Incerteza

Ano Lectivo 2010/2011- 1^o Semestre – Adaptado de http://aima.eecs.berkeley.edu/slides-tex/

Resumo

- Incerteza
- Probabilidade
- Sintaxe e semântica
- Inferência
- Independência e regra de Bayes

Incerteza

Seja a acção $A_t =$ sair para o aeroporto t minutos antes do vôo
Será que A_t me permite chegar a horas?

Problemas:

1) observações parciais (condições da estrada, outros condutores, etc.)
2) sensores com ruído (notícias de trânsito)
3) incerteza no resultado das acções (pneu vazio, etc.)
4) complexidade imensa da modelação e predição do trânsito

Logo numa aproximação lógica pura, das duas uma
1) arrisca-se a ser falsa: "A_{25} permite-me chegar a horas"
ou 2) leva a conclusões que são demasiado fracas para a tomada de decisões:
"A_{25} permitir-me-á chegar a horas se não existir acidente na ponte
e se não chover e se os pneus do meu carro não se furarem etc etc."

(A_{1440} pode garantir-me com alguma certeza que chegarei a tempo
mas terei de ficar durante a noite no aeroporto ...)

Métodos para lidar com incerteza

Lógicas por omissão ou não-monónotas:

Assumir que o meu carro não tem um pneu vazio

Assumir que A_{25} funciona a não ser que seja contradito pela evidência

Problemas: Quais as assunções razoáveis? Como lidar com a contradição?

Regras com factores inventados:

$A_{25} \mapsto_{0.3}$ chegar a tempo

$Sprinkler \mapsto_{0.99} WetGrass$

$WetGrass \mapsto_{0.7} Rain$

Problemas: como se combinam os factores, e.g., $Sprinkler$ provoca $Rain$??

Probabilidade

Dada a evidência existente,

A_{25} permitir-me-á chegar a horas com probabilidade 0.04

Mahaviracarya (9th C.), Cardamo (1565) teoria dos jogos de azar

(Lógica difusa trata *graus de verdade* NÃO incerteza e.g.,

$WetGrass$ é verdade com grau de 0.2)

Probabilidade

Asserções probabilísticas *sumariam* efeitos de

preguiça: incapacidade de enumerar excepções, qualificações, etc.
ignorância: desconhecimento de factos relevantes, condições iniciais, etc.

Probabilidades subjectivas ou Bayesianas:
Probabilidades relacionam proposições com o estado de conhecimento do agente, e.g., $P(A_{25}|\text{não há notícias de acidentes}) = 0.06$

Não são afirmações acerca da tendência probabilística na situação actual (mas podem ser aprendidas a partir de experiência do passado de situações semelhantes)

As probabilidades das proposições alteram-se com nova evidência:

e.g., $P(A_{25}|\text{não há notícias de acidentes, 5 a.m.}) = 0.15$

(Análogo ao estado de consequências lógicas em $KB \models \alpha$, não à noção verdade.)

Decidindo com incerteza

Suponha-se que eu acredito no seguinte:

$$
\begin{aligned}
P(A_{25} \text{ permite-me chegar a tempo}|\ldots) &= 0.04 \\
P(A_{90} \text{ permite-me chegar a tempo}|\ldots) &= 0.70 \\
P(A_{120} \text{ permite-me chegar a tempo}|\ldots) &= 0.95 \\
P(A_{1440} \text{ permite-me chegar a tempo}|\ldots) &= 0.9999
\end{aligned}
$$

Qual acção a escolher?

Depende das minhas preferências relativas a perder vôos vs. gastronomia aeroportuária, etc.

Teoria da utilidade é utilizada para representar e inferir preferências

Teoria da decisão $=$ teoria da utilidade $+$ teoria da probabilidade

Fundamentos de probabilidades

Comece-se com um conjunto Ω—o *espaço amostra*

$\omega \in \Omega$ é um ponto amostra/mundo possível/acontecimento atómico

Um acontecimento atómico é uma especificação completa do estado do mundo acerca do qual o agente tem incerteza. Os acontecimentos atómicos são mutuamente exclusivos e exaustivos (e.g. 6 lançamentos de um dado).

Um *espaço de probabilidade* ou *modelo de probabilidade* é um espaço amostra com uma atribuição $P(\omega)$ para todo o $\omega \in \Omega$ tal que

$0 \leq P(\omega) \leq 1$

$\sum_{\omega} P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

Um *acontecimento* A é qualquer subconjunto de Ω

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

E.g., $P(\text{lançamento do dado} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

Variáveis aleatórias

Uma *variável aleatória* é uma função de pontos amostra para algum contra-domínio. Os valores devem ser exaustivos e exclusivos.

O espaço de probabilidade P induz uma *distribuição de probabilidade* para qualquer v.a. X:

$$P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

e.g., $P(Par = true) = 1/6 + 1/6 + 1/6 = 1/2$

Nas aplicações de IA, os pontos amostra são normalmente *definidos* pelos valores de um conjunto de variáveis aleatórias, i.e., o espaço amostra é o produto Cartesiano dos contra-domínios das variáveis.

Sejam as v.as. $Dado \in \{1, 2\}$ e $Pintas \in \{1, 2, 3, 4, 5, 6\}$ os pontos amostra são (atenção isto não é lançamento simultâneo dos dados!):

$$< 1, 1 > \; < 1, 2 > \; < 1, 3 > \; < 1, 4 > \; < 1, 5 > \; < 1, 6 >$$
$$< 2, 1 > \; < 2, 2 > \; < 2, 3 > \; < 2, 4 > \; < 2, 5 > \; < 2, 6 >$$

Proposições

Entende-se uma proposição como sendo o acontecimento (conjunto de pontos amostra) onde a proposição é verdadeira

Dadas as variáveis aleatórias Booleanas A e B:

acontecimento a = conjunto de pontos amostra onde $A(\omega) = true$
acontecimento $\neg a$ = conjunto de pontos amostra onde $A(\omega) = false$
acontecimento $a \wedge b$ = pontos onde $A(\omega) = true$ e $B(\omega) = true$
acontecimento $a \vee b$ = pontos onde $A(\omega) = true$ ou $B(\omega) = true$

Com variáveis Booleanas, ponto amostra = mundo de lógica proposicional

e.g., $A = true$, $B = false$, ou $a \wedge \neg b$.

Proposição = disjunção de acontecimentos atómicos onde é verdadeira

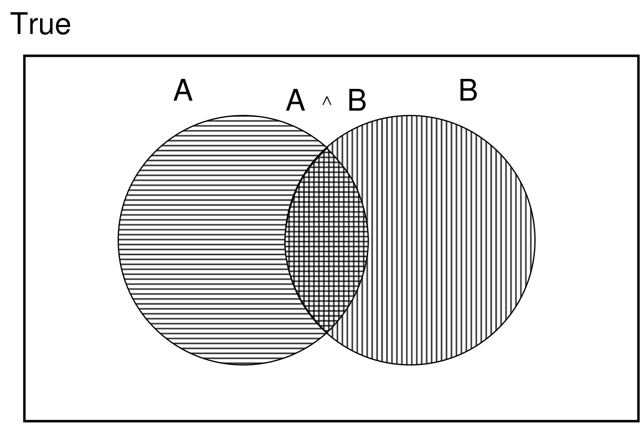
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Porquê utilizar probabilidades?

As definições implicam que certos acontecimentos lógicos relacionados devem ter probabilidades relacionadas

E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



de Finetti (1931): um agente que aposte com probabilidades que violem os axiomas da teoria das probabilidades pode ser forçado a apostar de maneira a perder dinheiro independentemente do resultado.

Sintaxe para as proposições

O elemento básico é a variável aleatória cujos valores devem ser exaustivos e mutuamente exclusivos.

◇ Variáveis aleatórias proposicionais ou Booleanas
$Cavity = true$ ou $Cavity = false$ (tenho/não tenho uma cárie?)
Estas proposições podem ser representadas por $cavity$ ou $\neg cavity$, resp.

◇ Variáveis aleatórias discretas (*finitas* ou *infinitas*)
V.a. $Weather$ toma valores em $< sunny, rainy, cloudy, snow >$
$Weather = rainy$ é uma proposição

◇ Variáveis aleatórias contínuas (*limitadas* ou *ilimitadas*)
e.g., $Temp = 21.6$; também permite, e.g. $Temp < 22.0$.

Combinações Booleanas arbitrárias de proposições básicas, por exemplo, $Cavity = false \vee Weather = sunny$.

Probabilidade a priori

Probabilidades a priori ou probabilidades incondicionais de proposições

e.g., $P(Cavity = true) = 0.1$ e $P(Weather = sunny) = 0.72$

corresponde à crença do agente antes da chegada de (nova) evidência

Distribuição de probabilidade dá valores a todas as atribuições possíveis:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*normalizado*, i.e., soma 1)

Distribuição de probabilidade conjunta para um conjunto de v.a.s. dá-nos a probabilidade de cada acontecimento atómico nessas v.a.s. (i.e., em cada ponto amostra)

$\mathbf{P}(Weather, Cavity) =$ uma matriz de 4×2 valores:

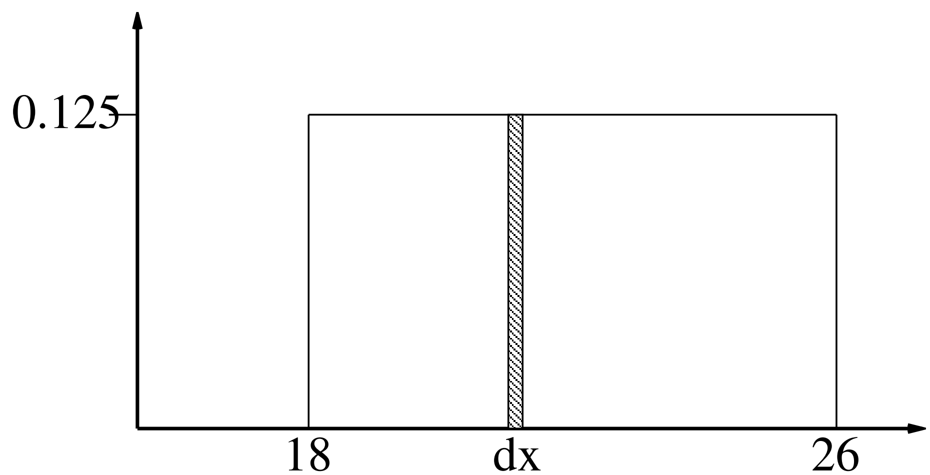
$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Toda a questão num domínio pode ser respondida pela distribuição conjunta porque todo o acontecimento é uma soma de pontos amostra

Probabilidade para variáveis contínuas

Exprime distribuição como uma função parametrizada de valor:

$P(X = x) = U[18, 26](x) =$ densidade uniforme entre 18 e 26

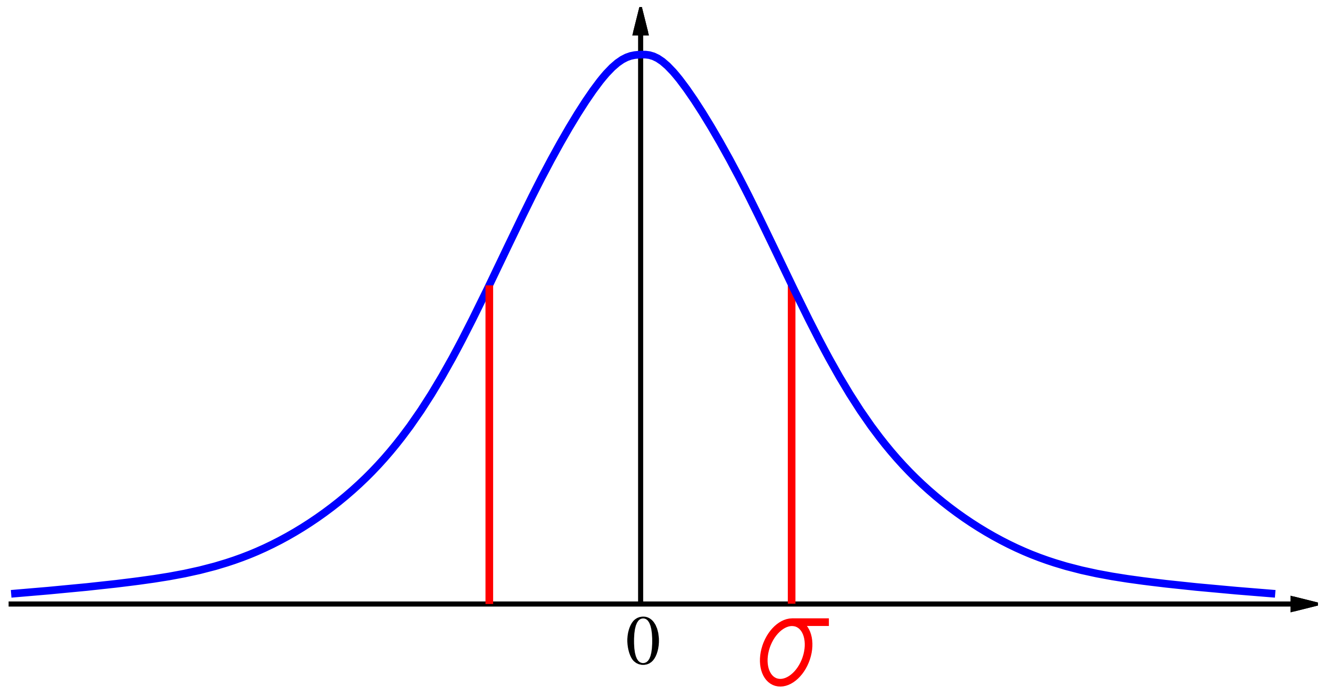


Em que P é uma *densidade* de probabilidade; integrando em todo o domínio dá 1.
$P(X = 20.5) = 0.125$ realmente significa

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

Função de densidade Gaussiana ou Normal

$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$



Probabilidade Condicional

Probabilidades a posteriori ou condicionais

e.g., $P(cavity|toothache) = 0.8$

i.e., $toothache$ é tudo o que sei

NÃO "se $toothache$ então 80% de hipótese de $cavity$"

Notação para distribuições condicionais:

$\mathbf{P}(Cavity|Toothache) =$ vector com 2 coordenadas com vector de 2 coordenadas

Se soubermos mais, e.g., $cavity$ também é dado, então temos

$P(cavity|toothache, cavity) = 1$

Nota: a crença menos específica *continua válida* após mais evidência chegar, mas nem sempre é *útil*

Nova evidência pode ser irrelevante, permitindo simplificação, e.g.,

$P(cavity|toothache, portoGanha) = P(cavity|toothache) = 0.8$

Esta forma de inferência, sancionada pelo conhecimento do domínio, é crucial

Probabilidade condicional

Definição de probabilidade condicional:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ se } P(b) \neq 0$$

Regra do Produto é uma formulação alternativa:

$$P(a \wedge b) = P(a|b)P(b) = p(b|a)P(a)$$

Uma versão genérica verifica-se para todas as distribuições, e.g.,

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

(Vista como um conjunto de 4×2 equações, *não* mult. de matrizes)

Regra da cadeia é derivada por aplicação sucessiva da regra do produto:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Inferência por enumeração

Comece-se com a distribuição conjunta:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Para qualquer proposição ϕ, somar os acontecimentos atómicos onde ϕ é verdade:
$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

Calcular $P(toothache)$ (probabilidade marginal de $toothache$)?

Inferência por enumeração

Comece-se com a distribuição conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Para qualquer proposição ϕ, somar os acontecimentos atómicos onde ϕ é verdade:
$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$

A probabilidade marginal de $toothace$ é:

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Calcular $P(cavity \vee toothache)$?

Inferência por enumeração

Comece-se com a distribuição conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Para qualquer proposição ϕ, somar os acontecimentos atómicos onde ϕ é verdade:
$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$

$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

E $P(\neg cavity | toothache)$?

Inferência por enumeração

Comece-se com a distribuição conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Também se podem calcular probabilidades condicionais:

$$\begin{aligned} P(\neg cavity | toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 \end{aligned}$$

Normalização

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	$\boxed{.108}$	$\boxed{.012}$	**.072**	**.008**
¬ cavity	$\boxed{.016}$	$\boxed{.064}$	**.144**	**.576**

Denominador pode ser visto como *constante de normalização* α

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)\\
&\quad= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]\\
&\quad= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]\\
&\quad= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

Ideia geral: calcular distribuição na variável interrogada
fixando as variáveis evidência e somando varáveis ocultas

Marginalização (método genérico)

Para quaisquer conjuntos de variáveis $\mathbf{Y}$, $\mathbf{Z}$ tem-se

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$

ou através de condicionalização

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}|\mathbf{z}) P(\mathbf{z})$$

Probabilidades condicionais (genérico)

Habitualmente, estamos interessados na
distribuição conjunta a posteriori das variáveis interrogadas $\mathbf{Y}$
dados valores específicos $\mathbf{e}$ para as variáveis evidência $\mathbf{E}$

Sejam as variáveis ocultas dadas por $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Logo o somatório das entradas conjuntas é obtida somando as variáveis ocultas:

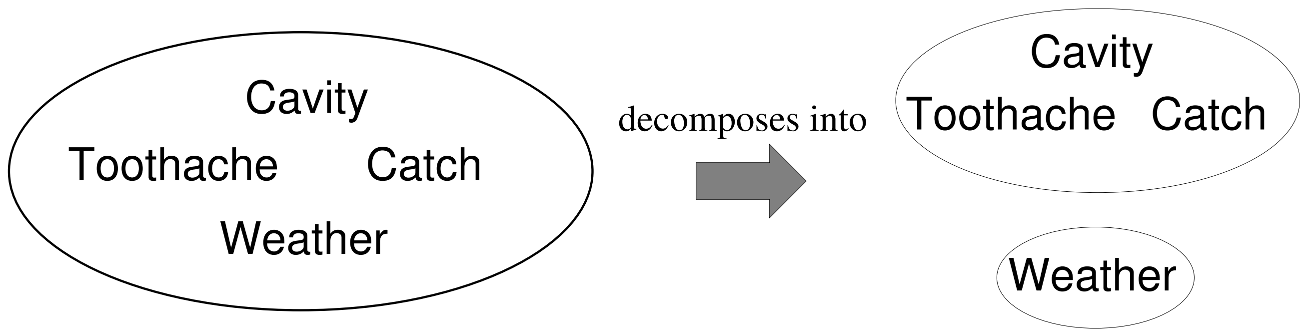
$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\sum_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

Os termos no somatório são as entradas da distribuição conjunta porque $\mathbf{Y}$, $\mathbf{E}$, e $\mathbf{H}$ é o conjunto de todas as variáveis aleatórias

Problemas óbvios:
1) Complexidade temporal pior caso $O(d^n)$ em que d é a maior aridade
2) Complexidade espacial $O(d^n)$ para armazenar a distribuição conjunta
3) Como encontrar os valores para $O(d^n)$ entradas???

Independência

A e B são independentes sse
$\mathbf{P}(A|B) = \mathbf{P}(A)$ ou $\mathbf{P}(B|A) = \mathbf{P}(B)$ ou $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entradas reduzidas a 12; para n moedas, $2^n \rightarrow n$

Independência absoluta poderosa mas rara

Medicina Dentária é um área do conhecimento com centenas de variáveis, nenhumas delas independentes. O que fazer?

Independência condicional

$\mathbf{P}(Toothache, Cavity, Catch)$ tem $2^3 - 1 =$ 7 entradas independentes

Se eu tenho uma cárie, a probabilidade que a broca toque no nervo é independente do facto de eu ter dor dentes:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

A mesma relação de independência se verifica se eu não tiver uma cárie:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ é *condicionalmente independente* de $Toothache$ dado $Cavity$:

$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Afirmações Equivalentes:

$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$

$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

Independência condicional (numericamente)

Sejam X, Y e Z três conjuntos de variáveis aleatórias.

Diz-se que X é condicionalmente independente de Y dado Z, quando para todo o acontecimento $Z = z$ com $P(Z = z) > 0$ e qualquer par de acontecimentos $X = x$ e $Y = y$ e z se tem:

$$P(X = x, Y = y, Z = z) = \frac{P(X = x, Z = z) \times P(Y = y, Z = z)}{P(Z = z)}$$

Independência condicional (cont)

Escrever a distribuição conjunta utilizando a regra da cadeia:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

I.e., $2 + 2 + 1 = 5$ valores independentes (equações 1 e 2 removem 2)

Na maioria das situações, a utilização da independência condicional reduz o tamanho da representação da distribuição conjunta de exponencial em n para linear em n.

A independência condicional é a mais básica e robusta forma de conhecimento acerca de ambientes incertos.

Desafio

Considere-se a seguinte situação real:

> 1% das mulheres com quarenta anos que efectuam regularmente rastreio têm cancro da mama. Cerca de 80% dessas mulheres obterão uma mamografia positiva. 9.6% das mulheres sem cancro da mama também têm mamografia positiva.

Uma mulher neste grupo etário teve uma mamografia positiva no rastreio regular. Qual é a probabilidade de ela ter cancro da mama?

(P.S: Só cerca de 15% dos médicos se aproxima da resposta correcta...)

Regra de Bayes

Regra do produto $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$\Rightarrow$ Regra de Bayes $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

ou na forma de distribuição

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$$

Variante condicionalizada em que evidência $\mathbf{e}$ fornecida:

$$\mathbf{P}(Y|X, \mathbf{e}) = \frac{\mathbf{P}(X|Y, \mathbf{e})\mathbf{P}(Y|\mathbf{e})}{\mathbf{P}(X|\mathbf{e})} = \alpha\mathbf{P}(X|Y, \mathbf{e})\mathbf{P}(Y|\mathbf{e})$$

Útil para obter a probabilidade do diagnóstico dada a probabilidade causal:

$$P(Causa|Efeito) = \frac{P(Efeito|Causa)P(Causa)}{P(Efeito)}$$

Exemplo de aplicação da regra de Bayes

Considere o seguinte conhecimento médico:

◇ Meningite provoca rigidez do pescoço em 70% dos casos

◇ A probabilidade a priori de um paciente ter meningite é de 1 em 50000.

◇ A probabilidade a priori de um paciente ter o pescoço rígido é 0.01.

E.g., seja M meningite, R pescoço rígido:

$$P(m|r) = \frac{P(r|m)P(m)}{P(r)} = \frac{0.7 \times 0.00002}{0.01} = 0.0014 \approx \text{ (1 em 7000)}$$

Nota: a probabilidade a posteriori de meningite é ainda muito pequena!

Já é capaz de responder ao desafio anterior? (Resposta: 7.7%!)

Regra de Bayes e independência condicional

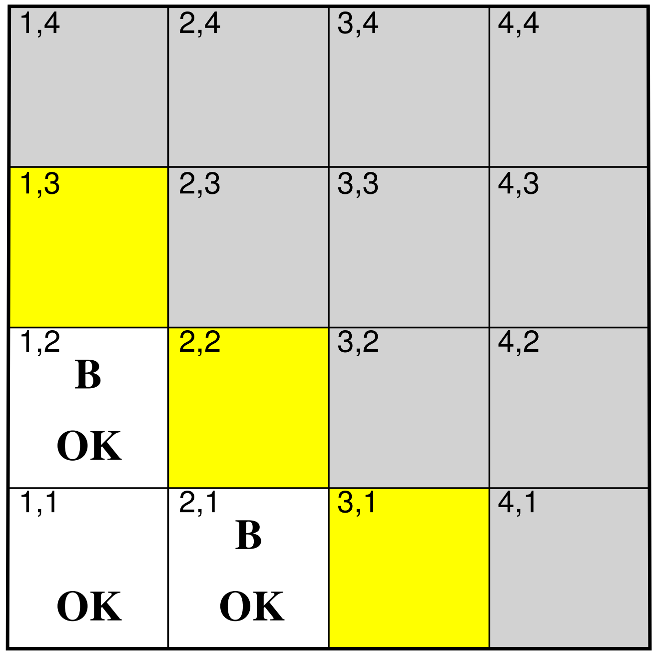
$$\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch) \\
&= \alpha\, \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\
&= \alpha\, \mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

Isto é um exemplo do modelo de *naife de Bayes*:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i|Cause)$$

O número total de parâmetros é *linear* em n

Mundo do Wumpus



$P_{ij} = true$ sse $[i, j]$ contém um poço

$B_{ij} = true$ sse $[i, j]$ é ventosa
Incluir apenas $B_{1,1}, B_{1,2}, B_{2,1}$ no modelo de probabilidade

Especificando o modelo de probabilidade

A distribuição conjunta total é $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Aplicando a regra do produto: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

(Faz-se assim para obter $P(Efeito|Causa)$.)

Primeiro termo: 1 se poços adjacentes a brisa, 0 caso contrário

Segundo termo: poços são dispostos aleatoriamente, probabilidade de 0.2 por casa:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} \mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

para n poços.

Observações e interrogação

Sabemos o seguinte:

$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$

$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$

Interrogação é $\mathbf{P}(P_{1,3}|known, b)$

Seja $Unknown = P_{ij}$s diferentes de $P_{1,3}$ e $Known$

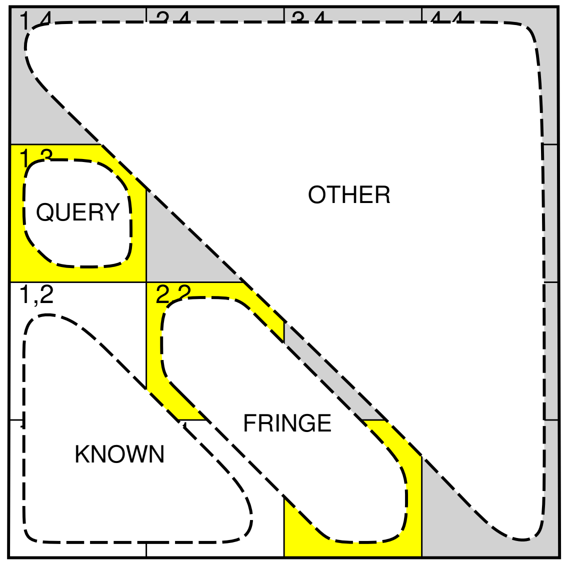
Por inferência por enumeração, obtemos

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Cresce exponencialmente com o número de casas!

Utilizando a independência condicional

Facto fundamental: observações são condicionalmente independentes de outras casas escondidas dadas as casas escondidas vizinhas



Definir $Unknown = Fringe \cup Other$
$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

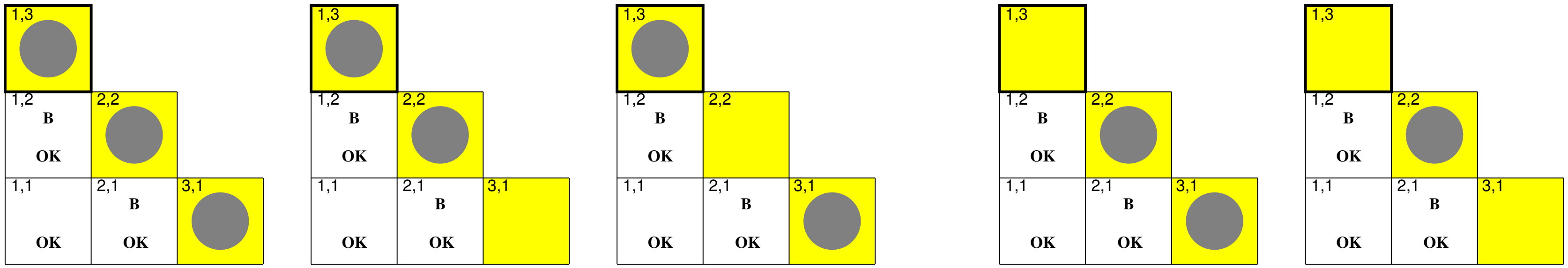
Manipular interrogação de maneira a que se possa utilizar a igualdade anterior!

Utilizando a independência condicional (cont.)

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
= \ & \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown) \\
= \ & \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
= \ & \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
= \ & \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
= \ & \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
= \ & \alpha \, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other)
\end{aligned}
$$

$$= \alpha' \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) P(fringe)$$

Utilizando a independência condicional (cont.)



0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16 | 0.8 x 0.2 = 0.16 | 0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha' \langle 0.2(0.04 + 0.16 + 0.16),\ \ 0.8(0.04 + 0.16) \rangle \\
&= \alpha' \langle 0.072, 0.16 \rangle \\
&= \langle \frac{0.072}{0.072 + 0.16}, \frac{0.16}{0.072 + 0.16} \rangle \\
&\approx \langle 0.31, 0.69 \rangle
\end{aligned}
$$

$$
\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14 \rangle
$$

Casos de Aplicação: Urnas

Considerem-se duas urnas, a urna A uma com bolas numeradas de 1 a 10, e a urna B com bolas numeradas de 1 a 1000.

- Escolha-se uma das urnas aleatoriamente
- Retire-se dela uma bola aleatoriamente
- A bola que saiu tem o número 4,
- Qual a probabilidade da bola ter sido retirada da urna A?

Solução para o problema das urnas

Pretendemos obter a probabilidade de

$$
\begin{aligned}
P(Urna = A|Bola = 4) &= \frac{P(Bola{=}4|Urna{=}A)\times P(Urna{=}A)}{P(Bola{=}4)} \\
&= \frac{P(Bola{=}4|Urna{=}A)\times P(Urna{=}A)}{\sum_{u\in\{A,B\}} P(Bola{=}4\wedge Urna{=}u)} \\
&= \frac{P(Bola{=}4|Urna{=}A)\times P(Urna{=}A)}{\sum_{u\in\{A,B\}} P(Bola{=}4|Urna{=}u)\times P(Urna{=}u)} \\
&= \frac{0.1\times 0.5}{0.1\times 0.5+0.001\times 0.5} = \frac{0.1}{0.1+0.001} = \frac{0.1}{0.101} \approx 0.99
\end{aligned}
$$

Outra maneira (não há variáveis escondidas)

$$
\begin{aligned}
P(Urna|Bola = 4) &= \alpha P(Urna \wedge Bola = 4) = \alpha P(Bola = 4|Urna) \times P(Urna) \\
&= \alpha \langle 0.1 \times 0.5, 0.001 \times 0.5\rangle = \alpha \langle 0.05, 0.0005\rangle \\
&=< \frac{0.05}{0.05+0.0005}, \frac{0.0005}{0.05+0.0005} >\approx< 0.99, 0.01 >
\end{aligned}
$$

Casos de Aplicação: Argumento do Juízo Final

Prólogo:

- Existem 100 cubículos numerados de 1 a 100.
- Os números estão pintados do lado de fora.
- É lançada uma moeda ao ar (não viciada).
- Se sair caras, então uma pessoa é colocada dentro de cada um dos cubículos.
- Se sair coroas, então uma pessoa é colocada dentro de cada um dos dez cubículos.
- Você encontra-se dentro de um dos cubículos.

Casos de Aplicação: Argumento do Juízo Final

Questões:

1. Qual é a probabilidade de ter saído caras?
2. Existem 10 ou 100 pessoas nos cubículos?
3. Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saíram caras?
4. Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saíram coroas?
5. Suponha que abre a porta e tem o número 7 escrito na porta. Qual é a probabalidade de terem saída coroas?

Casos de Aplicação: Argumento do Juízo Final

Questões:

1. Qual é a probabilidade de ter saído caras? $0.5\surd$.
2. Existem 10 ou 100 pessoas nos cubículos? $0.5\surd$ de hipóteses.
3. Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saíram caras? $0.1\surd$
4. Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saíram coroas? $1.0\surd$
5. Suponha que abre a porta e tem o número 7 escrito na porta. Qual é a probabalidade de terem saída coroas? $0.91\surd$

Casos de Aplicação: Argumento do Juízo Final

Consideremos agora outra situação "real", com as duas hipóteses rivais:

1. A raça humana extringuir-se-á no próximo século, com 200 mil milhões (200×10^9) o número total de humanos que nasceram.
2. A raça humana não se extinguirá no próximo século e colonizará a galáxia, sendo o número total de humanos nascidos 200 biliões (200×10^{12}).

Vamos supor que a probabilidade da hipótese 1 ocorrer a priori é de X%.

Suponha-se que se sabe que você é o humano número 60 mil milhões (60×10^9) (aproximadamente o número de pessoas que nasceram até hoje). Qual a probabilidade da raça humana se extinguir no próximo século dada esta informação?

Casos de Aplicação: Argumento do Juízo Final

A estocada final:

$$P(Hip = 1|\text{ você ser a pessoa } 60 \times 10^9) =$$

$$= \frac{P(\text{ você ser a pessoa } 60 \times 10^9 | Hip = 1) \times P(Hip = 1)}{P(\text{ você ser a pessoa } 60 \times 10^9)}$$

$$= \frac{\frac{1}{200 \times 10^9} X}{\frac{1}{200 \times 10^9} X + \frac{1}{200 \times 10^{12}}(1 - X)}$$

$$= \frac{X}{X + \frac{1-X}{1000}} = \frac{1000X}{999X + 1}$$

Casos de Aplicação: Argumento do Juízo Final

Alguns valores para X:

X	$P($ Extinção no próximo século | você ser a pessoa $60 \times 10^9)$
0.1%	50%
1%	90%
2%	95%
5%	98%
10%	99%

ATENÇÃO: Este argumento é alvo de grande discordância e debate filosófico intenso! Vejam a Wikipedia!

Casos de Aplicação: Monty Hall Problem

- Um apresentador honesto de um programa televisivo colocou um Ferrari atrás de uma de três portas numeradas.
- Existe uma cabra atrás das restantes duas portas.
- Um concorrente não tem qualquer informação que o permita escolher entre qualquer uma das portas.
- O apresentador informa-o das regras do jogo

  Primeiro aponte para uma porta. Depois, abrirei uma das outras portas que tem a cabra. Depois de eu lhe mostrar a cabra, terá de tomar a decisão final se mantém a hipótese inicial ou muda para a outra porta. Ganhará aquilo que estiver detrás da porta seleccionada no fim.

- Você começa para apontar para a porta número 1.
- O apresentador abre-lhe a porta 3, que tem uma cabra escondida.
- Qual a sua decisão?

Solução do Monty Hall Problem

Considerem-s as proposições:

C_i – carro encontra-se detrás da porta $i \in \{1, 2, 3\}$

A_{ij} – o apresentador abre a porta j depois do concorrente escolher a porta i com $i, j \in \{1, 2, 3\}$.

Vamos determinar $P(C_1|A_{13}) =$ recorrendo à Regra de Bayes:

$$= \frac{P(C_1 \wedge A_{13})}{P(A_{13})} = \frac{P(A_{13}|C_1) \times P(C_1)}{P(A_{13})}$$

Para o numerador temos:

$$P(A_{13}|C_1) \times P(C_1) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

O cálculo do denominador é um pouco mais difícil...

Solução do Monty Hall Problem (cont.)

Para o denominador podemos reescrever

$$
\begin{aligned}
P(A_{13}) &= \textstyle\sum_{i\in 1,2,3} P(A_{13}|C_i) \times P(C_i) \\
&= P(A_{13}|C_1) \times P(C_1) + P(A_{13}|C_2) \times P(C_2) + P(A_{13}|C_3) \times P(C_3) \\
&= \tfrac{1}{2} \times \tfrac{1}{3} + 1 \times \tfrac{1}{3} + 0 \times \tfrac{1}{3} \\
&= \tfrac{1}{2}
\end{aligned}
$$

Portanto $P(C_1|A_{13}) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$!

Logo, a decisão **RACIONAL** é trocar de porta! Pois a probabilidade de se encontrar aí o carro é de $\frac{2}{3}$.

Aplicações reais da Regra de Bayes

♢ Teoria da pesquisa Bayesiana
Utilizada pela Marinha dos EUA para localizar objectos perdidos no mar, particularmente uma bomba de Hidrogénio e um submarino...

♢ Classificadores de Bayes, nomeadamente para classificar documentos (e.g. detectar SPAM). Baseia-se no princípio:

$$Pr(spam|words) = \frac{Pr(words|spam) \times P(spam)}{P(words)}$$

♢ Diagnóstico médico

♢ Predição de tráfego

♢ Office Assistant do Microsoft Word

Sumário

A teoria das probabilidades é um formalismo rigoroso para lidar com conhecimento incerto

Distribuição de probabilidade conjunta especifica a probabilidade de todo o acontecimento atómico

As interrogações podem ser respondidas somando-se acontecimentos atómicos

Para domínios não triviais, é essencial reduzir o tamanho da distribuição conjunta

Independência e independência condicional fornecem as ferramentas