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:
  - $A_{25} \longmapsto_{0,3}$ chegar a tempo
  - Sprinkler $\longmapsto_{0,99}$ WetGrass
  - WetGrass $\longmapsto_{0,7}$ Rain
- Problemas: como se combinam os factores, e.g., Sprinkler provoca Rain??

- 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:
  - $P(A_{25}$ permite-me chegar a tempo | …) = 0.04
  - $P(A_{90}$ permite-me chegar a tempo | …) = 0.70
  - $P(A_{120}$ permite-me chegar a tempo | …) = 0.95
  - $P(A_{1440}$ permite-me chegar a tempo | …) = 0.9999
- Que ação escolher?
  - Depende de minhas preferências sobre perder o vôo vs. tempo de espera no aeroporto vs. gastronimia 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 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 (evento) A é qualquer subconjunto de Ω
  - $P(A)=\sum_{\{\omega \in A\}} P(\omega)$
  - E.g., P(lançamento do dado < 4) = 1/6 + 1/6 + 1/6 = 1/2

Variáveis Aleatórias

- Proposições descrevendo acontecimentos (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 ∈ {1,2} e Pintas ∈ {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!):
  - < 1, 1 > < 1, 2 > < 1, 3 > < 1, 4 > < 1, 5 > < 1, 6 >
  - < 2, 1 > < 2, 2 > < 2, 3 > < 2, 4 > < 2, 5 > < 2, 6 >

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 $< sunny, rainy, cloudy, snow >$
  - $Weather = rainy$ é uma proposição
  - Quando não há ambiguidade, podemos usar o valor da variável aleatória ($rainy$) em vez da proposição ($Weather = rainy$)
- Variáveis aleatórias continuas (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., $P(Cavity = true) = 0.1$ e $P(Weather = sunny) = 0.72$

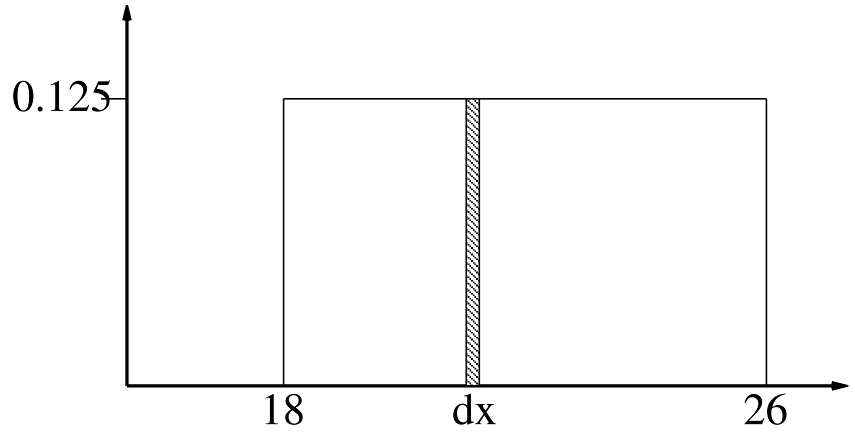
  correspondem à crença do agente antes da chegada de (nova) evidência
- Distribuição de probabilidade dá valores a todas as atribuições possíveis:
  - $\mathbf{P}(Weather) = \langle 0.72; 0.1; 0.08; 0.1 \rangle$ (normalizado, i.e., soma 1)
  - Bold (em $\mathbf{P}$) significa um vector, assumindo ordenação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

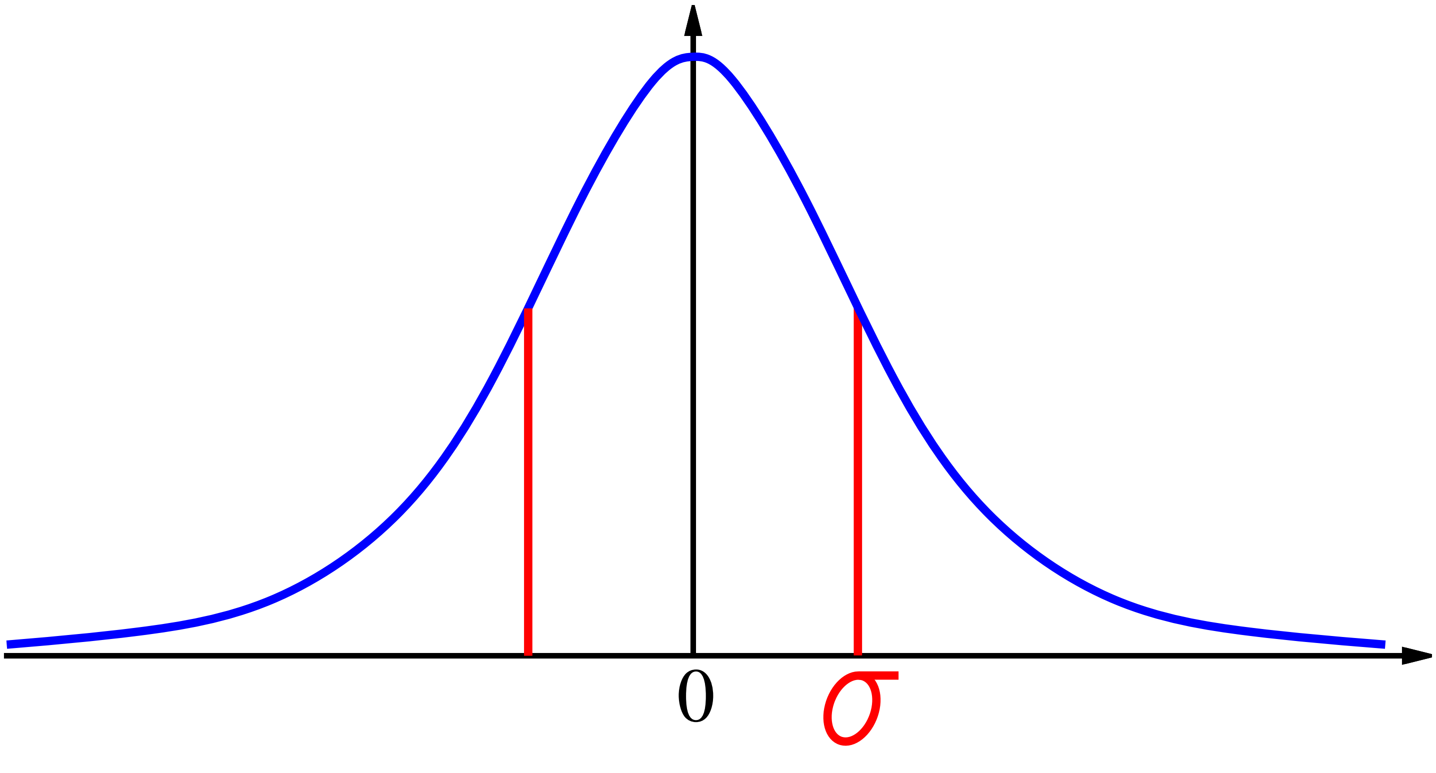


- Em que P é uma densidade de probabilidade; integrando em todo o domínio dá 1. $P(X = 20.5) = 0.125$ realmente signica

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

Função de densidade Gaussiana ou Normal

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



Probabilidade condicional ou à posteriori

- Probabilidades à posteriori ou condicionais
  - e.g., P(cavity | toothache) = 0.8
  - i.e., toothache é tudo o que sei
- Notação para distribuições condicionais:
  - **P**(X | 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(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, AcadémicaGanha) = P(cavity | toothache) = 0.8
- Esta forma de inferência, sancionada pelo conhecimento do domínio, é crucial

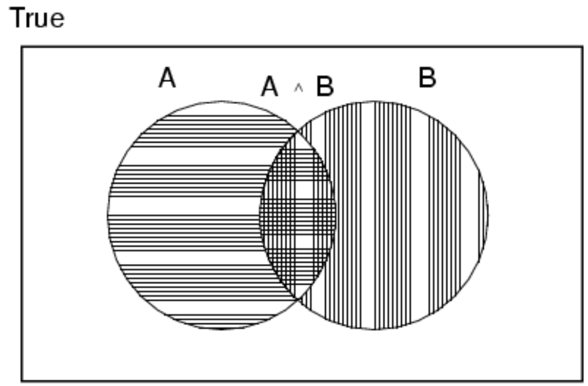
Probabilidade Condicional

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

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

Porquê usar probabilidades?

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



- 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.

Argumento de *de Finetti*

- Vamos agora assumir que o agente 1 tem crenças que violam os axiomas da probabilidade.
- Por exemplo, vamos considerar que o agente 1 tem uma crença em **a** de 0.4, uma crença em **b** de 0.3, e uma crença em **a**$\vee$**b** de 0.8 (que viola o axioma $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$).
- De acordo com de Finetti, será possível a um agente 2 construir uma aposta que o agente 1 aceitará (mostraremos o caso em que é indiferente), e que o fará perder dinheiro, independentemente do resultado.
- Neste caso, a aposta é composta por 3 apostas individuais, cada uma das quais é indiferente para o agente 1.

Agente 1		Agente 2		Resultado e payoff para o Agente 1			
Prop.	Crença	Aposta	Stakes	**a,b**	**a,¬b**	**¬a,b**	**¬a,¬b**
a	0.4						
b	0.3						
a$\vee$**b**	0.8						

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:

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

- Calcular *P(toothache)* (probabilidade marginal de *toothache*)?

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:

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

- A probabilidade marginal de *toothache* é:

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

- Calcular P(cavity ∨ toothache)?

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:

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

$$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.2$$

- Calcular P(¬cavity | toothache)?

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**

- Podemos calcular probabilidades condicionais:

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

$$P(\neg cavity \mid toothache) = P(\neg cavity \land toothache) \,/\, P(toothache)$$
$$= (0.016 + 0.064) \,/\, (0.108 + 0.012 + 0.016 + 0.064)$$
$$= 0.4$$

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**

- Denominador pode ser visto como constante de normalização α

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$$
$$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\, [<0.108, 0.016> + <0.012, 0.064>]$$
$$= \alpha <0.12, 0.08> = <0.6, 0.4>$$

- Ideia geral: calcular distribuição na variável interrogada, fixando as variaveis 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 probabilistica

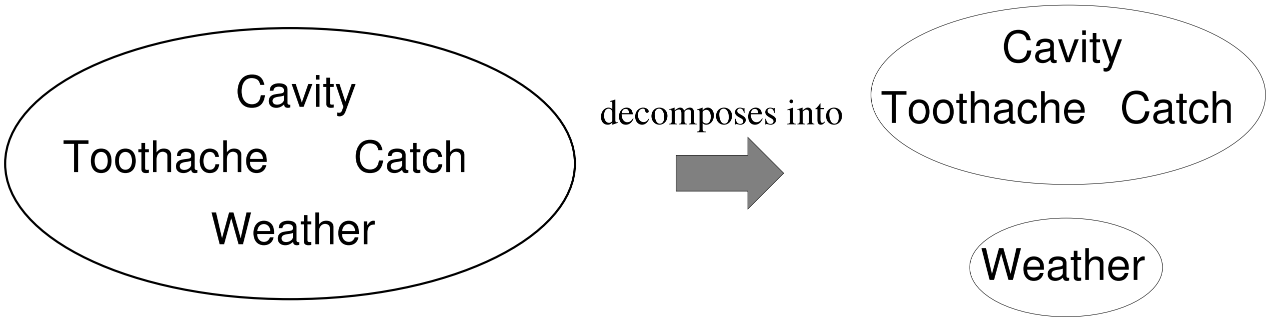
- Habitualmente, estamos interessados na
  - distribuição conjunta à posteriori das variáveis interrogadas $\mathbf{Y}$
  - dados valores específicos e para as variáveis evidência $\mathbf{E}$
- Sejam as variáveis ocultas dadas por $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$
- Logo, o somatório 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(Toothache,Catch,Cavity,Weather)
  = P(Toothache,Catch,Cavity)P(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, 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 sonda fique presa (catch) é independente do facto de eu ter dor de dentes:
  1. $P(catch \mid toothache, cavity) = P(catch \mid cavity)$
- A mesma relação de independência verifica-se se eu não tiver uma cárie:
  2. $P(catch \mid toothache, \neg cavity) = P(catch \mid \neg cavity)$
- *Catch* é condicionalmente independente de *Toothache* dado *Cavity*:
  - $\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$
- Afirmações Equivalentes:
  - $\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$
  - $\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\ \mathbf{P}(Catch \mid Cavity)$

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:

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

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:

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

- Ou na forma da distribuição conjunta:

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

- Variante condicionalizada na evidência $\mathbf{e}$ fornecida

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

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

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

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...)

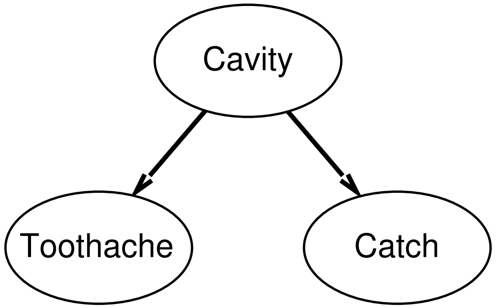
- Seja C="tem cancro da mama" e M="teve uma mamografia positiva"
- Problema: $P(c)=0.01;\ P(m|c)=0.8;\ P(m|\neg c)=0.096.\ P(c|m)=?$
- Regra de Bayes: $P(c|m)=P(m|c)P(c)/P(m)$
- Marginalização: $P(m)=P(m|c)P(c)+P(m|\neg c)P(\neg c)=.8*.01+.096*.99 = .10304$
- Resolvendo: $P(c|m)=P(m|c)P(c)/P(m)=.8*.01/.10304=.0776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

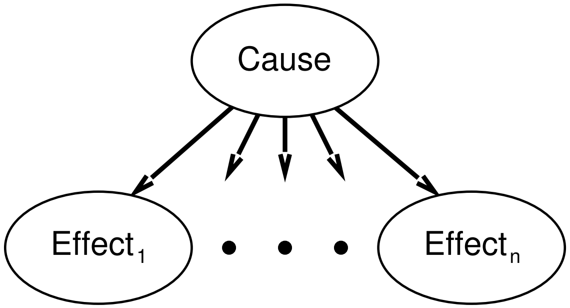
Regra de Bayes e independência condicional

- $\mathbf{P}(Cavity|toothache \wedge catch)$
  $= \alpha\, \mathbf{P}(toothache \wedge catch|Cavity)\ \mathbf{P}(Cavity)$
  $= \alpha\, \mathbf{P}(toothache|Cavity)\ \mathbf{P}(catch|Cavity)\ \mathbf{P}(Cavity)$

- 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:

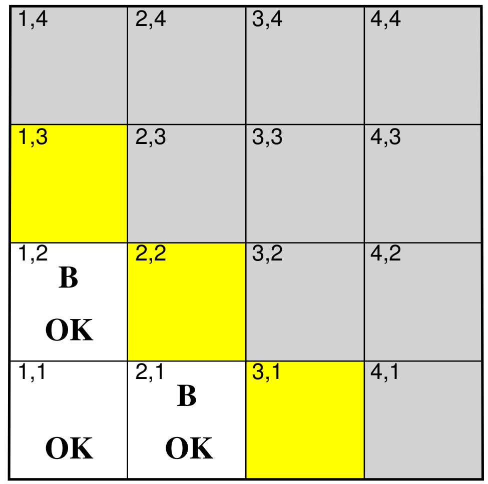
$$\mathbf{P}(Cause,Effect_1,\ldots,Effect_n)=\mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i|Cause)$$





- O numero 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

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 $\mathbf{P}(Efeito|Causa)$.)
- Primeiro termo: 1 se poços adjacentes a brisa, 0 caso contrario
- Segundo termo: poços são dispostos aleatoriamente, probabilidade de 0.2 por casa:

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

- Para uma configuração com n poços,

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

Observações e interrogações

- Sabemos o seguinte:

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

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

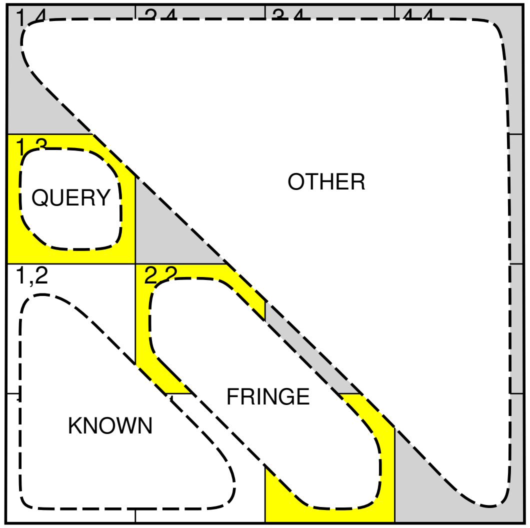
- Interrogação é $\mathbf{P}(P_{1,3}|known,b)$
- Seja $Unknown = P_{ij}$s diferentes de $P_{1,3}$ e $Known$
- Por inferência por enumeração (marginalização), obtemos

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

- 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

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

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

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

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

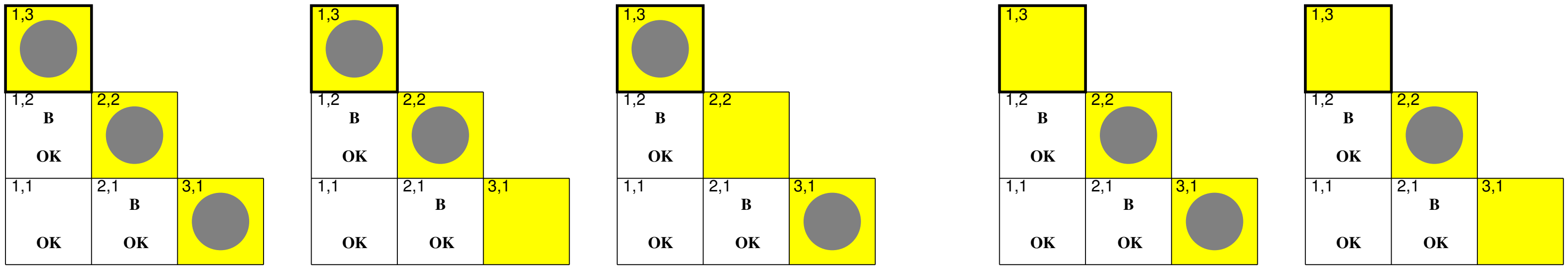
$$= \alpha \sum_{fringe} \mathbf{P}(b \mid known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

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

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

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

Utilizando a independência condicional



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

$$\mathbf{P}\left(P_{1,3} \mid known, b\right) = \alpha' \left\langle 0.2\left(0.04+0.16+0.16\right), 0.8\left(0.04+0.16\right)\right\rangle$$

$$= \alpha' \left\langle 0.072, 0.16 \right\rangle$$

$$= \alpha' \left\langle \frac{0.072}{0.072+0.16}, \frac{0.16}{0.072+0.16} \right\rangle$$

$$\approx \left\langle 0.31, 0.69 \right\rangle$$

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

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:

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

- Outra maneira (sem variáveis escondidas):

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

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 extinguira no próximo século e colonizara a galáxia, sendo o numero 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 numero 60 mil milhões (60×10^9) (aproximadamente o numero de pessoas que nasceram ate 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:

$$P\left(Hip=1 \mid \text{Você ser a pessoa } 60\times 10^{9}\right)=$$

$$=\frac{P\left(\text{Você ser a pessoa } 60\times 10^{9} \mid Hip=1\right)\times P\left(Hip=1\right)}{P\left(\text{Você ser a pessoa } 60\times 10^{9}\right)}=$$

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

$$=\frac{X}{X+\dfrac{1-X}{1000}}=\frac{1000}{900X+1}$$

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 mantem 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 difcil...

Casos de aplicação: Monty Hall

- Para o denomimnador 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 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