \sum_{l=1}^{L} \zeta_l [a^l]^T x \le b^0 + \sum_{l=1}^{L} \zeta_l b^l$

$\Leftrightarrow \max_{\|\zeta\|_2 \le \Omega} \left[\sum_{l=1}^{L} \zeta_l \left[[a^l]^T x - b^l \right] \right] \le b^0 - [a^0]^T x$

$\Leftrightarrow \Omega \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)^2} \le b^0 - [a^0]^T x$

△ Chance constraints robust optimization

△ recall

$\{ a^T x \le b \}_{[a;b] \in \mathcal{U}}$

where

$\mathcal{U} = \left\{ [a; b] = [a^0; b^0] + \sum_{l=1}^{L} \zeta_l [a^l; b^l] : \zeta \in \mathcal{Z} \right\}$ — random w/ known distribution.

△

$p(x) \equiv \mathrm{Prob}_{\zeta \sim P} \left\{ \zeta : [a^0]^T x + \sum_{l=1}^{L} \zeta_l [a^l]^T x > b^0 + \sum_{l=1}^{L} \zeta_l b^l \right\} \le \epsilon$

chance that ζ leads to disobeying the constraints is $\le \epsilon$

- difficult to evaluate w/ high accuracy
- non-convex
- sometimes, P is unknown/partial known, P is a part of given family distribution — Ambiguous Chance Constraint

△ safe convex approximation

❑ Safe Convex Approximation

$$p(x) \equiv \mathrm{Prob}_{\zeta \sim P} \left\{ \zeta : [a^0]^T x + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x > b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell \right\} \le \epsilon, \qquad (2.2.2)$$

Definition 2.1.1

Let $\{[a^\ell; b^\ell]\}_{\ell=0}^{L}$, P, ϵ be the data of chance constraint (2.2.2), and let S be a system of convex constraints on x and additional variables v. We say that S is a *safe convex approximation* of chance constraint (2.2.2), if the x component of every feasible solution (x, v) of S is feasible for the chance constraint.

A safe convex approximation S of (2.2.2) is called computationally tractable, if the convex constraints forming S are efficiently computable.

這東?發生機率 $< \epsilon$

△ $p(x) \equiv \mathrm{Prob}_{\zeta \sim P} \left\{ \zeta : [a^0]^T x + \sum_{l=1}^{L} \zeta_l [a^l]^T x > b^0 + \sum_{l=1}^{L} \zeta_l b^l \right\} \le \epsilon$

we define

- $\eta = \sum_{l=1}^{L} \left([a^l]^T x - b^l \right) \zeta_l \le b^0 - [a^0]^T x \quad$ —①

$\Rightarrow \mathrm{Prob}_{\zeta \sim P} \{ \zeta : \eta \le b^0 - [a^0]^T x \} \ge 1 - \epsilon$

& also

- $E\{\zeta_l\} = 0$ & $|\zeta_l| \le 1$
- std

$\sqrt{\left(\sum_{l=1}^{L} \left([a^l]^T x - b^l \right) \right) \left(\underbrace{E(\zeta_l^2) - E(\zeta_l)^2}_{=0} \right)}$

$\le \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)^2} \quad$ —②

tractable safe approximation

give a safe parameter w/ ① & ②

$\left(\Omega \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)} \le b^0 - [a^0]^T x \right)$ ③

(proper Ω, $P(③) \ge 1 - \epsilon$)

$\left(\Omega \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)} \ge b^0 - [a^0]^T x \right)$ ④

(proper Ω, $P(④) \le \epsilon$)

△ proposition

$\mathrm{Prob}\left\{ \zeta : \sum_{l=1}^{L} \sigma_l \zeta_l > \Omega \sqrt{\sum_{l=1}^{L} \sigma_l^2} \right\} \le \exp\{-\Omega^2/2\}$

$\Rightarrow \mathrm{prob}\left\{ \eta > \Omega \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)} \right\} \le \exp\{-\Omega^2/2\}$

from ④ we know

$\mathrm{prob}\left\{ \Omega \sqrt{\sum_{l=1}^{L} \left([a^l]^T x - b^l \right)} \ge b^0 - [a^0]^T x \right\} < \epsilon'$

$\Rightarrow \mathrm{prob}\{ \eta > \Omega \cdots \ge b^0 - [a^0]^T x \} \le \exp\{-\Omega^2/2\}$

$\Rightarrow \mathrm{prob}\left\{ \sum_{l=1}^{L} \left([a^l]^T x - b^l \right) \zeta_l \ge b^0 - [a^0]^T x \right\} \le \exp\{-\Omega^2/2\}$

$\Rightarrow \mathrm{prob}\left\{ [a^0]^T x + \sum_{l=1}^{L} \zeta_l [a^l]^T x > b^0 - \sum_{l=1}^{L} \zeta_l b^l \right\} \le \exp\{-\Omega^2/2\}$

$\Rightarrow \Omega \ge \sqrt{2 \ln(1/\epsilon)}$

$\Rightarrow$ ③ is a tractable safe approximation of safe constraint of ①

△ Statistical Estimation

△ Maximum Likelihood Estimation

Maximum A Posteriori

△ Liverpool won 30/38 last year

⇒ 試問 what is the winning probability in next year?

⇒ 答 30/38 , ⇒ this is MLE

△ How about,
the average winning percentage is 50%
actually 50% for the past 10 years.
Is your guess still 30/38 ?

⇒ 答 $\frac{1}{2} \sim \frac{30}{38}$ ⇒ this is MAP

△ MLE (what model best describes my data ⇔ that likelihood 最大)

△ assume binomial distribution

$$P(k/n \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

read : the prob of k win over n match given a winning model (probability)

△ let's say $\theta = 0.1$

⇒ $P(30/38 \mid \theta = 0.1) = 2.11 \times 10^{-21}$

我有 30/38 這數據，
我有 2.11×10^{-21} 的 likelihood 會有 $\theta=0.1$ 這個 model

我有 $\theta = 0.1$ 這個 model
我有 2.11×10^{-21} 的 likelihood 會有 30/38 這數據

△ maximize $P(D \mid \theta)$!

$$\frac{dP(D|\theta)}{d\theta} = \binom{n}{k}\left(k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}\right)$$
$$= \binom{n}{k}\theta^{k-1}(1-\theta)^{n-k-1}\left(k(1-\theta) - (n-k)\theta\right)$$
$$= 0$$

$\theta = 0, 1, k/n$ ⇒ 30/38 = 78.9%

△ MAP

- can be useful when data is little
  e.g. win 2/2 over 2

  MLE = 100%
  ↳ unlikely
  ↳ use prior data of 50% (avg winning perc.g.)

- $\operatorname{argmax}_\theta P(\theta \mid D)$

⇔ $\operatorname{argmax}_\theta \frac{P(D|\theta)P(\theta)}{P(D)}$

⇔ $\operatorname{argmax}_\theta P(D|\theta)P(\theta)$

- let $P(\theta)$ be beta distribution

$$P(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Conjugate prior of binomial is beta

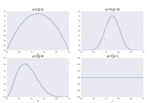


Figure 2. Visualisations of Beta distribution with different values of α and β

A conjugate prior is a prior distribution that, when combined with a likelihood function, results in a posterior distribution that belongs to the same family of distributions as the prior. In simpler terms, it's a prior distribution that, when updated with data using Bayes' theorem, conveniently leads to a posterior distribution with a known form.

For example, in Bayesian statistics, if you're estimating a parameter like the mean of a normal distribution and you choose a normal distribution as your prior, the posterior distribution will also be normal. This makes calculations much easier because you can update your prior beliefs with data in a straightforward manner.

Conjugate priors are particularly useful because they can simplify the Bayesian analysis process by allowing for closed-form solutions rather than requiring computationally intensive methods like Markov Chain Monte Carlo (MCMC) sampling. However, conjugate priors are not always available or appropriate for every problem, and in many cases, non-conjugate priors or computational methods like MCMC are necessary.

Using the Beta distribution as the prior for θ is advantageous because when combined with the binomial likelihood function (the data), the resulting posterior distribution for θ will also be a Beta distribution. This makes updating our beliefs about θ with observed data straightforward and analytically tractable.

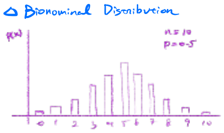
$$P(D|\theta)P(\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$
$$= \binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

$$\frac{dP(D|\theta)P(\theta)}{d\theta} = \binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left((k+\alpha-1)\theta^{k+\alpha-2}(1-\theta)^{n-k+\beta-1} - (n-k+\beta-1)\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-2}\right)$$
$$= \binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-2}(1-\theta)^{n-k+\beta-2}\left((k+\alpha-1)(1-\theta) - (n-k+\beta-1)\theta\right)$$
$$= 0$$

$$\theta = \frac{k+\alpha-1}{n+\alpha+\beta-2}$$

In this example, assuming we use $\alpha=10$ and $\beta=10$, then $\theta=(30+10-1)/(38+10+10-2) = 39/56 = 69.6\%$

△ Bionominal Distribution



△ Histogram density estimator

$B_1 = [0, \frac{1}{M})$
$B_2 = [\frac{1}{M}, \frac{2}{M})$
⋮
$B_{M-1} = [\frac{M-2}{M}, \frac{M-1}{M})$
$B_M = [\frac{M-1}{M}, 1]$

$\{B_\ell\}$ bins.

M = no. of bins
n = no. of data pts
w = width of each bin

$$\hat{P}_M(x) = \frac{|B_\ell|}{n\,w}$$

$\hat{P}_M(x) = f(x_0, w)$ — x_0: bin starting pt., w: width of each bin

Theorem I

△ suppose $\exists L$ s.t.

- $\sup_x |p'(x)| \le L$
- $\sup_x p(x) < \infty$ (p_{max})

⇒
$\text{bias}(\hat{P}_M(x)) \le \frac{L}{M}$
$\text{var}(\hat{P}_M(x)) \le M\frac{p_{max}}{n} + \frac{p_{max}^2}{n}$

proof.

- bias

$E(\hat{P}_M(x))$ — expectation of the estimation model of "ℓ" bin

$= M \cdot P(X_i \in B_\ell)$ → $\frac{P(x^*)}{M}$

$= M \cdot \int_{\frac{\ell-1}{M}}^{\frac{\ell}{M}} p(u)\,du$

$= M\left(F(\tfrac{\ell}{M}) - F(\tfrac{\ell-1}{M})\right)$ (CDF)

$= \frac{F(\frac{\ell}{M}) - F(\frac{\ell-1}{M})}{\frac{1}{M}}$

$= \frac{F(\frac{\ell}{M}) - F(\frac{\ell-1}{M})}{\frac{\ell}{M} - \frac{\ell-1}{M}}$

$= p(x^*) \quad x^* \in [\frac{\ell-1}{M}, \frac{\ell}{M}]$

∴ $\text{bias}(\hat{P}_M(x))$
$= E(\hat{P}_M(x)) - p(x)$
$= p(x^*) - p(x)$
$= p'(x^{**})\cdot(x^* - x)$ (mean value theorem)
$\le L \cdot \frac{1}{M}$ (width of a histogram)

- variance

$\text{Var}(\hat{P}_M(x))$

$= M^2 \,\text{var}\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i \in B_\ell)\right)$ ($1, 0$; $E(x^2) - [E(x)]^2$)

$= M^2 \frac{P(X_i \in B_\ell)(1 - P(X_i \in B_\ell))}{n}$

$= M^2 \frac{\frac{p(x^*)}{M}\cdot\left(1 - \frac{p(x^*)}{M}\right)}{n}$

$= M \cdot \frac{p(x^*)}{n} + \frac{p^2(x^*)}{n}$

$\le M \cdot \frac{p_{max}}{n} + \frac{p^2_{max}}{n}$

- lipschitz

$|p(x) - p(y)| \le L|x-y|$

⇒ $|\text{bias}(\hat{P}_M(x))| = |E(\hat{P}_M(x)) - p(x)|$
$= |p(x^*) - p(x)|$
$\le L|x^* - x|$
$\le \frac{L}{M}$

△ Mean Square Error

$MSE(\hat{P}_M(x))$
$= \text{bias}^2(\hat{P}_M(x)) + \text{Var}(\hat{P}_M(x))$
$\le \frac{L^2}{M^2} + M\frac{p_{max}}{n} + \frac{p^2_{max}}{n}$

- under smoothing
  $M\uparrow$
  bias ↓
  variance ↑
- over smoothing
  $M\downarrow$
  bias ↑
  variance ↓

M is the bandwidth selection

△ Kernel Density Estimator

$$\hat{P}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

└ kernel function, like Gaussian
└ $h > 0$, bandwidth

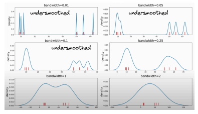
△ MISE

mean integrated squared error

$$MISE = E\left[\int (\hat{f}_h(x) - f(x))^2\,dx\right]$$

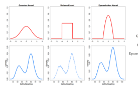
best h

⇒ $\frac{\partial}{\partial h} MISE(h) = 0$



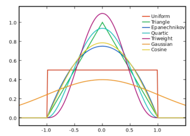
△ Kernel Density Estimator

K_1 $K(x)$ is symmetric
K_2 $\int K(x)\,dx = 1$
K_3 $\lim_{x\to-\infty} K(x) = \lim_{x\to\infty} K(x) = 0$



Gaussian $K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$,
Uniform $K(x) = \frac{1}{2}I(-1 \le x \le 1)$,
Epanechnikov $K(x) = \frac{3}{4}\cdot\max\{1-x^2, 0\}$.

The ***Epanechnikov*** is a special kernel that has the lowest (asymptotic) mean square error.



Theorem

- assume $p'''(x)$ is bounded
  K_1, K_2, K_3 hold
- $h \to 0$, $n \to \infty$ we have

$\text{bias}(\hat{P}_n(x_0)) = \frac{1}{2}h^2 p''(x_0)\mu_K + o(h^2)$
$\text{var}(\hat{P}_n(x_0)) = \frac{1}{nh}p(x_0)\sigma_K^2 + o\left(\frac{1}{nh}\right)$

- **Bias**
  - When we allow $h \to 0$, the bias is shrinking at a rate $o(h^2)$.
  - The bias of KDE is caused by the *curvature* (second derivative) of the density function! Namely, the bias will be very large at a point where the density function curves a lot (e.g., a very peaked bump).
  - This makes sense because for such a structure, KDE tends to smooth it too much, making the density function smoother (less curved) than it used to be.
- **Variance**
  - The variance shrinks at rate $O\left(\frac{1}{nh}\right)$ when $n \to \infty$ and $h \to 0$.
  - At point where the density value is large, the variance is also large!

❑ **Kernel Density Estimator**

- **MSE**

Asymptotic Mean Square Error (AMSE)

$$\begin{aligned}\mathbf{MSE}(\hat{p}_n(x_0)) &= \mathbf{bias}^2(\hat{p}_n(x_0)) + \mathsf{Var}(\hat{p}_n(x_0)) \\ &= \frac{1}{4}h^4|p''(x_0)|^2\mu_K^2 + \frac{1}{nh}p(x_0)\sigma_K^2 + o(h^4) + o\left(\frac{1}{nh}\right) \\ &= O(h^4) + O\left(\frac{1}{nh}\right).\end{aligned}$$

The bandwidth h minimizing the AMSE is:

$$h_{opt}(x_0) = \left(\frac{4}{n}\cdot\frac{p(x_0)}{|p''(x_0)|^2}\frac{\sigma_K^2}{\mu_K^2}\right)^{\frac{1}{5}} = C_1 \cdot n^{-\frac{1}{5}}$$

And this choice of smoothing bandwidth leads to a MSE at rate:

$$\mathbf{MSE}_{opt}(\hat{p}_n(x_0)) = O(n^{-\frac{4}{5}})$$

- **Discussion**
  - The optimal MSE of the KDE is at rate $O(n^{-\frac{4}{5}})$, which is faster than the optimal MSE of the histogram $O(n^{-\frac{2}{3}})$. However, both are slower than the MSE of a MLE ($O(n^{-1})$).
  - This reduction of error rate is the price we have to pay for a more flexible model.

△ K-nearest neighbour (k-NN)

- $X_1, X_2 \sim X_n$
  $X_i \in \mathbb{R}^d$
- for a given point x

  $R_k(x)$: distance
  $x \longleftrightarrow x$'s k^{th} nearest neighbour

- k-NN density estimator estimates:

$$\hat{P}_{kNN}(x) = \frac{k}{n}\cdot\frac{1}{V_d R_k^d(x)}$$

Volume of a d-dimensional ball w/ radius $R_k(x)$

where $V_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ (gamma) π

e.g. $d=1$
$X = \{1, 2, 6, 11, 13, 14, 20, 33\}$

$x=5$, $k=2$, $\hat{P}_{kNN}(5)$?

⇒ $R_k(5) = \{4, 3, 1, 6, 8, 9, 15, 28\}$
$R_2(5) = 3$

∴ $\hat{P}_{kNN}(5) = \frac{2}{8}\cdot\frac{1}{2 \cdot R^{d=1}_{k=2}(5)} = \frac{1}{24}$

$d=1$, $V_1 = 2$ $\left(\frac{\pi^{0.5}}{\Gamma(\frac{3}{2})} = 1.999\ldots\right)$
$d=2$, $V_1 = \pi$
$d=3$, $V_1 = \frac{4}{3}\pi$

△ Dynamical Systems

$$e = \sum_{n \to \infty} \left(1+\frac{1}{n}\right)^{n}$$

$$e^{A} = \exp A$$

$$= I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 \cdots$$

$$e^{x} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$f(x) = f(a) + \frac{df}{dx}(a)(x-a)$$

$$+ \frac{d^2f}{dx}(a)(x-a)^2$$

$$+ \frac{d^3f}{dx}(a)(x-a)^3$$

e.g.

$f(x) = \sin(x)$ @ $a=0$

$$= \sin(0) + \cos(0)x + -\sin(0)\cdot\frac{x^2}{2}$$

$$= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$f(x) = \cos(x)$ @ $a=0$

$$= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

$$e^{x} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \cdots$$

$$e^{ix} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!}$$

$$= \left(1 + \frac{-x^2}{2!} + \frac{x^4}{4} + \cdots\right)$$

$$+ i\left(ix - \frac{ix^3}{3!} + \frac{ix^5}{5!} + \cdots\right)$$

$$= \cos x + i\sin x$$