

INF623

2024/1



Inteligência Artificial

A14: Raciocínio Probabilístico I

Plano de aula

- ▶ Agentes probabilísticos
- ▶ Probabilidades
- ▶ Variáveis aleatórias
- ▶ Distribuições
- ▶ Probabilidades conjuntas
- ▶ Probabilidades condicionais
- ▶ Regras de probabilidade
- ▶ Inferência por enumeração

Incerteza

Problemas do mundo real geralmente envolvem incerteza, o que pode ser muito difícil de representar usando lógica, por exemplo:

DorDente \Rightarrow *Carie*

Regra falsa! Nem todos os pacientes com dor de dentes têm cáries.

DorDente \Rightarrow *Carie* \vee *Gengivite* \vee *Abscessos* ...

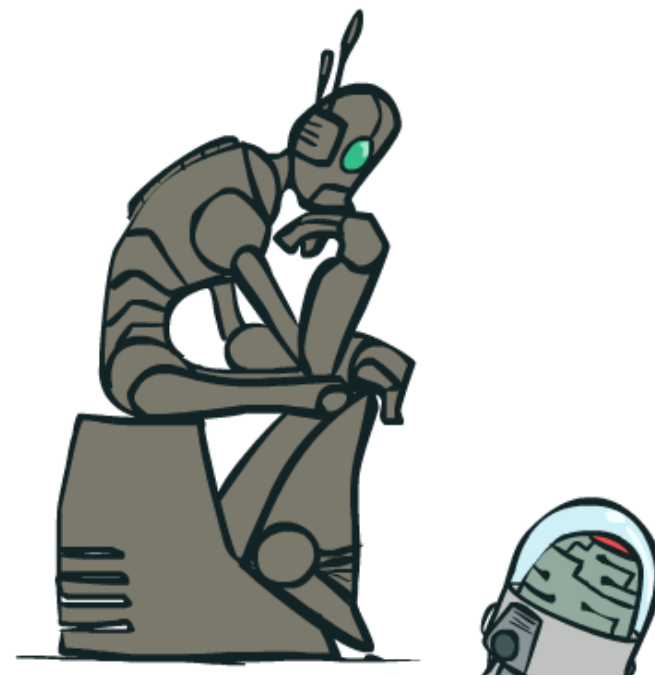
Para tornar Verdadeira, teríamos que adicionar uma lista quase ilimitada de causas possíveis.

Carie \Rightarrow *DorDente*

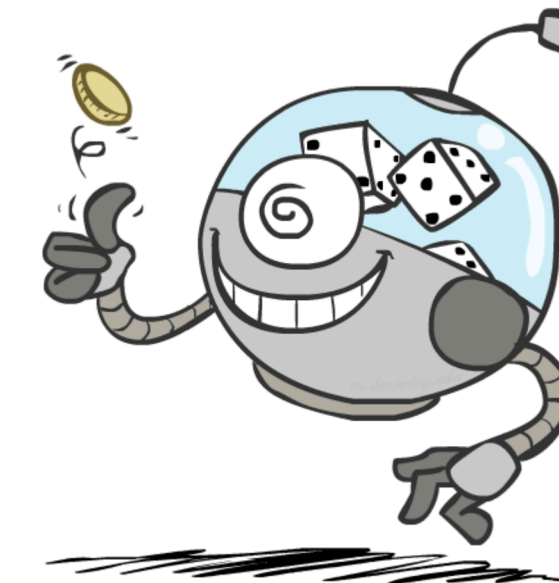
Poderíamos tentar transformar essa regra em uma regra causal, mas também não estaria correta; nem todas as cáries causam dor.

Teoria de probabilidades

A teoria de probabilidades tem o mesmo objetivo da lógica, representar conhecimento do mundo em sentenças que são válidas ou não



A teoria de probabilidades nos permite expressar incerteza sobre uma sentença!

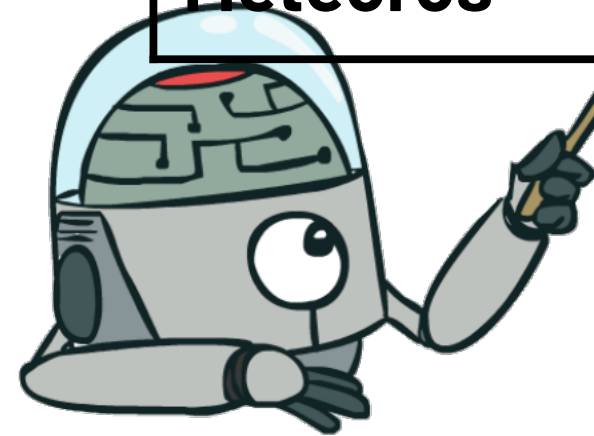


- ▶ Um agente lógico acredita que uma sentença seja:
 - ▶ Falsa
 - ▶ Verdadeira
 - ▶ "Se dor de dente, então cárie"
- ▶ Um agente probabilístico pode ter um grau de crença numérico entre 0 e 1:
 - ▶ 0 para sentenças certamente falsas
 - ▶ 1 para sentenças certamente verdadeiras
 - ▶ "A probabilidade de um paciente com dor de dente ter cárie é de 0,8"

Agentes probabilísticos

Agentes probabilísticos representam o mundo por um **modelo de probabilidades** e derivam conclusões a partir de **consultas** utilizando **inferência probabilística**.

	Quente	Frio
Ensolarado	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00



- ▶ **Modelo de probabilidades** é representado por uma distribuição conjunta $P(X_1, X_2, \dots, X_n)$
- ▶ Uma **variável aleatória** X_i representa um aspecto do mundo do qual temos incerteza
- ▶ **Consultas** são probabilidades condicionais $P(Q | E)$ que representam a crença do agente em Q (query) dado que E (evidence) ocorreu .

Probabilidades

Uma **probabilidade** $P(A)$ representa a chance de um determinado **evento** A ocorrer dentro de um **espaço amostral** Ω

0 **espaço amostral** representa todos os eventos elementares ω (em IA, mundos) de um experimento

$$\Omega = \left\{ \begin{array}{c} \square \bullet \\ \square \bullet \bullet \\ \square \bullet \bullet \bullet \\ \square \bullet \bullet \bullet \bullet \\ \square \bullet \bullet \bullet \bullet \bullet \\ \square \bullet \bullet \bullet \bullet \bullet \bullet \end{array} \right\}$$

$P(\omega = 1)$	$P(\omega = 2)$	$P(\omega = 3)$	$P(\omega = 4)$	$P(\omega = 5)$	$P(\omega = 6)$
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Um **modelo de probabilidade** associa uma probabilidade $P(\omega)$ para cada mundo $\omega \in \Omega$, seguindo as seguintes regras:

1. $0 \leq P(\omega) \leq 1$: A probabilidade $P(\omega)$ de um mundo $\omega \in \Omega$ é um valor real entre 0 e 1
2. $\sum_{\omega \in \Omega} P(\omega) = 1$: A probabilidade total do conjunto de mundos possíveis é 1

Eventos

Um **evento** ϕ é um subconjunto de Ω :

- ▶ "Resultado da rolagem de um dado regular ser igual a 1" é o subconjunto $A = \{1\}$
- ▶ "Resultado da rolagem de um dado regular ser menor que 4" é o subconjunto $B = \{1,2,3\}$
- ▶ "Resultado da rolagem de um dado regular ser um número ímpar" é o subconjunto $C = \{1,3,5\}$

A probabilidade $P(\phi)$ de um **evento** ϕ é dada por $P(\phi) = \sum_{\omega \in \phi} P(\omega)$:

- ▶ $P(A) = P(\omega = 1) = 1/6$
- ▶ $P(B) = P(\omega = 1) + P(\omega = 2) + P(\omega = 3) = 1/6 + 1/6 + 1/6 = 3/6$
- ▶ $P(C) = P(\omega = 1) + P(\omega = 3) + P(\omega = 5) = 1/6 + 1/6 + 1/6 = 3/6$

Variáveis aleatórias

Informalmente

Um aspecto do mundo sob o qual temos incerteza:

- ▶ X – “O resultado do dado é ímpar?”
- ▶ T – “Está frio ou quente?”
- ▶ D – “Quanto tempo para chegar em BH?”
- ▶ L – “Onde está o poço (Wumpus)?”

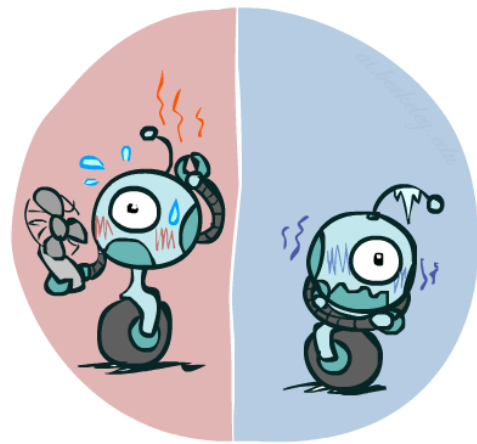
Formalmente

Uma função que mapeia os mundos possíveis Ω em um valor dentro de um domínio:

- ▶ $X \in \{V, F\}$
 - ▶ $X(1) = V$
 - ▶ $X(6) = F$
- ▶ $T \in \{quente, frio\}$
- ▶ $D \in [0, \infty]$
- ▶ $L \in \{(1,1), (1,2), \dots\}$

Distribuições de probabilidades

A **distribuição de probabilidades** de uma variável aleatória X é uma atribuição de probabilidade $P(X = x)$ para cada valor x de seu domínio:



T – “Está frio ou quente?”

$P(T)$

Temperatura (T)	P
Quente	0,5
Frio	0,5



C – “Como está o clima?”

$P(C)$

Clima (C)	P
Ensolarado	0,6
Chuvoso	0,1
Nublado	0,3
Meteoros	0,0

► Uma distribuição deve respeitar:

► $0 \leq P(X = x) \leq 1$

► $\sum_x P(X = x) = 1$

$P(X)$ denota a distribuição inteira (tabela)

Quando não for ambíguo, usaremos $P(x)$ para denominar $P(X = x)$

Distribuições conjuntas

Uma **distribuição conjunta** é uma atribuição de probabilidade para todas as combinações de ocorrências de um conjunto de variáveis aleatórias $\{X_1, X_2, \dots, X_n\}$:

	Quente	Frio
Ensolarado	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00

$P(T, C)$ – distribuição conjunta de T e C

▶ Uma distribuição conjunta deve respeitar:

▶ $0 \leq P(x_1, x_2, \dots, x_n) \leq 1$

▶ $\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n) = 1$

▶ Qual o tamanho de uma distribuição conjunta com n variáveis com domínio de tamanho d ?

▶ d^n valores na tabela

Distribuições conjuntas

[Lembrete]

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

Dada uma distribuição conjunta, podemos calcular a probabilidade de qualquer evento relacionado a elas, por exemplo:

- ▶ A probabilidade de estar quente E ensolarado?
- ▶ A probabilidade de estar quente?
- ▶ A probabilidade de estar quente OU não nublado?

Em geral, os eventos que nos interessam são atribuições parciais, por exemplo $P(T = Quente)$

P(T, C)

	Quente	Frio
Ensolarado	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00

Distribuições marginais

Distribuições marginais são sub-tabelas que eliminam algumas variáveis de uma distribuição conjunta. Podemos extrair as distribuições marginais somando dimensões da distribuição conjunta:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

$P(T, C)$

$$P(C) = \sum_t P(C, T = t)$$

$$P(T) = \sum_c P(T, C = c)$$

	Quente	Frio
Ensolarado	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00

somar as linhas

Clima (C)	P
Ensolarado	0,6
Chuvoso	0,1
Nublado	0,3
Meteoros	0,0

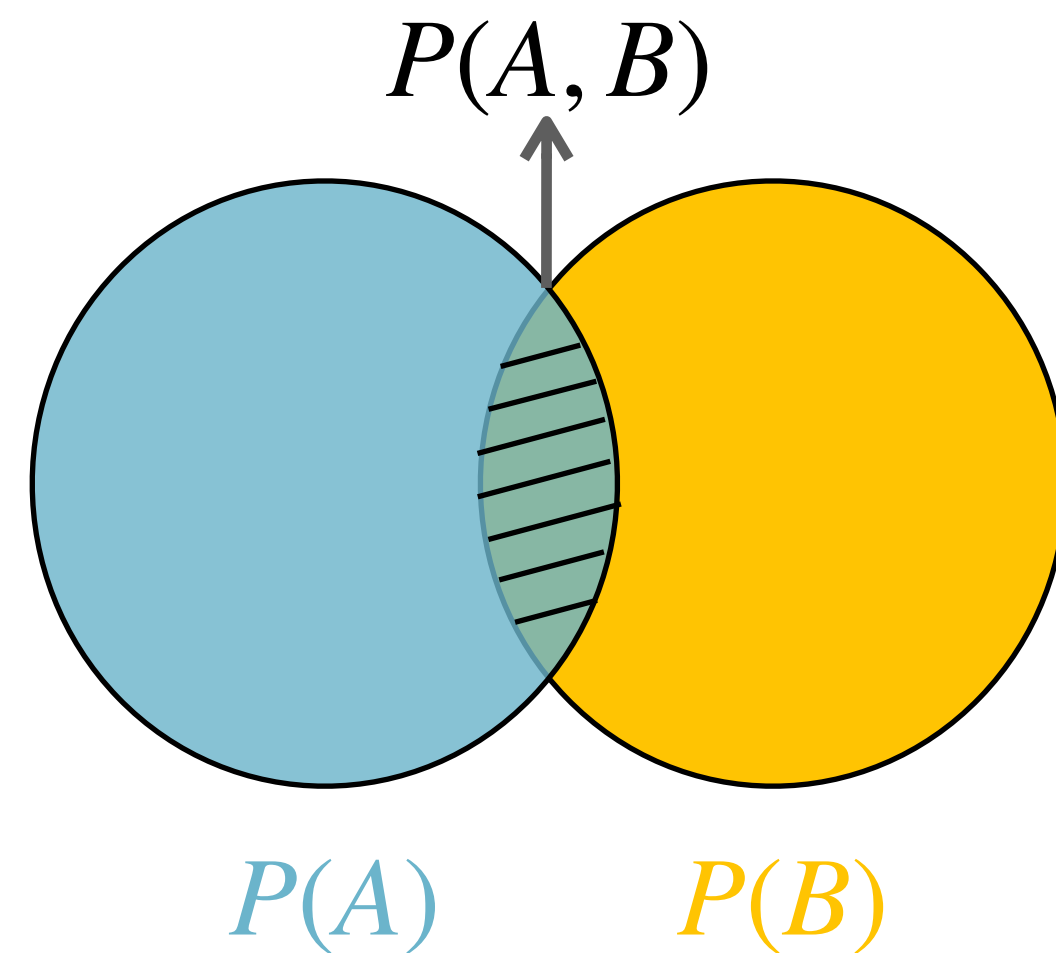
somar as colunas

Temperatura (T)	P
Quente	0,5
Frio	0,5

- ▶ Podemos deduzir as marginais da conjunta
- ▶ **Geralmente não podemos deduzir a conjunta a partir das marginais!**

Probabilidades condicionais

A **probabilidade condicional** $P(A | B)$ representa a chance de um determinado evento A ocorrer dado que B ocorreu (probabilidade de A dado B).



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

Exercício

[Lembrete]

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

Qual a probabilidade da temperatura estar fria dado que o clima está ensolarado?

$P(T, C)$	Quente	Frio (f)
Ensolarado (e)	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00

$$P(C = e | T = f)$$

$$P(C = e | T = f) = \frac{P(C = e, T = f)}{P(T = f)}$$

$$P(C = e | T = f) = \frac{0,15}{0,15 + 0,08 + 0,27 + 0}$$

$$P(C = e | T = f) = \frac{0,15}{0,5} = 0,3$$

Distribuições condicionais

Dadas duas variáveis aleatórias X e Y distribuídas conjuntamente, a **distribuição condicional** de X dado Y é a distribuição de X quando Y tem um determinado valor conhecido.

$P(T, C)$

	Quente	Frio
Ensolarado	0,45	0,15
Chuvoso	0,02	0,08
Nublado	0,03	0,27
Meteoros	0,00	0,00

$P(C | T = quente)$

Clima (C)	P
Ensolarado	0,90
Chuvoso	0,04
Nublado	0,06
Meteoros	0,00

$P(C | T = frio)$

Clima (C)	P
Ensolarado	0,30
Chuvoso	0,16
Nublado	0,54
Meteoros	0,00

As distribuições condicionais são consequências das distribuições conjuntas

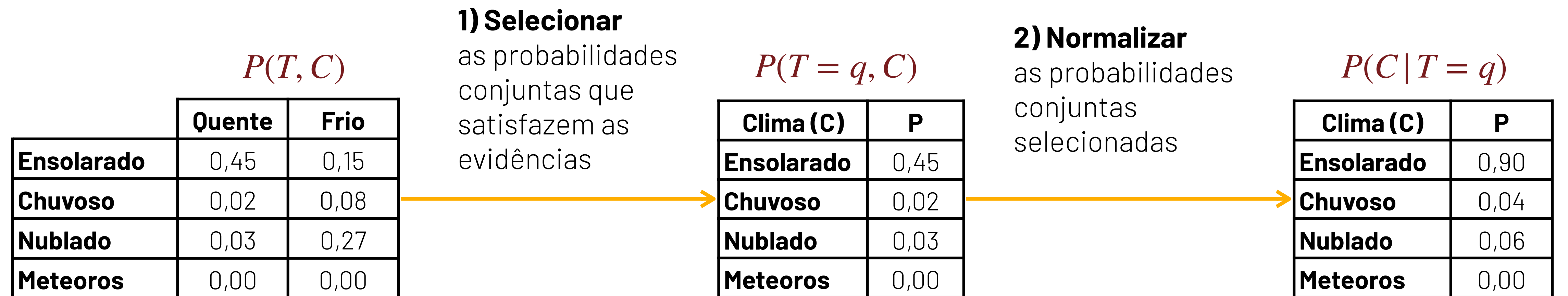
$P(C | T)$

Essa notação é usada para se referir ao conjunto de distribuições condicionais para todos os valores de T .

Distribuições condicionais

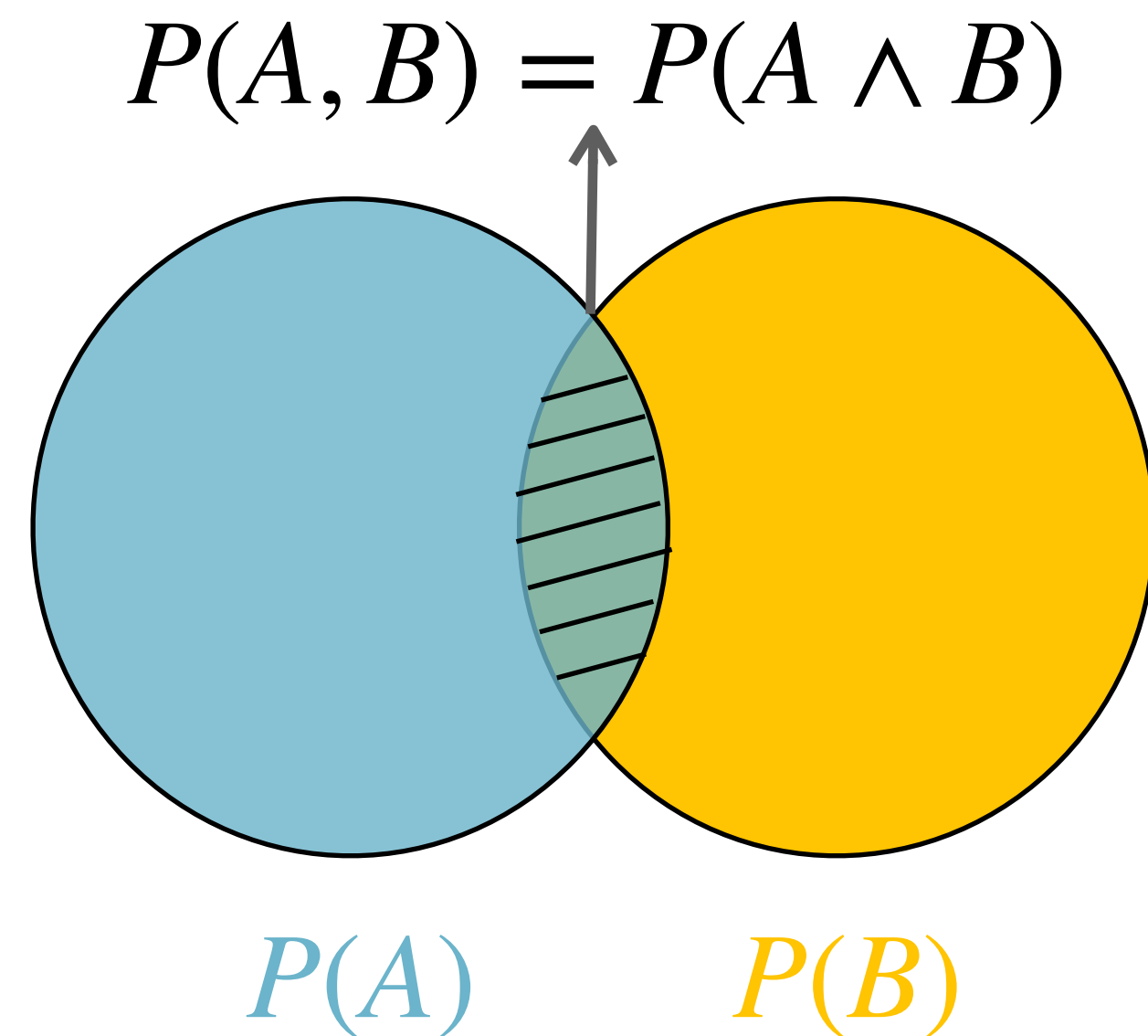
Extraindo distribuições condicionais $P(X | Y = y)$ de distribuições conjuntas $P(X, Y)$.

Por exemplo, extrair $P(C | T = q)$ de $P(T, C)$



$$\begin{aligned}
 P(x_1 | x_2) &= \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)} \longrightarrow P(C = e | T = q) = \frac{P(C = e, T = q)}{P(T = q)} \\
 &= \frac{P(C = e, T = q)}{P(C = e, T = q) + P(C = c, T = q) + P(C = n, T = q) + P(C = m, T = q)} \\
 &= \frac{0,45}{0,45 + 0,02 + 0,03 + 0,00} = 0,9
 \end{aligned}$$

Regras de probabilidades



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

▶ **Negação**

$$P(\neg A) = 1 - P(A)$$

▶ **Inclusão e exclusão**

$$P(A \vee B) = P(A) + P(B) - P(A, B)$$

▶ **Produto**

$$P(A, B) = P(A | B)P(B)$$

▶ **Marginalização**

$$P(A) = P(A, B) + P(A, \neg B)$$

$$P(X = x) = \sum_y P(X = x, Y = y)$$

▶ **Condicionamento**

$$P(A) = P(A | B)P(B) + P(A | \neg B)P(\neg B)$$

$$P(X = x) = \sum_y P(X = x | Y = y)P(Y = y)$$

Regra da cadeia

[Lembrete]

$$P(A, B) = P(A | B)P(B)$$

Uma distribuição conjunta pode ser escrita como um produto de distribuições condicionais pela aplicação repetida da regra do produto:

- ▶ Para 3 variáveis:

$$P(X_1, X_2, X_3) = P(X_3 | X_1, X_2)P(X_1, X_2) = P(X_3 | X_1, X_2)P(X_2 | X_1)P(X_1)$$

- ▶ Generalizando para N variáveis:

$$P(X_1, X_2, \dots, X_n) = \prod_i^n P(X_i | X_1, \dots, X_{i-1})$$

Inferência probabilística

Inferência probabilística consiste em calcular novas probabilidades a partir de probabilidades conhecidas

- ▶ Em IA, estamos interessados em representar a crença do agente a partir de evidências
 - ▶ Essa crença é modelada com probabilidades condicionais
 - ▶ Exemplo: $P(\text{chegar no tempo} \mid \text{sem acidentes}) = 0,90$
- ▶ A crença do agente pode mudar com mais evidências:
 - ▶ $P(\text{chegar no tempo} \mid \text{sem acidentes}, 5:00\text{h}) = 0,95$
 - ▶ $P(\text{chegar no tempo} \mid \text{sem acidentes}, 5:00\text{h}, \text{chuva}) = 0,80$



Inferência por enumeração

Dado um modelo de probabilidades $P(X_1, X_2, \dots, X_n)$, queremos calcular a probabilidade condicional $P(Q | e_1, \dots, e_k)$, onde:

- ▶ Q é uma variável aleatória definindo uma consulta
- ▶ $E_1, \dots, E_k = e_1, \dots, e_k$ é um conjunto de variáveis aleatórias definindo a evidência
- ▶ Além disso, H_1, \dots, H_r é o conjunto de variáveis "escondidas" do modelo, não inclusas na consulta nem na evidência

Inferência por enumeração:

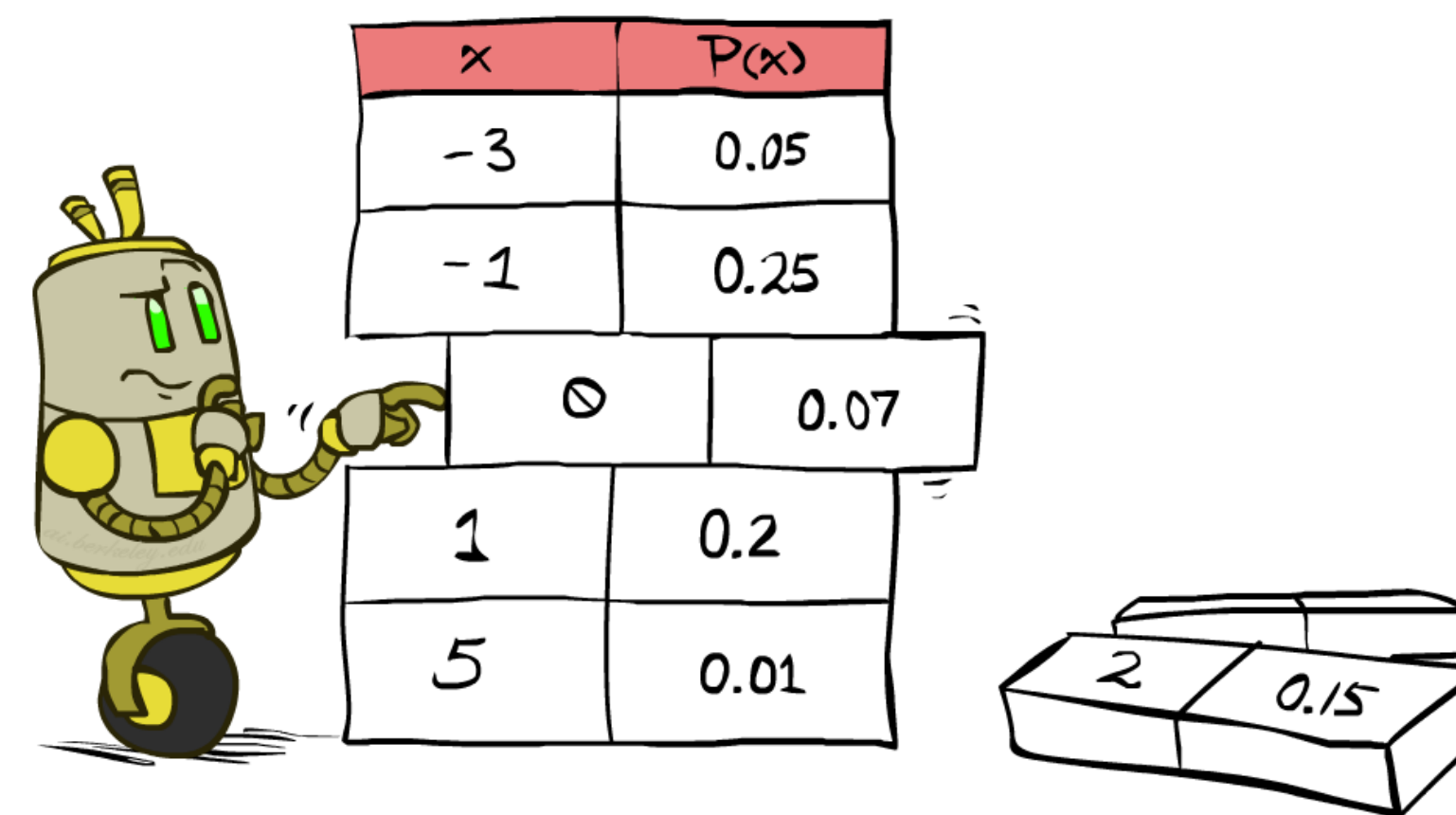
1) **Selecionar** as entradas consistentes com a evidência

2) **Somar** as variáveis em H para obter $P(Q, e_1, \dots, e_k)$

$$P(Q, e_1, \dots, e_k) = \sum_{h_1, \dots, h_r} P(Q, h_1, \dots, h_r, e_1, \dots, e_k)$$

3) **Normalizar** $P(Q, e_1, \dots, e_k)$ para obter $P(Q | e_1, \dots, e_k)$

$$P(Q | e_1, \dots, e_k) = \frac{P(Q, e_1, \dots, e_k)}{\sum_q P(Q, e_1, \dots, e_k)}$$



Exercícios

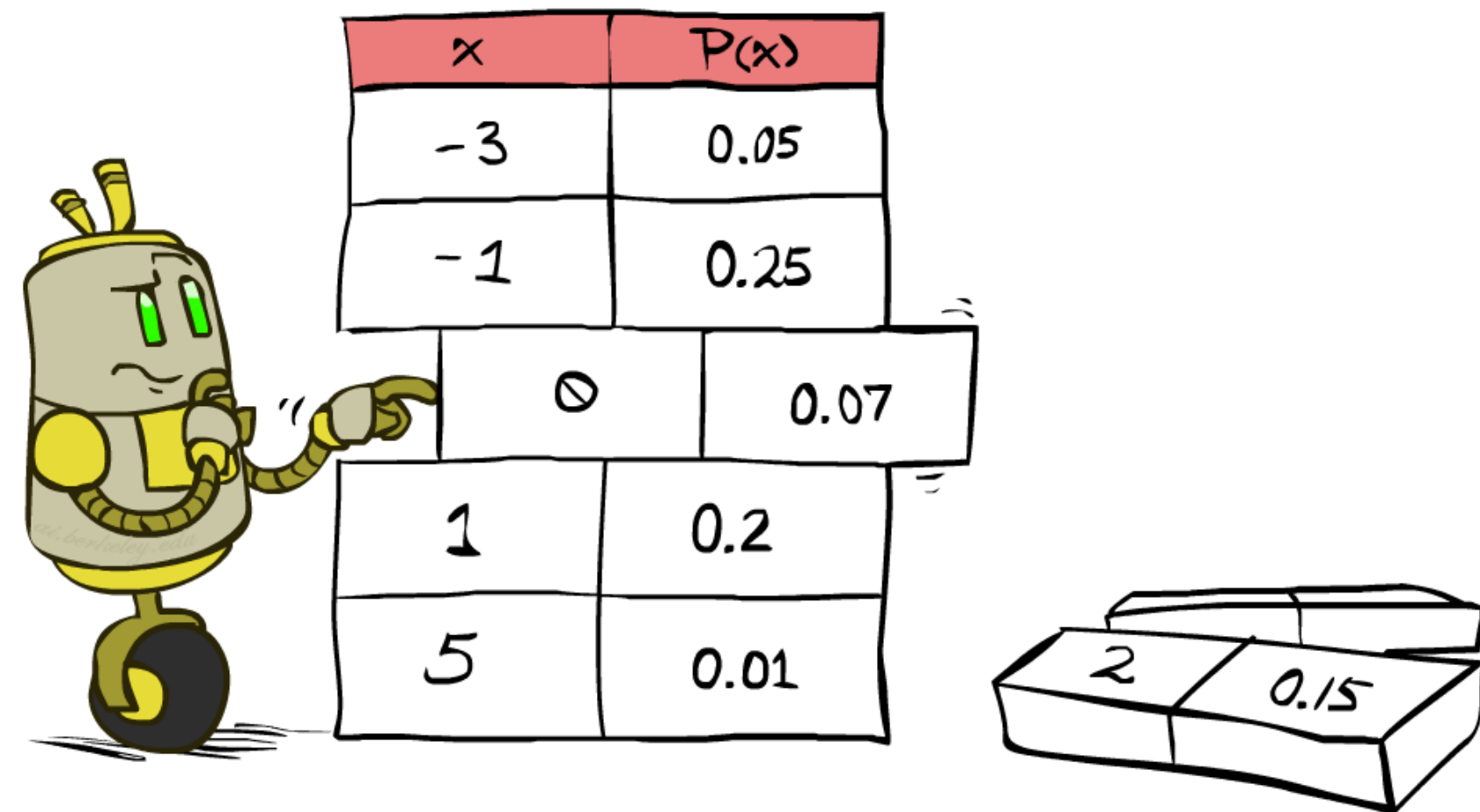
▶ $P(W)$?

▶ $P(W | inverno)$?

▶ $P(W | inverno, calor)$?

S	T	C	P
verão	calor	ensolarado	0,30
verão	calor	chuvoso	0,05
verão	frio	ensolarado	0,10
verão	frio	chuvoso	0,05
inverno	calor	ensolarado	0,10
inverno	calor	chuvoso	0,05
inverno	frio	ensolarado	0,15
inverno	frio	chuvoso	0,20

Inferência por enumeração



x	$P(x)$
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01

Problemas?

- ▶ Complexidade de tempo $O(d^n)$
- ▶ Complexidade de espaço para armazenar a distribuição conjunta $O(d^n)$
- ▶ Obter $O(d^n)$ exemplos para estimar as entradas da distribuição conjunta

Próxima aula

A15: Raciocínio Probabilístico II

Teorema de Bayes, independência, independência condicional, redes bayesianas