

INCERTEZA

CAP 13

Parcialmente adaptado de
<http://aima.eecs.berkeley.edu>

Resumo

- Incerteza
- Probabilidade
- Sintaxe e Semântica
- Inferência
- Independência e Regra de Bayes

Incerteza

- Seja a ação A_t = sair para o aeroporto t minutos antes do vôo.
- Será que A_t me permite chegar ao aeroporto a tempo?
- Problemas:
 - Estados parcialmente observáveis (condições da estrada, outros condutores, etc.)
 - Sensores com ruído (notícias de trânsito)
 - Incerteza quanto aos efeitos das ações (pneu furado, etc.)
 - Grande complexidade da modelação e previsão do trânsito
- Logo, numa aproximação puramente lógica, das duas uma:
 1. Arrisca-se a ser falsa a proposição “ A_{25} permite-me chegar ao aeroporto a tempo”, ou
 2. Leva a conclusões que são demasiado fracas para a tomada de decisões: “ A_{25} permitir-me-a chegar a tempo, se não houver qualquer acidente na ponte, e se não chover, e se nenhum pneu furar, etc., etc.”
- A_{1440} pode garantir-me com alguma certeza que chegarei a tempo, mas terei que pernoitar no aeroporto

Métodos para lidar com incerteza

- **Lógicas por omissão ou não-monótonas:**
 - 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
 - $\text{Aspersor} \mapsto_{0,99}$ RelvaMolhada
 - $\text{RelvaMolhada} \mapsto_{0,7}$ Chuva
- Problemas: como se combinam os factores, e.g., Aspersor provoca Chuva??

- **Probabilidade**
 - Modela o grau de crença de um agente
 - Dada a evidência existente,
 - A_5 permitir-me-á chegar a tempo com probabilidade 0.04

Probabilidade

- Asserções probabilísticas **sumariam** efeitos de
 - **preguiça**: incapacidade de enumerar exceções, qualificações, etc.
 - **ignorância**: desconhecimento de teoria sobre o domínio, factos relevantes, condições iniciais, etc.
- **Probabilidades subjetivas 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 proposições sobre o mundo** – o mundo real não é incerto.
- **São proposições sobre o estado de conhecimento/crença do agente sobre o mundo**.
 - As probabilidades das proposições alteram-se com nova evidência:
 - e.g., $P(A_{25} | \text{não há notícias de acidentes, 5 a.m.}) = 0,15$
- Compromisso ontológico da probabilidade é igual ao da lógica. O compromisso epistemológico é que é diferente: na lógica é verdadeiro/falso enquanto que na probabilidade é numérico.

Decidindo sob condições de incerteza

- Suponha o seguinte conjunto de crenças:
 - $P(A_{25} \text{ permite-me chegar a tempo} \mid \dots) = 0.04$
 - $P(A_{90} \text{ permite-me chegar a tempo} \mid \dots) = 0.70$
 - $P(A_{120} \text{ permite-me chegar a tempo} \mid \dots) = 0.95$
 - $P(A_{1440} \text{ permite-me chegar a tempo} \mid \dots) = 0.9999$
- Que ação escolher?
 - Depende de minhas **preferências** sobre perder o voo vs. tempo de espera no aeroporto vs. gastronomia aeroportuária, etc.
 - **Teoria da utilidade** é utilizada para representar e inferir preferências (todo estado tem um grau de utilidade)
 - **Teoria da Decisão** = teoria da probabilidade + teoria da utilidade

Fundamentos de Probabilidades

- Comece-se com um conjunto Ω , o **espaço amostra**
- $\omega \in \Omega$ é um **ponto amostra**/mundo possível/acontecimento atômico
- Um **acontecimento atômico** é uma especificação completa do estado do mundo acerca do qual o agente tem incerteza. Os acontecimentos atômicos 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 (evento)** 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

- Proposições descrevendo acontecimentos (eventos) (conjuntos de pontos de amostra) usam uma notação que combina a lógica proposicional e a notação de satisfação de restrições.
- As variáveis em teoria da probabilidade são designadas por **variáveis aleatórias**.
- Cada variável aleatória tem um **domínio**.
 - E.g. o domínio de *Weather* é $\{sunny, rainy, cloudy, snow\}$.
- Uma variável aleatória booleana tem o domínio $\{verdadeiro, falso\}$.
- 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 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, mas sim o lançamento de um dos dados!):
 - $\langle 1, 1 \rangle \langle 1, 2 \rangle \langle 1, 3 \rangle \langle 1, 4 \rangle \langle 1, 5 \rangle \langle 1, 6 \rangle$
 - $\langle 2, 1 \rangle \langle 2, 2 \rangle \langle 2, 3 \rangle \langle 2, 4 \rangle \langle 2, 5 \rangle \langle 2, 6 \rangle$

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 $\neg cavity$.
- Variáveis aleatórias **discretas** (**finitas** ou **infinitas**)
 - V.a. $Weather$ toma valores em $\langle sunny, rainy, cloudy, snow \rangle$
 - $Weather = rainy$ é uma proposição
 - Quando não há ambiguidade, podemos usar o valor da variável aleatória ($rainy$) em vez da proposição ($Weather = rainy$)
- Variáveis aleatórias **contínuas** (**limitadas** ou **ilimitadas**)
 - e.g., $Temp = 21.6$; também permite, e.g. $Temp < 22.0$.
- Podemos combinar proposições (usando os conectivos da lógica proposicional) para obter outras proposições, por exemplo, $Cavity = false \vee Weather = sunny$.
- Entende-se uma proposição como sendo o acontecimento (conjunto de pontos amostra) onde a proposição é verdadeira

Probabilidade incondicional ou à priori

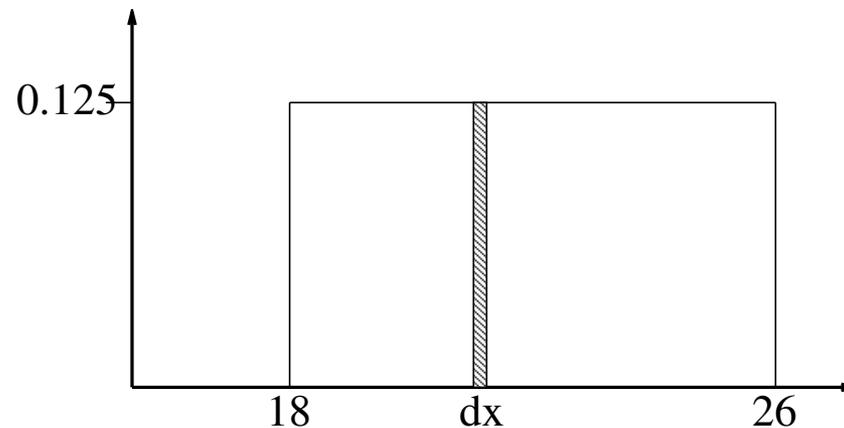
- Probabilidades à **priori** ou **probabilidades incondicionais** de proposições
 - e.g., $P(\text{Cavity} = \text{true}) = 0.1$ e $P(\text{Weather} = \text{sunny}) = 0.72$correspondem à crença do agente antes da chegada de (nova) evidência
- **Distribuição de probabilidade** dá valores a todas as atribuições possíveis:
 - $\mathbf{P}(\text{Weather}) = \langle 0.72; 0.1; 0.08; 0.1 \rangle$ (**normalizado**, i.e., soma 1)
 - Bold (em **P**) significa um vector, assumindo ordenação fixa do domínio da variável.
- **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)
 - $\mathbf{P}(\text{Weather}, \text{Cavity}) =$ 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 função parametrizada de valor:
- $P(X=x) = U[18,26](x) =$ densidade uniforme entre 18 e 26

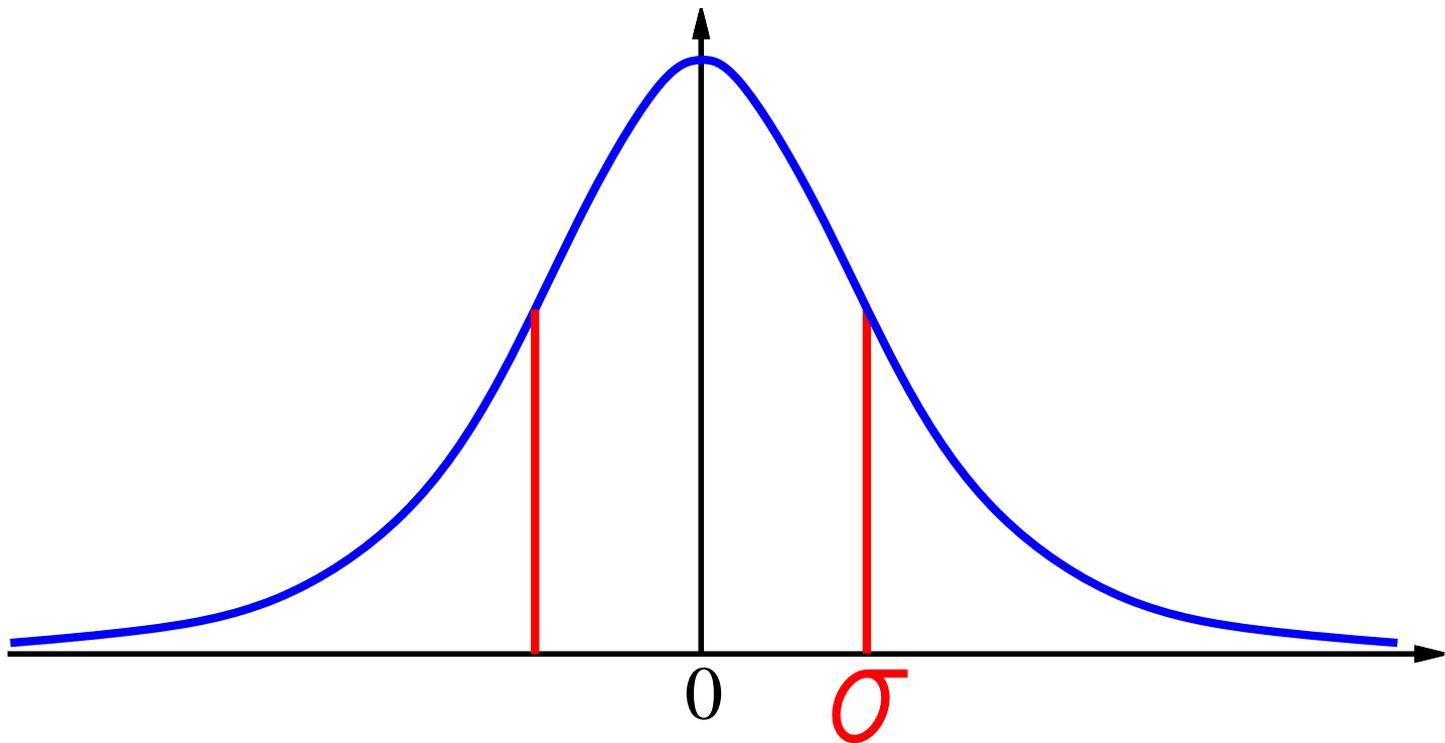


- Em que P é uma densidade de probabilidade; integrando em todo o domínio dá 1. $P(X=20.5) = 0.125$ realmente significa

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

Função de densidade Gaussiana ou Normal

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



Probabilidade condicional ou à posteriori

- Probabilidades à posteriori ou condicionais
 - e.g., $P(\text{cavity} \mid \text{toothache}) = 0.8$
 - i.e., **toothache é tudo o que sei**
- Notação para **distribuições condicionais**:
 - $\mathbf{P}(X \mid Y)$ = conjunto de valores de $P(X=x_i \mid Y=y_j)$ para cada i, j possível
- Se soubermos mais, e.g., **cavity também é dado**, então temos
 - $P(\text{cavity} \mid \text{toothache}, \text{cavity}) = 1$
- 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.,
 - $P(\text{cavity} \mid \text{toothache}, \text{AcadémicaGanha}) = P(\text{cavity} \mid \text{toothache}) = 0.8$
- Esta forma de inferência, sancionada pelo conhecimento do domínio, é crucial

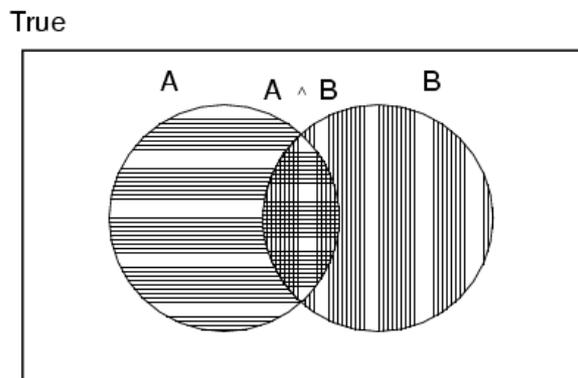
Probabilidade Condicional

- Probabilidade condicional pode ser definida em termos das probabilidades incondicionais:
 - $P(a | b) = P(a \wedge b) / P(b)$, se $P(b) \neq 0$
- **Regra do produto** é uma formulação alternativa:
 - $P(a \wedge b) = P(a | b) P(b) = P(b | a) P(a)$
- Pode ser generalizado para distribuições totais:
 - $\mathbf{P}(\textit{Weather}, \textit{Cavity}) = \mathbf{P}(\textit{Weather} | \textit{Cavity}) \mathbf{P}(\textit{Cavity})$
 - (vista como um conjunto de 4×2 equações, **não** multiplicação de matrizes)
- **Regra da cadeia** é obtida a partir de aplicações sucessivas da regra do produto:

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

Porquê usar probabilidades?

- **Axiomas da probabilidade.** Para quaisquer proposições A e B:
 - $0 \leq P(A) \leq 1$
 - $P(\text{true}) = 1$ e $P(\text{false}) = 0$
 - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



- 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.

Argumento de de Finetti

- **Preâmbulo:** se um agente tem um grau de crença numa proposição, deveria ser capaz de indicar a chance (odds) na qual lhe seria indiferente apostar a favor ou contra essa proposição.
- Por exemplo, se um agente 1 tiver um grau de crença num evento a de 0.4, um agente 2 pode escolher se pretende apostar a favor ou contra a , com valores consistentes com o grau de crença.
- O agente 2 poderia aceitar a aposta de 1 em a , oferecendo e.g. 6€ contra os 4€ do agente 1, ou apostar a favor de a , oferecendo 4€ em troca de 6€ do agente 1.
- Para o agente 1, as duas apostas seriam igualmente aceitáveis i.e. para o agente 1 é indiferente apostar a favor ou contra a com 4 para 6, ou 6 para 4, respectivamente.

Inferência por enumeração

- Comece-se com a distribuição conjunta:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Para qualquer proposição φ , somar os acontecimentos atômicos onde φ é verdade:

$$P(\varphi) = \sum_{\omega:\omega \models \varphi} P(\omega)$$

- Calcular $P(\textit{toothache})$ (probabilidade marginal de *toothache*)?

Inferência por enumeração

- Comece-se com a distribuição conjunta:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Para qualquer proposição φ , somar os acontecimentos atômicos onde φ é verdade:

$$P(\varphi) = \sum_{\omega:\omega \models \varphi} P(\omega)$$

- A probabilidade marginal de *toothache* é:

$$P(\textit{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

- Calcular $P(\textit{cavity} \vee \textit{toothache})$?

Inferência por enumeração

- Comece-se com a distribuição conjunta:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Para qualquer proposição φ , somar os acontecimentos atômicos onde φ é verdade:

$$P(\varphi) = \sum_{\omega:\omega\models\varphi} P(\omega)$$

$$P(\text{cavity} \vee \text{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.2$$

- Calcular $P(\neg\text{cavity} \mid \text{toothache})$?

Inferência por enumeração

- Comece-se com a distribuição conjunta:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Podemos calcular probabilidades condicionais:

$$P(\varphi) = \sum_{\omega:\omega \models \varphi} P(\omega)$$

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

Inferência por enumeração

- Comece-se com a distribuição conjunta:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Denominador pode ser visto como **constante de normalização** α

$$\mathbf{P}(Cavity \mid toothache) = \alpha \mathbf{P}(Cavity, toothache)$$

$$= \alpha [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$

$$= \alpha [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle]$$

$$= \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle$$

- Ideia geral: calcular distribuição na **variável interrogada**, fixando as **variáveis evidência** e somando **variáveis ocultas**

Marginalização (método genérico)

- Para quaisquer conjuntos de variáveis \mathbf{Y} , \mathbf{Z} , temos:
- $\mathbf{P}(\mathbf{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, z)$
- Ou através da condicionalização
- $\mathbf{P}(\mathbf{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}|z)P(z)$

Inferência probabilística

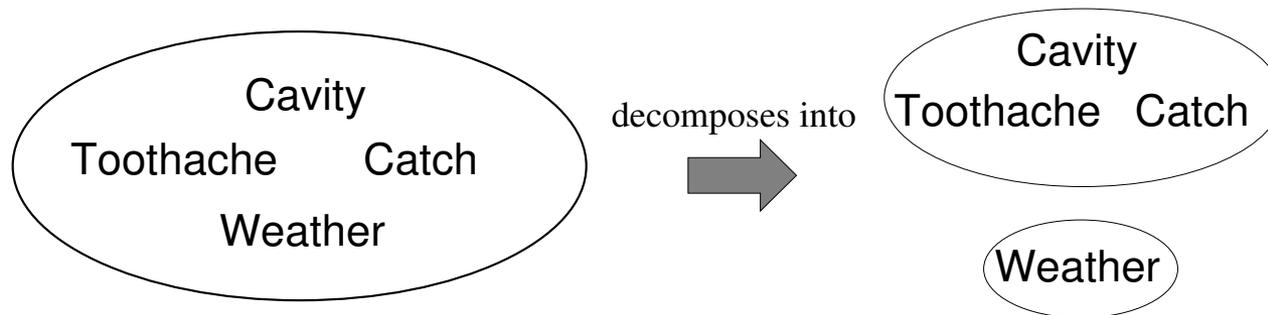
- Habitualmente, estamos interessados na
 - distribuição conjunta à posteriori das **variáveis interrogadas** \mathbf{Y}
 - dados valores específicos e para as **variáveis evidência** \mathbf{E}
- Sejam as **variáveis ocultas** dadas por $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$
- Logo, o somatório 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
- $P(A|B) = P(A)$ ou $P(B|A) = P(B)$ ou $P(A,B)=P(A)P(B)$



- $P(\text{Toothache}, \text{Catch}, \text{Cavity}, \text{Weather})$
= $P(\text{Toothache}, \text{Catch}, \text{Cavity})P(\text{Weather})$
- 32 entradas reduzidas a 12, para n moedas, $2^n \rightarrow n$
- Independência absoluta é poderosa, mas rara
- Medicina Dentária é uma área do conhecimento com centenas de variáveis, nenhuma delas independentes. O que fazer?

Independência Condicional

- $\mathbf{P}(\textit{Toothache}, \textit{Cavity}, \textit{Catch})$ tem $2^3 - 1 = 7$ entradas independentes
- Se eu tenho uma cárie, a probabilidade que a sonda fique presa (catch) é independente do facto de eu ter dor de dentes:
 1. $P(\textit{catch} \mid \textit{toothache}, \textit{cavity}) = P(\textit{catch} \mid \textit{cavity})$
- A mesma relação de independência verifica-se se eu não tiver uma cárie:
 2. $P(\textit{catch} \mid \textit{toothache}, \neg\textit{cavity}) = P(\textit{catch} \mid \neg\textit{cavity})$
- *Catch* é **condicionalmente independente** de *Toothache* dado *Cavity*:
 - $\mathbf{P}(\textit{Catch} \mid \textit{Toothache}, \textit{Cavity}) = \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$
- Afirmações Equivalentes:
 - $\mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity})$
 - $\mathbf{P}(\textit{Toothache}, \textit{Catch} \mid \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$

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 $Y=y$ se tem:

$$P(X=x, Y=y, Z=z) = P(X=x, Z=z) P(Y=y, Z=z) / P(Z=z)$$

Independência Condicional

- Escrevendo a distribuição conjunta utilizando a regra da cadeia:

$$P(\textit{Toothache}, \textit{Catch}, \textit{Cavity})$$

$$= P(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) P(\textit{Catch}, \textit{Cavity})$$

$$= P(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) P(\textit{Catch} \mid \textit{Cavity}) P(\textit{Cavity})$$

$$= P(\textit{Toothache} \mid \textit{Cavity}) P(\textit{Catch} \mid \textit{Cavity}) P(\textit{Cavity})$$

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 efetuam 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?

- (Só cerca de 15% dos médicos se aproxima da resposta correta...)

Regra de Bayes

- Da regra do produto $P(a \wedge b) = P(a|b)P(b) = P(b|a) P(a)$ obtemos a Regra de Bayes:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

- Ou na forma da distribuição conjunta:

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

- Variante condicionalizada na evidência 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

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

Regra de Bayes: exemplo

- Considere o seguinte conhecimento médico:
 - Meningite provoca rigidez do pescoço em 70% dos casos
 - A probabilidade à priori de um paciente ter meningite é de 1 em 50000.
 - A probabilidade à priori de um paciente ter o pescoço rígido é 0.01.
- Seja M meningite, R pescoço rígido:
- $P(m|r) = P(r|m)P(m)/P(r) = .7 * .00002 / .01 = .0014 \approx (1 \text{ em } 7000)$
- Nota: a probabilidade à posteriori de meningite é ainda muito pequena!
- Já é capaz de responder ao desafio anterior?

Desafio

- Considere-se a seguinte situação real:
 - 1% das mulheres com quarenta anos que efetuam 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?

- (Só cerca de 15% dos médicos se aproxima da resposta correta...)
- Seja C=“tem cancro da mama” e M=“teve uma mamografia positiva”
- Problema: $P(c)=0.01$; $P(m|c)=0.8$; $P(m|\neg c)=0.096$. $P(c|m)=?$
- Regra de Bayes: $P(c|m)=P(m|c)P(c)/P(m)$
- Marginalização: $P(m)=P(m|c)P(c)+P(m|\neg c)P(\neg c)=.8*.01+.096*.99 = .10304$
- Resolvendo: $P(c|m)=P(m|c)P(c)/P(m)=.8*.01/.10304=.0776 \approx 7.8\%$

Regra de Bayes e independência condicional

- $\mathbf{P}(Cavity|toothache \wedge catch)$
= $\alpha \mathbf{P}(toothache \wedge catch|Cavity) \mathbf{P}(Cavity)$
= $\alpha \mathbf{P}(toothache|Cavity) \mathbf{P}(catch|Cavity) \mathbf{P}(Cavity)$
- Isto é um exemplo do modelo naive de Bayes (uma causa única influencia um número de efeitos, todos independentes dada a causa).
- A distribuição conjunta total é dada por:

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



- O numero total de parâmetros é linear em n.

Mundo do Wumpus

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 B OK	2,2	3,2	4,2
1,1 OK	2,1 B OK	3,1	4,1

- $P_{ij}=true$ 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

Modelo de Probabilidade

- A distribuição conjunta total é

$$\mathbf{P}(P_{1,1}, \dots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$$

- Aplicando a regra do produto:

- $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} | P_{1,1}, \dots, P_{4,4}) \mathbf{P}(P_{1,1}, \dots, P_{4,4})$

(Faz-se assim para obter $\mathbf{P}(\text{Efeito} | \text{Causa})$.)

- Primeiro termo: 1 se poços adjacentes a brisa, 0 caso contrario
- Segundo termo: poços são dispostos aleatoriamente, probabilidade de 0.2 por casa:

$$\mathbf{P}(P_{1,1}, \dots, P_{4,4}) = \prod_{ij=1,1..4,4} \mathbf{P}(P_{i,j})$$

- Para uma configuração com n poços,

$$P(P_{1,1}, \dots, P_{4,4}) = 0.2^n \times 0.8^{16-n}$$

Observações e interrogações

- Sabemos o seguinte:

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

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

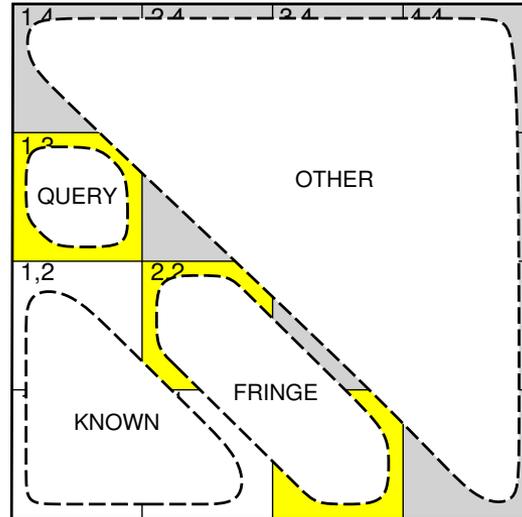
- Interrogação é $\mathbf{P}(P_{1,3}|known, b)$
- Seja $Unknown = P_{ij}$ s diferentes de $P_{1,3}$ e $Known$
- Por inferência por enumeração (marginalização), obtemos

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

- 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 $Unknown = Fringe \cup Other$

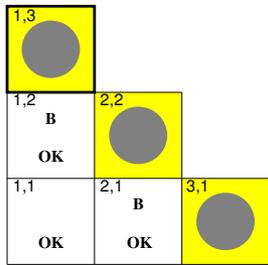
$$P(b|P_{1,3}, Known, Unknown) = P(b|P_{1,3}, Known, Fringe)$$

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

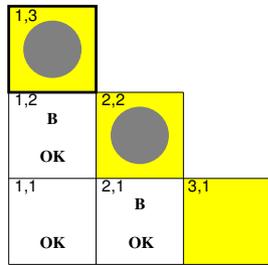
Utilizando a independência condicional

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

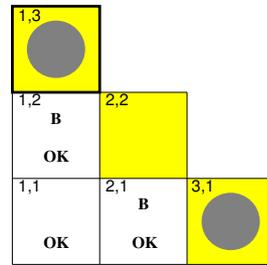
Utilizando a independência condicional



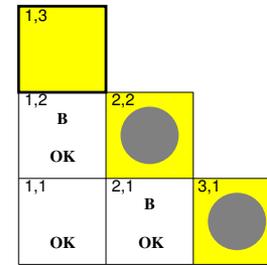
$$0.2 \times 0.2 = 0.04$$



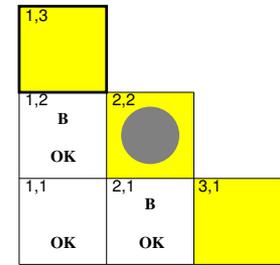
$$0.2 \times 0.8 = 0.16$$



$$0.8 \times 0.2 = 0.16$$



$$0.2 \times 0.2 = 0.04$$



$$0.2 \times 0.8 = 0.16$$

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

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

Casos de Aplicação

Casos de aplicação: Urnas

- Considerem-se duas urnas, a urna A 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 numero 4,
- Qual a probabilidade da bola ter sido retirada da urna A?

Casos de aplicação: Urnas

- Pretendemos obter a probabilidade de:

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

- Outra maneira (sem variáveis escondidas):

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

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 cara, então uma pessoa é colocada dentro de cada um dos 100 cubículos.
 - Se sair coroa, 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

- Qual é a probabilidade de ter saído cara? **0.5**
- Existem 10 ou 100 pessoas nos cubículos? **0.5**
- Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saiu cara? **0.1**
- Qual é a probabilidade de estar pintado um número de 1 a 10 no seu cubículo dado que saiu coroa? **1.0**
- Suponha que abre a porta e tem o número 7 escrito na porta. Qual é a probabilidade de ter saído coroa? **0.91**

Casos de aplicação: Argumento do Juízo Final

- Consideremos agora outra situação “real”, com as duas hipóteses rivais:
 1. A raça humana extinguir-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 extinguiu no próximo século e colonizara a galáxia, sendo o número total de humanos nascidos 200 biliões (200×10^{12}).
- Vamos supor que a probabilidade à priori da hipótese 1 ocorrer é de $X\%$.
- Suponha-se que se sabe que você e 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:

$$\begin{aligned} P(Hip = 1 | \text{Você ser a pessoa } 60 \times 10^9) &= \\ &= \frac{P(\text{Você ser a pessoa } 60 \times 10^9 | Hip = 1) \times P(Hip = 1)}{P(\text{Você ser a pessoa } 60 \times 10^9)} = \\ &= \frac{\frac{1}{200 \times 10^9} X}{\frac{1}{200 \times 10^9} X + \frac{1}{200 \times 10^{12}} (1 - X)} = \\ &= \frac{X}{X + \frac{1 - X}{1000}} = \frac{1000 X}{900 X + 1} \end{aligned}$$

Casos de aplicação: Argumento do Juízo Final

- Alguns valores para X:

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

- **ATENÇÃO:** Este argumento é alvo de grande discordância e debate filosófico intenso! Ver e.g. a Wikipédia! (Doomsday Argument)

Casos de aplicação: Monty Hall

- 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 que tomar a decisão final se mantém a hipótese inicial ou muda para a outra porta. Ganhara aquilo que estiver atrás da porta selecionada no fim.
- Você começa por apontar para a porta numero 1.
- O apresentador abre-lhe a porta 3, que tem uma cabra escondida.
- Qual a sua decisão?

Casos de aplicação: Monty Hall

- Considerem-se as proposições:
- C_i – carro encontra-se atrás 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:
- $P(C_1|A_{13}) = P(C_1 \wedge A_{13})/P(A_{13}) = P(A_{13}|C_1) \times P(C_1)/P(A_{13})$.
- Para o numerador temos:
- $P(A_{13}|C_1) \times P(C_1) = 1/2 \times 1/3 = 1/6$
- O calculo do denominador é um pouco mais difícil...

Casos de aplicação: Monty Hall

- Para o denominador podemos reescrever:
- $P(A_{13}) = \sum_{i \in \{1,2,3\}} P(A_{13}|C_i) \times P(C_i) =$
- $= P(A_{13}|C_1) \times P(C_1) + P(A_{13}|C_2) \times P(C_2) + P(A_{13}|C_3) \times P(C_3)$
- $= 1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3 = 1/2$
- Portanto, $P(C_1|A_{13}) = (1/6)/(1/2) = 1/3$
- Logo, a decisão racional é trocar de porta pois a probabilidade de aí se encontrar o carro é de $2/3$.

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 atômico
- As interrogações podem ser respondidas somando-se acontecimentos atômicos
- Para domínios não triviais, é essencial reduzir o tamanho da distribuição conjunta
- Independência e independência condicional fornecem as ferramentas