Conhecimento e Raciocínio

Aula 8

Árvores Bayesianas, Dempster Shaffer, Cadeias de Markov

Viriato A.P. Marinho Marques

DEIS - ISEC

2020 / 2021

8. Incerteza

7. Métodos de Representação

1. **Factores de Certeza** (*Certainty Factors - CF*)
    1. MYCIN
    2. CLIPS
    3. EXSYS

2. **Probabilidade**
    1. Redes Bayesianas
    2. Árvores de Decisão Bayesianas
    3. Cadeias de Markov

3. **Teoria de Dempster-Shaffer**

4. **Lógica Difusa**
    1. Inferência de Mandani / *JFS freeware*
    2. Regra Composicional de Inferência / *Sistema CADIAG-2*
    3. *Fuzzy Pattern Matching / Fuzzy CLIPS*

8.1 Redes Bayesianas (resumo)

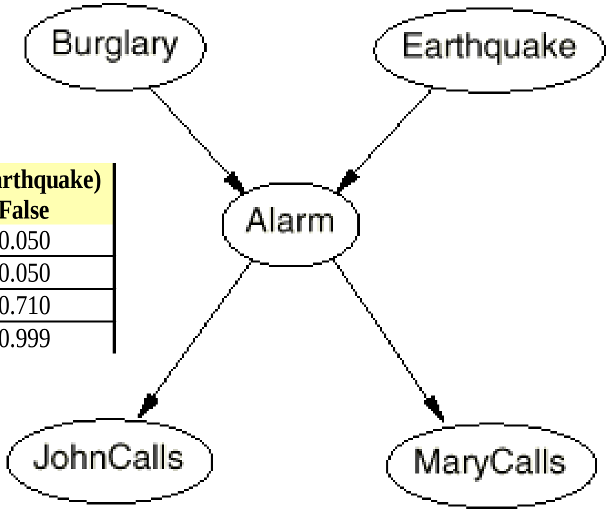
8.1 Redes Bayesianas - Resumo

- Já abordadas
- Expressam a dependência de efeitos a partir de causas, representando entre eles apenas as ligações consideradas relevantes
- As tabelas que figuram nos nós são de Probabilidade Condicionada ou *a posteriori*

Burglary
0.001

Heartquake
0.002

Burglary	**Hearthquake**	**P(Alarm | Burglary, Earthquake)**	
		True	**False**
True	**True**	0.950	0.050
True	**False**	0.950	0.050
False	**True**	0.290	0.710
False	**False**	0.001	0.999



Alarm	**P(JohnCalls|Alarm)**	
	True	**False**
True	0.90	0.10
False	0.05	0.95

Alarm	**P(MaryCalls|Alarm)**	
	True	**False**
True	0.70	0.30
False	0.01	0.99

8.1 Redes Bayesianas (resumo)

Nota sobre Probabilidade Conjunta e Condicionada:

	X	¬X	
Crash	0.6	0.1	0.7
¬Crash	0.2	0.1	0.3
	0.8	0.2	1.00

Exemplo

A tabela ao lado é de Probabilidade Conjunta e exprime a probabilidade de

- Escolher um disco avariado **E** ele ser de marca X (0.6) (ou não ser de marca X (0.1))
- Escolher um disco bom **E** ele ser de marca X (0.2) (ou não ser marca X (0.1))

	X	¬X	
Crash	$p(C \cap X)$	$p(C \cap \neg X)$	p(Crash)
¬Crash	$p(\neg C \cap X)$	$p(\neg C \cap \neg X)$	p(¬Crash)
	p(X)	p(¬X)	1.00

- A soma das linhas e das colunas dá a probabilidade *a priori* do acontecimento X, ¬X, Crash ou ¬Crash
- Estas somas chamam-se probabilidades marginais e **não são** 1 mas...
- A soma das probabilidades marginais de uma linha ou coluna é 1

8.1 Redes Bayesianas (resumo)

A partir destas tabelas podem calcular-se probabilidades condicionadas de *Crash* ou *Not Crash* suposto que *Marca é X* ou *não é X*, segundo a Fórmula de Bayes:

	X	¬X	
Crash	0.6	0.1	0.7
¬Crash	0.2	0.1	0.3
	p(X) = 0.8	p(¬X) = 0.2	1.00

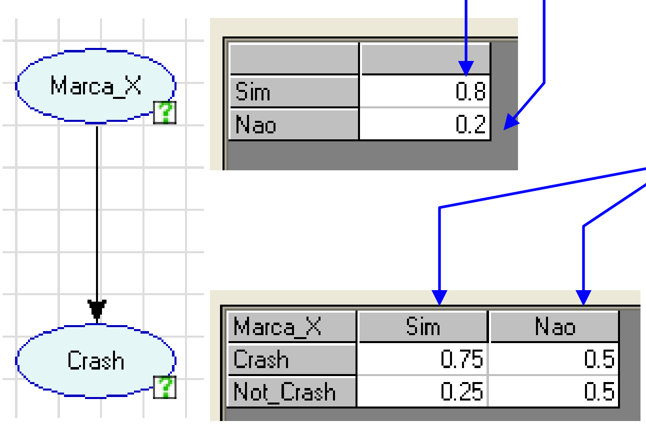
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(Crash|X) = \frac{P(Crash \cap X)}{P(X)} = \frac{0.6}{0.8} = 0.75$$

$$P(\neg Crash|X) = \frac{P(\neg Crash \cap X)}{P(X)} = \frac{0.2}{0.8} = 0.25$$

$$P(Crash|\neg X) = \frac{P(Crash \cap \neg X)}{P(\neg X)} = \frac{0.1}{0.2} = 0.5$$

$$P(\neg Crash|\neg X) = \frac{P(\neg Crash \cap \neg X)}{P(\neg X)} = \frac{0.1}{0.2} = 0.5$$

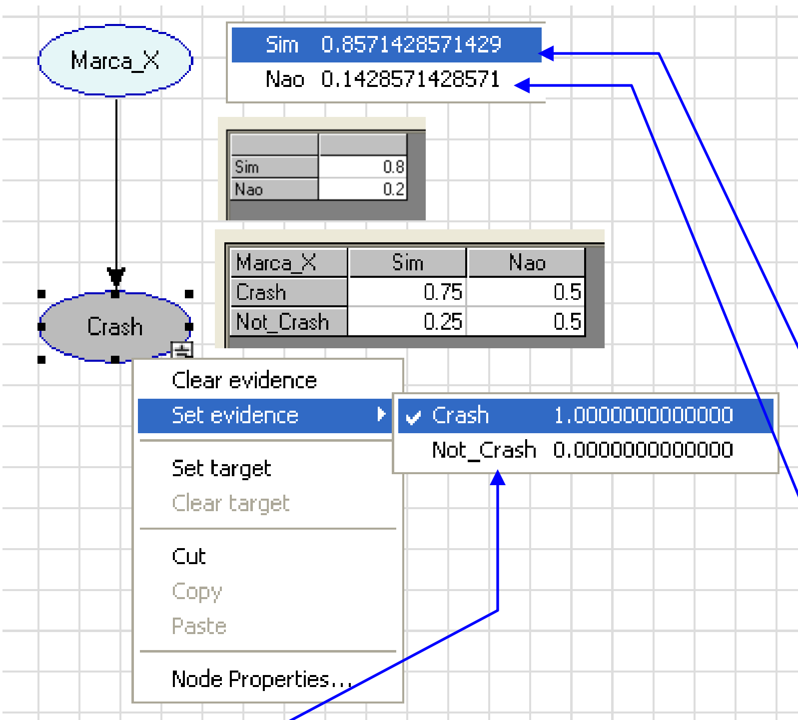


Estas probabilidades figuram na Rede Bayesiana:

- Agora as colunas têm de somar 1 (porque suposto que um disco é de marca X, ele *Crasha* ou *Não Crasha*)
- As probabilidades não condicionadas também somam 1 (porque um disco é X ou não é) e o seu valor é igual à soma das colunas na Tabela de Probabilidade Conjunta

8.1 Redes Bayesianas (resumo)

A rede permite calcular as probabilidades inversas *Ser Marca X* (ou *Não ser Marca X*) suposto *Crash*



Estes resultados também podem ser obtidos recorrendo à Fórmula de Bayes no seguinte formato:

$$P(B|A) = \frac{P(A|B).P(B)}{P(A|B).P(B) + P(A|\neg B).P(\neg B)}$$

$$P(X\,|\,Crash) = \frac{P(0.75).P(0.8)}{P(0.75).P(0.8) + P(0.5).P(0.2)} = 0.8571$$

$$P(\neg X\,|\,Crash) = \frac{P(0.5).P(0.2)}{P(0.5).P(0.2) + P(0.75).P(0.8)} = 0.1429$$

Se em vez de *Crash* se colocasse aqui a evidência *Not_Crash=1* a rede permitiria obter a probabilidade de *Ser Marca X* (ou *não ser Marca X*) suposto *Not Crash*

$$P(X\,|\,\neg Crash) = \ldots = 0.6667$$

$$P(\neg X\,|\,\neg Crash) = \ldots = 0.3333$$

8.1 Redes Bayesianas (resumo)

Note-se que todas estas probabilidades poderiam ainda ser obtidas directamente da Tabela de Probabilidade Conjunta: bastaria criar a rede "ao contrário":

	X	¬X	
Crash	0.6	0.1	0.7
¬Crash	0.2	0.1	0.3
	p(X) = 0.8	p(¬X) = 0.2	1.00

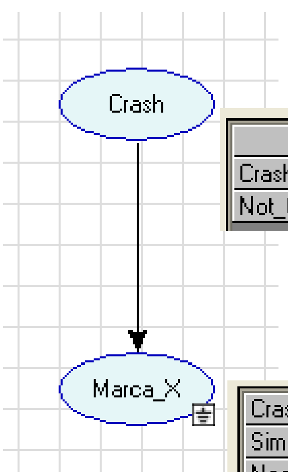
$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

$$P(X \mid Crash) = \frac{P(X \cap Crash)}{P(Crash)} = \frac{0.6}{0.7} = 0.8571$$

$$P(\neg X \mid Crash) = \frac{P(\neg X \cap Crash)}{P(Crash)} = \frac{0.1}{0.7} = 0.1429$$

$$P(X \mid \neg Crash) = \frac{P(X \cap \neg Crash)}{P(\neg Crash)} = \frac{0.2}{0.3} = 0.6667$$

$$P(\neg X \mid \neg Crash) = \frac{P(\neg X \cap \neg Crash)}{P(\neg Crash)} = \frac{0.1}{0.3} = 0.3333$$



Crash	0.7
Not_Crash	0.3

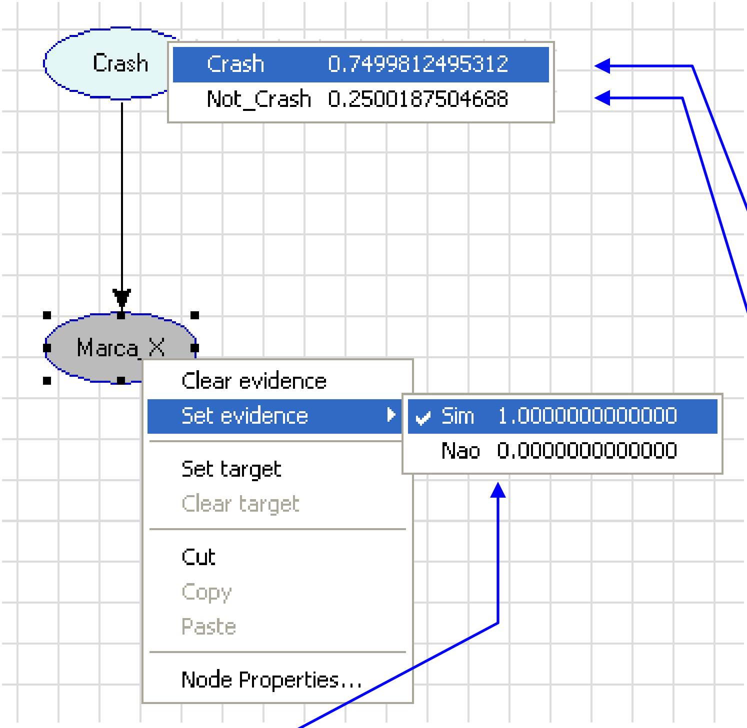
Crash	Crash	Not_Crash
Sim	0.8571	0.6667
Nao	0.1429	0.3333

Como anteriormente, estas probabilidades figuram na Rede Bayesiana:

- As colunas continuam a somar 1
- As probabilidades não condicionadas também somam 1 (porque um disco *crasha* ou não *crasha*) e o seu valor é igual à soma das **linhas** na Tabela de Probabilidade Conjunta

8.1 Redes Bayesianas (resumo)

Agora, o cálculo das probabilidades inversas regenera os valores da 1ª Rede Bayesiana:



Estes resultados também podem ser obtidos recorrendo à Fórmula de Bayes no formato já utilizado:

$$P(B|A) = \frac{P(A|B).P(B)}{P(A|B).P(B) + P(A|\neg B).P(\neg B)}$$

$$P(Crash \mid X) = \ldots = 0.75$$

$$P(\neg Crash \mid X) = \ldots = 0.25$$

Se em vez de *X* se colocasse aqui a evidência *Not_X=1* a rede permitiria obter a probabilidade de *Crashar* (ou *não ser Crashar*) suposto *Marca X*

$$P(Crash \mid \neg X) = \ldots = 0.5$$

$$P(\neg Crash \mid \neg X) = \ldots = 0.5$$

8.2 Ávores de Decisão Bayesianas

8.2 Árvores de Decisão e Raciocínio Hipotético

Numa Árvore de Decisão Bayesiana

- Rectângulos representam decisões
- Círculos representam acontecimentos

Tomemos o exemplo do disco marca X ocorrendo ou não *Crash*:

	X	¬X	
Crash	0.6	0.1	0.7
¬Crash	0.2	0.1	0.3
	p(X) = 0.8	p(¬X) = 0.2	1.00

$$P(Crash|X)=\frac{P(Crash\cap X)}{P(X)}=\frac{0.6}{0.8}=0.75$$

$$P(\neg Crash|X)=\frac{P(\neg Crash\cap X)}{P(X)}=\frac{0.2}{0.8}=0.25$$

$$P(Crash|\neg X)=\frac{P(Crash\cap\neg X)}{P(\neg X)}=\frac{0.1}{0.2}=0.5$$

$$P(\neg Crash|\neg X)=\frac{P(\neg Crash\cap\neg X)}{P(\neg X)}=\frac{0.1}{0.2}=0.5$$

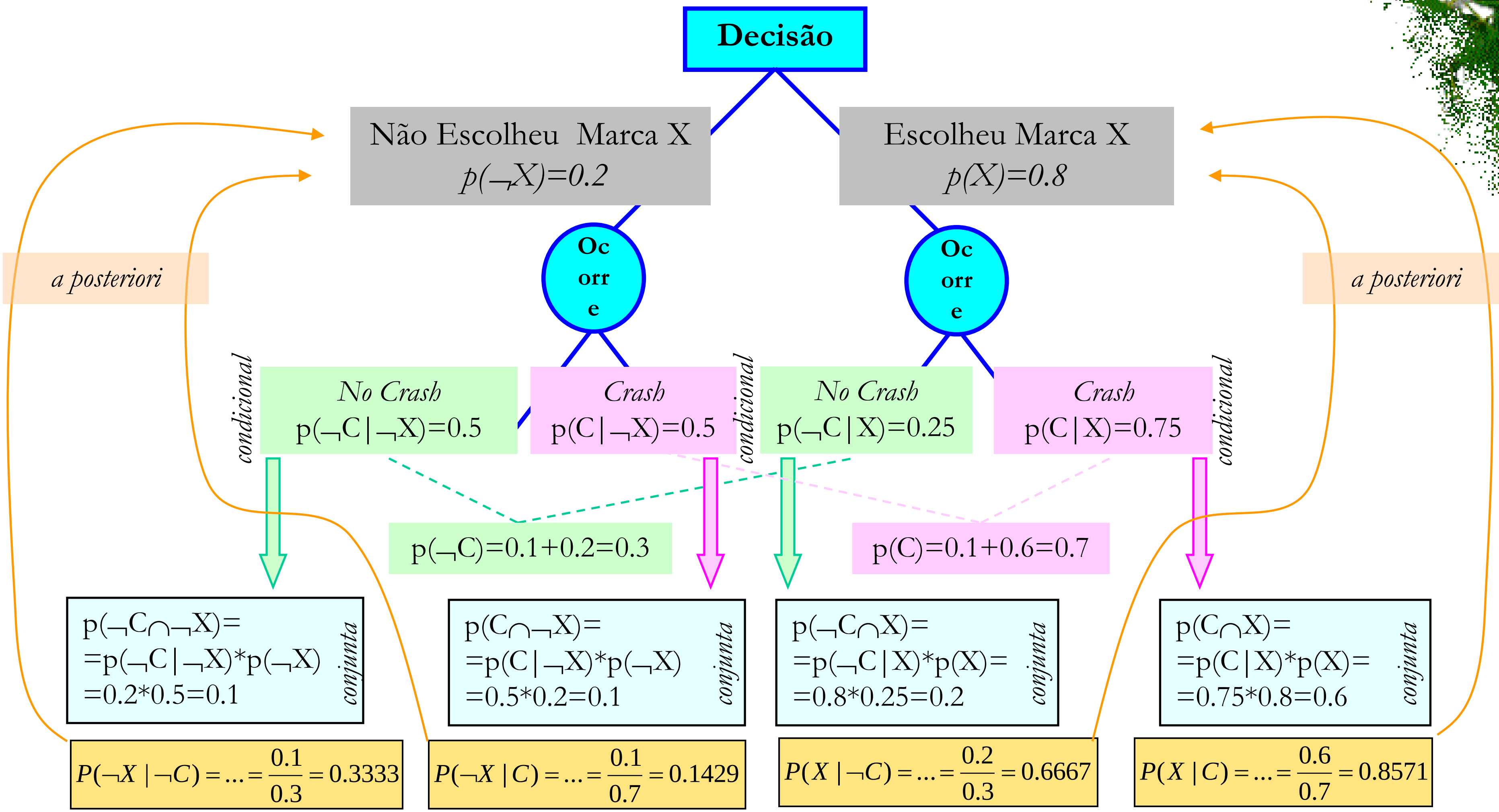
$$P(X\mid Crash)=\frac{P(X\cap Crash)}{P(Crash)}=\frac{0.6}{0.7}=0.8571$$

$$P(X\mid\neg Crash)=\frac{P(X\cap\neg Crash)}{P(\neg Crash)}=\frac{0.2}{0.3}=0.6667$$

$$P(\neg X\mid Crash)=\frac{P(\neg X\cap Crash)}{P(Crash)}=\frac{0.1}{0.7}=0.1429$$

$$P(\neg X\mid\neg Crash)=\frac{P(\neg X\cap\neg Crash)}{P(\neg Crash)}=\frac{0.1}{0.3}=0.3333$$

8.2 Ávores de Decisão Bayesianas



8.2 Ávores de Decisão Bayesianas

PROSPECTOR: para aconselhamento de prospecção mineira, baseia-se em árvores de decisão Bayesianas:

- A sua fama provém de ter sido o primeiro SP a descobrir uma jazida de molibdénio que se verificou ser de grande valor.

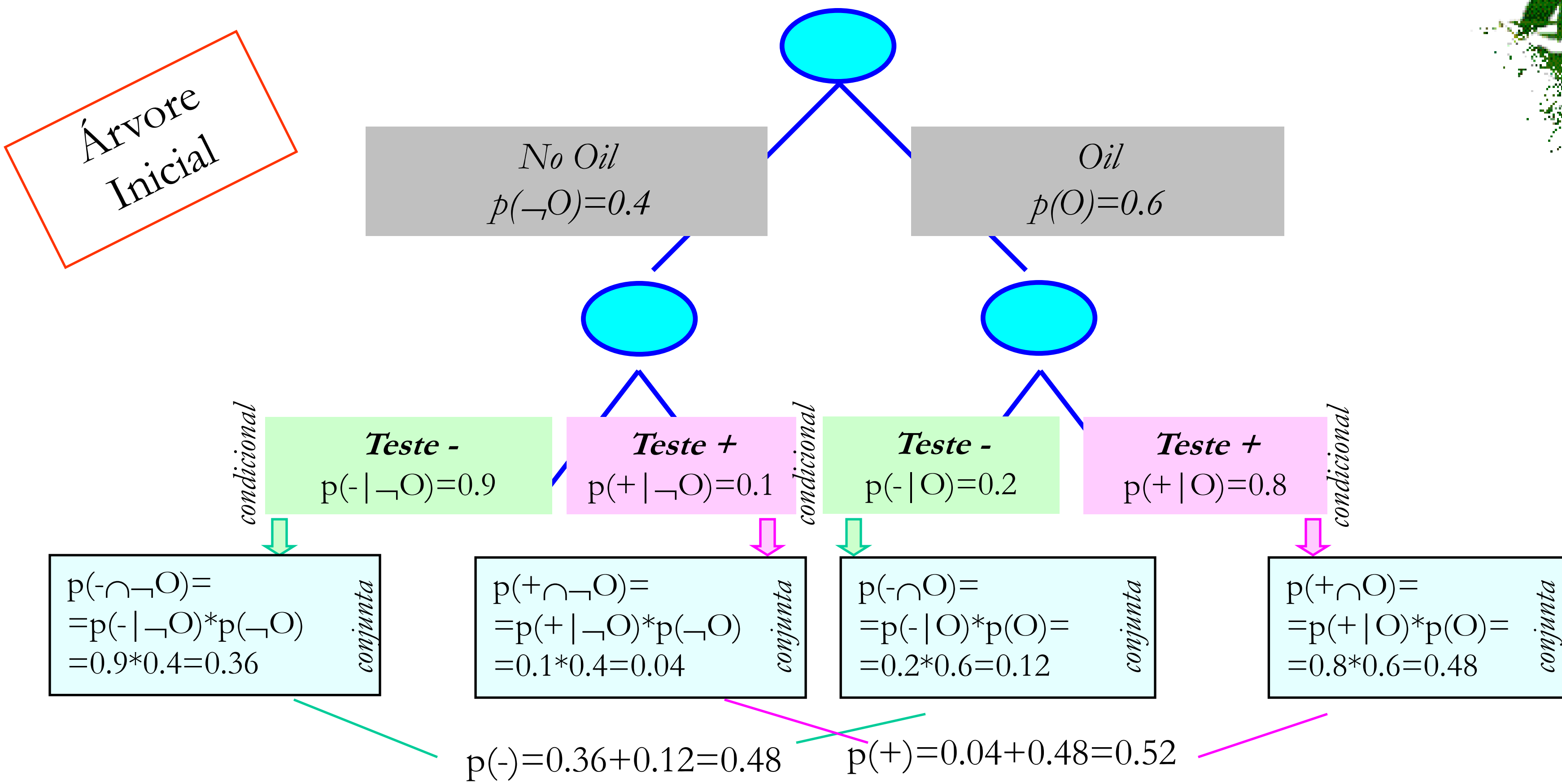
Considere-se a decisão *perfurar ou não perfurar* para procurar petróleo:

- Sem qualquer evidência, pode partir-se de $p(O)=p(\neg O)=0.5$
- Porém, na prática, como se está a procurar petróleo num dado local, então há algumas razões para crer que ele lá pode estar. Partamos então de $\mathbf{p(O)=0.6}$ e $\mathbf{p(\neg O)=0.4}$

Na procura de petróleo usam-se testes sísmicos, i.e. explodem-se cargas e analisa-se o tipo de ondas recebidas num local pré-determinado. Mas

- As ondas estão sujeitas a interferências causadas pelo subsolo:
  - Podem indicar presença de petróleo onde de facto não existe: seja a probabilidade destas ocorrências $\mathbf{p(+|O)=0.8}$ e $\mathbf{p(+|\neg O)=0.1}$
  - Podem também não indicar petróleo onde afinal ele existe: seja a probabilidade destas ocorrências $\mathbf{p(-|O)=0.2}$ e $\mathbf{p(-|\neg O)=0.9}$

8.2 Ávores de Decisão Bayesianas



Com base nas probabilidades conjuntas podemos calcular as *a posteriori*

$p(\neg O|-) = \frac{0.36}{0.36+0.12} = 3/4$

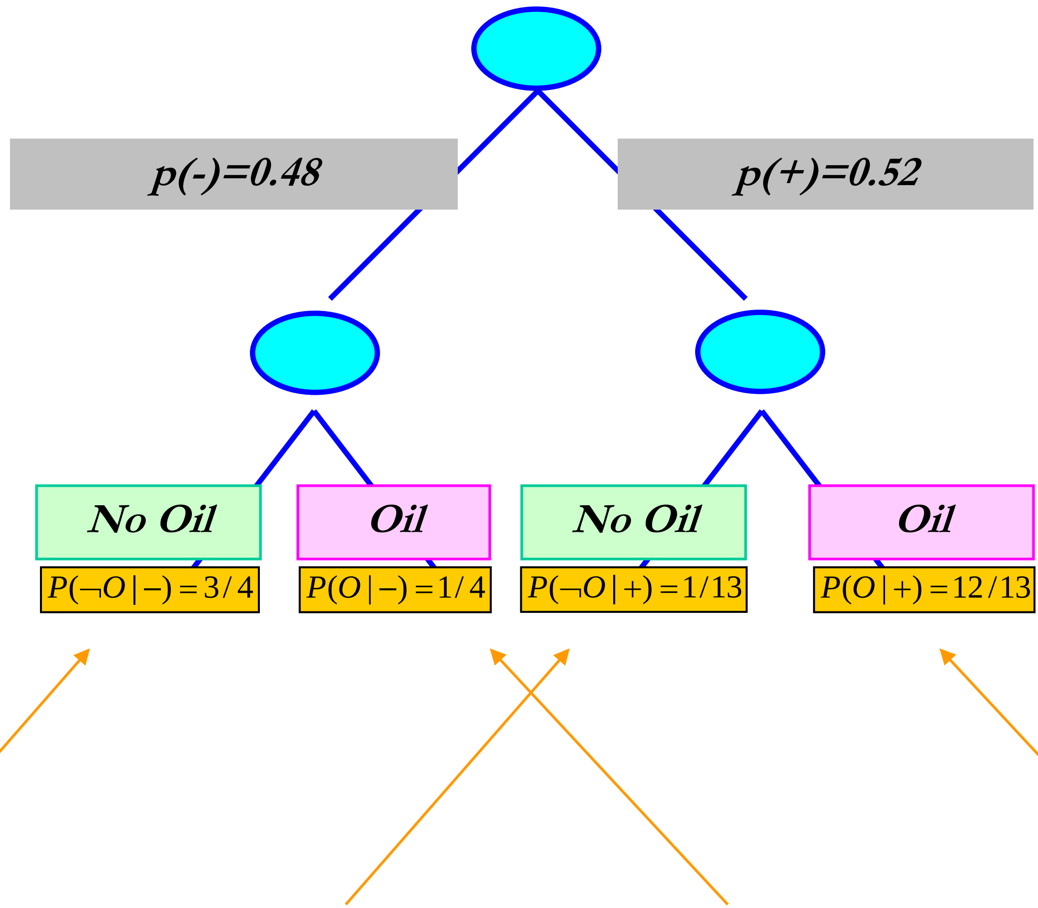
$P(\neg O|+) = \frac{0.04}{0.04+0.48} = 1/13$

$P(O|-) = \ldots = \frac{0.12}{0.36+0.12} = 1/4$

$P(O|+) = \frac{0.8}{0.48+0.04} = 12/13$

8.2 Ávores de Decisão Bayesianas

Árvore Revista



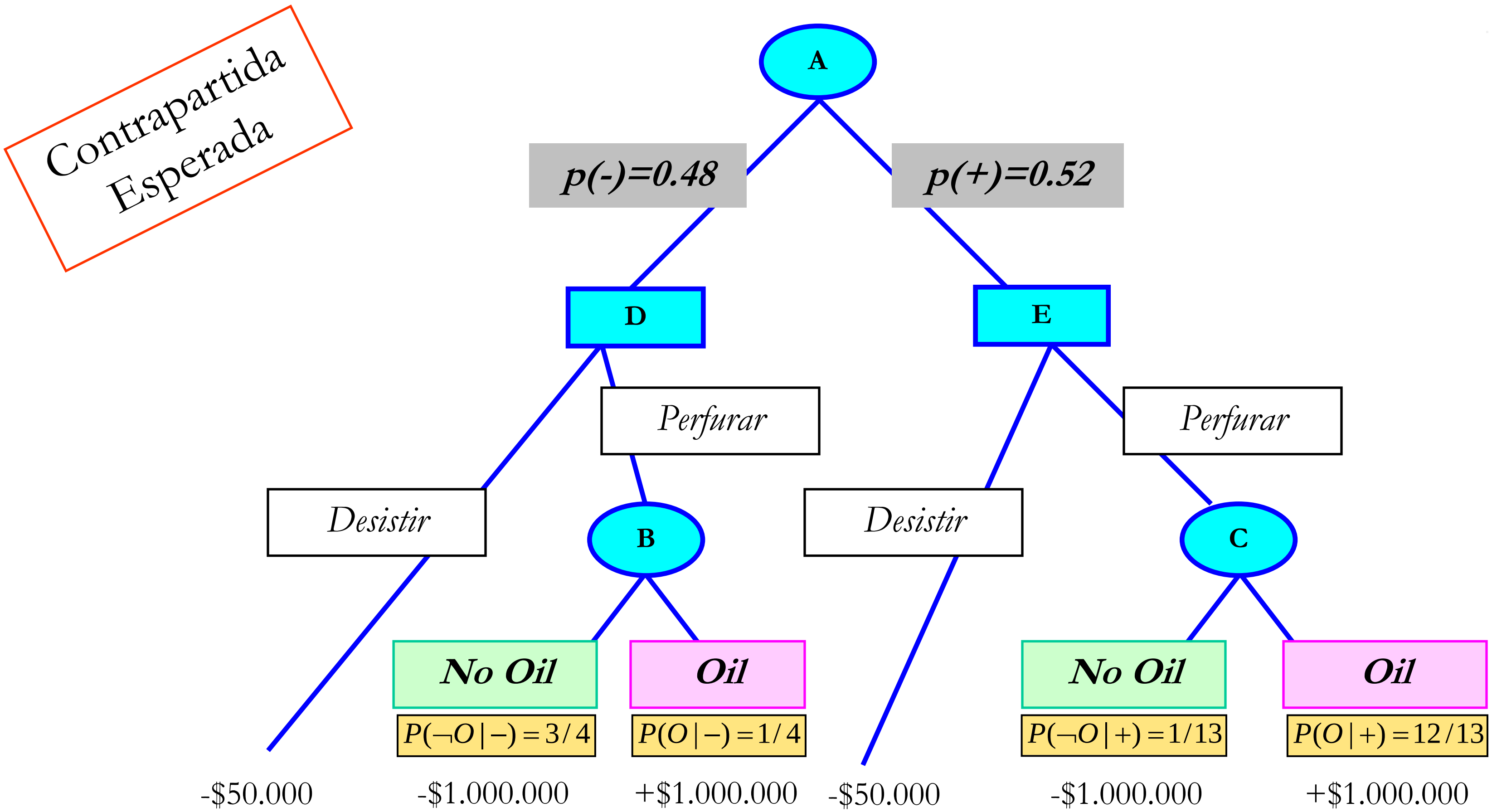
$$p(\neg O\,|\,-) = \frac{0.36}{0.36+0.12} = 3/4$$

$$P(\neg O\,|\,+) = \frac{0.04}{0.04+0.48} = 1/13$$

$$P(O\,|\,-) = \ldots = \frac{0.12}{0.36+0.12} = 1/4$$

$$P(O\,|\,+) = \frac{0.8}{0.48+0.04} = 12/13$$

8.2 Ávores de Decisão Bayesianas

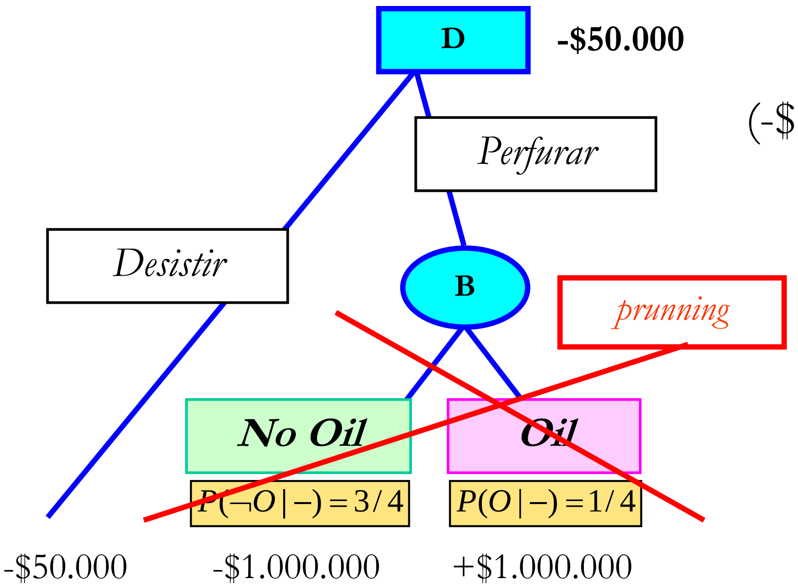


Petróleo: +$1.250.000
Perfuração: -$200.000
Estudo Sísmico: -$50.000

Nestas condições:

Deve perfurar-se ou não ?

8.2 Ávores de Decisão Bayesianas



Calcular esperança no nó B

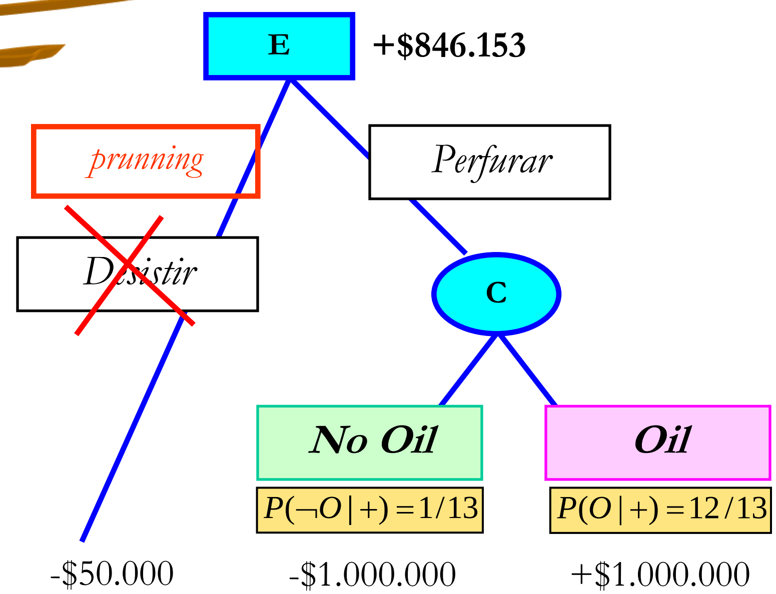
(-$1.000.000*3/4)+($1.000.000*1/4)=-$500.000

Em D há agora uma decisão entre perder $50.000 ou perder $500.000.
A melhor escolha é perder apenas $50.000
Este valor escreve-se em D

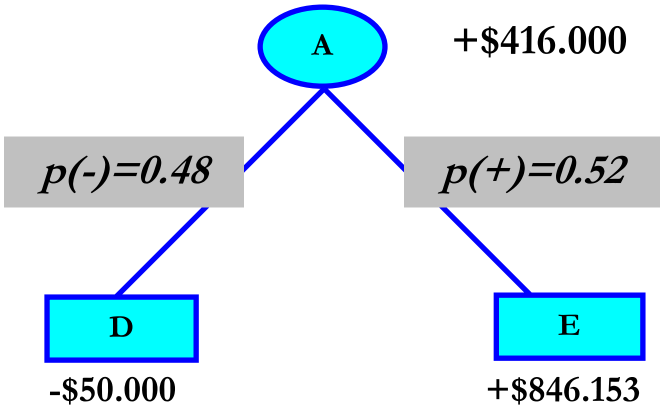
Calcular esperança no nó C

(-$1.000.000*1/13)+($1.000.000*12/13)=$846.153

Em E há agora uma decisão entre perder $50.000 ou ganhar $846.153.
A melhor escolha é ganhar $846.153
Este valor escreve-se em E



8.2 Ávores de Decisão Bayesianas



Calcular esperança no nó A
(-$50.000*0.48)+($846.133*0.52)=+$416.000

Este valor é escrito em A.
Como este nó é a raiz da árvore, nada mais há a fazer.

Resposta:

- Se o teste for +, deve perfurar-se (contrapartida esperada = $846.153)
- Se o teste for -, o local deve ser abandonado

Uma árvore de decisão:

- É essencialmente um diagrama da estratégia óptima para solucionar um problema
- .É um exemplo de **Raciocínio Hipotético** *(What If...)*.

A MindBox (antiga Inference Coporation) incorpora no Art*Enterprise mecanismos de raciocínio hipotético.

Art Enterprise 10.0

8.3 Cadeias de Markov

8.3 Cadeias de Markov e Raciocínio Temporal

Além do Hipotético, alguns SP utilizam **Raciocínio Temporal**:

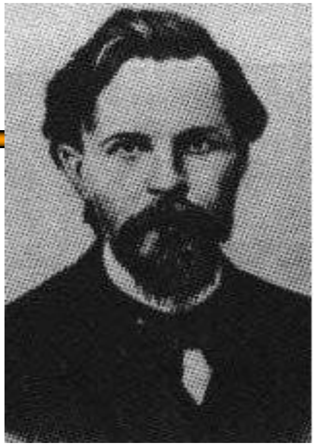
- O Raciocínio Temporal é complexo de implementar
- Os humanos usam-no com certa facilidade para gerarem possíveis estados futuros de um sistema em função do estado actual
- **Exemplo:** um controlador de tráfego aéreo

No campo da medicina foram desenvolvidos alguns SP incorporando raciocínio temporal:

- VM - controlo de ventilação assistida
- CASNET - tratamento do glaucoma ocular
- DIGITALIS - Conselheiro de administração de digitalina a pacientes cardíacos

As **Cadeias de Markov** são um suporte matemático para a implementação de Raciocínio Temporal baseado em probabilidade.

8.3 Cadeias de Markov



Andreevich Markov
(1856-1922)

Imaginemos um sistema que evolui ao longo do tempo, passando de um estado a outro, em que

- A mudança de estado é determinada pela probabilidades de ocorrência de certos eventos
- A progressão de um sistema através de uma sequência de estados determinados probabilisticamente chama-se **Processo Estocástico**
- **Exemplos**: stocks, votações, meteorologia, doenças e falhas de equipamento em que os sintomas evoluem com o tempo

8.3 Cadeias de Markov

Um processo estocástico pode representar-se por matrizes:

Exemplo:

- p(comprar disco de marca X)=0.8
- p(ter disco de marca X e comprar de novo marca X)=0.1
- p(ter disco de outra marca e comprar outro de marca X)=0.6

Estado inicial $S_1 = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix}$

A Matriz (probabilística) de Transição é um

- **Vector Probabilísico**: As linhas somam 1 e os elementos são todos positivos

$$T = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

Estado seguinte: Qual a % de população que na "geração seguinte" terá discos de marca X e de outra marca?

- 0.2 serão marca X
- 0.8 serão de outra marca

$$S_2 = S_1.T =$$
$$= \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} . \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix} =$$
$$= \begin{bmatrix} (0.8)(0.1) + (0.2)(0.6) & (0.8)(0.9) + (0.2)(0.4) \end{bmatrix} =$$
$$= \begin{bmatrix} 0.2 & 0.8 \end{bmatrix}$$

8.3 Cadeias de Markov

A sucessão de estados pode continuar a calcular-se da mesma forma

Os estados convergem para [0.4   0.6]

Esta matriz chama-se *Matriz de Steady-State (equilíbrio)*

$$S_3 = S_2.T = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$
$$S_4 = S_3.T = \begin{bmatrix} 0.35 & 0.65 \end{bmatrix}$$
$$S_5 = S_4.T = \begin{bmatrix} 0.425 & 0.575 \end{bmatrix}$$
$$S_6 = S_5.T = \begin{bmatrix} 0.3875 & 0.6125 \end{bmatrix}$$
$$S_7 = S_6.T = \begin{bmatrix} 0.40625 & 0.59375 \end{bmatrix}$$
$$S_8 = S_7.T = \begin{bmatrix} 0.396875 & 0.602125 \end{bmatrix}$$

A Matriz de Equilíbrio não depende do Estado Inicial, mas apenas da Matriz T:

- Portanto, se S_1 fosse qualquer outro estado, o equilíbrio seria ainda **[0.4   0.6]**

Uma **Cadeia de Markov** é definida da seguinte forma:

- Tem um número finito de estados
- Num dado momento, o processo pode estar num só estado
- O processo move-se de estado para estado ao longo do tempo
- O estado seguinte depende apenas do anterior e não de qualquer outro estado que já tenha sido visitado (*memoryless property*)

8.3 Cadeias de Markov

Um *vector probabilístico S* é uma *matriz de steady-state* para a *matriz de transição* ***T*** quando ***S=T.S***

Exemplo

Calcular a *matriz de steady-state* para a *matriz de transição*

$$T = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

De acordo com a definição teremos de ter

$$S = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix} .S$$

De acordo com a regra de multiplicação de matrizes, S terá de ter 1 linha e 2 colunas. Logo

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix} . \begin{bmatrix} x & y \end{bmatrix}$$

$$\begin{cases} 0.1x + 0.6y = x \\ 0.9x + 0.4y = y \end{cases} \Rightarrow \begin{cases} -0.9x + 0.6y = 0 \\ 0.9x - 0.6y = 0 \end{cases} \Rightarrow$$

Para resolver tem de usar outra equação. Qual ? x+y=1

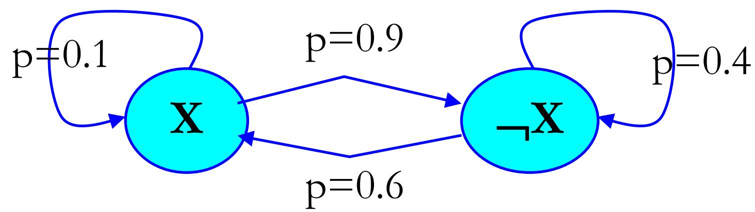
$$\begin{cases} -9x + 6y = 0 \\ x + y = 1 \end{cases} \qquad \begin{cases} -9x + 6y = 0 \\ 9x + 9y = 9 \end{cases} \qquad \begin{cases} 15y = 9 \\ ---- \end{cases} \qquad \begin{cases} y = \frac{9}{15} = \frac{3}{5} = 0.6 \\ x = 1 - y = 0.4 \end{cases} \Rightarrow S = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$$

Note-se que esta é a *matriz de steady-state* que atrás pareceu de facto ser a convergência da sucessão de estados

8.3 Cadeias de Markov

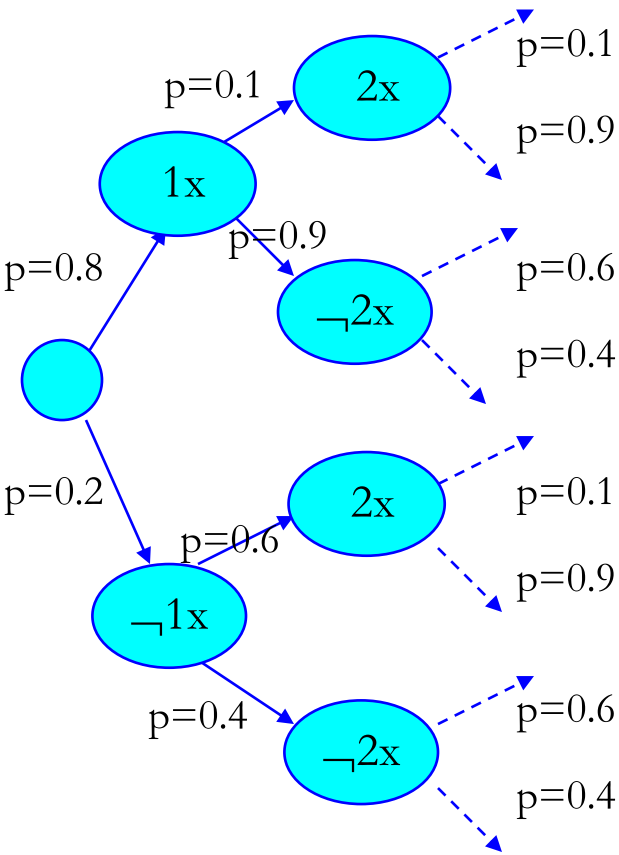
Os processos estocásticos podem também ser representados por:

- **Diagramas de Estado**

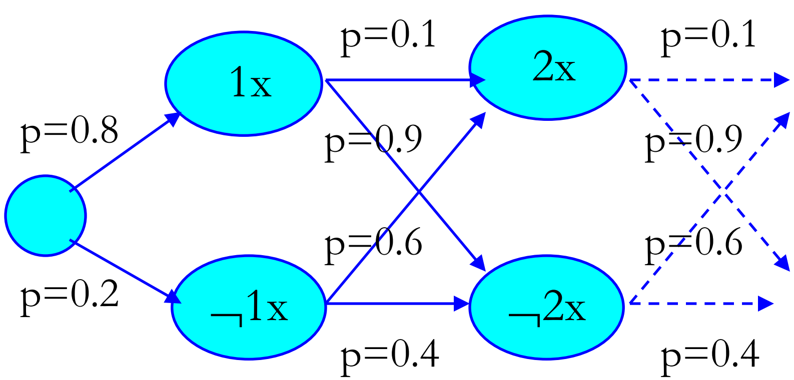


$$T = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

- **Árvores**



- ***Lattices***



8.4 Teoria de Dempster-Shaffer

8.4 Teoria de Dempster-Shaffer

Dado um acontecimento aleatório, segue uma interpretação diferente daquela em que se baseia a Teoria da Probabilidade

- Define uma função Crença (*Believe*) e uma função Dúvida (*Doubt*) acerca da veracidade de uma proposição A, representadas por ***Bel(A)*** e ***D(A)=Bel(¬A)***.
- Em vez de calcular *p(A)* e *p(¬A)* dá como resultado um intervalo chamado ***Intervalo de Confiança = [Bel(A), 1-D(A)]***

Exemplos de interpretação dos resultados:

- [0,1] Representa nenhuma crença na veracidade de uma proposição (não há qualquer evidência que permita crer num dado resultado)
- [0,0] Representa a crença de que uma proposição é falsa
- [1,1] Representa a crença de que uma proposição é verdadeira
- [0.3,1] Representa uma crença parcial (entre 0,3 e 1) numa determinada proposição

8.4 Teoria de Dempster-Shafer

Exemplo

Suponhamos que alguém deseja apostar que *o lançamento de uma moeda dará caras* (**proposição A**). A teoria de Dempster-Shafer diz que, como não há qualquer evidência acerca da veracidade do acontecimento (p.e. não se sabe se a moeda estará viciada...) o grau de crença na proposição deve ser 0: **$Bel(A)=0$** ; e o grau de descrença também deve ser 0: **$D(A)=Bel(\neg A)$**. Portanto:

Confiança em "vai sair caras" = **$[Bel(A), 1 - D(A)] = [0,1 - 0] = [0,1]$**
(significa nenhuma crença na veracidade de "vai sair caras")

Suponhamos agora que um perito afiança que com 90% de probabilidade que a moeda não é viciada. Neste caso

$Bel(A) = 0.9 \times 0.5 = 0.45$

$D(A) = Bel(\neg A) = 0.9 \times 0.5 = 0.45$

Confiança em *vai sair caras* = **$[Bel(A), 1 - D(A)] = [0.45, 1-0.45] = [0.45, 0.55]$**

(significa uma crença parcial, entre 0.45 e 0.55, na veracidade de *vai sair caras*)

8.4 Teoria de Dempster-Shaffer

Em suma:

- A largura do intervalo de confiança pode ser interpretada como uma medida acerca da necessidade de adquirir ou não mais evidencias que permitam crer ou não crer num determinado resultado (desfecho) em consideração (p.e. a evidência "moeda justa" alterou o intervalo de confiança de *"sair caras"* de [0,1] para [0.45, 0.55])

Características:

- É apelativa, porque parece traduzir melhor certos factos reais (p.e. na ausência de qualquer prova, a probabilidade de haver ou não haver petróleo no subsolo do ISEC é de 50% (=1/n pelo Princípio da Indiferença)! Já a Dempster-Shaffer dará [0,1] significando que na ausência de uma evidência, nada se pode dizer)
- Em vez de calcular probabilidades, a teoria diferencia incerteza e ignorância (quando *nada se sabe*, existe ignorância e nada se pode dizer; à medida que se *vai sabendo mais*, a incerteza acerca de um certo desfecho vai diminuindo...)
- Contudo pode não fornecer suporte para decisão onde a probabilidade o oferece
- Parte de acontecimentos exclusivos (p.e. cara ou coroa) o que na realidade geralmente não é aplicável

Conhecimento e Raciocínio

2010/2011 - LEIS (Ramos)

FIM

Escola de Atenas (Grécia)

Viriato A.P. Marinho Marques
DEIS - ISEC
2010 / 2011