Comparing solutions : Independence & Dependence

We will discuss how the solutions to stochastic optimization problems with independent distributions & a distributionally robust approach can be compared. Note that, as we have seen, in general stochastic optimization problems with independent discrete distributions are #P-hard. On the other hand, we can sample from this distribution. In comparison, the distributionally robust problems can be solved efficiently in some special cases, but in other cases it is not clear how to find the worst-case distribution & hence sample from it.

<u>Problem</u> : Consider a collection of random variables $\tilde{\xi}_1, \ldots, \tilde{\xi}_n$ with marginal distributions $F_1, \ldots, F_n$ where $\Theta(F_1, \ldots, F_n)$ denotes the Frechet class of joint distributions with fixed marginals.

Compare the objective & optimal solutions from:

$$(*) \quad \min_{x \in \overline{X}} E_{\theta_{ind}}[g(x, \tilde{\xi})] \qquad \text{where } \theta_{ind} = \underbrace{\theta_1 \times \theta_2 \times \ldots \times \theta_n}_{\text{product measure}}$$

$$(**) \quad \min_{x \in \overline{X}} \sup_{\theta \in \Theta} E_{\theta}[g(x, \tilde{\xi})]$$

(we use cdf & probability measures interchangeably)

Let $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ where Ω_i is the support of random variable $\tilde{\xi}_i$.

For a function $f(\xi) : \Omega \to R_+$, define

$$\text{Correlation gap} = \frac{\sup_{\theta \in \Theta(F_1, \dots, F_n)} E_\theta[f(\tilde{\xi})]}{E_{\theta_{ind}}[f(\tilde{\xi})]}.$$

Assume $g(x, \xi) : X \times \Omega \to R_+$. Let,

$$\underbrace{\text{Correlation gap of function } g(\cdot,\cdot) \text{ at } x}_{K(x)} = \frac{\sup_{\theta \in \Theta(F_1, \dots F_n)} E_\theta[g(x, \tilde{\xi})]}{E_{\theta_{ind}}[g(x, \tilde{\xi})]} \quad \forall x \in X$$

Let,

$$x_I = \arg\min \left\{ E_{\theta_{ind}}[g(x, \tilde{\xi})] : x \in X \right\}$$

$$x_R = \arg\min \left\{ \sup_{\theta \in \Theta(F_1 \dots F_n)} E_\theta[g(x, \tilde{\xi})] : x \in X \right\}$$

where I = independent, R = robust.

The correlation gap can be used to compare the performance of solutions.

For example, define

$$\underbrace{\text{Price of correlation}}_{POC} = \frac{\sup_{\theta \in \Theta(F_1, \dots F_n)} E_\theta[g(x_I, \tilde{\xi})]}{\sup_{\theta \in \Theta(F_1, \dots F_n)} E_\theta[g(x_R, \tilde{\xi})]}$$

$POC \geq 1$ where $POC = 1$ implies the performance of the optimal solution from independent distribution model does as well as robust solution in worst case distributions.

Proposition

Suppose $K(x) \leq K \;\; \forall x \in \mathcal{X}$. Then, $POC \leq K$.

Proof

$$POC = \frac{\sup_{\Theta \in \Theta(F_1, \ldots F_n)} E_\Theta[g(x_I, \xi)]}{\min_{x \in \mathcal{X}} \sup_{\Theta \in \Theta(F_1, \ldots F_n)} E_\Theta[g(x, \xi)]}$$

$$\leq \frac{\sup_{\Theta \in \Theta(F_1, \ldots F_n)} E_\Theta[g(x_I, \xi)]}{E_{\Theta_{ind}}[g(x_R, \xi)]}$$

(since x_R is robust solution & $\sup_{\Theta \in \Theta} E[g(x_R, \xi)] \geq E_{\Theta_{ind}}[g(x_R, \xi)]$)

$$\leq \frac{\sup_{\Theta \in \Theta(F_1, \ldots F_n)} E_\Theta[g(x_I, \xi)]}{E_{\Theta_{ind}}[g(x_I, \xi)]}$$

(since x_I is optimal for Θ_{ind} & hence $E_{\Theta_{ind}}[g(x_I, \xi)] \leq E_{\Theta_{ind}}[g(x_R, \xi)]$)

$$= K(x_I)$$

$$\leq K$$

Hence an uniform bound on the correlation gap implies an uniform bound on POC.

K is small implies, solution from independent distribution is close to robust solution in worst-case.

The correlation gap can be similarly used to compare the performance of robust solution to the optimal solution from the independent distribution under nicely behaved distributions such as independence.

$$\frac{E_{\Theta_{ind}}[g(x_R, \tilde{\xi})]}{E_{\Theta_{ind}}[g(x_I, \tilde{\xi})]} \leq \frac{\sup\limits_{\Theta \in \Theta(F_1, \dots F_n)} E_{\Theta}[g(x_R, \tilde{\xi})]}{E_{\Theta_{ind}}[g(x_I, \tilde{\xi})]}$$

(Since the worst-case distribution will have higher cost than Θ_{ind} for fixed x_R)

$$\leq \frac{\sup\limits_{\Theta \in \Theta(F_1, \dots F_n)} E_{\Theta}[g(x_I, \tilde{\xi})]}{E_{\Theta_{ind}}[g(x_I, \tilde{\xi})]}$$

(x_R is optimum under worst case while x_I is only feasible)

$$= K(x_I)$$

$$\leq K$$

Here the correlation gap provides an upper bound on the performance of the robust solution for the independent distribution compared to the optimal solution x_I.

We focus on monotone, nonnegative submodular functions & introduce another useful bound.

Let $\hat{f}_4(p) = \min_{S \subseteq N} f(S) + \sum_{i \in N} f_S(i) p_i$

Proposition For monotone, submodular functions

$$\hat{f}_2(p) \leq \hat{f}_4(p) \quad \forall p \in [0,1]^n$$

Proof Consider any feasible solution to the LP in $\hat{f}_2(p)$ denoted by p_S for $S \subseteq N$.

Then,

$$\sum_{S \subseteq N} p_S f(S) \leq \sum_{S \subseteq N} p_S \left(f(T) + \sum_{i \in S} f_T(i) \right) \quad \forall T \subseteq N$$

(since $p_S \geq 0 \ \forall S$ & f is nondecreasing & submodular)

$$= f(T) + \sum_{S \subseteq N} \sum_{i \in S} p_S f_T(i) \quad \forall T \subseteq N$$

(since $\sum_{S \subseteq N} p_S = 1$)

$$= f(T) + \sum_{i \in N} \left(\sum_{\substack{S \subseteq N \\ S \ni i}} p_S \right) f_T(i) \quad \forall T \subseteq N$$

(interchanging order of summations)

$$= f(T) + \sum_{i \in N} p_i f_T(i) \quad \forall T \subseteq N$$

(from feasibility conditions in linear program)

Taking minimum over all $T \subseteq N$ & then maximizing over all feasible $p_S \ \forall S \subseteq N$ on the LHS subject to marginals, we get the upper bound.

We will focus on functions of the form

$f(\xi) : \{0,1\}^n \rightarrow R$. where $N = \{1,\dots,n\}$

and are interested in computing the expectation

$E[f(\tilde{\xi})]$ with information on marginals given by

$P(\tilde{\xi}_i = 1) = p_i$ for $i = 1,\dots,n$.

This can be viewed as Bernoulli random variables.

Note for a given $\xi \in \{0,1\}^n$, we equivalently use the set $S = \{i \in \{1,\dots,n\} : \xi_i = 1\}$.

We consider the following extensions of the functions from $\{0,1\}^n$ to $[0,1]^n$.

1) <u>Independence (multilinear extension)</u>

$$\hat{f}_1(p) = E_{\theta_{ind}}[f(\tilde{\xi})] = \sum_{S \subseteq N} \left(\prod_{i \in S} p_i\right)\left(\prod_{i \in N \setminus S}(1-p_i)\right) f(S)$$

(Compute expectation where each ξ_i is included with probability p_i & not with probability $1-p_i$ independently)

2) <u>Maximum expectation with marginals (~~concave~~ closure)</u>

$$\underbrace{\hat{f}_2(p) = \max_{\theta \in \Theta(F_1,\dots F_n)} E_\theta[f(\tilde{\xi})]}_{\text{concave in } p} = \max \sum_{S \subseteq N} p_S f(S)$$

s.t $\sum_{S \subseteq N} p_S = 1$, $P(S) \geq 0 \ \forall S \subseteq N$

$\sum_{S \subseteq N : S \ni i} p_S = p_i \ \forall i \in N$

3) <u>Minimum expectation with marginals (convex closure)</u>

$$\hat{f}_3(p) = \min_{\theta \in \Theta(F_1,\dots F_n)} E_\theta[f(\tilde{\xi})] = \min \sum_{S \subseteq N} p_S f(S)$$

s.t $\sum_{S \subseteq N} p_S = 1$, $P(S) \geq 0 \ \forall S \subseteq N$

$\sum_{S \subseteq N : S \ni i} p_S = p_i \ \forall i \in N$

<u>$\hat{f}_3(p) \leq \hat{f}_1(p) \leq \hat{f}_2(p)$</u>

<u>Proposition</u> For nonnegative, monotone, submodular function $f(\cdot)$, we have

$$\hat{f}_1(p) \geq \left(1 - \frac{1}{e}\right) \hat{f}_4(p) \quad \text{for all } p \in [0,1]^n$$

<u>Proof</u>. We create a random process which starts with an empty set at time 0, $S(0) = \emptyset$. For each $i \in N$, we have an independent Poisson clock C_i that sends signals at rate p_i. If C_i sends a signal, we include i to S. This increases the objective value by $f_S(i)$ (where if i is already present in the set, contribution $= 0$).

Let $S(t)$ = Random set of i at time t included in the set

At time 1, $S(1)$ contains element i with probability given by

$$1 - P\left(N_i(1) = 0\right) = 1 - e^{-p_i} \leq p_i \quad \left(\text{since } e^x \geq 1+x\right) \quad (*)$$

where $P(N_i(t) = n) = \dfrac{e^{-p_i t}(p_i t)^n}{n!}$ for $n = 0, 1, 2, \ldots$

is the probability of n events happening in time $[0, t]$.

Since these happen independently & $f(\cdot)$ is monotone & nonnegative, we have from $(*)$

$$\hat{f}_1(p) \geq E[f(S(1))]$$

as the random set has a lower probability of including the element.

The expected increase in objective in interval dt, conditioned on $S(t) = S$ is given by.

$$E\left[f(S(t+dt)) - f(S(t)) \mid S(t) = S \right]$$

$$= \sum_{i \in N} f_S(i)\, p_i \, dt$$ (effectively disjoint events for small dt in Poisson clocks & ignoring dt^2 terms)

$$\geq \left(\hat{f}_4(p) - f(S) \right) dt$$

(since $\hat{f}_4(p) = \min_{S \subseteq N} f(S) + \sum_{i \in N} f_S(i)\, p_i$)

Divide by dt & take expectation over S,

$$\frac{E[f(S(t+dt)) - f(S(t))]}{dt} \geq \hat{f}_4(p) - E[f(S(t))]$$

Let $\phi(t) = E[f(S(t))]$. Then,

$$\frac{d\phi}{dt} \geq \hat{f}_4(p) - \phi$$

Let $\psi(t) = e^t \phi(t)$. Then, $\frac{d\psi}{dt} = e^t \frac{d\phi}{dt} + e^t \phi(t)$.

$$\therefore \frac{d\psi}{dt} = e^t\left(\phi(t) + \frac{d\phi}{dt} \right) \geq e^t \hat{f}_4(p)$$

Here $\psi(0) = \phi(0) = E[f(S(0))] = E[f(\emptyset)] = 0$

$$\therefore \psi(t) \geq \int_0^t e^x \hat{f}_4(p)\, dx = (e^t - 1)\hat{f}_4(p)$$

Hence, $E[f(S(t))] = e^{-t}\psi(t)$

$$\geq e^{-t}(e^t - 1)\, \hat{f}_4(p)$$

$$= \left(1 - \frac{1}{e^t}\right) \hat{f}_4(p)$$

Finally, $\hat{f}_1(p) \geq E[f(S(1))] \geq \left(1 - \frac{1}{e}\right)\hat{f}_4(p)$.

This implies, for nonnegative monotone submodular functions, we have for all $p \in [0,1]^n$

$$\underbrace{\hat{f}_1(p)}_{\text{ind}} \le \underbrace{\hat{f}_2(p)}_{\text{worst case}} \le \hat{f}_4(p) \le \left(\frac{e}{e-1}\right) \underbrace{\hat{f}_1(p)}_{\text{ind}}$$

Then,

$$\text{Correlation gap} = \frac{\hat{f}_2(p)}{\hat{f}_1(p)} \le \frac{e}{e-1}$$

$$= 1.5819.$$

This implies when dealing with such functions, the price of correlation can be bounded by $e/e-1$.

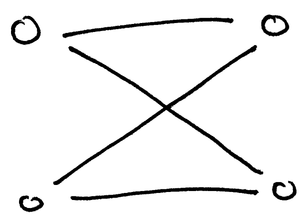
<u>Example</u>: Bottleneck assignment or matching problem.

$$\min \underbrace{\max_{i,j} c_{ij} x_{ij}}$$

$$\sum_j x_{ij} = 1 \ \forall i$$

$$\sum_i x_{ij} = 1 \ \forall j$$

$$x_{ij} \in \{0,1\} \ \forall i,j$$



The bottleneck objective is submodular & nondecreasing & nonnegative & hence correlation gap $\le \frac{e}{e-1}$

Here, we can solve the distributionally robust problem efficiently using methods from previous lectures.

Worst-case distribution (Lower bound - submodular)

Assume $f(\cdot)$ is submodular with $f(\emptyset) = 0$.

The convex closure is defined as:

$$\hat{f}_3(P) = \min \sum_{S \subseteq N} P_S f(S)$$

$$\text{s.t.} \sum_{S \subseteq N : S \ni i} P(S) = P_i \quad \forall i \in N \qquad (P)$$

$$\sum_{S \subseteq N} P(S) = 1$$

$$P(S) \geq 0 \quad \forall S \subseteq N$$

These equivalently correspond to finding upper bounds on the expected value of supermodular functions with given marginals.

In this case, the worst-case distribution has a simple chain structure (also referred to as comonotone distributions).

We assume WLOG that $1 \geq P_1 \geq P_2 \geq \cdots \geq P_{n-1} \geq P_n \geq 0$.

Consider the dual formulation of (P) given as:

$$\hat{f}_3^D(P) = \max \; Y_0 + \sum_{i \in N} Y_i P_i$$

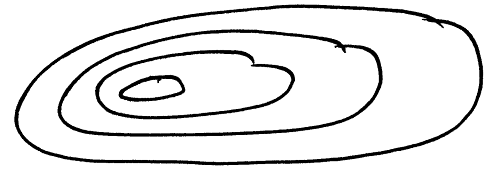
$$\text{s.t.} \; Y_0 + \sum_{i \in S} Y_i \leq f(S) \quad \forall S \subseteq N.$$

We construct an optimal primal & dual solution as follows:

<u>Primal</u>: Consider the sets $\emptyset = S_0 \subset S_1 \subset S_2 \subset \cdots \subseteq S_n = N$

where $1 \geq p_1 \geq p_2 \cdots \geq p_n \geq 0$

	S	p_S
S_0	$\emptyset$	$1-p_1$
S_1	$\{1\}$	$p_1 - p_2$
S_2	$\{1,2\}$	$p_2 - p_3$
$\vdots$		
S_{n-1}	$\{1,2,\ldots n-1\}$	$p_{n-1} - p_n$
S_n	$\{1,2,\ldots,n\}$	p_n



Objective $= (1-p_1) f(\emptyset) + (p_1 - p_2) f(\{1\}) + \cdots + p_n f(\{1,\ldots n\})$

$= \sum_i f(S_i)(p_i - p_{i+1})$

All other sets S, make $p_S = 0$. This is a feasible solution to the primal.

<u>Dual</u>: Set $y_0 = 0$, $y_i = f(S_i) - f(S_{i-1})$ for $i \in N$.

Then, $y_0 + \sum_{i \in S} y_i = \sum_{i \in S}\left[f(S_i) - f(S_{i-1})\right]$

$\leq \sum_{i \in S}\left[f((S_{i-1} \cap S) \cup \{i\}) - f(S_{i-1} \cap S)\right]$

($\because f$ is submodular & add i to S_{i-1} & $S_{i-1} \cap S$)

$= f(S) \Rightarrow$ dual feasible

(For example, $S = \{1,2,4,6\}$, then LHS =

$f(1) - f(\emptyset) + f(1,2) - f(1) + \underbrace{f(1,2,3,4) - f(1,2,3)} + \underline{f(1,2,3,4,5,6)}$

$- \underline{f(1,2,3,4,5)}$

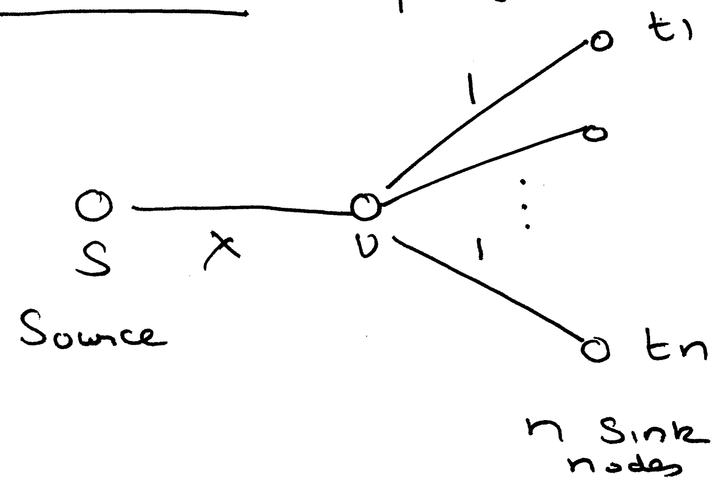
$\leq f(\cancel{1}) - f(\emptyset) + f(\cancel{1,2}) - f(\cancel{1}) + f(1,\cancel{2},4) - f(\cancel{1,2}) + f(1,2,4,6) - f(\cancel{1,2,4})$

$= f(1,2,4,6) - f(\emptyset)$)

Note (P) & (D) objective are equal & hence optimal since dual objective $= \sum_i p_i (f(S_i) - f(S_{i-1}))$

Example

Capacity decision x



Demand at each node

$\underbrace{P(\tilde{\xi}_i = 0) = 1/2 \quad P(\tilde{\xi}_i = 1) = 1/2}$

Total demand lies in $\{0, 1, 2, \ldots, n\}$ which needs to be satisfied

$\underbrace{\text{First stage decision}}$

Choose capacity on edge (s, v) denoted by x

$$c_1(x) = \begin{cases} x & x \le n-1 \\ n+1 & x = n \end{cases}$$

$\underbrace{\text{Second stage}}$

You can buy capacity after demands are known but at high cost

$$c_2(x) = 2^n x$$

$$\min_{x \ge 0} \; c_1(x) + E[h(x, \tilde{\xi})]$$
$$h(x, \tilde{\xi}) = c_2\left(\sum_{i=1}^{n} \tilde{\xi}_i - x\right)^+$$

Note $h(x, \xi) = h(x, S)$ is supermodular & nondecreasing in S for a given x. for $c_2 \ge 0$.

$$h(\cdot, S) = c_2(|S| - x)^+ \quad \forall S \subseteq N$$

- Nondecreasing is obvious since increasing in $|S|$.
- Supermodularity can be checked by verifying

$$(|S \cup i| - x)^+ + (|S| - x)^+ \le (|T \cup i| - x)^+ - (|T| - x)^+$$
$$\forall S \subseteq T \subseteq N, \; i \in N \setminus T.$$

The worst case distribution is given by the chain distribution, namely

$$(\tilde{\xi}_1, \ldots, \tilde{\xi}_n) = \begin{cases} (0, \ldots 0) & \text{w.p } 1/2 \\ (1, \ldots, 1) & \text{w.p } 1/2 \end{cases}$$

Then,

$$\min_{x \geq 0} c_1(x) + \sup_{Q \in \mathcal{Q}} E_Q \left[c_2 \left(\sum_i \tilde{\xi}_i - x \right)^+ \right]$$

$$\Leftrightarrow \min_{x \geq 0} c_1(x) + \frac{1}{2} c_2 (n-x)^+ = \min_{x \geq 0} c_1(x) + \frac{2^n (n-x)^+}{2}$$

Set $x = n$ & objective $= n+1$.

If we choose any $x < n$, then we pick up exponential cost in the second stage which cannot be optimal.

On the other hand, for the independent distribution

$$\min_{x \geq 0} c_1(x) + 2^n E \big(\underbrace{Bin(n, 1/2)}_{} - x \big)^+$$

Binomial = sum of Bernoulli i.i.d

This random variable takes values $\{0, 1, 2, \ldots k, \ldots n\}$ where it takes value k with probability ${}^nC_k \left(\frac{1}{2}\right)^n$.

Here $x = n-1$ is optimal with objective given by $n-1 + \frac{2^n}{2^n} = n$. On the other hand this solution behaves very poorly under the worst-case with cost of $n-1 + \frac{2^n}{2} = 2^{n-1} + n-1$ which is exponentially large.