

Ciência dos Dados

Teoria da Probabilidade

Magalhães e Lima, Capítulo 2

Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- Traduzir informações descritas em problemas práticos fazendo uso da teoria da probabilidade.
- Aplicar o Teorema de Bayes dentro de contextos práticos.

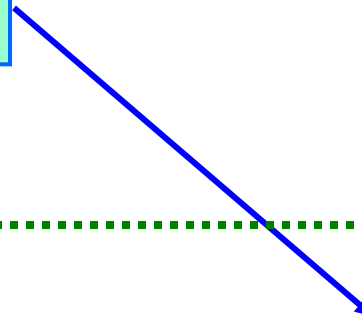
**Acompanhe, previamente, o PLANO DE AULA
no BLACKBOARD!**

Novos passos...

Coleta de dados



Análise Exploratória



Probabilidades



Variáveis
Aleatórias

Resultados básicos

Probabilidade Condicional

Independência

Teorema de Bayes

Conceitos básicos de probabilidade

Denomina-se **fenômeno** (ou experimento) **aleatório** à situação ou acontecimento cujos resultados não podem ser previstos com certeza.

A teoria das probabilidades estabelece modelos matemáticos para os fenômenos aleatórios, tais como:

- ✓ Quais as chances das vendas crescerem se abaixarmos os preços?
- ✓ Qual a probabilidade do projeto terminar no prazo?
- ✓ Quais as chances de um novo investimento ser lucrativo?

Nestes casos, modelos podem ser estabelecidos para quantificar as incertezas das diversas ocorrências.

Conceitos básicos de probabilidade

- ☆ **Espaço amostral (Ω ou S)** - conjunto de resultados possíveis de um experimento aleatório.
- ☆ **Evento (A, B, \dots)** - qualquer subconjunto do espaço amostral.

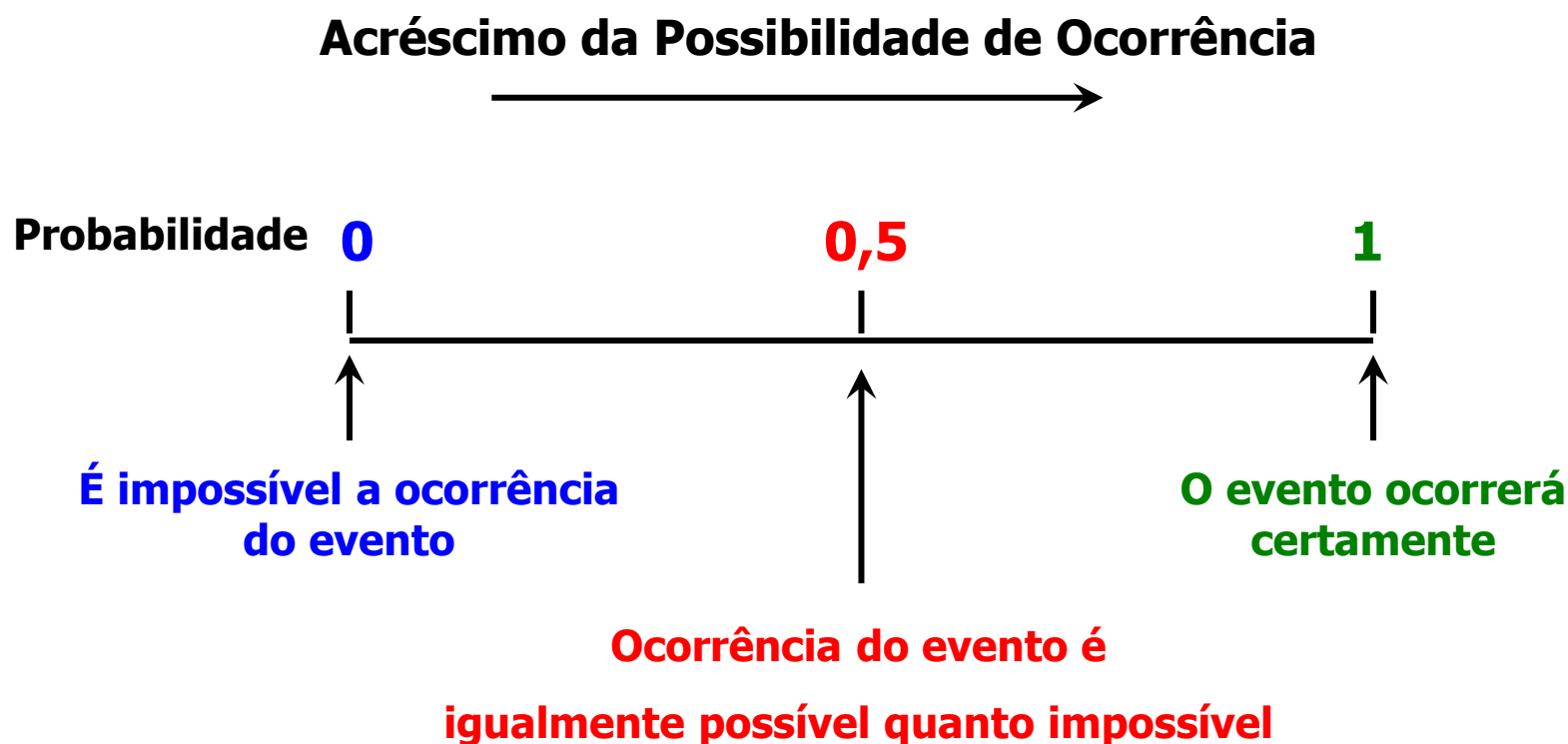
Exemplo: Estudar o preço de uma determinada passagem de avião daqui uma semana. Qual a probabilidade do preço da passagem não aumentar?

Espaço amostral = {aumento de preço, preço estável, diminuição no preço}

Evento = {preço estável, diminuição no preço}

Conceitos básicos de probabilidade

Probabilidade: medida da incerteza da ocorrência de um evento



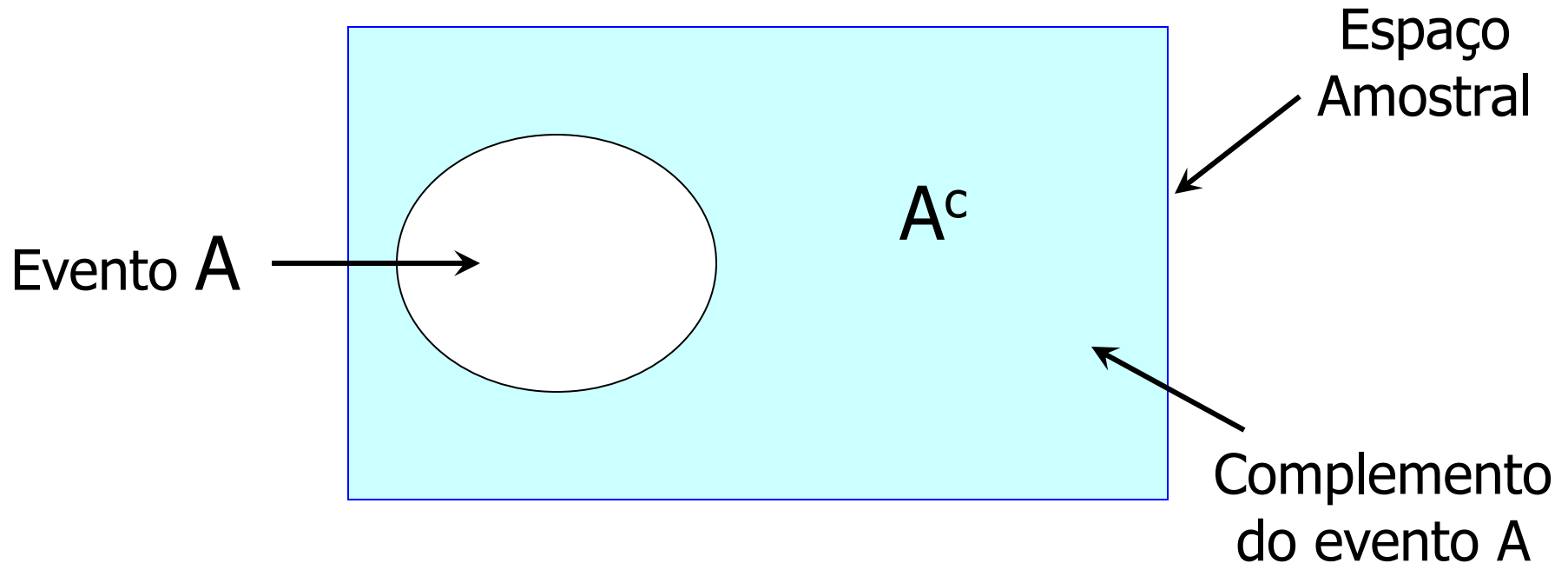
Os principais métodos de estabelecer probabilidades aos eventos analisados são os métodos clássico, frequentista e subjetivo.

Conceitos básicos de probabilidade

- $P(\Omega) = 1$ (evento certo)
- A : evento qualquer pertencente a Ω
 $0 \leq P(A) \leq 1$

Definição:

Evento Complementar

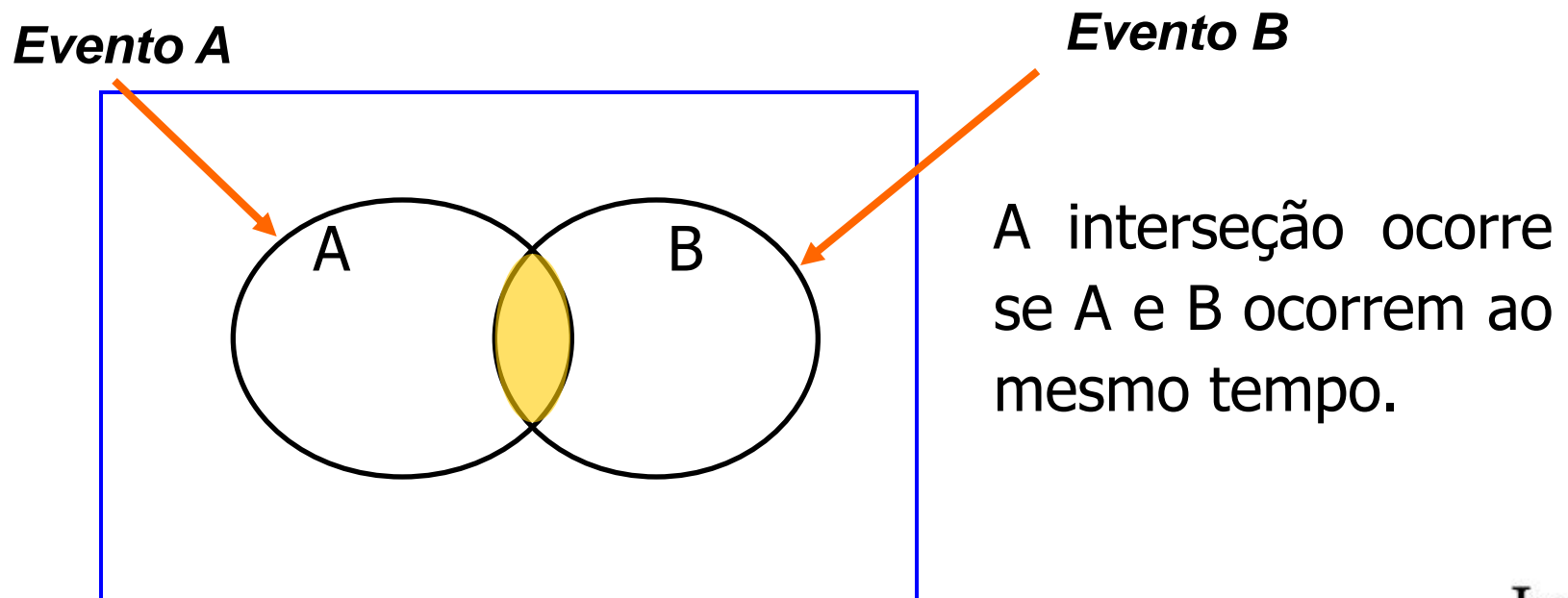


$$P(A^c) = 1 - P(A)$$

Definição:

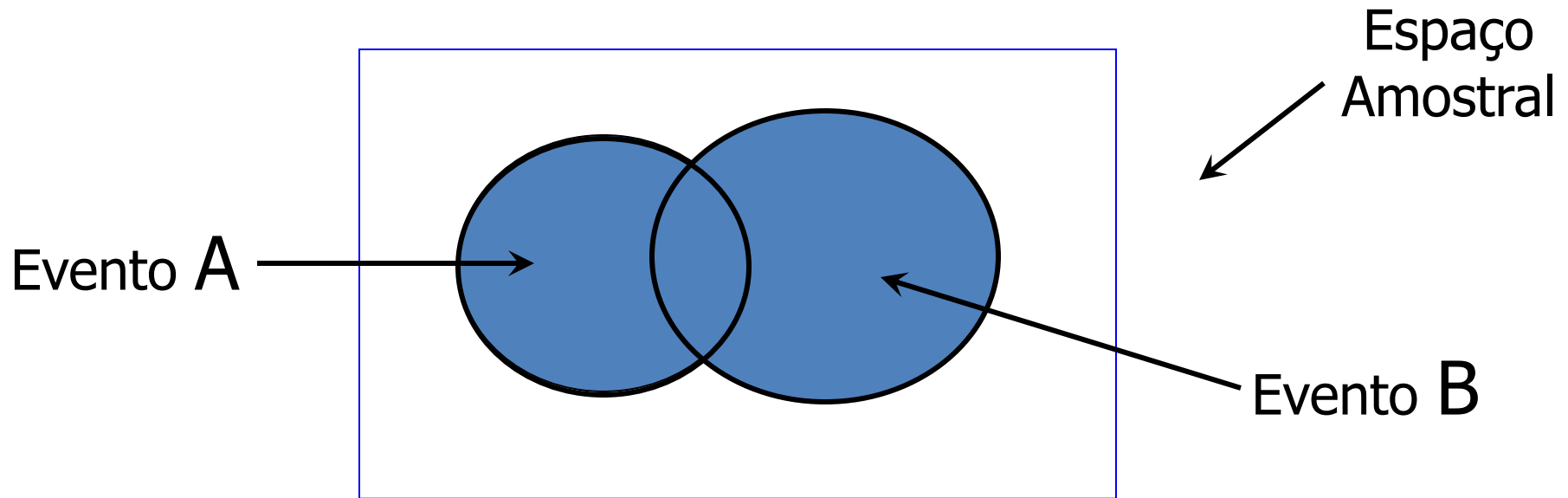
Interseção

$\{A \cap B\}$: conjunto dos pontos do espaço amostral que pertencem, simultaneamente, aos eventos A e B.



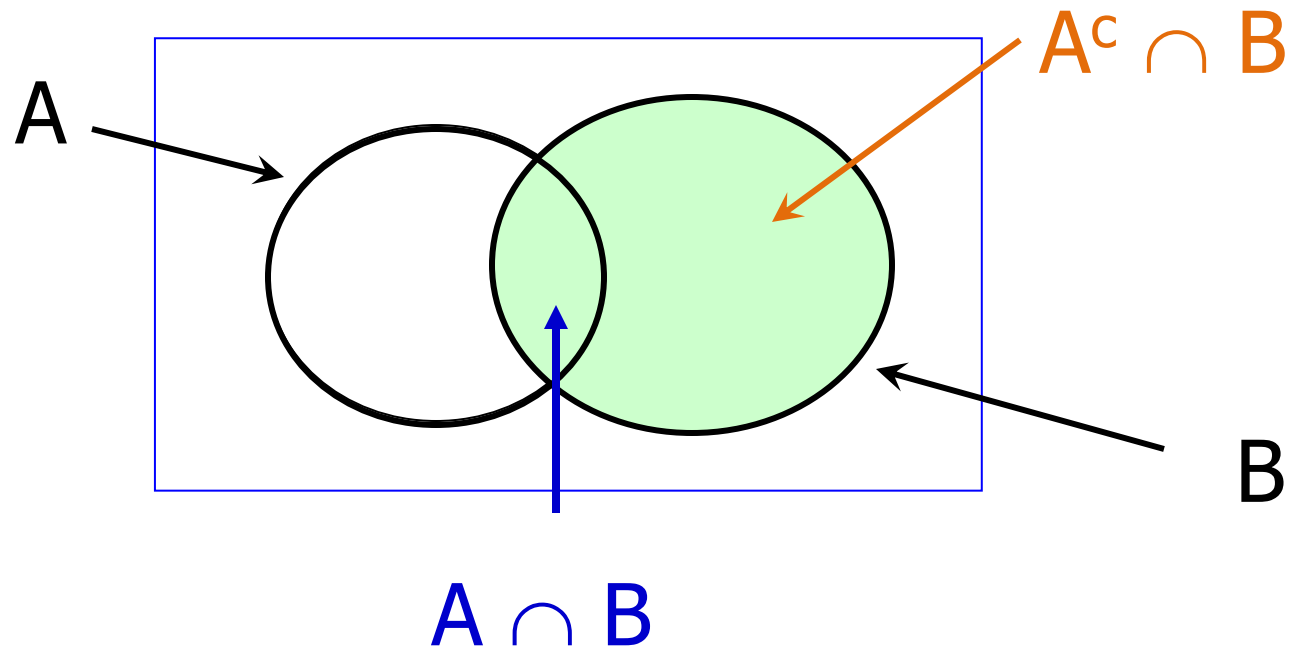
Definição:

União de 2 eventos

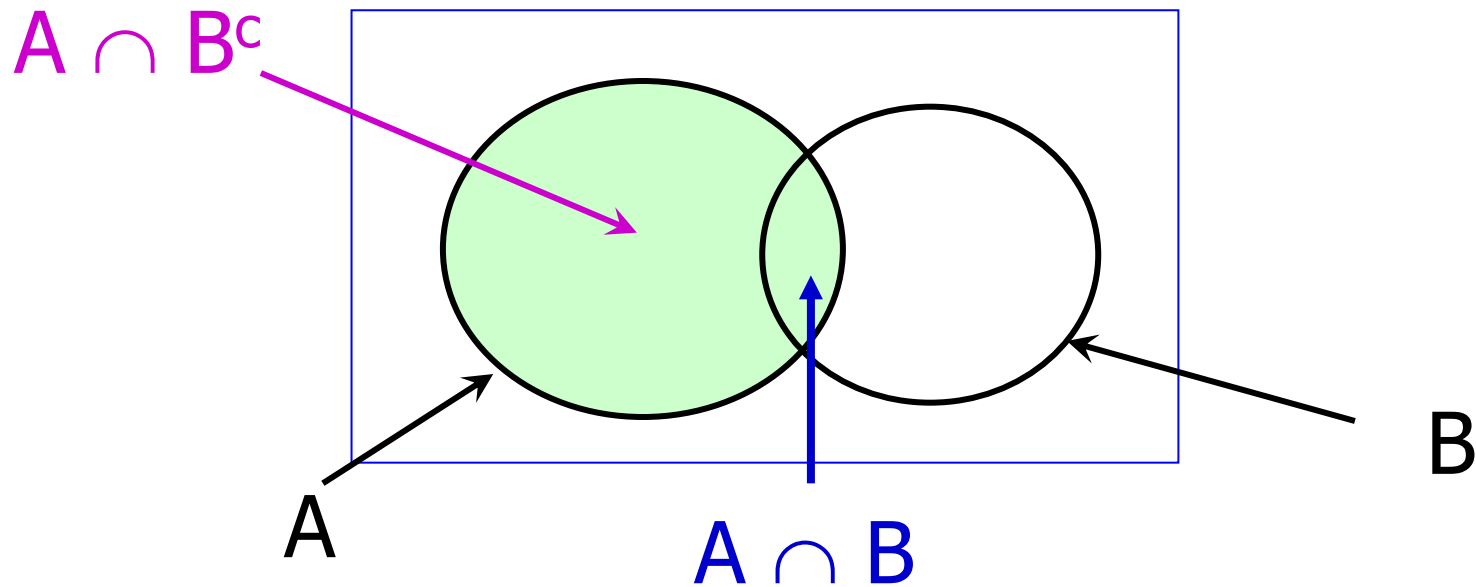


$$P(A \text{ ou } B) =$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



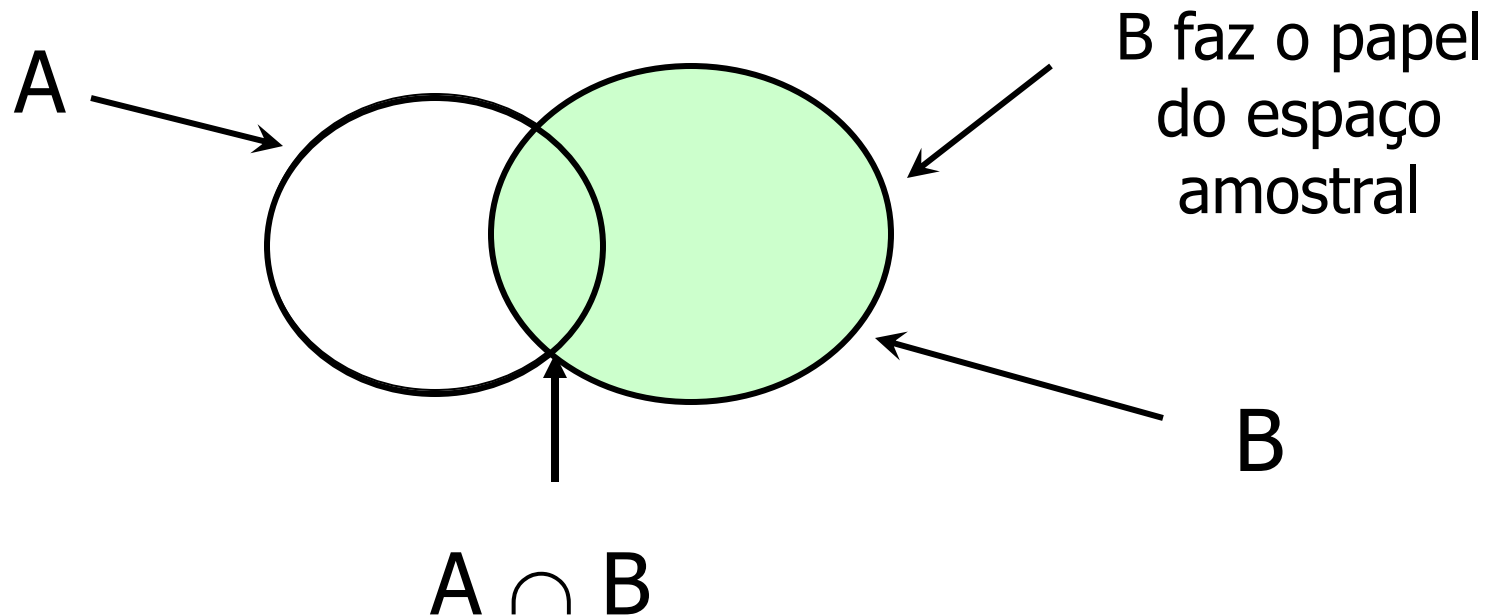
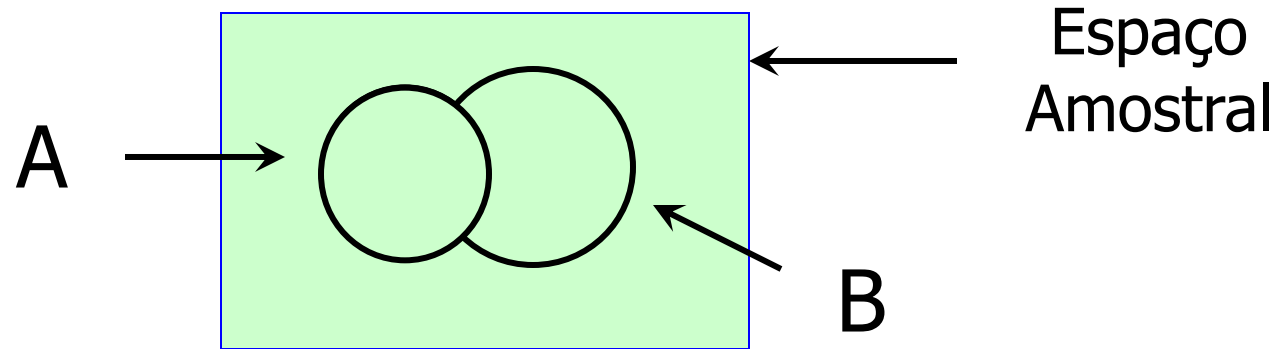
$$P(B) = P(A \cap B) + P(A^c \cap B)$$



$$P(A) = P(A \cap B) + P(A \cap B^c)$$

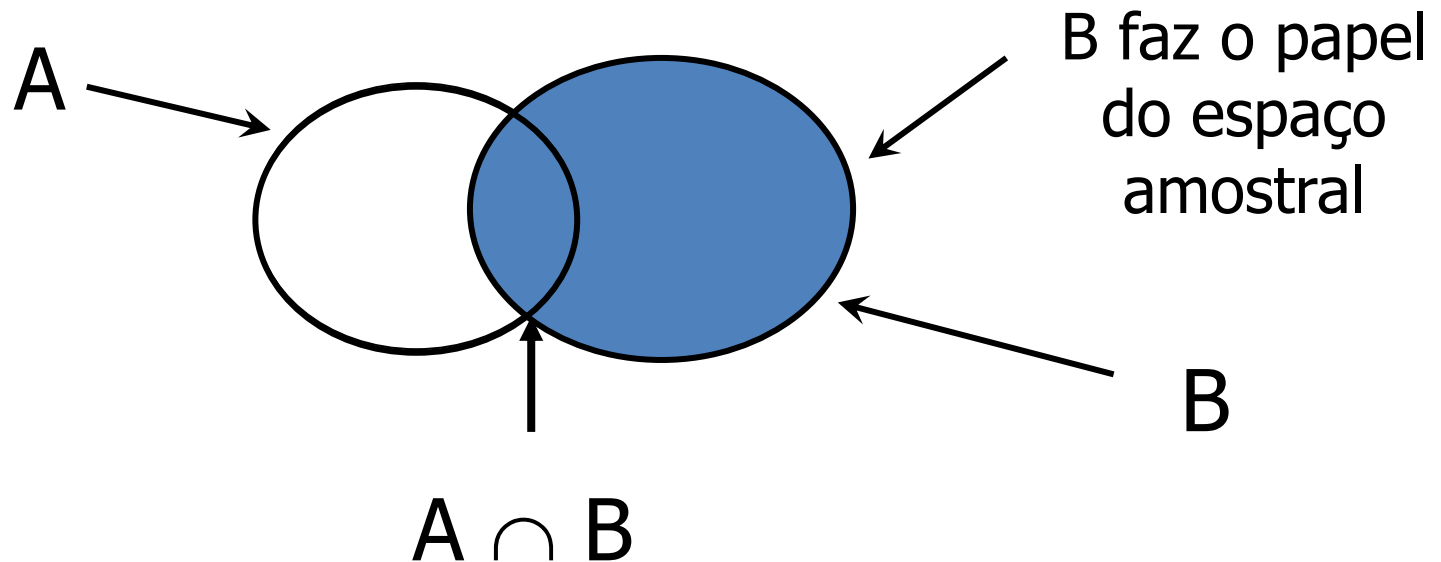
Definição:

Probabilidade Condicional



Definição:

Probabilidade Condicional

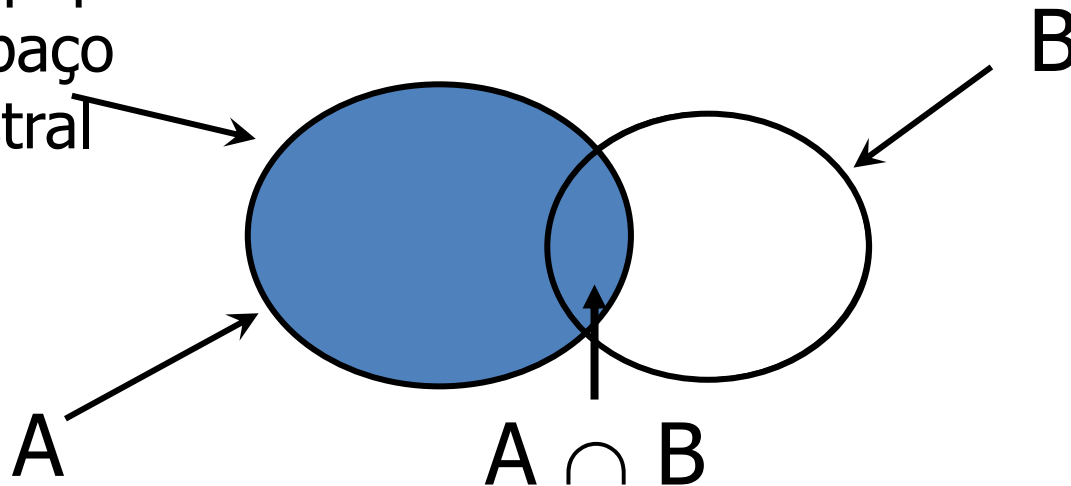


$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definição:

Probabilidade Condicional

A faz o papel
do espaço
amostral



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

Probabilidade Condicional

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

ou

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

Exemplo da Netflix

A Netflix considera diversas variáveis antes de sugerir uma lista de filmes a um usuário. Entre elas, considere o período que usuário acessa a conta (Diurno e Noturno) e, por consequência, o gênero sugerido ao usuário (Romance e Ação).

A tabela abaixo apresenta as **frequências absolutas** que relacionam essas duas variáveis.

Período	Gênero		Total
	Romance	Ação	
Diurno	35	15	50
Noturno	45	105	150
Total	80	120	200

Exemplo da Netflix

Distribuição conjunta: avaliação do comportamento conjunto de 2 variáveis.

Distribuição marginal: avaliação do comportamento individual de cada uma das ações.

Período	Gênero		Total
	Romance	Ação	
Diurno	35	15	50
Noturno	45	105	150
Total	80	120	200

Distribuição conjunta das ações

Distribuição marginal das ações

Exemplo da Netflix

Período	Gênero		Total
	Romance	Ação	
Diurno	35	15	50
Noturno	45	105	150
Total	80	120	200

Perguntas:

- Qual a probabilidade de um acesso ser no período diurno?
- Qual a probabilidade de um acesso não ser no período diurno?
- Qual a probabilidade de ser sugerido filme de Romance?
- Qual a probabilidade de ser sugerido filme de Romance e período ser diurno?

Exemplo da Netflix

Período	Gênero		Total
	Romance	Ação	
Diurno	0,175	0,075	0,250
Noturno	0,225	0,525	0,750
Total	0,400	0,600	1,000

Perguntas:

- Qual a probabilidade de acesso ser no período diurno ou ser sugerido filme de romance?
- No período diurno, qual a probabilidade de sugerir Romance?
- Se filme sugerido é de Romance, qual a probabilidade do acesso ter sido diurno?



EXERCÍCIOS

Exercício 1

A probabilidade de que o preço dos combustíveis aumente no mês vindouro é estimada em 0,4.

Se isto ocorrer, a probabilidade de que os preços dos transportes coletivos também aumentem é de 0,5; caso contrário, esta probabilidade é de 0,1.

Se naquele mês o preço das passagens de fato subirem, qual a probabilidade de os preços dos combustíveis não terem sofrido majoração?

Exercício 2

Uma revendedora de veículos trabalha com duas marcas de automóveis: A e B. 70% dos que adquirem carros populares, escolhem a marca A. Dentre os que adquirem carros não populares, 80% compram A. Sabe-se que 60% das vendas são de carros populares.

a. Qual é a probabilidade de um consumidor comprar um carro da marca A? 0,74

b. Sabendo que uma pessoa comprou um carro da marca A, qual é a probabilidade de ter sido um carro popular? 0,568

Exercício 3

80% dos funcionários de uma agência financeira são homens.

Nos últimos 2 anos, dos funcionários promovidos, 89% eram homens. Das mulheres, 15% foram promovidas.

Qual é a probabilidade de alguém ser promovido num prazo de 2 anos? 0,273

Se uma pessoa é do sexo masculino, qual é a probabilidade de ser promovida? E se for do sexo feminino? 0,30 e 0,15

Um comitê sindical levantou a questão de que estava havendo discriminação.

A acusação procede?

Exercício 4

Analizando-se o histórico de acidentes provocados por clientes de uma seguradora, sabe-se que 10% dos contratos resultam em sinistros. A probabilidade de um homem causar um sinistro é 12%, enquanto que a probabilidade de uma mulher causar um sinistro é de 0,06.

- a. Sorteado um contrato ao acaso, qual é a probabilidade de ser de um homem? 0,667
- b. Sorteado um contrato no qual ocorreu um sinistro, qual é a probabilidade de ter sido causado por uma mulher? 0,20

Exercício 5

Uma pesquisa de assinantes de revista mostrou que 45,8% alugaram um carro nos últimos 12 meses por razões comerciais, 54% alugaram um carro durante os últimos 12 meses por razões pessoais e 30% alugaram um carro nos últimos 12 meses tanto por razões comerciais como por razões pessoais.

- a. Qual é a probabilidade de um assinante ter alugado um carro durante os últimos 12 meses por razões comerciais ou pessoais? 0,698
- b. Qual é a probabilidade de um assinante não ter alugado um carro durante os últimos 12 meses por razões comerciais e nem pessoais? 0,302

Exercício 6

O gerente de um posto de gasolina sabe da sua experiência que 80% dos clientes usam cartão de crédito quando compram gasolina.

Admita independência entre clientes.

Qual é a probabilidade dos dois próximos clientes comprarem gasolina usando cartão de crédito ?

E de pelo menos 1 usar cartão de crédito?

A = evento de que o 1o. cliente use CC

B = evento de que o 2o. cliente use CC

Resposta: 0,64 e 0,96, respectivamente.

Exercício 7

Considerando apenas os países que ganharam pelo menos uma medalha nos Jogos Olímpicos de 2012, as seguintes probabilidades foram observadas ao analisar, conjuntamente, o índice de desenvolvimento humano ($IDH \geq 0,75$ ou $IDH < 0,75$) de 2011 e o tipo de medalha olímpica (ouro, prata ou bronze).

- ✓ A probabilidade de um atleta, que pertence a um destes países, ganhar uma medalha de ouro é de 31,5%. **$P(O)=0,315$**
- ✓ Se um país tem IDH igual ou superior a 0,75, então a probabilidade de um atleta ganhar uma medalha de ouro é 35,0% e, na mesma condição, a de um atleta ganhar uma medalha de prata é de 32,9%. **$P(O|A)=0,35$ e $P(P|A)=0,329$**
- ✓ Entre países com IDH inferior a 0,75, a probabilidade de um atleta ganhar uma medalha de ouro é 23,3%. **$P(O|A^c)=0,233$**
- ✓ Por fim, se um atleta ganhou uma medalha de bronze, a probabilidade de que ele pertença a um país com IDH inferior a 0,75 é de 38,4%. **$P(A^c|B)=0,384$**

Considerando apenas os países que ganharam pelo menos uma medalha nos Jogos Olímpicos de 2012, responda:

a) Qual a probabilidade de um atleta ser de um país com IDH igual ou superior a 0,75?

$P(A)=0,701$

b) Qual a probabilidade de uma medalha de prata pertencer a um atleta que representa um país com IDH igual ou superior a 0,75? **$P(A|P)=0,722$**

c) Com base nos resultados mencionados que consideram apenas países que ganharam pelo menos uma medalha nas Olimpíadas de Londres, o IDH de um destes países pode influenciar o tipo de medalha que um atleta pode ganhar? Justifique sua resposta considerando **informações numéricas**. **Como $P(O)$ é diferente de $P(O|A)$, já é o suficiente para mostrar que o tipo de medalha é dependente do IDH de um país.**

Exercício 8

Uma fábrica de parafusos tem exatamente três máquinas trabalhando na linha de produção, aqui, nomeadas de: M1, M2 e M3. A máquina M2 produz 28% do total de parafusos e a M3 é responsável por 39% da produção total. Para o setor de Controle de Qualidade, cada parafuso (produzido por uma das três máquinas) pode ser classificado como: perfeito, defeituoso recuperável ou defeituoso descartável.

Do total de parafusos produzidos, 81% são classificados como perfeitos.

Da produção total, 2% são produzidos pela máquina M2 e classificados como defeituosos descartáveis.

Dos parafusos produzidos pela máquina M1, 75,8% são classificados como perfeitos. Da produção da máquina M2, 92,8% são classificados como perfeitos.

Apenas considerando o que é produzido pela máquina M3, 17,9% são classificados como defeituosos recuperáveis. Entretanto, de todos os parafusos defeituosos recuperáveis, 41,7% são produzidos pela máquina M1.

Exercício 8

Responda:

- a) Calcule a probabilidade de um parafuso ser produzido pela máquina M2 e classificado como perfeito. **=0,260**
- b) Considerando que um parafuso é classificado como perfeito, qual a probabilidade dele ter sido produzido pela máquina M1? **=0,309**
- c) Qual a probabilidade de um parafuso ter sido produzido pela máquina M3 dado que esse foi classificado como perfeito? **=0,370**
- d) Calcule a probabilidade de um parafuso ser classificado como defeituoso recuperável. **=0,120**

APS 3 - Check...

- Ler a base de dados disponível
- Separar a base de dados em Treinamento e Teste na fração 3 para 1.
- Ter iniciado o código que guardar as palavras contidas das mensagens SPAM considerando o *dataset* Treinamento. Idem para as mensagens HAM.