Probabilidade e Estatística

Aula 4

Probabilidade: Conceitos Básicos - parte 2

Leituras:

Obrigatória: Devore, Capítulo 2

Complementar: Bertsekas e Tsitsiklis, Capítulo 1

Objetivos

Nesta aula, aprenderemos:

- Probabilidade condicional
- Regra do produto
- Teorema da probabilidade total
- Teorema de Bayes
- Independência

Probabilidade Condicional

- **Como incorporar novas informações à probabilidade de um evento?**

- A probabilidade condicional é a probabilidade de um evento levando em conta que já sabemos que outro evento ocorreu.

- **Exemplos:**
  - Qual é a probabilidade de chover amanhã? Qual é a probabilidade de chover amanhã dado que choveu hj?
  - Em um jogo de chute de palavras, qual a probabilidade de a segunda letra ser u? E se a primeira letra de uma palavra é t, qual é a probabilidade de que a segunda seja u? E se a primeira letra for q?
  - Qual a probabilidade de uma pessoa ter uma doença se um exame médico deu negativo?

Probabilidade Condicional

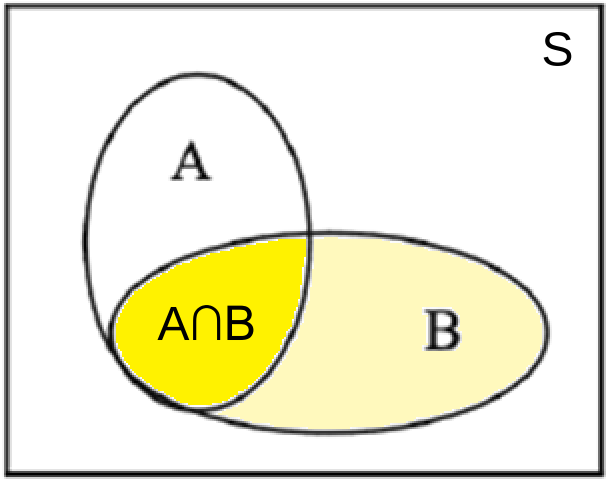
- A probabilidade condicional é a probabilidade de um evento dado que outro evento ocorreu.

$$P(A|B)$$

Probabilidade de A dado B ou probabilidade de A condicional a B.

- B se torna o novo conjunto universo (espaço amostral)! **Renormalização!**

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



Probabilidade Condicional

- A probabilidade condicional do evento A, dado o evento B é definida por:

Definição!

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

desde que $P(B) > 0$. Se $P(B) = 0$, então $P(A|B) = 0$

Probabilidade Condicional

Exercício: Dos carros de um feirão de carros usados, 70% tem ar condicionado (AC) e 40% tem um tocador de CD (CD). 20% dos carros possuem ambos.

Qual é a probabilidade de que o carro tenha um tocador de CD, dado que o carro tem ar condicionado?

Queremos determinar: $P(CD|AC)$.

Probabildade Condicional

Solução:

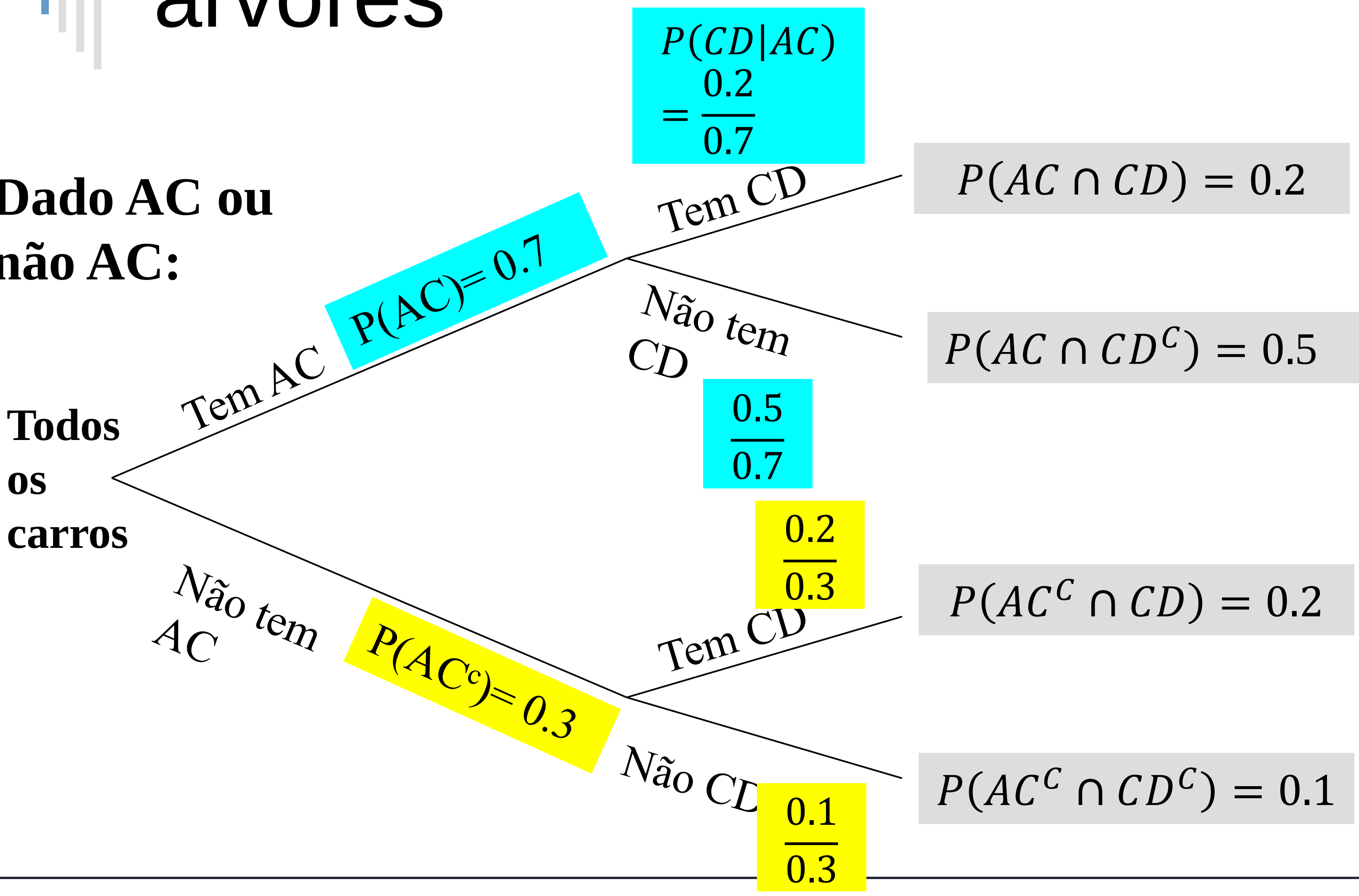
	CD	Sem CD	Total
AC	0.2	0.5	0.7
Sem AC	0.2	0.1	0.3
Total	0.4	0.6	1.0

$$P(CD|AC) = \frac{P(CD \cap AC)}{P(AC)} = \frac{0.2}{0.7} = 0.2857$$

Dado AC, nós apenas consideramos a linha de cima (70% dos carros). Destes, 20% tem um tocador de CD. 20% de 70% é aproximadamente 28.57%.

Probabilidade Condicional: árvores

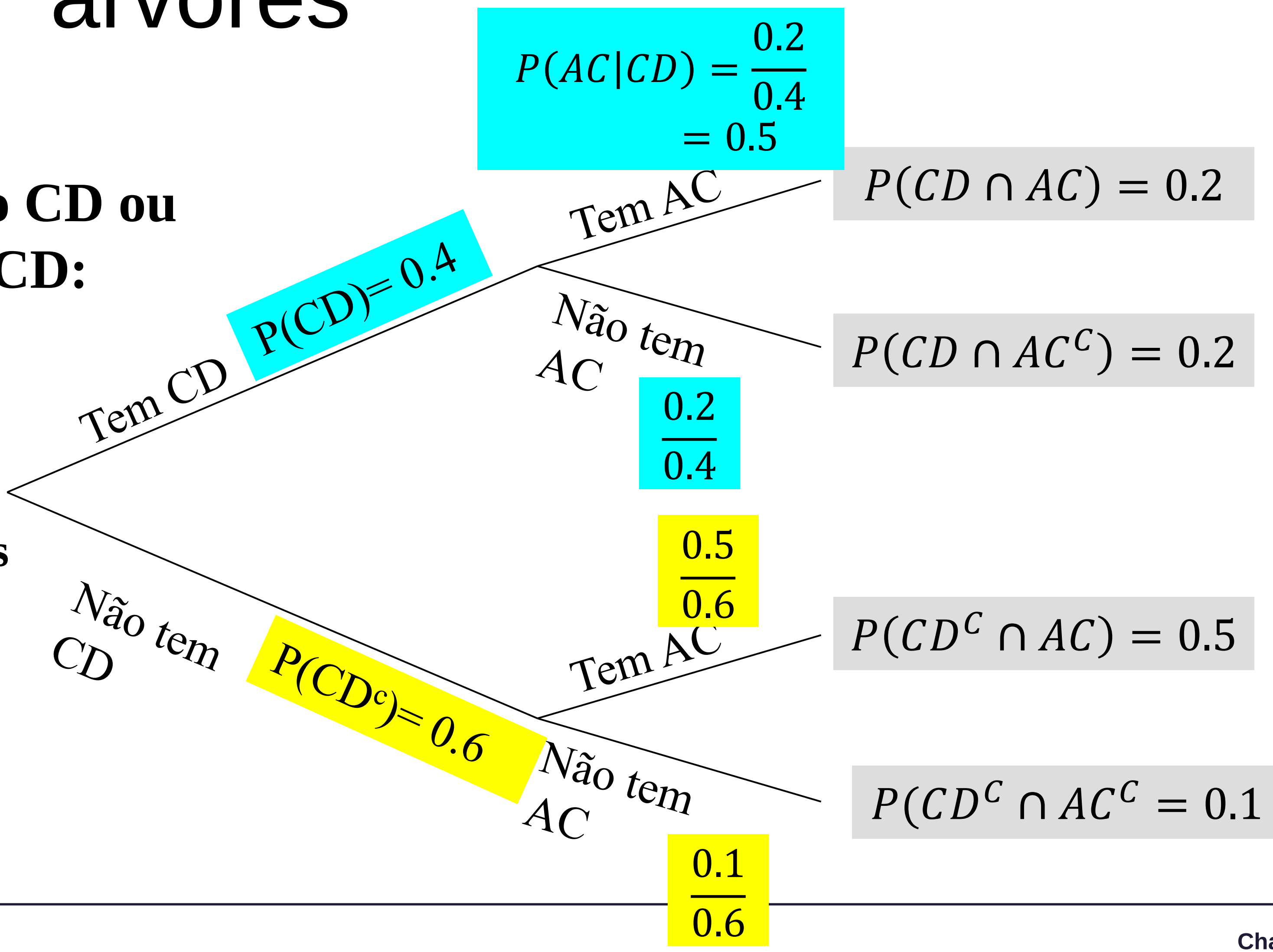
Dado AC ou não AC:



Probabilidade Condicional: árvores

Dado CD ou sem CD:

Todos os carros



Probabilidade Condicional

Exercício: Discos de policarbonato plástico, provenientes de um fornecedor, são analisados com relação à resistência a arranhões e choques. Os resultados de 200 discos estão resumidos a seguir:

		Resistência a Choque	
		Alta (A)	**Baixa**
Resistência a arranhão	**Alta (B)**	160	18
	Baixa	12	10

Seja A o evento em que o disco tenha alta resistência a choque e B o evento em que um disco tenha alta resistência a arranhão. Determine as seguintes probabilidades:

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

A Regra do Produto

- Outra forma de calcular a probabilidade de uma interseção de eventos:

  Para a interseção de dois eventos A e B:

Corolário

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

- Generalizando para a interseção de n eventos $A_1, \ldots, A_n$:

$$P(A_1 \cap A_2 \cap A_3 \cap \cdots A_n) = P(A_n|A_{n-1}) * P(A_{n-1}|A_{n-2}) * \cdots * P(A_2|A_1) * P(A_1)$$

Regra do Produto

Exercício: Suponha que o conselho da universidade seja composto de 5 brancos, 4 negros e 3 orientais. Qual a probabilidade de selecionarmos aleatoriamente um branco seguido por um oriental?

Regra do Produto

Exercício: Suponha que o conselho da universidade seja composto de 5 brancos, 4 negros e 3 orientais. Qual a probabilidade de selecionarmos aleatoriamente um branco seguido por um oriental?

Solução: Note que uma vez que a pessoa branca é escolhida (de um total de 12 membros do conselho), apenas 11 restam no espaço amostral do segundo sorteio, dos quais 3 são orientais.

$$P(A_1 = B \cap A_2 = O) = P(A_2 = O \mid A_1 = B) * P(A_1 = B)$$
$$= \frac{3}{11} * \frac{5}{12} = 0.114$$

O Teorema da Probabilidade Total

Exercício 1: Você entra em um torneio de xadrez em que a probabilidade de ganhar um jogo é de

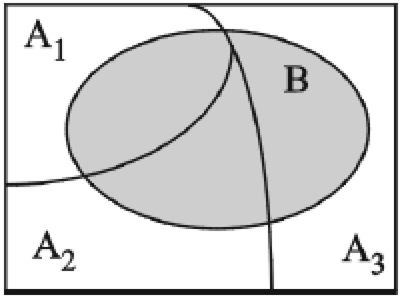
- 0.3 contra a metade (0.5) dos jogadores (tipo 1)
- 0.4 contra um quarto (0.25) dos jogadores (tipo 2)
- 0.5 contra um quarto (0.25) dos jogadores (tipo 3)

Você joga contra um oponente escolhido aleatoriamente. Qual a probabilidade de você ganhar o jogo?

O Teorema da Probalidade Total

Divida e Conquiste

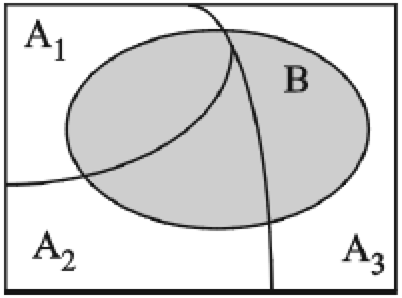
- Estamos interessados em determinar a probabilidade do evento B.
- Dividimos o espaço amostral em uma série de cenários possíveis, $A_1, \dots, A_n$ (que formam uma **partição** do espaço amostral, S).



- $\{A_i\}_{i=1}^{n}$ são tais que, para cada cenário A_i:
  - $P(A_i)$: probabilidade de estar em cada cenário é conhecida!
  - $P(B|A_i)$ são conhecidos ou fáceis de calcular!

O Teorema da Probalidade Total

- Estamos interessados em determinar a probabilidade do evento B.



- $P(B)$: média ponderada de B ocorrer em cada cenário, $P(B|A_i)$.
- Cada cenário é ponderado por sua probabilidade de ocorrência:

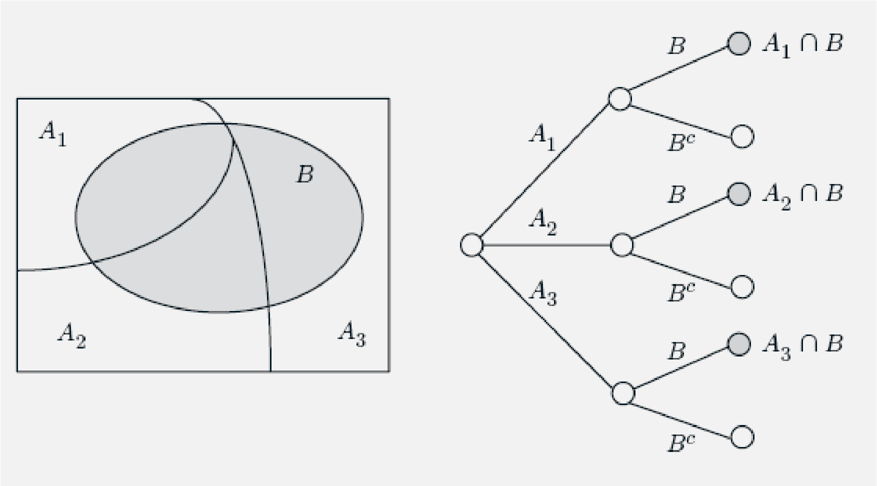
$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$$
$$= P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + \cdots + P(B|A_n) * P(A_n)$$

Teorema

O Teorema da Probabilidade Total

Divida e Conquiste

- Vizualizando teorema da Probabildade Total:



$$
\begin{aligned}
P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\
&= P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + P(B|A_3) * P(A_3)
\end{aligned}
$$

- A chave é a escolha da partição A_i!!

O Teorema da Probalidade Total

Exercício 1: Você entra em um torneio de xadrez em que a Probabilidade de ganhar um jogo é de

- 0.3 contra a metade dos jogadores (tipo 1)
- 0.4 contra um quarto dos jogadores (tipo 2)
- 0.5 contra um quarto dos jogadores (tipo 3)

Você joga contra um oponente escolhido aleatoriamente. Qual a probabilidade de você ganhar o jogo?

O Teorema da Probalidade Total

Exercício: 1 Solução

- Partição: 3 tipos dos oponentes:

$$P(A_1) = 0.5, P(A_2) = 0.25 \; e \; P(A_3) = 0.25$$

- Para cada um dos cenários, vc sabe a sua pr de ganhar (B):

$$P(B|A_1) = 0.3, P(B|A_2) = 0.4 \; e \; P(B|A_3) = 0.5$$

- Probabilidade de ganhar, usando o teorema da probabilidade total:

$$P(B) = P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + P(B|A_3) * P(A_3)$$
$$P(B) = 0.3 * 0.5 + 0.4 * 0.25 + 0.5 * 0.25$$
$$P(B) = 0.375$$

O Teorema da Probalidade Total

Exercício 2: Você joga um dado de 4 lados. Se o resultado for 1 ou 2, você joga o dado mais uma vez, caso contrário, você para. Qual a probabilidade da soma dos dados ser de pelo menos 4?

O Teorema da Probalidade Total

- Exercício 2: Solução

  Partição:

  - A_1: 1° dado = 1, $P(A_1) = 0.25$

    Dado $A_1 \rightarrow$ A soma total $\geq$4 se 2° dado = 3,4: $P(B|A_1) = 0.5$
  - A_2: 1°dado = 2, $P(A_2) = 0.25$

    Dado $A_2 \rightarrow$ A soma total $\geq$ 4 se 2° dado = 2,3,4: $P(B|A_2) = 0.75$
  - A_3: 1° dado=3, $P(A_3) = 0.25$

    Dado $A_3 \rightarrow$ A soma total $\geq$ 4: nunca! $P(B|A_3)=0$
  - A_4: 1° dado=4, $P(A_4) = 0.25$

    Dado $A_4 \rightarrow$ A soma total $\geq$ 4 se 1° dado = 4. Sempre! $P(B|A_4)=1$

    $P(B) = 0.5 * 0.25 + 0.75 * 0.25 + 0 * 0.25 + 1 * 0.25$

    $P(B) = 0.5625$

Teorema de Bayes

- O teorema de Bayes é usado para rever probabilidades previamente calculadas quando novas informações aparecem. Antes da informação nova, tínhamos a probabilidade "a priori", depois de incorporar a nova informação, temos a probabilidade "a posteriori".

- Matematicamente, nada mais é do que uma extensão da probabilidade condicional e teorema da probilidade total.

- Usada para inferência: existem várias causas diferentes para um evento (efeito observado). Estamos interessados em apenas uma das causas...

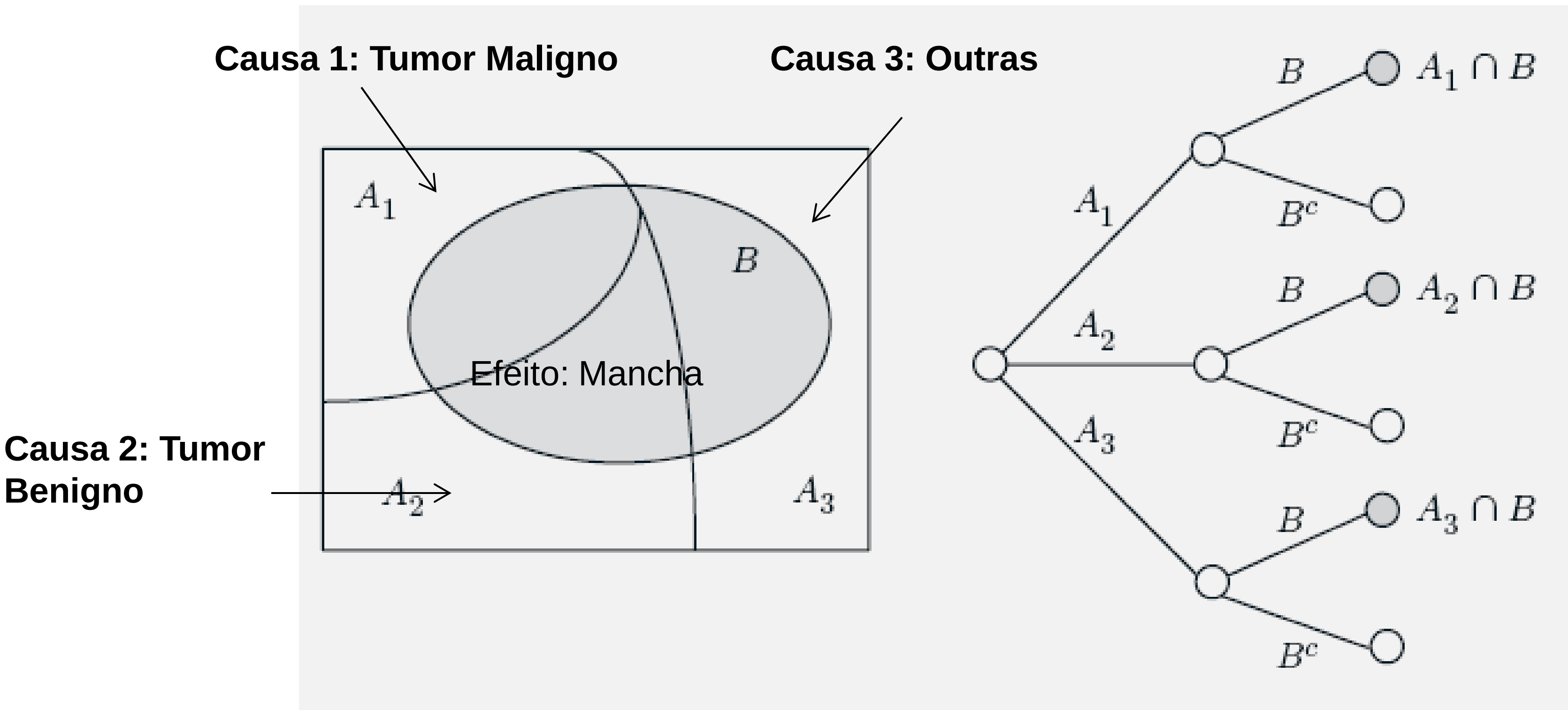
  Observo o evento (efeito). Qual a probabilidade de o efeito ser devido a causa de interesse?

Teorema de Bayes

- Observamos um mancha no raio-x do pulmão de um paciente.
- Sabemos que a mancha pode ser causada por diversos fatores:
  - Tumor maligno
  - Tumor benigno
  - Outras causas (falha de medição)
- Queremos investigar a chance do paciente ter um tumor maligno, dado o resultado do exame.

Teorema de Bayes

- P(Tumor maligno dado mancha no raio-x)?



Teorema de Bayes

- Nós temos uma probabilidade « a priori » para eventos $\{A_i\}_{i=1}^{n}$ partição do espaço amostral:
  - $P(A_i)$: probabilidades das causas (maligno, benigno, etc…)

- Para cada A_i, nós sabemos $P(B|A_i)$: para cada causa, sabemos a probabilidade de observarmos o efeito.

- Mas na verdade nós queremos computar $P(A_i|B)$: probabilidade da causa (maligno), se efeito (mancha) foi observado.

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

Teorema de Bayes

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

$$P(A_i|B) = \frac{P(B|A_i) * P(A_i)}{P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + \cdots + P(B|A_n) * P(A_n)}$$

Em que:

A_i = i^{esimo} evento de n eventos mutuamente excludentes e coletivamente exaustivos (partição).

B = novo evento observado que talvez mude a chance de A_i ocorrer.

$P(A_i)$ é chamada de probabilidade "a priori". $P(A_i|B)$ é chamada de probabilidade "a posteriori"

Teorema de Bayes

Exercício: O puzzle do falso negativo

Um teste para uma doença rara (ex AIDS) dá o resultado "correto" 95% das vezes:

- Se a pessoa tem a doença, o resultado é positivo 95% das vezes
- Se a pessoa não tem a doença, o resultado é negativo 95% das vezes

A incidência da doença na população é de 0.001 (1 em mil), ou seja, uma pessoa em mil tem a doença

Dado que uma pessoa fez o teste e o resultado deu positivo, qual a probabilidade de a pessoa estar realmente doente?

Teorema de Bayes

- Exercício: O puzzle do falso negativo

Temos informações sobre probabilidade de resultados de teste dado que observamos a saúde do paciente…

Queremos saber sobre a saúde de um paciente, dado o resultado de um teste.

Sejam P: resultado do teste é positivo

D: o paciente está doente

$P(D|P) = ?$

$$P(D|P) = \frac{P(P|D) * P(D)}{P(P)} = \frac{0.95 * 0.001}{P(P)} = \frac{0.95 * 0.001}{0.05 * 0.999 + 0.95 * 0.001} = 0.0187$$
$$= 1.87\%!!$$

Veja que o resultado positivo do teste fez com que revisássemos para cima a probabilidade de o paciente estar doente: $P(D|P) = 1.87\%$ em comparação a $P(D) = 0.1\%$. Mas este valor está longe de 95%!!

Teorema de Bayes

Exercício: Uma loja vende 3 marcas diferentes de aparelhos de som (AS). Dessas vendas, 50% são da marca 1 (a mais barata), 30% são da marca 2 e 20% são da marca 3.

Cada fabricante oferece um ano de garantia. Sabe-se que 25% dos AS da marca 1 necessitam de reparos de garantia, enquanto os percentuais correspondentes para as marcas 2 e 3 são 20% e 10%, respectivamente.

a) Qual é a probabilidade de que um comprador selecionado aleatoriamente compre um AS da marca 1 que precise de reparo durante a garantia?

b) Qual é a probabilidade de que um comprador selecionado aleatoriamente posssua um aparelho que necessite de reparos durante a garantia?

c) Se um cliente voltar à loja com um AS que precise de reparos em garantia, qual é a prob. de ele ser da marca 1? E da marca 2? E da marca 3?

Independência

- Dois eventos, A *e* B, são independentes se e somente se:

$$P(A|B) = P(A) \text{ ou } P(B|A) = P(B)$$

Definição!

- Os eventos A e B são independentes quando *a probabilidade de ocorrência de um evento não é afetada pela ocorrência do outro evento.*

Teorema

- Pela definição de prob. condicional, A é independente de B se:

$$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A) * P(B)$$

Esta formulação é mais usada para testar se dois eventos são independentes.

Independência

- Algumas vezes é fácil intuir quando eventos são independentes:
  - A ocorrência dos eventos é governada por processos físicos distintos que não interagem entre si. Ex: Faces superiores do lançamento de 2 dados balanceados.

- É difícil visualizar eventos independentes no espaço amostral. Confusão frequente: achar que eventos mutuamente excludentes são independentes. Isto não é verdade: eventos mutuamente excludentes são necessariamente dependentes:

$$P(A \cap B) = 0 \neq P(A) * P(B).$$

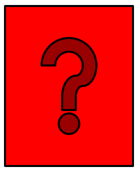
Independência

Exercício 1: Lançamento sucessivo de 2 dados de 4 lados.

16 resultados equiprováveis: prob 1/16.

Os eventos A_i = {Resultado do primeiro dado é i} e
B_j = {Resultado do segundo dado é j} são independentes?

Independência



- Solução:

temos,

$P(A_i \text{ e } B_j) = P(\text{ resultado: } (i,j)\ ) = 1/16$

$P(A_i) = P(\text{ resultados } (i,1), (i,2), (i,3) \text{ e } (i,4)\ ) = 4/16$

$P(B_j) = P(\text{ resultados } (1,j), (2,j), (3,j) \text{ e } (4,j)\ ) = 4/16$

Assim, como

$$P(A_i) * P(B_j) = \frac{4}{16} * \frac{4}{16} = \frac{1}{16} = P(A_i \cap B_j)$$

podemos afirmar que os eventos A_i e B_j são independentes.

Independência

Exercício 2: Lançamento sucessivo de 2 dados de 4 lados.

16 resultados equiprováveis: prob 1/16.

Os eventos A = {Resultado do primeiro dado é 1} e

B= {Soma dos dois dados é 5} são independentes?

Independência

- Solução:

temos ,

P(A e B) = P(resultado: *(1,4)*) = 1/16

P(A) = P(resultados *(1,1)*, *(1,2)*, *(1,3)* e *(1,4)*) = 4/16

P(B) = P(resultados *(1,4)*, *(2,3)*, *(3,2)* e *(4,1)*) = 4/16

Como

$$P(A) * P(B) = \frac{4}{16} * \frac{4}{16} = \frac{1}{16} = P(A \cap B)$$

São independentes? SIM!

Independência

Exercício 3: Lançamento sucessivo de 2 dados de 4 lados.

16 resultados equiprováveis: prob 1/16.

Os eventos A = {Máximo valor das faces é 2} e

B= {Mínimo valor das faces é 2} são independentes?

Independência

- Solução:

temos,

P(A e B) = P(resultado: *(2,2)*) = 1/16

P(A: Max=2) = P(resultados *(1,2), (2,1) e (2,2)*) = 3/16

P(B: Min=2) = P(resultados *(2,2), (2,3), (2,4), (3,2) e (4,2)*) = 5/16

como

$$P(A) * P(B) = \frac{3}{16} * \frac{5}{16} = \frac{15}{16^2} < \frac{1}{16} = P(A_i \cap B_j)$$

são independentes? NÃO!

Independência

- Podemos extender para indepêndencia entre mais de dois eventos…

- Os eventos $A_1, A_2, \ldots, A_n$ são independentes se e somente se:
  Para qualquer subconjunto K de {1, 2, …, n}:

Definição!

$$P\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} P(A_i)$$

- Então temos que verificar todas a combinações possíveis!

Independência

- Para 3 eventos A, B e C, precisamos verificar se:

$$\left.\begin{aligned} P(A \cap B) &= P(A) * P(B) \\ P(A \cap C) &= P(A) * P(C) \\ P(B \cap C) &= P(B) * P(C) \end{aligned}\right\} \quad \text{Independência 2 a 2}$$

e

$$P(A \cap B \cap C) = P(A) * P(B) * P(C)$$

Independência

Exercício: Um atirador acerta 80% de seus disparos e outro (na mesmas condições de tiro), 70%.

Qual é a probabilidade de algum dos 2 atiradores acertarem o alvo se ambos disparam simultaneamente e de forma independente?

Considere que o alvo foi acertado quando pelo menos uma das duas balas tenha feito impacto no alvo.

Independência

Exercício: Um atirador acerta 80% de seus disparos e outro (na mesmas condições de tiro), 70%.

Solução:

Pelo menos 1 acerta: A_1 acerta ou A_2 acerta ou ambos.

P(ao menos 1) = 0.8 + 0.7 – P(os dois acertam)

= 0.8 + 0.7 – 0.8*0.7 = 0.94

ou

P(ao menos 1) = 1 – P(nenhum dos 2 acerta)

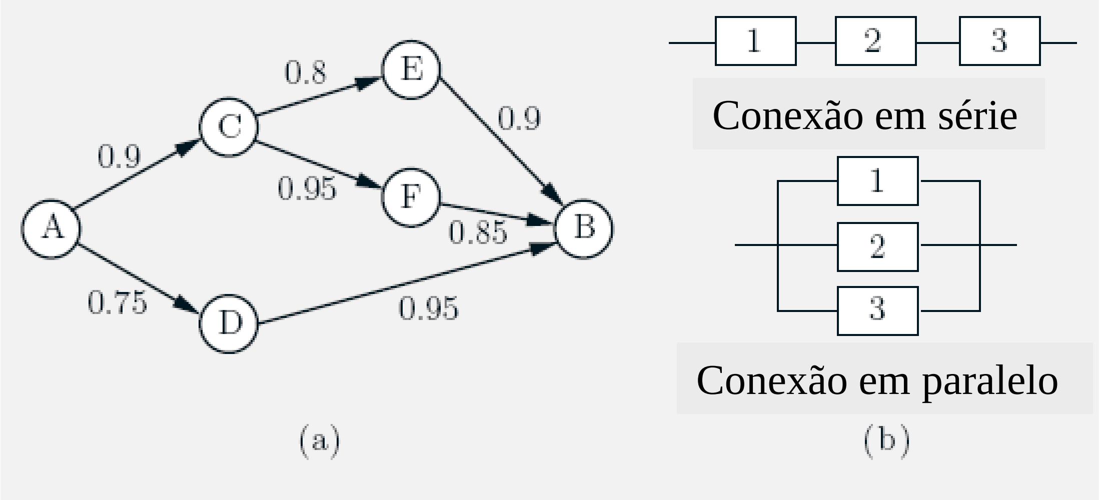
= 1 – 0.2*0.3 = 0.94

Atenção!

Este truque ($P(A \cup B) = 1 - P(A^c \cap B^c) = 1 - P(A^c) * P(B^c)$) é muito útil!!

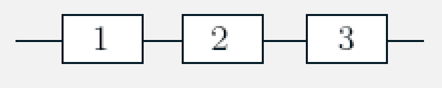
Independência

- Confiabilidade de um sistema conectando A a B:

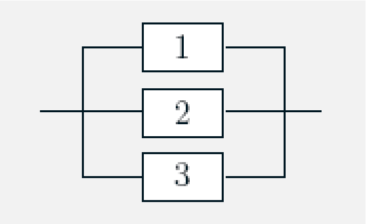


- As falhas nos links são independentes
- Números indicam que probabilidades do link funcionar.

Independência

(1) Para componentes em série  :

- P(subsistema funcione) = $p_1 * p_2 * \ldots * p_n$

(2) Para componentes em paralelo  :

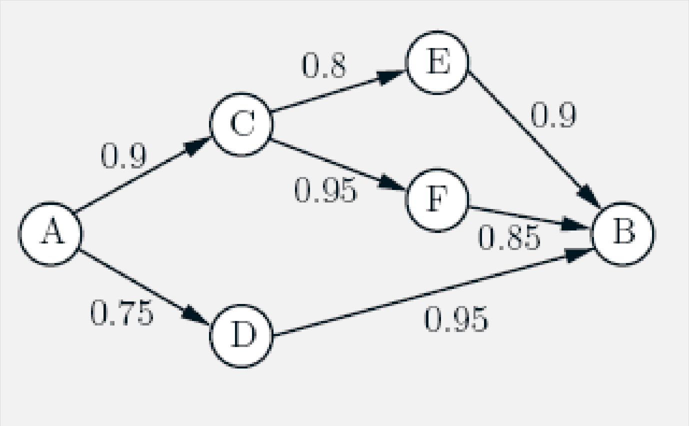
P(subsistema funcione) = 1 – P(falhe) = 1 – P(todos falham)

= 1 –P(1 falha e 2 falha e ... n falha)

= $1 - (1-p_1)*(1-p_2)*\ldots*(1-p_n)$

Independência

Exercício 1: Qual a probabilidade que o sistema abaixo funcione?



Independência

Exercício 2: Amostras de uma peça de alumínio fundido são classicadas com base no acabamento (em micropolegadas) da superfície e nas medidas de comprimento. Os resultados de 100 peças são resumidos a seguir:

		Comprimento	
		Excelente (B)	**Bom**
Acabamento Superfície	**Excelente (A)**	75	7
	Bom	10	8

Seja A: evento em que disco tem excelente acabamento e B: evento em que disco tem comprimento considerado excelente.

Os eventos A e B são independentes?

Independência

Exercício 3: A probabilidade de que um espécime de laboratório contenha altos níveis de contaminação é de 10%. Cinco amostras são verificadas, sendo elas independentes:

- Qual é a probabilidade de que nenhum espécime contenha altos níveis de contaminação?
- Qual é a probabilidade de que pelo menos um espécime contenha altos níveis de contaminação?
- Qual é a probabilidade de que exatamente um espécime contenha altos níveis de contaminação?

Resumo

Nesta parte, vimos:

- Conceitos básicos de probabilidade.
  - Espaço amostral e eventos, propriedades de conjuntos.
- Funções probabilidade:
  - Probabilidade clássica, probabilidade frequentista
  - Regra da adição
  - Técnicas de contagem
- Probabilidade Condicional
  - Regra do Produto e Teorema da Probabilidade Total
  - Teorema de Bayes
  - Indepedência

Perguntas Recapitulativas

- O que é um experimento aleatório?
- O que é espaço amostral?
- Qual a diferença entre um evento simples e um evento composto?
- Qual é a diferença entre eventos mutuamente excludentes e eventos coletivamente exaustivos?
- O que é uma partição do espaço amostral?
- Qual a diferença entre a probabilidade clássica e a probabilidade frequentista?
- Como você pode usar a regra da adição para encontrar a ocorrência do evento A ou B?

Perguntas Recapitulativas

- Quando o Teorema da probabilidade total é útil para calcular a probabilidade de um evento B?
- No teorema de Bayes, de que modo a probabilidade "a priori" difere da probabilidade revisitada ("a posteriori")?
- De que modo a probabilidade condicional se relaciona ao conceito de independência Estatística?