Propriedades de Programas Estocásticos (Multi-estágios) com Recurso

Prof. Alexandre Salles da Cunha

Universidade Federal de Minas Gerais
Departamento de Ciência da Computação
Belo Horizonte, Brasil

acunha@dcc.ufmg.br

2021/2

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA
COMPUTAÇÃO

Referências bibliográficas

Relativas à esta apresentação

1. J. R. Birge & F. Louveaux. Introduction to Stochastic Programming, Springer, 2011, 2a Edição [Caps 2, 3 e 4]

Conteúdo

1. Revisão de probabilidades
2. Problemas multi-períodos vs multi-estágios, um exemplo.
3. Propriedades de problemas de Programação Estocástica Lineares (e inteiros) com dois ou múltiplos estágios com recurso.
4. Desigualdades envolvendo EVPI, VSS, RP.

1. Incerteza é representada por meio de experimentos aleatórios. Os resultados destes experimentos são indicados por ω.
2. O conjunto de todos os resultados é Ω.
3. Os resultados podem ser combinados em subconjuntos de Ω, denominados eventos. Exemplo: se $\Omega = \{1, 2, 3, 4, 5, 6\}$ representa os possíveis resultados de se jogar um dado, o subconjunto de eventos $\mathcal{A}$ pode representar os elementos ímpares de Ω.
4. Para cada elemento $A \in \mathcal{A}$ é atribuída uma probabilidade $P(A)$, satisfazendo $0 \leq P(A) \leq 1, P(\emptyset) = 0, P(\Omega) = 1$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ se $A_1 \cap A_2 = \emptyset$.
5. É comum em Programação Estocástica, que os elementos de Ω sejam empregados para representar estados possíveis finitos do mundo ou cenários. Neste caso, todos os elementos incertos dependem destes cenários finitos.

É uma descrição do quão provável são os valores que podem ser assumidos por uma ou um conjunto de variáveis aleatórias. A forma como apresentamos a distribuição de probabilidades depende das variáveis aleatórias serem discretas ou contínuas.

1. Para uma variável aleatória $\boldsymbol{\xi}$, definimos sua distribuição de probabilidades acumulada como:

$$F_{\boldsymbol{\xi}}(x) = P(\boldsymbol{\xi} \leq x) = P(\{\omega | \boldsymbol{\xi} \leq x\}).$$

2. Dois casos devem ser considerados:
   1. Variáveis aleatórias discretas (ou em número contável de estados). Estas variáveis são descritas por distribuições de probabilidades (probability mass functions), correspondendo à lista de possíveis valores $\xi^k, k \in K$, com probabilidades:

   $$0 \leq f(\xi^k) = P(\xi = \xi^k), \text{ satisfazendo } \sum_{k \in K} f(\xi^k) = 1.$$

   2. Variáveis aleatórias contínuas. São descritas por funções densidade de probabilidade $f(\xi)$. A probabilidade de ξ percenter a um intervalo $[a, b]$ é dada por:

   $$P(a \leq \boldsymbol{\xi} \leq b) = \int_a^b f(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_a^b dF(\boldsymbol{\xi})$$

   No caso contínuo, $P(\boldsymbol{\xi} = a) = 0$, diferentemente do caso discreto.

1. A expectativa de uma variável aleatória é calculada como $\mu = \sum_{k \in K} \xi^k f(\xi^k)$, no caso discreto, ou $\mu = \int_{-\infty}^{\infty} \boldsymbol{\xi} dF(\boldsymbol{\xi})$, no caso contínuo.

2. A variância σ^2 de uma variável aleatória é definida como $\mathbb{E}[(\boldsymbol{\xi} - \mu)^2]$.

3. A expectativa de $\boldsymbol{\xi}^r$ é denominada de $r-$ésimo momento de $\boldsymbol{\xi}$, sendo designada por $\bar{\xi}^r = E(\boldsymbol{\xi}^r)$.

4. Momentos sobre a média $\mathbb{E}[(\boldsymbol{\xi} - \mu)^r]$.

5. A média é o primeiro momento, a variância é o segundo momento sobre a média.

Discreta $U[1, n]$

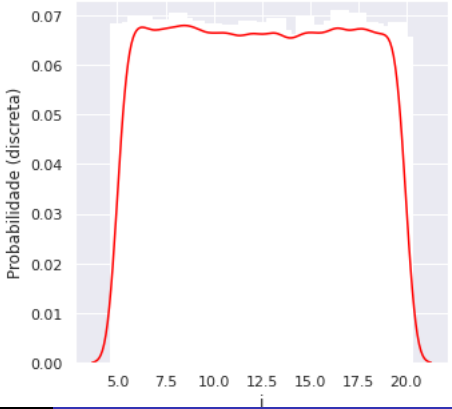
$P(\boldsymbol{\xi} = i) = \frac{1}{n}$ para qualquer $i = 1, \ldots, n, n \geq 1$.
$\mathbb{E}[\boldsymbol{\xi}] = \frac{n+1}{2}, \sigma^2 = \frac{n^2-1}{12}$.

Contínua $U[0, a]$

$f(\xi) = \frac{1}{a}$, $a > 0, 0 \leq \xi \leq a$. $\mathbb{E}[\boldsymbol{\xi}] = \frac{a}{2}, \sigma^2 = \frac{a^2}{12}$.

Geração de amostra de dados segundo distribuição uniforme

```
from scipy.stats import uniform
n = 100000
a = 5
delta = 15
data_uniform = uniform.rvs(size=n, loc = a, scale=delta)
ax = sns.distplot(data_uniform,bins=50,kde=True,color='red',
                  hist_kws={"linewidth": 10,'alpha':1})
ax.set(xlabel='i', ylabel='Probabilidade (discreta)')
print("%5.2f" % data_uniform.mean(),"%5.2f" % data_uniform.var())
```



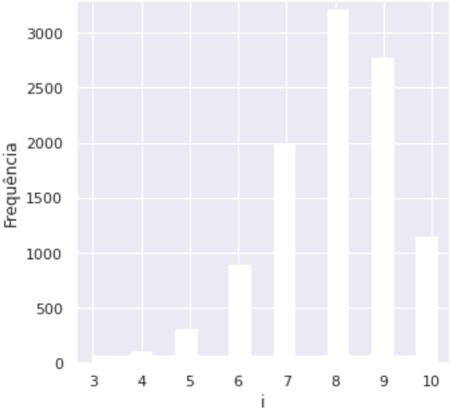
Binomial: $Bi(n, p)$

Dois resultados possíveis: sucesso ou fracasso, p denota a probabilidade de sucesso. $P(i, n)$ fornece a probabilidade de i sucessos em n tentativas. $\mathbb{E}[\xi] = np$ e $\sigma^2 = np(1-p)$.

$$P(\xi = i, n) = \binom{n}{i} p^i (1-p)^{n-i}$$

A distribuição binomial é uma extensão da distribuição de Bernoulli. Nesta, temos dois estados possíveis: fracasso e sucesso, probabilidade de sucesso é p, probabilidade de fracasso é $1-p$. Ou seja, a distribuição binomial é a distribuição de probabilidades para o número de sucessos em uma sequência independente de tentativas que seguem uma distribuição de Bernoulli.

```
from scipy.stats import binom
data_binom = binom.rvs(n=10,p=0.8,size=10000)
ax = sns.distplot(data_binom,kde=False,color='red',\
                  hist_kws={"linewidth": 10,'alpha':1})
ax.set(xlabel='i', ylabel='Frequência')
print("%5.2f" % data_binom.mean(),"%5.2f" % data_binom.var())
8.00  1.62
```



Poisson

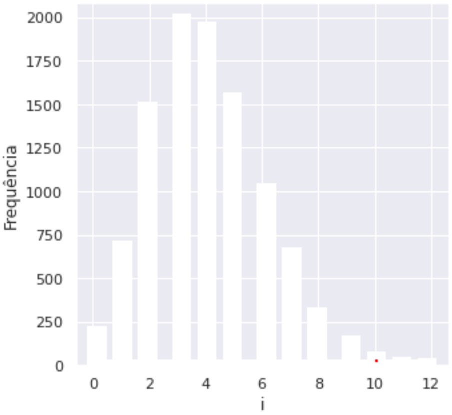
Bastante empregada para modelar o número de vezes que um evento ocorre em um determinado período de tempo. Por exemplo, número de vezes que um serviço de emergência é chamado em uma hora de atendimento. A distribuição é caracterizada pela taxa λ através dos quais o evento ocorre. $P(i)$ fornece a probabilidade de que o evento ocorra i vezes no período.

$$P(\boldsymbol{\xi} = i) = e^{-\lambda}\frac{\lambda^i}{i!}$$

$\mathbb{E}[\boldsymbol{\xi}] = \lambda, \sigma^2 = \lambda.$

Geração de amostra de dados segundo distribuição de Poisson

```
from scipy.stats import poisson
data_poisson = poisson.rvs(mu=4, size=10000)
ax = sns.distplot(data_poisson,bins=40,kde=False,color='red',
                  hist_kws={"linewidth": 8,'alpha':1})
ax.set(xlabel='i', ylabel='Frequência')
print("%5.2f" % data_poisson.mean(),"%5.2f" %  data_poisson.var())
4.01  3.97
```



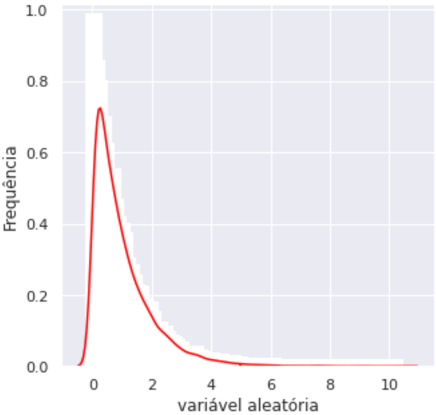
Exponencial

$$f(\xi) = \lambda e^{-\lambda\xi}, \lambda > 0, 0 \leq \xi$$

$\mathbb{E} = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$.

Geração de amostra de dados segundo distribuição exponencial

```
from scipy.stats import expon
data_expon = expon.rvs(scale=1,loc=0,size=10000)
ax = sns.distplot(data_expon,kde=True,bins=100,color='red',
                  hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='variável aleatória', ylabel='Frequência')
print("%5.2f" % data_expon.mean(),"%5.2f" % data_expon.var())
1.00  1.02
```



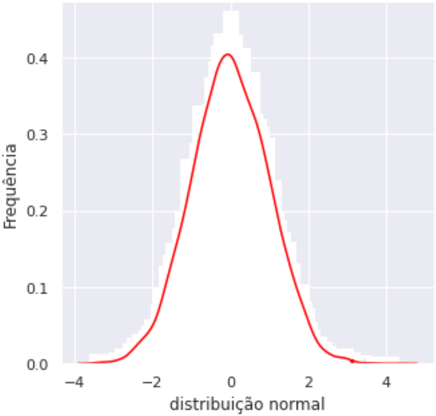
Normal

$$f(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}}, \sigma > 0$$

Parâmetros da distribuição: expectativa μ e variância σ^2.

Geração de amostra de dados segundo distribuição normal

```
from scipy.stats import norm
# N(0,1)
data_normal = norm.rvs(size=10000,loc=0,scale=1)
ax = sns.distplot(data_normal,bins=100,
                  kde=True,color='red',
                  hist_kws={"linewidth": 10,'alpha':1})
ax.set(xlabel='distribuição normal', ylabel='Frequência')
-0.00  1.01
```



Teorema

Se executarmos um grande número de tentativas de um processo aleatório, então a distribuição de probabilidades para a soma (ou para a média) do resultado é aproximadamente uma distribuição gaussiana. Quanto maior o número de tentativas, melhor a aproximação com a distribuição de Gauss.

Exemplo

Evento aleatório: jogar uma moeda não adulterada n vezes, retornando o número de caras divido por n. O evento base em si segue uma distribuição binomial com $p = \frac{1}{2}$. Porém, se executarmos este experimento por muitas vezes, a distribuição dos resultados retornados no experimento será aproximadamente uma gaussiana com média $\frac{1}{2}$.

É um resultado fundamental para os métodos baseados em Simulação de Monte-Carlo.

- A principal classe de problemas de otimização a ser investigada no curso é de Problemas de Programação Estocástica Lineares com Estágios, com recurso.
- Nesta grande classe incluimos as variantes onde as variáveis de decisão de primeiro estágio e de estágios subsequentes podem assumir valores inteiros apenas.
- O termo recurso significa que o modelo de otimização prevê ações a serem tomadas após a incerteza se revelar, uma vez que alguns dados que definem o problema são representados por meio de variáveis aleatórias $\boldsymbol{\xi} = \boldsymbol{\xi}(\omega)$.

As decisões são dividas em dois grupos, relativos à sua ordem cronológica:

- 1o. estágio (*here-and-now*). Precisam ser tomadas antes do experimento ou da revelação dos dados incertos. Normalmente, são indicadas pelo vetor x. Uma vez escolhida a decisão de primeiro estágio, x, não pode ser alterada.
- 2o. estágio (*wait-and-see*). São tomadas após as variáveis aleatórias serem definidas, considerando também as decisões x do estágio anterior. São normalmente indicadas pelo vetor y.

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

$$
\begin{aligned}
\min \quad z = & c^T x + \mathbb{E}_{\xi}\left[q(\omega)^T y(\omega)\right] \\
s.t. \quad & Ax = b \\
& T(\omega)x + Wy(\omega) = h(\omega) \\
& x \geq 0, y(\omega) \geq 0
\end{aligned}
$$

1. Dados determinísticos relativos às decisões de 1o. estágio: c, b, A de dimensões $n_1, m_1, m_1 \times n_1$, respectivamente.
2. No segundo estágio, um conjunto de eventos aleatórios $\omega \in \Omega$ se realizam. Para uma dada realização ω, os dados do problema de segundo estágio são: $q(\omega), h(\omega)$ e $T(\omega)$ se tornam conhecidos, com dimensões (fixas, independentes de ω) $n_2, m_2 e m_2 \times n_1$, respectivamente.
3. $\xi(\omega)^T = (q(\omega)^T, h(\omega)^T, T_1(\omega), \ldots, T_{m_2}(\omega))$ possui potencialmente até $N = n_2 + m_2 + m_2 n_1$ entradas. $T_i(\omega)$ denota a i-ésima linha de $T(\omega)$;

- $\Xi \subset \mathbb{R}^N$ denota o suporte de $\boldsymbol{\xi}$, isto é, o menor subconjunto fechado do $\mathbb{R}^N$ que satisfaz $\mathrm{P}(\Xi) = 1$.
- A dependência de $y(\omega, x)$ em função de ω difere da dependência de $q(\omega), h(\omega), T(\omega)$ em relação a ω. No caso de y, indica que uma vez que uma realização acontece, a decisão y pode ser tomada. Também indica que a decisão y seria possivelmente diferente diante de realizações distintas de ω. No caso de $q(\omega), h(\omega), T(\omega)$, indica que as grandezas não são conhecidas até a realização do experimento.

Two-stage SP with recourse

$$\begin{aligned} \min \quad z = & c^T x + \mathbb{E}_\xi \left[q(\omega)^T y(\omega) \right] \\ s.t. \quad & Ax = b \\ & T(\omega)x + Wy(\omega) = h(\omega) \\ & x \geq 0, y(\omega) \geq 0 \end{aligned}$$

PDE

$$\begin{aligned} \min \quad z = & c^T x + \mathcal{Q}(x) \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- Para uma dada realização ω, definimos:

$$Q(x, \xi(\omega)) = \min_y \left\{ q(\omega)^T y : Wy = h(\omega) - T(\omega)x, y \geq 0 \right\}$$

  como o valor da parcela da fo relativa ao segundo estágio.
- Definimos $\mathcal{Q}(x) = \mathbb{E}_\xi Q(x, \xi)$ como a *second stage value function* e o Problema Determinístico Equivalente (PDE).
- Nos modelos acima, podemos impor a integralidade das variáveis, restringindo, por exemplo: $x \in X = \mathbb{Z}^{n_1}$, $y(\omega) \in Y = \mathbb{Z}^{n_2}$.

Problema de Localização de Facilidades - Determinístico

- Versão não capacitada.
- $I = \{1, \ldots, m\}$: conjunto de clientes. Demandas $d_i : i \in I$ de uma dada mercadoria. r_i denota a receita obtida ao se atender 100% da demanda de i.
- $J = \{1, \ldots, n\}$: conjunto de potenciais localidades onde instalar um armazém, a partir de onde a mercadoria será distribuída. $c_j : j \in J$ é o custo fixo de abrir a localidade j. $v_j : j \in J$ representa o custo variável de operar a localidade j e t_{ij} é o custo de atender 100% da demanda de i pela localidade j.
- $q_{ij} := (r_i - v_j - t_{ij})d_i$ é a margem de contribuição de atender a demanda de i por j.

$$
\begin{aligned}
\max_{x,y} z(x,y) = -\sum_{j\in J} c_j x_j &+ \sum_{i\in I}\sum_{j\in J} q_{ij} y_{ij} && \\
s.t. \sum_{j\in J} y_{ij} &\leq 1 && i \in I \\
y_{ij} &\leq x_j && i \in I, j \in J \\
y_{ij} &\geq 0 && i \in I, j \in J \\
x_j &\in \{0,1\} && j \in J
\end{aligned}
$$

- $x_j \in \{0,1\}$ indica se a localidade $j \in J$ foi aberta.
- $y_{ij} \in [0,1]$ indica a fração de $d_i : i \in I$ atendida por $j \in J$.

Variante estocástica #1

- A distribuição será fixa no primeiro estágio, isto é, agora y_{ij} indica a quantidade transportada de j para i, independentemente da demanda que não é conhecida quando esta decisão é tomada.
- Os elementos de custos, v_j, t_{ij}, r_i podem ou não ser estocásticos. Independentemente disso, y_{ij} é uma decisão de primeiro estágio.
- Variáveis de decisão de segundo estágio: $w_i^+(\omega), w_i^-(\omega)$, respectivamente indicando a falta e o excesso, no atendimento da demanda de $i \in I$ quando $d_i(\omega)$ se revelar. Quando houver falta

  no atendimento da demanda, paga-se q_i^+ e quando houver excesso, paga-se q_i^-, proporcionais à quantidade.
- Para fins de acoplamento, usaremos um big-M, representando uma capacidade máxima fictícia para uma localidade qualquer.

$$\max -\sum_{j\in J} c_j x_j + \sum_{i\in I}\sum_{j\in J} \mathbb{E}_{\xi}(-v_j - t_{ij})y_{ij} +$$

$$\mathbb{E}_{\xi}\left[-\sum_{i\in I} q_i^+ w_i^+(\omega) - \sum_{i\in I} q_i^- w_i^-(\omega) - \sum_{i\in I} r_i d_i(\omega)\right]$$

$$\sum_{i\in I} y_{ij} \leq M x_j \qquad j \in J$$

$$w_i^+(\omega) - w_i^-(\omega) = d_i(\omega) - \sum_{j\in J} y_{ij} \qquad i \in I$$

$$y_{ij} \geq 0 \qquad i \in I, j \in J$$

$$x_j \in \{0,1\} \qquad j \in J$$

$$w_i^+(\omega), w_i^-(\omega) \geq 0 \qquad i \in I$$

Este modelo força a demanda a ser atendida...

Variante estocástica #2

- A distribuição será definida no segundo estágio, se ajustando à realização da variável aleatória.
- Porém, precisamos definir as capacidades instaladas nas facilidades abertas, no primeiro estágio.
- A variável $y_{ij}(\omega)$ indica o percentual de $d_i(\omega)$ atendido por j quando a demanda se revelar.
- Vamos introduzir uma variável w_j, de primeiro estágio, que indicará, no caso da facilidade $j \in J$ ser aberta, a capacidade a ser instalada em j. O custo por unidade de capacidade instalada é $g_j : j \in J$, determinístico.
- $q_{ij}(\omega) := (r_i - v_j - t_{ij})d_i(\omega)$ denota o custo de atender a demanda de i em uma realização $\omega \in \Omega$.

$$
\begin{aligned}
\max \; & -\sum_{j \in J} c_j x_j - \sum_{j \in J} g_j w_j + \mathbb{E}_{\boldsymbol{\xi}} \left[\sum_{i \in I} \sum_{j \in J} q_{ij}(\omega) y_{ij}(\omega) \right] \\
& \sum_{j \in J} y_{ij}(\omega) \leq 1 && i \in I \\
& \sum_{i \in I} d_i(\omega) y_{ij}(\omega) \leq w_j && j \in J \\
& y_{ij} \geq 0 && i \in I, j \in J \\
& x_j \in \{0, 1\} && j \in J \\
& w_j \geq 0 && j \in J
\end{aligned}
$$

Neste modelo, a demanda pode não ser atendida completamente.
Poderíamos impor uma penalidade, incorrida na parcela não atendida.

Variante estocástica #2 do UFLP e as seguintes observações.

1. Planejamento para os próximos 36 meses.
2. A decisão de primeiro estágio, instalação das facilidades, leva cerca de 6 meses para ser concluída. As decisões de segundo estágio, como distribuir de j para i, ocorrem ao longo dos 30 meses seguintes.
3. Embora possa se pensar com mais estágios, o número de estágios necessários, dois no caso considerado nesta discussão, está ligado a quando a decisão de instalar as facilidades é tomada. Ela é que muda a estrutura possível de distribuição.
4. Assim, faz sentido um modelo de dois estágios, no qual o segundo estágio é representado pelas decisões ao longo de 30 períodos consecutivos.
5. Podemos até ter as demandas sendo reveladas mês a mês, o que não muda a natureza do problema ser de dois estágios. O recurso permite não atender a demanda.

Quando haveria a necessidade ou faria sentido em incorporar mais estágios no caso anterior ? Por exemplo, quando se prevê a possibilidade de, ao longo do horizonte de planejamento, reavaliar as localidades abertas, abrindo algumas novas ou fechando outras que foram abertas no passado.

1. Vamos considerar que após 12 meses (isto é, após 6 meses de operação da distribuição), a empresa possa abrir novas localizações.
2. Sob as condições acima, o primeiro estágio compreenderia a decisão do que abrir hoje, compreendendo o mês 1 ao 6, o segundo estágio, a operação do mês 7 ao 18, bem como a decisão de quais novas localidades abrir no mês 12 e, finalmente, o terceiro estágio compreenderia a operação do mês 19 ao 36.
3. ξ_2 e ξ_3 representam vetores de variáveis aleatórias associadas ao segundo e terceiro estágios, neste exemplo.

Versão 3 estágios do problema.

1. x_j^1 e $x_j^2(\omega_2)$ representam variáveis binárias, relativas à abertura ou não da localidade j no primeiro e no segundo estágio de decisão, respectivamente.

2. $y^2(\omega_2), y^3(\omega_3)$ representam os vetores de distribuição da mercadoria, durante o segundo e terceiro estágios, respectivamente.

3. $\omega_2 \in \Omega_2, \omega_3 \in \Omega_3$ representam possíveis estados do mundo, no estágio 2 e 3.

4. O modelo apresentado na sequência envolve restrições puras de cada estágio e acoplamento entre as decisões em estágios distintos. Também envolve expectativas condicionais.

$$\max -\sum_{j\in J} c_j x_j^1 + \mathbb{E}_{\boldsymbol{\xi}_2} \max \left\{ \sum_{i\in I}\sum_{j\in J} q_{ij}^2(\omega_2) y_{ij}^2(\omega_2) - \sum_{j\in J} c_j^2(\omega_2) x_j^2(\omega_2) \right.$$

$$\left. + \mathbb{E}_{\boldsymbol{\xi}_3|\boldsymbol{\xi}_2} \max \left\{ \sum_{i\in I}\sum_{j\in J} q_{ij}^3(\omega_3) y_{ij}^3(\omega_3) \right\} \right\}$$

$$\sum_{j\in J} y_{ij}^2(\omega_2) \leq 1 \qquad i \in I$$

$$\sum_{i\in I} d_i(\omega_2) y_{ij}^2(\omega_2) \leq M x_j^1 \qquad j \in J$$

$$\sum_{j\in J} y_{ij}^3(\omega_3) \leq 1 \qquad i \in I$$

$$\sum_{i\in I} d_i(\omega_3) y_{ij}^3(\omega_3) \leq M(x_j^1 + x_j^2(\omega_2)) \qquad j \in J$$

$$x_j^1 + x_j^2(\omega_2) \leq 1 \qquad j \in J$$

$$y_{ij}^2(\omega_2), y_{ij}^3(\omega_3) \geq 0 \qquad i \in I, j \in J$$

$$x_j^1, x_j^2(\omega_2) \in \{0,1\} \qquad j \in J$$

A frequência dos eventos incertos é relevante. Dois grandes grupos:

1. Eventos incertos ocorrem com grande frequência, de forma recorrente, no curto prazo, e decisões precisam serem tomadas na mesma frequência. Por exemplo, o UFLP, em que devemos atender a demanda incerta durante todo o horizonte de planejamento. Há vários cenários de demanda, e ao longo dos períodos, muitos deles serão realizados.

2. Poucos cenários efetivamente serão realizados, uma vez que poucas decisões serão tomadas. Exemplo: precificação de ingressos para um grande evento esporádico como Copa do Mundo. Nestes casos deseja-se maximizar o lucro, mas também se proteger contra cenários potencialmente desastrosos. Nestes casos, faz muito sentido incorporar elementos de Programação Robusta ou Chance Constrained Models, por meio de restrições no Programa Estocástico.

- Modelos em que se maximiza uma função de utilidade, para um dado perfil de risco estabelecido. Por exemplo, modelo de Markowitz para construção da fronteira eficiente de portfolios.
- Risco pode ser representado pela probabilidade de se obter um retorno abaixo de um dado patamar estabelecido, *downside risk*. Incorpora-se uma restrição no modelo, estabelecendo que esta probabilidade deva ser inferior ao valor desejado.
- Abordagem puramente robusta. Por exemplo, são definidos intervalos para as variáveis incertas e busca-se uma solução ótima que permaneça viável ainda que um número de eventos incertos ocorra da maneira mais desfavorável possível.

Two-stage SPWR

$$\begin{aligned} \min \quad z = & c^T x + \mathbb{E}_{\boldsymbol{\xi}} \left[q(\omega)^T y(\omega) \right] \\ s.t. \quad & Ax = b \\ & T(\omega)x + Wy(\omega) = h(\omega) \\ & x \geq 0, y(\omega) \geq 0 \end{aligned}$$

$$Q(x, \xi(\omega)) = \min_y \left\{ q(\omega)^T y : Wy = h(\omega) - T(\omega)x, y \geq 0 \right\}$$

$$\mathcal{Q}(x) = \mathbb{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$$

Precisamos dispor de representação adequada e computacionalmente tratável para a função recurso $\mathcal{Q}(x)$:

- ou dispomos da forma analítica de $\mathcal{Q}(x)$ (raro).
- ou conseguimos computar $\mathcal{Q}(x)$, para uma dada decisão x de 1o. estágio fixa.

Problema Determinístico Equivalente ao SPWR

$$\begin{aligned} \min \quad z = & c^T x + \mathcal{Q}(x) \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

$$Q(x, \xi(\omega)) = \min_y \left\{ q(\omega)^T y : Wy = h(\omega) - T(\omega)x, y \geq 0 \right\}$$

$$\mathcal{Q}(x) = \mathbb{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$$

Hipóteses usuais:

- A matriz de recurso W é fixa.
- Ξ é um conjunto finito, ou equivalentemente $\boldsymbol{\xi}$ é uma variável aleatória finita, seja porque de fato dispomos de um conjunto finito de cenários ou porque utilizamos técnicas de amostragem para gerar cenários a partir de distribuições contínuas.

- Conjunto de viabilidade de 1o. estágio:

$$K_1 = \{x \in \mathbb{R}^{n_1} : Ax = b, x \geq 0\}$$

- Para um determinado ξ, definimos

$$K_2(\boldsymbol{\xi}) = \{x \in \mathbb{R}^{n_1} : \exists y \geq 0, Wy = h(\omega) - T(\omega)x\}$$

- $K_2 = \bigcup_{\xi \in \Xi} K_2(\xi)$
- Reformulação do SPWR: $\min c^T x + \mathcal{Q} : x \in K_1 \cap K_2$.
- $\text{pos}(W) = \{t | Wy = t, y \geq 0\}$ indica o conjunto dos vetores de termos independentes t que podem ser escritos como combinação linear não negativa das colunas de W.

Vamos supor que ξ_1 e ξ_2 só possam assumir valores nos conjuntos $\{2,3,4\}$ e $\{1,4,7\}$. Considere o seguinte problema de 2o. estágio.

$$
\begin{aligned}
\min\ & 2y_1 + y_2 \\
& y_1 + 2y_2 \geq \xi_1 - x_1 \\
& y_1 + y_2 \geq \xi_2 - x_1 - x_2 \\
& 0 \leq y_1, y_2 \leq 1
\end{aligned}
$$

- Usando os limites superiores para os valores admissíveis de y, temos:

$$K_2(\xi) = \{x \in \mathbb{R}^2 : x_1 \geq \xi_1 - 3, x_1 + x_2 \geq \xi_2 - 2\}.$$

- Considerando a distribuição de ξ que é discreta,

$$K_2 = \{x \in \mathbb{R}^2 : x_1 \geq 1, x_1 + x_2 \geq 5\}$$

Teorema

1. Para um dado ξ, $K_2(\xi)$ é um conjunto poliedral.
2. Se $\boldsymbol{\xi}$ é uma variável aleatória discreta, K_2 é um conjunto poliedral, e portanto, convexo.

Teorema

Para um dado ξ fixo, a função $Q(x, \xi)$ é:

1. linear por partes e convexa em (h, T).
2. linear por partes côncava em q.
3. linear por partes e convexa em x para todo $x \in K_2$.

Quando $\boldsymbol{\xi}$ é uma variável aleatória discreta, $\mathcal{Q}$ é linear por partes e convexa em K_2. Como K_1 é poliedral, o problema é convexo.

Exemplo

Considere o seguinte problema de 2o. estágio e assuma $0 \leq x_1, x_2 \leq 1$.

$$
\begin{aligned}
\min\ & 2y_1 + y_2 \\
& y_1 + y_2 \geq 1 - x_1 \\
& y_1 \geq \boldsymbol{\xi} - x_1 - x_2 \\
& 0 \leq y_1, y_2
\end{aligned}
$$

- A decisão ótima de 2o. estágio é:
  - Se $\xi \leq x_1 + x_2 \rightarrow y_1 = 0,\ \ y_2 = 1 - x_1$
  - Se $\xi > x_1 + x_2 \rightarrow\ y_1 = \xi - x_1 - x_2, \quad y_2 = \max\{0, 1 - \xi + x_2\}$

- A função recurso é linear por partes em x (resta provar convexidade).

$$
Q(x, \xi) = \begin{cases} 1 - x_1 & 0 \leq \xi \leq x_1 + x_2 \\ \xi + 1 - 2x_1 - x_2 & x_1 + x_2 \leq \xi \leq 1 + x_2 \\ 2(\xi - x_1 - x_2) & 1 + x_2 \leq \xi \end{cases}
$$

- Recurso relativamente completo: quando toda solução viável no primeiro estágio possui uma realização viável no segundo estágio, isto é, $K_1 \subseteq K_2$. Não é fácil de ser caracterizado.
- Recurso completo: caso particular do recurso relativamente completo, mais fácil de ser caracterizado pois depende da estrutura de W. Quando para qualquer $t \in \mathbb{R}^{m_2}$ existir $y \geq 0, Wy = t$, temos o recurso completo. Em outras palavras: $\text{pos}(W) = \mathbb{R}^{m_2}$.
- Recurso simples. Neste caso $W = [I, -I]$, y é particionada em y^+, y^- com custos $\text{q} = \text{q}^+ - \text{q}^-$, onde $\text{q}_i^+ - \text{q}_i^- \geq 0$ para todo i, com probabilidade 1.

Integer Two-stage SPWR

$$
\begin{aligned}
\min\ \ z = & c^T x + \mathbb{E}_{\boldsymbol{\xi}}\left[q(\omega)^T y(\omega)\right] \\
s.t.\ \ & Ax = b \\
& T(\omega)x + Wy(\omega) = h(\omega) \\
& x \in X, y(\omega) \in Y
\end{aligned}
$$

$$Q(x, \xi(\omega)) = \min_{y \in Y} \left\{ q(\omega)^T y : Wy = h(\omega) - T(\omega)x \right\}$$

$$\mathcal{Q}(x) = \mathbb{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$$

Os conjuntos X e Y podem ser, por exemplo, $\mathbb{Z}^{n_1}$ e $\mathbb{Z}^{n_2}$, respectivamente. Mesmas definições anteriores de $c, b, \boldsymbol{\xi}, A, W, T$.

1. Se a integralidade é imposta penas em x, as propriedades para $\mathcal{Q}(x)$ e K_2 para o caso SPWR de dois estágios linear são preservadas.
2. Assumimos portanto que a integralidade seja imposta nas variáveis y de segundo estágio.

Teorema

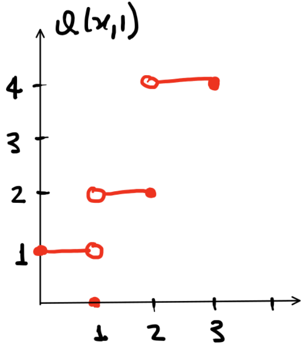
A função valor esperado do recurso $\mathcal{Q}(x)$ de um programa inteiro é, no caso geral, semi-contínua inferior, não convexa e descontínua.

$Q(x,\xi) = \min\{2y_1 + y_2 : y_1 \geq x - \xi, y_2 \geq \xi - x, y \geq 0, y \in \mathbb{Z}\}$.
Vamos também assumir que ξ possa valer 1 ou 2 com probabilidade $\frac{1}{2}$ e que $x \geq 0$.

Caso $\xi = 1$

$$y_1 \geq x - 1$$
$$y_2 \geq 1 - x$$

- Se $x \leq 1 \rightarrow y_1 = 0,$ e $y_2 = \lceil 1 - x \rceil$
- Se $x > 1 \rightarrow y_1 = \lceil x - 1 \rceil$ e $y_2 = 0$.

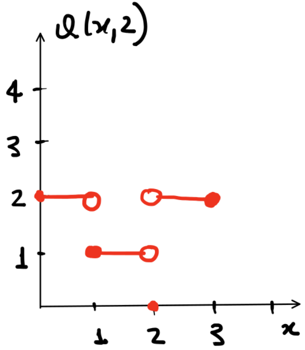


Exemplo (continuação)

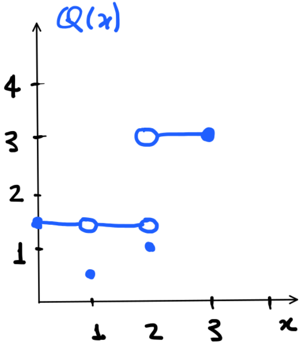
Caso $\xi = 2$

$$y_1 \geq x - 2$$
$$y_2 \geq 2 - x$$

- Se $x \leq 2 \rightarrow y_1 = 0$, e $y_2 = \lceil 2 - x \rceil$
- Se $x > 2 \rightarrow y_1 = \lceil x - 2 \rceil$ e $y_2 = 0$.



$$\mathcal{Q}(x) = \frac{1}{2}(Q(x,1) + Q(x,2))$$



Veja que $\mathcal{Q}(x)$ é:

- Descontínua nos x inteiros.
- Não convexa.
  Considere $\lambda = \frac{1}{2}, x^1 = 1, x^2 = 2$, e $x(\lambda) = \lambda x^1 + (1-\lambda)x^2 = \frac{1}{2}1 + \frac{1}{2}2 = 1.5$. Temos que $\mathcal{Q}(1.5) = 1.5 >$ $0.5\mathcal{Q}(1) + 0.5\mathcal{Q}(2) = 0.75$
- Não é fácil encontrar $\arg\min \; c^T x + \mathcal{Q}(x)$.

Continuidade de $\mathcal{Q}$

Proposição

A função recurso $\mathcal{Q}(x)$ de um programa inteiro com variáveis de segundo estágio inteiras é contínua quando a variável aleatória é absolutamente contínua.

Exemplo anterior modificado, distribuição contínua

$$Q(x,\xi) = \min\{2y_1 + y_2 : y_1 \geq x - \xi, y_2 \geq \xi - x, y \geq 0, y \in \mathbb{Z}\}.$$

Assumimos: $x \geq 0$, e a distribuição acumulada para $\boldsymbol{\xi}$ é $F(t) = \mathrm{P}(\boldsymbol{\xi} \leq t) = 2 - \frac{2}{t}$ para $t \in [1,2]$.

- $1 < x < 2$ e $1 \leq \xi < x \rightarrow y_1 = 1, y_2 = 0$
- $1 < x < 2$ e $x < \xi \leq 2 \rightarrow y_1 = 0, y_2 = 1$

$$\begin{aligned}\mathcal{Q}(x) &= \int_1^x 2dF(t) + \int_x^2 1dF(t) = 2F(x) + 1 - F(x) \\ &= F(x) + 1 \\ &= 3 - \frac{2}{x}\end{aligned}$$

Assim como nos casos anteriores, assumimos que $y \in Y$ é inteiro.

- As propriedades do conjunto de viabilidade de 2o. estágio não são muito melhores que as da função recurso $\mathcal{Q}(x)$.
- Para um valor fixo de $\boldsymbol{\xi}$, definimos:

$$K_2(\xi(\omega)) = \{x \in \mathbb{R}^{n_1} : \exists y \in Y, Wy = h(\omega) - T(\omega)x\}$$

onde $\xi(\omega)$ compreende as componentes estocásticas de $h(\omega)$ e $T(\omega)$.

Proposição

Geralmente, o conjunto de viabilidade de segundo eságio $K_2(\xi)$ é um conjunto não convexo.

Consideramos que $\boldsymbol{\xi}$ assuma valores $\{1, 2\}$ equiprováveis.

$$-y_1 + y_2 \leq \boldsymbol{\xi} - x_1 \tag{1}$$

$$y_1 + y_2 \leq 2 - x_2 \tag{2}$$

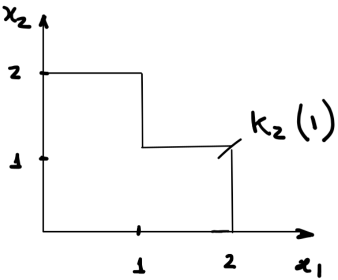
$$y \in \mathbb{Z}^2_+ \tag{3}$$

$K_2(1)$

- (2) $\rightarrow$ $x_2 \leq 2$ é necessária para viabilidade no 2o. estágio.
- Para $x_2 \in (1, 2]$, o único ponto inteiro satisfazendo (2) é $y_1 = y_2 = 0$.
- $y_1 = y_2 = 0$ também satisfaz (1) se $x_1 \leq \xi = 1$.
- Para $x_2 \in (0, 1]$, valores de y que satisfazem (2) são $(0, 0), (0, 1), (1, 0)$, sendo que $(1, 0)$ produz o menor lado esquerdo, mais provável de produzir pontos em $K_2(1)$.
- Para $y = (1, 0)$, (1) requer $x_1 \leq 2$.

Veja que a interseção de $K_2(1)$ com o quadrante positivo de $\mathbb{R}^2$ pode ser formulada da seguinte forma:

$$K_2(1) = \{x \in \mathbb{R}^2_+ : \min\{x_1 - 1, x_2 - 1\} \leq 0, x_1 \in [0, 2], x_2 \in [0, 2]\}.$$



Variáveis de decisão contínuas

- H estágios
- x^1 : variável de decisão de 1o. estágio.
- $x^t(\omega^t)$: $t = 2, \ldots, H$ variável de decisão do estágio t, dependendo da realização ω^t.

- $\xi^t(\omega)^T = (c^t(\omega)^T, h^t(\omega)^T, T_1^{t-1}(\omega), \ldots, T_{m_t}^{t-1}(\omega))$ é um vetor N_t dimensional. $\xi^t(\omega)$ é independente de $\xi^{t-1}(\omega), \cdots, \xi^1(\omega)$.
- $T^{t-1}(\omega^t)$ é uma matriz $m_t \times n_{t-1}$ dimensional.
- W^t é uma matriz $m_t \times n_t$ dimensional, fixa, independented de $\boldsymbol{\xi}$.
- As decisões x dependem da história até o instante t, que indicamos por ω^t.
- Ξ^t é o suporte de $\boldsymbol{\xi}^t$.
- No modelo seguinte, a notação transposta foi intencionalmente suprimida.

$$\min c^1 x^1 + \mathrm{E}_{\xi^2}\left[\min c^2(\omega^2)x^2(\omega^2) + \cdots + \mathbb{E}_{\xi^H}\left[\min c^H(\omega^H)x^H(\omega^H)\right]\cdots\right]$$

$$\begin{aligned}
h^1 &= W^1 x^1 \\
h^2(\omega^2) &= T^1(\omega^2)x^1 + W^2 x^2(\omega^2) \\
\ldots &\vdots \\
h^H(\omega^H) &= T^{H-1}(\omega^H)x^{H-1} + W^H x^H(\omega^H)
\end{aligned}$$

$$\begin{aligned}
x_1 &\geq 0 \\
x^t(\omega^t) &\geq 0, \quad t = 2, \ldots, H.
\end{aligned}$$

Estágios vs estados da PD

- Estágios: estágios do programa estocástico $t = 1, \ldots, H$.
- Estados: $x^t(\omega^t) : t = 1, \ldots, H$

Formulação recursiva

Como de costume em PD, vamos formular o problema de forma recursiva, detalhando o caso base, quanto $t = H$, e os demais, quando $t \in \{1, \ldots, H-1\}$.

Caso base da recursão, $t = H$: observe que assume-se dispor da decisão do estágio anterior, x^{H-1}, e da realização da incerteza em $t = H$.

$$
\begin{aligned}
Q^H(x^{H-1}, \xi^H(\omega)) =& \min c^H(\omega) x^H(\omega) \\
W^H x^H(\omega) =& h^H(\omega) - T^{H-1}(\omega) x^{H-1} \\
x^H(\omega) \geq& 0
\end{aligned}
$$

Observe que se dispusermos de x^{H-1}, basta resolver o PPL acima. Ou seja, podemos ter uma política ótima para qualquer x^{H-1} de entrada.

Para $t = 1, \ldots, H-1$, desejamos conhecer o estado $x^t(\omega)$.

$$\begin{aligned} Q^t(x^{t-1}, \xi^t(\omega)) = &\min c^t(\omega) x^t(\omega) + \mathcal{Q}^{t+1}(x^t) \\ W^t x^t(\omega) = &h^t(\omega) - T^{t-1}(\omega) x^{t-1} \\ x^t(\omega) \geq &0 \end{aligned} \tag{4}$$

onde a função recurso é definida como:

$$\mathcal{Q}^{t+1}(x^t) = \mathbb{E}_{\boldsymbol{\xi}^{t+1}} \left[Q^{t+1}(x^t, \xi^{t+1}(\omega)) \right]$$

Estrutura ótima explorada na recursão (4) (Princípio de Otimalidade): Não sabemos se x^{t-1} é o estado ótimo do estágio $t-1$. Porém, ao minimizar $c^t(\omega)x^t(\omega) + \mathcal{Q}^{t+1}(x^t)$ (note a natureza das duas parcelas), tomamos a decisão ótima caso a política ótima para o estágio anterior seja de fato x^{t-1}...

Desejamos resolver o programa abaixo, que tem a forma do PDE de um Problema Estocástico Linear em dois estágios.

PDE do problema estocástico linear multi-estágio

$$
\begin{aligned}
\min \quad z &= \min c^1 x^1 + \mathcal{Q}^2(x^1) \\
W^1 x^1(\omega) &= h^1 \\
x^1 &\geq 0
\end{aligned}
$$

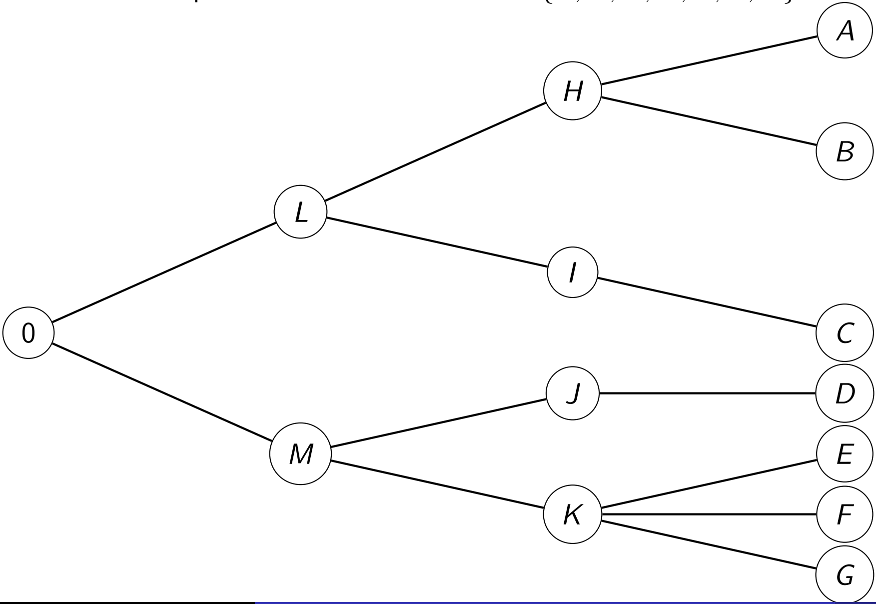
De forma análoga ao que fizemos para o caso de 2 estágios, vamos definir o conjunto de viabilidade do $t-$ésimo estágio como:

$$K^t = \{x^t \in \mathbb{R}^{n_t} : \mathcal{Q}^{t+1}(x^t) < \infty\}$$

Teorema

Os conjuntos K^t e as funções $\mathcal{Q}^{t+1}(x^t)$ são convexas para $t = 1, \ldots, H-1$ e, se se o conjunto suporte Ξ^t é finito para todo $t = 1, \ldots, H$, então K^t e $\mathcal{Q}^{t+1}(x^t)$ são poliedrais.

7 cenários possíveis - folhas da árvore: $\{A, B, C, D, E, F, G\}$



"Primeira parte"da matriz (colunas/linhas do 1o. ao 3o. estágio)

		1o.	2o.		3o.			
Est.	Nó	x^1	$x^2(\omega^L)$	$x^2(\omega^M)$	$x^3(\omega^H)$	$x^3(\omega^I)$	$x^3(\omega^J)$	$x^3(\omega^K)$
1o.	O	W^1						
2o.	L	T^1	W^2					
2o.	M	T^1		W^2				
3o.	H		$T^2(\omega^L)$		W^3			
3o.	I		$T^2(\omega^L)$			W^3		
3o.	J			$T^2(\omega^M)$			W^3	
3o.	K			$T^2(\omega^M)$				W^3

"Segunda parte"da matriz (colunas do 3o. ao 4o. estágio), linhas do 4o. estágio.

		3o.				4o.						
Est.	Nó	$x^3(\omega^H)$	$x^3(\omega^I)$	$x^3(\omega^J)$	$x^3(\omega^K)$	$x^4(\omega^A)$	$x^4(\omega^B)$	$x^4(\omega^C)$	$x^4(\omega^D)$	$x^4(\omega^E)$	$x^4(\omega^F)$	$x^4(\omega^G)$
4o.	A	$T^3(\omega^H)$				W^4						
4o.	B	$T^3(\omega^H)$					W^4					
4o.	C		$T^3(\omega^I)$					W^4				
4o.	D			$T^3(\omega^J)$					W^4			
4o.	E				$T^3(\omega^K)$					W^4		
4o.	F				$T^3(\omega^K)$						W^4	
4o.	G				$T^3(\omega^K)$							W^4

Estrutura da matriz de restrições do programa

- Os problemas de Programação Estocástica Multi-estágios com suporte finito são programas lineares muito grandes. Dependendo do número de cenários, torna-se difícil carregar estes modelos no computador. Assim, algum algoritmo de decomposição precisa ser utilizado para resolver tais problemas.
- A decomposição explora a estrutura bloco diagonal da matriz de restrições.

Impacto da estrutura bloco-diagonal

Assuma que v seja um nó no estágio $e(v)$ da árvore de cenários e que seus sucessores na árvore sejam $c_i(v) : i = 1, \ldots, nc_v$. Obter o estado $x^{e(v)+1}(w^{c_i(v)})$ para $i = 1, \ldots, nc_v$ dado o estado $x^{e(v)}(w^{e(v)})$, pode ocorrer em paralelo, uma vez que a função objetivo também é desacoplada.

Um conjunto de políticas ótimas a serem adotadas, para cada estado do mundo ao longo dos H estágios.

- Programas estocásticos são computacionalmente caros.
- É comum resolver alternativas mais baratas para tentar evitar este custo:
  - Resolver um problema determinístico formulado com a expectativa das variáveis aleatórias.
  - Resolver vários problemas determinísticos, cada um deles formulado com um dos cenários possíveis, combinando as soluções obtidas por meio de alguma regra heurística.

Questão que se coloca - nosso objeto de estudo.

- Estas soluções alternativas são de boa qualidade ou são absolutamente imprecisas e desinformadas ?
- A resposta para estas questões qualitativas é dada por duas grandezas principais: O *valor esperado da informação perfeita* (EVPI) e o *valor da solução estocástica.* (VSS)

A incerteza é representada por um vetor de variáveis aleatórias $\boldsymbol{\xi}$.

Para um dado $\xi \in \Xi$

$$\min z(x, \xi) = c^T x + \min\{q^T y : Wy = h - Tx, y \geq 0\} \tag{5}$$
$$s.t. \;\; Ax = b$$
$$x \geq 0$$

Hipóteses:

- Para todo $\xi \in \Xi$, $K_1 \cap K_2(\xi) \neq \emptyset$.
- Vamos designar por $\bar{x}(\xi)$ a solução ótima de (5) para um $\xi \in \Xi$.

Recordando

$$RP = \min_x \mathbb{E}_{\boldsymbol{\xi}}[z(x, \boldsymbol{\xi})]$$

$$WS = \mathbb{E}_{\boldsymbol{\xi}}[z(\bar{x}(\xi), \xi]$$

RP é o valor ótimo do Programa Estocástico. WS é o valor esperado das soluções *wait-and-see*.

Expected Value of Perfect Information (min)

Mede quanto um tomador de decisão racional se disporia a pagar por informação completa e precisa sobre o futuro.

$$EVPI = RP - WS$$

Recordando

$$\bar{\xi} = \mathbb{E}(\boldsymbol{\xi})$$

$$EV = \min_{x} z(x, \bar{\xi})$$

$$EEV = \mathbb{E}_{\boldsymbol{\xi}}[z(\bar{x}(\bar{\xi}), \boldsymbol{\xi})]$$

$$VSS = EEV - RP$$

Valor da solução estocástica (min)

EEV mede o desempenho solução do problema determinístico formulado com a expectativa da variável aleatória, $\bar{x}(\bar{x})$). Quanto o desempenho é muito bom, *VSS* é pequeno, ou seja, não resolver o problema estocástico e obter RP não traz tanto valor adicional.

Proposição

$$WS \leq RP \leq EEV$$

Prova

Denote por x^* a solução ótima de RP. Para um dado $\xi \in \Xi$ temos: $z(\bar{x}(\xi), \xi) \leq z(x^*, \xi)$. Tome a expectativa em ambos os lados e a desigualdade esquerda segue. Para mostrar a desigualdade à direita, basta lembrar que x^* é a solução ótima de RP enquanto $\bar{x}(\bar{\xi})$ é apenas uma solução viável para RP.

Os exemplos que discutimos na introdução sugerem que EVPI e VSS normalmente sejam distintos. Vamos discutir relações entre estas grandezas.

Proposição

1. Para qualquer programa estocástico, $EVPI \geq 0$, $VSS \geq 0$.
2. Para programas estocásticos com matriz de recurso fixa (W) e função objetivo fixa (q não é estocástico), temos:

$$EVPI \leq EEV - EV$$
$$VSS \leq EEV - EV.$$

Observe que apresentamos o mesmo limites superior, $EEV - EV$ para $EVPI$ e VSS. Naturalmente, se $EEV = EV$, $VSS = EVPI = 0$.

$EEV = EV \to VSS = EVPI = 0$

$$\mathbb{E}_{\boldsymbol{\xi}}[\min z(\bar{x}(\bar{\xi}), \boldsymbol{\xi})] = \min_{x} z(x, \bar{\xi})$$

- Uma condição suficiente (óbvia) para que isto ocorra é que $\bar{x}(\boldsymbol{\xi})$ seja independente de $\boldsymbol{\xi}$, ou seja, a solução ótima é insenível aos cenários.
- Neste caso, obter a solução ótima $x(\xi)$ para um cenário $\xi \in \Xi$ ou para $\xi = \bar{\xi}$ fornece o mesmo resultado e não há necessidade de se resolver o problema RP. É bastante raro isto acontecer.
- Possibilidades de investigação: Relações entre *EVPI* e *VSS* e classes de problemas para os quais se espera que *EVPI* seja pequeno.

Considere o problema

$$
\begin{aligned}
z(x,\xi) = {} & x_1 + 4x_2 + \min \{ y_1 + 10y_2^+ + 10y_2^- \\
& y_1 + (y_2^+ - y_2^-) = \xi + x_1 - 2x_2 \\
& y_1 \leq 2 \\
& y \geq 0 \} \\
s.t.\ & x_1 + x_2 = 1 \\
& x \geq 0
\end{aligned} \tag{6}
$$

e que ξ assuma uma distribuição uniforme em $[1, 3]$ e, assim $\bar{\xi} = 2$. Veja que $y_2 = (y_2^+ - y_2^-)$ denota a (variável artificial) violação da restrição.

Para um dado $x \in K_1$ e $\xi \in U[1,3]$, podemos concluir:

$$y^*(x,\xi) = \begin{cases} y_1 = \xi + x_1 - 2x_2, \;\; y_2 = 0 & \text{se } 0 \le \xi + x_1 - 2x_2 \le 2 \\ y_1 = 2, \;\; y_2^+ = \xi + x_1 - 2x_2 & \text{se } \xi + x_1 - 2x_2 > 2 \\ y_1 = 0, \;\; y_2^- = 2x_2 - \xi - x_1 & \text{se } \xi + x_1 - 2x_2 < 0 \end{cases}$$

$$z(x,\xi) = \begin{cases} 2x_1 + 2x_2 + \xi & \text{se } 0 \le \xi + x_1 - 2x_2 \le 2 \\ -18 + 11x_1 - 16x_2 + 10\xi & \text{se } \xi + x_1 - 2x_2 > 2 \\ -9x_1 + 24x_2 - 10\xi & \text{se } \xi + x_1 - 2x_2 < 0 \end{cases}$$

Observações, lembrando que $\xi \in U[1,3], \bar{\xi} = 2$:

- Usando o fato de que $x \in K_1 \iff x_1 + x_2 = 1, x \geq 0$, temos que $z(x,\xi) = 2 + \xi$ para o primeiro caso ($y_2 = 0$) e $z(x,\xi) \geq 2 + \xi$ para os demais ($y_2 \neq 0$).
- O mínimo ocorre em $z(x,\xi) = 2 + \xi$ (para o caso $y_2 = 0$).
- Assim, qualquer $x \in K_1$ é uma solução ótima de (6), para $-x_1 + 2x_2 \leq \xi \leq 2 - x_1 + 2x_2$ ou equivalentemente $2 - 3x_1 \leq \xi \leq 4 - 3x_1$.
- Observe a multiplicidade de soluções ótimas:
  $(\frac{1}{3}, \frac{2}{3})$ é ótimo para qualquer $\xi \in [1,3]$,
  $(0,1)$ é ótimo para qualquer $\xi \in [2,3]$ e
  $(1,0)$ é ótimo para $\xi = \{1\}$.
- Tomando $(\frac{1}{3}, \frac{2}{3})$ como ótimo para qualquer ξ, temos:
  WS $=$ RP $= 4$ e EVPI $= 0$.

- Ao resolver $z(x, \xi = 2)$, podemos tomar outra solução ótima, por exemplo, $\bar{x}(2) = (0, 1)$ (claro, também com EV $= 4$).

- Com a solução $\bar{x}(2)$, calculamos EEV e VSS:

$$\begin{aligned} EEV &= \mathbb{E}_{\boldsymbol{\xi}<2}(24 - 10\xi) + \mathbb{E}_{\boldsymbol{\xi}\geq 2}(2 + \xi) \\ &= \frac{24 - 10(1.5)}{2} + \frac{2 + 2.5}{2} = 6.75 \\ VSS &= 6.75 - 4 = 2.75 \end{aligned}$$

- A existência de múltiplas soluções ótimas para PLLs torna este tipo de caso comum.

Considere o mesmo problema definido anteriormente

$$
\begin{aligned}
z(x,\xi) = & x_1 + 4x_2 + \min \{ y_1 + 10y_2^+ + 10y_2^- \\
& y_1 + (y_2^+ - y_2^-) = \xi + x_1 - 2x_2 \\
& y_1 \leq 2 \\
& y \geq 0 \} \\
s.t. \; & x_1 + x_2 = 1 \\
& x \geq 0
\end{aligned}
$$

onde ξ assume uma distribuição discreta em $\{0, \frac{3}{2}, 3\}$, equiprováveis, e assim $\bar{\xi} = 1\frac{1}{2}$.

$$z(x,\xi) = \begin{cases} 2x_1 + 2x_2 + \xi & \text{se } 0 \leq \xi + x_1 - 2x_2 \leq 2 \\ -18 + 11x_1 - 16x_2 + 10\xi & \text{se } \xi + x_1 - 2x_2 > 2 \\ -9x_1 + 24x_2 - 10\xi & \text{se } \xi + x_1 - 2x_2 < 0 \end{cases}$$

- $\bar{x}(0) = \{x : x_1 + x_2 = 1, \frac{2}{3} \leq x_1 \leq 1\}$.
- $\bar{x}(\frac{3}{2}) = \{x : x_1 + x_2 = 1, \frac{1}{6} \leq x_1 \leq \frac{5}{6}\}$.
- $\bar{x}(3) = \{x : x_1 + x_2 = 1, 0 \leq x_1 \leq \frac{1}{3}\}$.

- Tomando $\bar{x}(\frac{3}{2}) = (\frac{2}{3}, \frac{1}{3})$, EV $= z(\bar{x}, \frac{3}{2}) = 2 + \frac{3}{2} = 3.5$ e $EEV = \frac{1}{3}(2 + 3.5 + 14) = 6.5$.

- Observe que não há interseção para os três intervalos acima. Logo, como não há uma solução que seja ótima para todos os casos, esperamos observar EVPI $\neq 0$.

$$z(x,\xi) = \begin{cases} 2x_1 + 2x_2 + \xi & \text{se } 0 \leq \xi + x_1 - 2x_2 \leq 2 \\ -18 + 11x_1 - 16x_2 + 10\xi & \text{se } \xi + x_1 - 2x_2 > 2 \\ -9x_1 + 24x_2 - 10\xi & \text{se } \xi + x_1 - 2x_2 < 0 \end{cases}$$

- Obtendo o valor de WS:

$$WS = \frac{1}{3}(2+0) + \frac{1}{3}(\frac{7}{2}) + \frac{1}{3}(1+4) = 3.5$$

- Resolvendo o RP, obtemos: $x^* = (\frac{2}{3}, \frac{1}{3})$ e $RP = 6.5$.
- Logo: $EV = WS = 3.5 \leq RP = 6.5$, EVPI = 6.5-3.5 = 3.0 e VSS = 0.

Quando *EVPI* ou *VSS* assumem valores grandes.

Se ambos os valores são pequenos, não há estímulo para se obter RP, através da resolução de um programa estocástico, de elevado custo computacional:

1. Se *VSS* é alto, usar a solução de EV é muito caro, comparada à solução do programa estocástico, RP. Por outro lado, se VSS é pequeno, a solução com a incerteza média ξ funciona bem.
2. Se EVPI é baixo, a incerteza é pouco significante, uma vez que esperar para ver o que acontece e então decidir não é muito melhor do que decidir com base nos cenários que se vislumbram.

Intuitivamente...

1. Resolver um problema estocástico parece fazer mais sentido quando há mais aleatoriedade nos dados.
2. Isto nos faz pensar que EVPI e VSS devem crescer quando a variância das variáveis aleatórias do problema cresce.
3. Mas isso pode não ser o caso, como o exemplo a seguir demonstra.

- $x \in \mathbb{R}$ é a única variável do problema
- $\boldsymbol{\xi}$ assume apenas dois valores ξ_1, ξ_2 de forma que $\bar{\xi} = \frac{1}{2}$.
- Considere o seguinte problema:

$$\begin{aligned} \min \;\; & 6x + 10\mathbb{E}_{\boldsymbol{\xi}}|x - \boldsymbol{\xi}| \\ & x \geq 0 \end{aligned}$$

Caso base

$\xi_1 = \frac{1}{3}, \xi_2 = \frac{2}{3}, p_1 = p_2 = \frac{1}{2}$. $\mathrm{Var}(\boldsymbol{\xi}) = \frac{1}{36}$

Problema Determinístico Equivalente

$$\min \quad f(x) = 6x + 10\frac{1}{2}(|x - \frac{1}{3}| + |x - \frac{2}{3}|) : x \geq 0$$

Reformulando o PDE

$$\min \quad f(x) = \begin{cases} -4x + 5 & x \in [0, \frac{1}{3}] \\ 6x + \frac{5}{3} & x \in [\frac{1}{3}, \frac{2}{3}] \\ 16x - 5 & x \geq \frac{2}{3} \end{cases}$$

$$\min \quad f(x) = \begin{cases} -4x + 5 & x \in [0, \frac{1}{3}] \\ 6x + \frac{5}{3} & x \in [\frac{1}{3}, \frac{2}{3}] \\ 16x - 5 & x \geq \frac{2}{3} \end{cases}$$ Solução do PDE:

$x^* = \frac{1}{3}$, $RP = 3\frac{2}{3}$.

- $\bar{x}(\frac{1}{3}) = \frac{1}{3}, f(\bar{x}(\frac{1}{3})) = 2$ e $\bar{x}(\frac{2}{3}), f(\bar{x}(\frac{2}{3})) = 4$, WS $= \frac{1}{2}(2 + 4) = 3$
- EVPI $=$ RP - WS $= \frac{2}{3}$
- $\bar{\xi} = \frac{1}{2}$, $\bar{x}(1/2) = 1/2, EV = 3.$
- EEV $= \frac{1}{2}(\frac{28}{6} + \frac{28}{6}) = \frac{28}{6}$
- VSS $= 3\frac{2}{3} - \frac{28}{6} = 1$

Caso #2 - cresce variância em relação ao caso base.

$\xi_1 = 0, \xi_2 = 1, p_1 = p_2 = \frac{1}{2}$. Var$(\boldsymbol{\xi}) = \frac{1}{4}$

- EVPI $= 2$
- VSS $= 3$

Tanto EVPI quanto VSS crescem com o aumento de variância, em relação ao caso base.

Caso #3 - decresce variância em relação ao Caso #2

$\xi_1 = 0, \xi_2 = \frac{5}{8}$ e $p_1 = 0.2, p_2 = 0.8$. Var$(\boldsymbol{\xi}) = \frac{1}{16}$

- EVPI $= 2$
- VSS $= 0$

O EVPI cresceu em relação ao caso base, com variância maior. Porém, $VSS = 0$ e se este resultado fosse previamente conhecido, a resolução do problema estocástico poderia ser evitada, resolvendo-se o determístico formulado para $\bar{\xi} = \mathbb{E}_{\boldsymbol{\xi}}$.

Caso #4 - decresce variância em relação ao caso base.

$\xi_1 = 0.4, \xi_2 = 0.8, p_1 = 0.75, p_2 = 0.25$. $\text{Var}(\boldsymbol{\xi}) = 0.03$

- EVPI $= 0.4$
- VSS $= 1.1$

Comportamento inverso ao observado no caso imediatamente anterior, em relação ao caso base: redução em EVPI e incremento em VSS, com a redução de variância.

Observações:

- Um programa estocástico mais difícil (envolvendo variáveis inteiras no primeiro e/ou segundo estágio) não necessariamente produz valores maiores de EVPI e VSS - Exercício 3 §1.1.
- Não é fácil obter uma regra geral para o comportamento de EVPI e VSS.

- Vamos assumir que a única variável aleatória é o vetor h, isto é, $\boldsymbol{\xi} = h(\omega)$ e que Ξ é finito. $\xi^1, \ldots, \xi^K$ indicam as possíveis realizações de $\boldsymbol{\xi}$, com probabilidades $p^k : k = 1, \ldots, K$.
- Dizemos que o cenário k corresponde à realização ξ^k de $\boldsymbol{\xi}$.

Cenários de referência, ξ^r

- Para refinar os limites superiores, vamos empregar o conceito de cenários de referência. Dois possíveis são:
  - $\bar{\xi} = \mathbb{E}_{\boldsymbol{\xi}}$, a esperança de $\boldsymbol{\xi}$.
  - O cenário de pior caso, por exemplo, o cenário no qual a maior demanda deve ser atendida. O cenário de pior caso é fácil de ser determinado se as entradas de $\boldsymbol{\xi}$ são independentes. Caso contrário, não é trivial encontrá-lo.
- Observe que nos dois casos, os cenários de referência podem não ser cenários em Ξ.
- Definimos $p^r = P(\boldsymbol{\xi} = \xi^r)$. Se o cenário de referência não é um cenário no suporte Ξ, $p^r = 0$.

Dado um cenário de referência ξ^r e um cenário $\xi^k \in \Xi$, definimos o subproblema correspondente ao par de cenários:

$$\begin{aligned}
\min z^P(x, \xi^r, \xi^k) &= c^T x + p^r q^T y(\xi^r) + (1 - p^r) q^T y(\xi^k) \\
s.t.\ \ Ax &= b \\
Wy(\xi^r) &= \xi^r - Tx \\
Wy(\xi^k) &= \xi^k - Tx \\
x, y &\geq 0
\end{aligned}$$

1. O programa acima pode ser entendido como um PDE com dois cenários apenas, com probabilidades p^r e $1 - p^r$.
2. Assumimos que $(\bar{x}^k, \bar{y}^k, y(\xi^k))$ represente uma solução ótima para o subproblema associado ao par ξ^r, ξ^k e $z_k = z^P(\bar{x}^k, \bar{y}^k, y(\xi^k))$, sua fo ótima.

Casos de interesse para o problema de par de cenários

1. $z^P(x, \xi^r, \xi^r) = z(x, \xi^r)$, isto é, quando $\xi^r = \xi^k$, o problema de par de cenários é o PDE em que há apenas o cenário de referência.

2. Se $\xi^r \notin \Xi$, $P(\boldsymbol{\xi} = \xi^r) = 0$ e $z^P(x, \xi^r, \xi^k) = z(x, \xi^k)$.

Definição

Vamos definir SPEV como a *soma dos pares de valores esperados*:

$$SPEV = \frac{1}{1 - p^r} \sum_{k=1, \xi^k \neq \xi^r}^{k} p^k \min z^P(x, \xi^r, \xi^k)$$

Veja que SPEV é um conceito que faz sentido mesmo se $\xi^r \notin \Xi$, neste caso, não sendo um conceito novo....

Proposição

Se $\xi^r \notin \Xi$, SPEV = WS.

Prova:

$\xi^r \notin \Xi \rightarrow p^r = 0$ e $z^P(x, \xi^r, \xi^k) = z(x, \xi^k)$. Então:
$SPEV = \sum_{k=1, \xi^k \neq \xi^r}^{K} p^k \min z^P(x, \xi^r, \xi^k) =$
$\sum_{k=1, \xi^k \neq \xi^r}^{K} p^k \min z(x, \xi^k) = WS$, pela definição de WS.

Proposição

$WS \leq SPEV \leq RP$.

Prova - primeira parte: $WS \leq SPEV$

- $SPEV = \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k \left(\frac{c^T \bar{x}^k + p^r q^T \bar{y}^k + (1-p^r) q^T y(\xi^k)}{1-p^r} \right)$ onde $(\bar{x}^k, \bar{y}^k, y(\xi^k))$ é uma solução ótima para o subproblema do par ξ^r, ξ^k.
- Pela definição do par de subproblemas, o vetor $(\bar{x}^k, \bar{y}^k)$ é viável para o problema $z(x, \xi^r)$, de forma que: $c^T \bar{x} + q^T \bar{y}^k \geq \min z(x, \xi^r) = z_r^*$
- Vamos escrever $c^T \bar{x} = p^r c^T \bar{x} + (1 - p^r) c^T \bar{x}$ e reescrever $SPEV$ como $SPEV = \sum_{k=1,\xi^k \neq \xi^r}^{K} \frac{p^k \left[p^r (c^T \bar{x}^k + q^T \bar{y}^k) + (1-p^r)(c^T \bar{x}^k + q^T y(\xi^k)) \right]}{1-p^r}$
- Definimos $z_k^* = \min z(x, \xi^k)$.

Prova - primeira parte: $WS \leq SPEV$ (continua)

$$SPEV = \sum_{k=1,\xi^k \neq \xi^r}^{K} \frac{p^k \left[p^r(c^T\bar{x}^k + q^T\bar{y}^k) + (1-p^r)(c^T\bar{x}^k + q^T y(\xi^k))\right]}{1-p^r}$$

$$\geq \sum_{k=1,\xi^k \neq \xi^r}^{K} \frac{p^k p^r z_r^*}{1-p^r} + \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k(c^T\bar{x}^k + q^T y(\xi^k))$$

$$= p^r z_r^* + \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k(c^T\bar{x}^k + q^T y(\xi^k))$$

$$\geq p^r z_r^* + \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k z_k^*$$

$$= WS$$

Prova - segunda parte: $SPEV \leq RP$

- Definimos $x^*, y^*(\xi^k) : k = 1, \ldots, K$ a solução ótima de RP.
- Por simplicidade, assumimos que $\xi^r \in \Xi$ (isso não reduz a generalidade).
- $(x^*, y^*(\xi^r), y^*(\xi^k))$ é viável para o par ξ^r, ξ^k de subproblemas. Logo:

$$c^T\bar{x}^k + p^r q^T \bar{y}^k + (1-p^r)q^T y(\xi^k) \leq c^T x^* + p^r q^T y^*(\xi^r) + (1-p^r)q^T y^*(\xi^k)$$

- Multiplicando cada desigualdade acima para todo $k : \xi^k \neq \xi^r$ com peso p^k e somando, obtemos a desigualdade

$$(1-p^r)SPEV \leq \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k \left(c^T x^* + p^r q^T y^*(\xi^r) + (1-p^r)q^T y^*(\xi^k)\right)$$

Prova - segunda parte: $SPEV \leq RP$ (continua)

$$
\begin{aligned}
(1-p^r)SPEV \leq & \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k \left(c^T x^* + p^r q^T y^*(\xi^r) + (1-p^r) q^T y^*(\xi^k)\right) \\
=& (c^T x^* + p^r q^T y^*(\xi^r)) \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k + \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k (1-p^r) q^T y^*(\xi^k) \\
=& (1-p^r)(c^T x^* + p^r q^T y^*(\xi^r)) + \sum_{k=1,\xi^k \neq \xi^r}^{K} p^k (1-p^r) q^T y^*(\xi^k) \\
=& (1-p^r)(c^T x^* + \sum_{k=1}^{K} p^k q^T y^*(\xi^k)) \\
=& (1-p^r) RP \quad \square
\end{aligned}
$$

- Seja $z(x, \xi^r)$ o problema determinístico formulado com o cenário de referência que, como discutimos, pode não ser um cenário possível (no suporte).
- Definimos $\bar{x}^r \in \arg\min_x z(x, \xi^r)$

EVRS e (re)definição de VSS

O expected value of reference scenario (EVRS) é definido como

$$EVRS = \mathbb{E}_{\boldsymbol{\xi}} z(\bar{x}^r, \boldsymbol{\xi})$$

e a partir dele generalizamos o VSS como

$$VSS = EVRS - RP.$$

Veja que o conceito de VSS coincide com o anteriormente dado caso $\xi^r = \bar{\xi}$. Independentemente disso, $VSS \geq 0$, pois se $\bar{x}^r$ for viável (para ξ^k), $z(\bar{x}^r, \xi^k)$ fornece um limite superior para o recurso associado a k e $EVRS \geq RP$. Caso contrário, sendo inviável, $EVRS = \infty$.

- Assim como antes, $(\bar{x}^k, \bar{y}^k, y(\xi^k))$ denota uma solução ótima para o par de subproblema ξ^r, ξ^k.
- Definimos (*expectations of pairs expected value*)

$$EPEV := \min_{k=1,\dots,K\cup\{r\}} \mathbb{E}_{\boldsymbol{\xi}} z(\bar{x}^k, \boldsymbol{\xi})$$

Proposição

$$RP \leq EPEV \leq EVRS$$

Prova

Os três valores são o resultado ótimo para a função recurso $\min_x \mathbb{E}_{\boldsymbol{\xi}} z(x, \boldsymbol{\xi})$ em sucessivas relaxações do domínio. Veja que o domínio de RP é $x \in K_1 \cap K_2$, o domínio de EPEV é $\{\bar{x}^k : k = 1, \dots, K \cup \{r\}\} \subseteq K_1 \cap K$ e, o domínio de EVRS é $\bar{x}^r$.

Corolário

$$0 \leq EVRS - EPEV \leq VSS \leq EVRS - SPEV \leq EVRS - WS$$

Exemplo ilustrando caso onde WS e EEV são inconclusivos.

$$\begin{aligned}
\min\ & 3x_1 + 2x_2 + \mathbb{E}_{\xi} \min(-15y_1 - 12y_2) \\
s.t.\ & 3y_1 + 2y_2 \leq x_1 \\
& 2y_1 + 5y_2 \leq x_2 \\
& .8\xi_1 \leq y_1 \leq \xi_1 \\
& .8\xi_2 \leq y_2 \leq \xi_2 \\
& x, y \geq 0
\end{aligned}$$

onde $\xi_1 \in \{4, 6\}$ e $\xi_2 \in \{4, 8\}$, com distribuições independentes, com probabilidade 0.5 para cada um dos casos.

Interpretação do problema

x_1, x_2 podem ser vistos como decisões de investimentos em dois ativos distintos, tomados hoje, que serão necessários em um segundo estágio, para cobrir 80% de uma certa demanda incerta, ξ_1, ξ_2.

Cenário	Solução		fo ótima
	x : 1o. estágio	y : 2o. estágio	$z(\bar{x}(\xi), \xi)$
1. (4,4)	(18.4,24)	(4,3.2)	4.8
2. (6,4)	(24.4,28)	(6,3.2)	0.8
3. (4,8)	(24.8,40)	(4,6.4)	17.6
4. (6,8)	(30.8,44)	(6,6.4)	13.6
$\bar{\xi} = (5, 6)$	(24.6,34)	(5,4.8)	EV $= 9.2$

- A solução $\bar{x}(\bar{\xi}) = (24.6, 34)$ é inviável para o programa estocástico (cenário 4, $\xi_1 = 6, \xi_2 = 8$). Logo $EEV = +\infty$.
- $WS = \frac{1}{4}(4.8 + 0.8 + 17.6 + 13.6) = 9.2$
- $WS = EV = 9.2$
- $EV = WS = 9.2 \leq RP \leq EEV \leq +\infty$.
- Conclusões óbvias (inúteis): $0 \leq EVPI \leq +\infty$ e $0 \leq VSS \leq +\infty$

- Em casos como o do exemplo, a investigação dos pares de subproblemas gera informação relevante.
- Vamos tomar como caso de referência o pior caso, $\xi^r = (6, 8)$, quando a demanda é máxima (e não a demanda média).
- Este cenário foi escolhido uma vez que trata-se de um problema de atendimento de demanda, faz mais sentido para a aplicação.
- O caso de referência é um dos cenários previstos, o cenário 4. Portanto, $p^r = \frac{1}{4}$.

$$
\begin{aligned}
\min\ & 3x_1 + 2x_2 - \frac{1}{4}(15y_1^r + 12y_2^r) - \frac{3}{4}(15y_1^k + 12y_2^k) \\
& 3y_1^r + 2y_2^r \leq x_1 \\
& 2y_1^r + 5y_2^r \leq x_2 \\
& 3y_1^k + 2y_2^k \leq x_1 \\
& 2y_1^k + 5y_2^k \leq x_2 \\
& 4.8 \leq y_1^r \leq 6 \\
& 6.4 \leq y_2^r \leq 8 \\
& 0.8\xi_1^k \leq y_1^k \leq \xi_1^k \\
& 0.8\xi_2^k \leq y_2^k \leq \xi_2^k \\
& x, y^r, y^k \geq 0
\end{aligned}
$$

Pares de cenários	Solução x :	y^r :	$y(\xi^k)$:	fo ótima z^P
1. (4,4) e ξ^r	(27.2,41.6)	(4.8, 6.4)	(4,4)	46.6
2. (6,4) e ξ^r	(27.2,41.6)	(4.8, 6.4)	(6,4)	24.1
3. (4,8) e ξ^r	(27.2,41.6)	(4.8, 6.4)	(4,6.72)	22.12

- $SPEV = \frac{4}{3}\frac{1}{4}(46.6 + 24.1 + 22.12) = 30.94$
- $EPEV = \min_k \mathbb{E}_{\boldsymbol{\xi}} z(\bar{x}^k(\xi^k), \boldsymbol{\xi}) = \mathbb{E}_{\boldsymbol{\xi}} z((27.2, 41.6), \boldsymbol{\xi}) = 30.94$
- $EVRS = \mathbb{E}_{\boldsymbol{\xi}} z((30.8, 44), \boldsymbol{\xi}) = 40.6$

Então temos:
$WS = 9.2 \leq SPEV = 30.94 \leq RP \leq EPEV = 30.94 \leq EVRS = 40.6$.
Ou seja, $RP = 30.94$ com $x = (27.2, 41.6)$ como solução ótima do programa estocástico. Observe que resolvemos RP indiretamente, resolvendo vários PPLs determinísticos, de dimensões muito menores que o PDE.