Probabilidade discreta

Mário S. Alvim
(msalvim@dcc.ufmg.br)

Information Theory

DCC-UFMG
(2017/02)

Teoria de probabilidade: Introdução

- A **Teoria de Probabilidade** nasceu junto com a Análise Combinatória, ambas aplicadas ao estudo dos jogos de azar.
- A Teoria de Probabilidade é hoje fundamental em diversas áreas, incluindo:
  1. estatística e estudo de fenômenos complexos,
  2. aprendizado de máquina,
  3. inteligência artificial,
  4. física (termodinâmica, quântica),
  5. algoritmos probabilísticos,
  6. protocolos seguros,
  7. teoria da informação,
  8. ...
- Aqui nos concentraremos na **Teoria de Probabilidade Discreta**, que lida com **conjuntos enumeráveis**.

Introdução

Probabilidade discreta: Conjuntos enumeráveis

- A **probabilidade discreta** lida apenas com conjuntos enumeráveis.
- Um conjunto é um **conjunto enumerável** se
    1. ele é finito,

       ou

    2. existe uma bijeção entre seus elementos e o conjunto dos números naturais $\mathbb{N}$.
- Um conjunto é enumerável se é possível enumerar seus elementos, ou seja, se é possível produzir uma lista (possivelmente infinita) que contenha todos seus elementos.

  Por exemplo:

    1. São conjuntos enumeráveis: os naturais $\mathbb{N}$, os inteiros $\mathbb{Z}$ e os racionais $\mathbb{Q}$.
    2. São conjuntos não-enumeráveis: os reais $\mathbb{R}$ e qualquer intervalo contínuo de $\mathbb{R}$.

Probabilidade finita

- Um **experimento** ou **ensaio** é um procedimento que produz um resultado dentre vários resultados possíveis.
- O **espaço amostral** de um experimento é o conjunto de todos os resultados possíveis para o experimento.
- Um **evento** é um subconjunto do espaço amostral.
- A definição de Laplace para probabilidade é como se segue.

  Seja S um espaço amostral finito de resultados igualmente prováveis, e seja $E \subseteq S$ um evento.

  A **probabilidade (de Laplace) do evento** E é dada por

  $$p(E) = \frac{|E|}{|S|}.$$

  Ou seja, a probabilidade de um evento E é a razão entre o número de resultados que satisfazem E e o total de resultados possíveis.

Probabilidade finita

- $\boxed{\text{Exemplo 1:}}$ Uma urna contém 5 bolas vermelhas e 3 bolas azuis. Qual a probabilidade de uma bola retirada aleatoriamente da urna ser azul?

  Solução: O experimento consiste em retirar uma bola da urna.

  O espaço amostral S do experimento é o conjunto de todos os resultados possíveis, ou seja, um conjunto de 8 elementos (um para cada bola que pode ser retirada da urna).

  Estamos interessados no evento E em que uma bola azul é retirada. Este evento corresponde ao subconjunto de S contendo todas as bolas azuis, e E possui 3 elementos.

  Pela definição de Laplace, se todos os resultados do espaço amostral são igualmente prováveis, a probabilidade de E é dada por

$$P(E) = \frac{|E|}{|S|} = \frac{3}{8}.$$

◁

Probabilidade finita

- $\boxed{\text{Exemplo 2:}}$ Quando um dado de seis faces é rolado, qual a probabilidade de se obter a face 5 voltada para cima?

  Solução: O experimento consiste em jogar rolar um dado de seis faces.

  O espaço amostral S do experimento é o conjunto de todas as faces do dado que podem estar voltadas para cima, ou seja, o espaço amostral é o conjunto $S = \{1, 2, 3, 4, 5, 6\}$.

  Estamos interessados no evento $E = \{5\}$ correspondendo à face 5 cair voltada para cima.

  Pela definição de Laplace, se todos os resultados do espaço amostral são igualmente prováveis, a probabilidade de E é dada por

$$P(E) = \frac{|E|}{|S|} = \frac{|\{5\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{1}{6}.$$

$\triangleleft$

- $\boxed{\text{Exemplo 3:}}$ Quando dois dados de seis faces são rolados, qual a probabilidade da soma dos números de cada dado ser 7?

  Solução: O experimento consiste em rolar dois dados de seis faces.

  O espaço amostral S do experimento é o conjunto de todos os pares de dois números possíveis, sendo cada número do par o resultado de um dos dados. Como há 6 possibilidades para o resultado de cada dado, pela regra do produto o espaço amostral tem $6 \cdot 6 = 36$ possibilidades no total.

  Estamos interessados no evento E em que a soma dos números em cada dado é 7. Ou seja, $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$.

  Pela definição de Laplace, se todos os resultados do espaço amostral são igualmente prováveis, a probabilidade de E é dada por

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}.$$

$\triangleleft$

Probabilidade finita

- $\boxed{\text{Exemplo 4:}}$ Na Mega Sena, 6 dezenas são sorteadas aleatoriamente de um conjunto de dezenas de 01 a 60, e uma aposta é premiada se ela contiver as 6 dezenas sorteadas.

  Qual a probabilidade de uma aposta de 6 dezenas ser premiada?

  Solução: O experimento consiste em sortear 6 dezenas dentre 60.

  O espaço amostral S do experimento é o conjunto de todas combinações de 6 dezenas de um conjunto de 60, ou seja,

$$|S| = C(60, 6) = \frac{60!}{6!(60-6)!} = 50\,063\,860.$$

  Estamos interessados no evento E em que o sorteio coincide com uma aposta específica de 6 dezenas.

  Se todos os resultados do espaço amostral são igualmente prováveis, temos

$$P(E) = \frac{|E|}{|S|} = \frac{1}{50\,063\,860} \approx 0.00000002.$$

◁

- $\boxed{\text{Exemplo 5:}}$ Qual a probabilidade de que uma mão de Pôquer de 5 cartas seja um *full house*, ou seja, 3 cartas de um mesmo tipo e 2 cartas de um outro tipo?

  (Lembre-se de que num baralho normal há 52 cartas divididas em 4 naipes (copas, espadas, ouros e paus), e cada naipe contém uma carta de cada um dos 13 tipos existentes (ás, valete, dama, rei, ou um número entre 2 e 10).)

  Solução: O experimento consiste em retirar 5 cartas de um conjunto de 52 cartas do baralho.

  O espaço amostral S do experimento é o conjunto de todas combinações de 5 cartas retiradas de um conjunto de 52, ou seja,

$$|S| = C(52, 5) = \frac{52!}{5!(52-5)!} = 2\,598\,960.$$

Probabilidade finita

- $\boxed{\text{Exemplo 5:}}$ (Continuação)

Estamos interessados no evento E em que a mão consiste em um *full house*.

Para contar quantos resultados do experimento são um *full house*, note que um *full house* pode ser obtido através das seguintes etapas consecutivas:

1. Escolhem-se os 2 tipos de cartas dentre os 13 possíveis. Já que a ordem dos tipos importa (2 rainhas e 3 azes é diferente de 3 rainhas e 2 azes), há $P(13, 2) = 13!/(13 - 2)! = 156$ maneiras de se fazer isto.
2. Escolhem-se 3 naipes do primeiro tipo entre os 4 possíveis. Há $C(4, 3) = 4!/(3!(4 - 3)!) = 4$ maneiras de se fazer isto.
3. Escolhem-se 2 naipes do segundo tipo entre os 4 possíveis. Há $C(4, 2) = 4!/(2!(4 - 2)!) = 6$ maneiras de se fazer isto.

Logo, pela regra da multiplicação, o número total de maneiras de se obter um *full house* é

$$P(13, 2) \cdot C(4, 3) \cdot C(4, 2) = 3\,744.$$

Probabilidade finita

- $\boxed{\text{Exemplo 5:}}$ (Continuação)

  Assim, pela definição de Laplace, se todos os resultados do espaço amostral são igualmente prováveis, a probabilidade de E é dada por

  $$P(E) = \frac{|E|}{|S|} = \frac{3\,744}{2\,598\,960} \approx 0.0014.$$

  $\triangleleft$

Probabilidade finita: Combinação de eventos

- Segundo a definição de Laplace, um evento nada mais é que um subconjunto do espaço amostral.
- Portanto, é natural que a probabilidade da combinação de eventos possa ser calculada utilizando operações sobre conjuntos (união, interseção, complemento, etc.),

Probabilidade finita: Combinação de eventos

- O seguinte resultado associa a probabilidade de um evento à probabilidade de seu evento complementar.

- **Teorema:** Seja E um evento de um espaço amostral S. A probabilidade do evento $\overline{E}$, o complemento do evento E, é dada por

$$p(\overline{E}) = 1 - p(E).$$

Prova. Note que $|\overline{E}| = |S| - |E|$. Logo

$$p(\overline{E}) = \frac{|\overline{E}|}{|S|} = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E). \quad \square$$

- **Corolário:** Se E é um evento num espaço amostral S, então

$$p(E) + p(\overline{E}) = 1.$$

Probabilidade finita: Combinação de eventos

- $\boxed{\text{Exemplo 6:}}$ Uma sequência de 10 bits é aleatoriamente gerada.

  Qual a probabilidade de ao menos um dos bits ser 0?

  Solução: O espaço amostral S consiste no conjunto de todas as strings de 10 bits, logo $|S| = 2^{10} = 1\,024$.

  Estamos interessados no evento E em que ao menos um bit da string é 0.

  O evento oposto, $\overline{E}$, é aquele em que nenhum bit da string é 0. Como só há uma string sem bits 0 (a string em que todos os bits são 1), temos que $|\overline{E}| = 1$.

  Logo, podemos calcular $P(E)$ como

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{1\,024} = \frac{1\,023}{1\,024}.$$

◁

Probabilidade finita: Combinação de eventos

- O seguinte resultado associa a probabilidade da união de eventos às probabilidades de seus componentes.

- **Teorema:** Sejam E_1 e E_2 eventos de um espaço amostral S. Então

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2).$$

Prova. Pelo princípio da inclusão-exclusão de conjuntos,

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|.$$

Logo

$$\begin{aligned}
p(E_1 \cup E_2) &= \frac{|E_1 \cup E_2|}{|S|} \\
&= \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} \\
&= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} \\
&= p(E_1) + p(E_2) - p(E_1 \cap E_2). \quad \square
\end{aligned}$$

Probabilidade da combinação de eventos

- $\boxed{\text{Exemplo 7:}}$ Qual a probabilidade de um inteiro selecionado aleatoriamente entre os 100 primeiros inteiros ser divisível por 2 ou por 5?

  Solução: O espaço amostral S consiste no conjunto dos inteiros positivos menores ou iguais a 100, logo $|S| = 100$.

  Estamos interessados no evento E em que um inteiro é divisível por 2 ou por 5.

  Podemos escrever $E = E_1 \cup E_2$, onde E_1 representa o evento em que o número é divisível por 2, e E_2 representa o evento em que o número é divisível por 5.

  É fácil calcular que

  - $|E_1| = 50$ (pois há 50 números divisíveis por 2 no intervalo), e
  - $|E_2| = 20$ (pois há 20 números divisíveis por 5 no intervalo).

Probabilidade da combinação de eventos

- Exemplo 7: (Continuação)

Note que, nesse caso, $E_1 \cap E_2$ representa o evento em que o número é divisível por 2 e por 5 ao mesmo tempo, ou seja, o evento em que o número é divisível por 10.

É fácil calcular que

- $|E_1 \cap E_2| = 10$ (pois há 10 números divisíveis por 10 no intervalo).

Logo, podemos calcular $P(E)$ como

$$\begin{aligned} P(E) &= P(E_1 \cup E_2) \\ &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\ &= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} \\ &= \frac{3}{5}. \end{aligned}$$

Teoria de probabilidades

Teoria de probabilidades: Introdução

- Anteriormente definimos a probabilidade de um evento E em um espaço amostral S como Laplace fez, ou seja, como

$$P(E) = \frac{|E|}{|S|}.$$

  Esta definição de probabilidade assume que os resultados de um experimento são todos igualmente prováveis.

  (Além disso, ela só funciona quando o espaço amostral é finito.)

- Nesta seção vamos estudar como definir probabilidades em situações em que os resultados de um evento não são todos igualmente prováveis.

- Além disso, vamos estudar como as probabilidades de eventos se relacionam através de **condicionamento** e **independência**, e vamos introduzir o conceito de **variáveis aleatórias**.

Atribuindo probabilidades

- Seja S o espaço amostral de um experimento, tal que S é um conjunto enumerável.

  Podemos associar uma probabilidade $p(s)$ para cada resultado $s \in S$, desde que as seguintes condições sejam satisfeitas:

  1. a probabilidade de qualquer resultado é um real não negativo menor ou igual a 1:

  $$0 \leq p(s) \leq 1 \qquad \text{para todo } s \in S;$$

  2. com certeza um dentre os resultados deve ocorrer:

  $$\sum_{s \in S} p(s) = 1.$$

- Podemos ver p como uma função que mapeia elementos do espaço amostral para os reais no intervalo $[0, 1]$.

  Chamamos tal função de **distribuição de probabilidade**.

Atribuindo probabilidades

- A atribuição de probabilidades aos resultados de um experimento é normalmente feita segundo a **interpretação frequentista de probabilidade**.

- Segundo a interpretação frequentista, a probabilidade $p(s)$ associada a um resultado s do espaço amostral deve refletir a frequência com que o resultado s é esperado em relação o total de resultados possíveis para o experimento.

  Mais formalmente, $p(s)$ deve ser igual ao limite do número de vezes que o resultado s acontece, dividido pelo número de vezes que o experimento é realizado, quando o número de experimentos tende a infinito.

- Existem outras interpretações para probabilidades (e.g., Bayesiana), que estudaremos mais adiante no curso.

Atribuindo probabilidades

- Seja S um conjunto de n elementos.

  A **distribuição uniforme de probabilidade** atribui probabilidade $1/n$ para cada elemento de S.

Atribuindo probabilidades

- $\boxed{\text{Exemplo 8:}}$ Um dado justo (não-viciado) de 6 faces é rolado.

  Qual a distribuição probabilidade devemos atribuir aos possíveis resultados do experimento?

  Solução: Os resultados possíveis do experimento de se rolar um dado de 6 faces são dados pelo espaço amostral $S = \{1, 2, 3, 4, 5, 6\}$.

  Se o dado não está viciado, espera-se que, se rolarmos o dado infinitas vezes, obtenhamos cada uma das face em $1/6$ das vezes.

  Logo, atribuímos a distribuição de probabilidade uniforme aos resultados do experimento:

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

$\triangleleft$

Atribuindo probabilidades

- $\boxed{\text{Exemplo 9:}}$ Um dado viciado de 6 faces é rolado. Neste dado, a probabilidade da face de número 3 cair para cima é o dobro da probabilidade de qualquer outra face cair para cima.

  Qual a distribuição probabilidade devemos atribuir aos possíveis resultados do experimento?

  Solução: Os resultados possíveis do experimento de se rolar um dado de 6 faces são dados pelo espaço amostral $S = \{1, 2, 3, 4, 5, 6\}$.

  Para se atribuir probabilidade a cada um dos resultados possíveis, notamos que a face 3 ocorre com o dobro da frequência que qualquer outra face, ou seja:

$$p(1) = p(2) = p(4) = p(5) = p(6) = \frac{1}{2}p(3).$$

Atribuindo probabilidades

- $\boxed{\text{Exemplo 9:}}$ (Continuação)

Como a distribuição de probabilidade deve satisfazer

$$\sum_{s \in S} p(s) = 1,$$

temos que

$$p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1,$$

o que implica que

$$\frac{1}{2}p(3) + \frac{1}{2}p(3) + p(3) + \frac{1}{2}p(3) + \frac{1}{2}p(3) + \frac{1}{2}p(3) = 1,$$

de onde concluímos que $p(3) = 2/7$, e a distribuição deve ser

$$p(3) = \frac{2}{7} \quad \text{e} \quad p(1) = p(2) = p(4) = p(5) = p(6) = \frac{1}{7}.$$

$\triangleleft$

Atribuindo probabilidades

- Vimos como atribuir probabilidade a cada resultado do espaço amostral de um experimento.

  Falta estender a noção de probabilidade para eventos, que são subconjuntos do espaço amostral.

- A **probabilidade de um evento** E é a soma das probabilidades dos resultados em E:

$$p(E) = \sum_{s \in E} p(s).$$

Atribuindo probabilidades

- $\boxed{\text{Exemplo 10:}}$ Qual a probabilidade de obtermos um número par ao rolar um dado de 6 faces não-viciado?

  Solução: Em um exemplo anterior atribuímos a distribuição de probabilidade uniforme ao espaço amostral do experimento, ou seja:

  $$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

  Estamos interessados no evento $E = \{2, 4, 6\}$, ou seja, o evento de um número par ser obtido no lançamento do dado.

  A probabilidade deste evento é dada por:

  $$p(\{2, 4, 6\}) = p(2) + p(4) + p(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

  $\triangleleft$

Atribuindo probabilidades

- $\boxed{\text{Exemplo 11:}}$ Um dado viciado de 6 faces é rolado. Neste dado, a probabilidade da face de número 3 cair para cima é o dobro da probabilidade de qualquer outra face cair para cima.

  Qual a probabilidade de obtermos um número primo no lançamento deste dado?

  Solução: Em um exemplo anterior atribuímos a seguinte distribuição de probabilidade ao espaço amostral do experimento:

  $$p(3) = \frac{2}{7} \quad \text{e} \quad p(1) = p(2) = p(4) = p(5) = p(6) = \frac{1}{7}.$$

  Estamos interessados no evento $E = \{2, 3, 5\}$, ou seja, o evento de um número primo ser obtido no lançamento do dado (1 não é um número primo!).

  A probabilidade deste evento é dada por:

  $$p(\{2, 3, 5\}) = p(2) + p(3) + p(5) = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} = \frac{4}{7}.$$

Combinação de eventos

- Uma vez definida a probabilidade de um evento, podemos definir a probabilidade da combinação de eventos.

  Como eventos são subconjuntos do espaço amostral, é natural que a probabilidade da combinação de eventos seja definida em termos de operações sobre conjuntos (união, interseção, complemento, etc.)

- Como veremos, as definições de Laplace ainda são válidas para a combinação de eventos.

Combinação de eventos

- **Teorema:** Seja E um evento de um espaço amostral S. A probabilidade do evento $\overline{E}$, o complemento do evento E, é dada por

$$p(\overline{E}) = 1 - p(E)$$

Prova. Sabemos que

$$\sum_{s \in S} p(s) = 1.$$

Como cada resultado $s \in S$ ou pertence a E ou pertence a $\overline{E}$, temos que

$$\sum_{s \in E} p(s) + \sum_{s \in \overline{E}} p(s) = 1 \qquad \Rightarrow \qquad p(E) + p(\overline{E}) = 1,$$

de onde concluímos que

$$p(E) = 1 - p(\overline{E}).$$

□

Combinação de eventos

- **Teorema:** Sejam E_1 e E_2 eventos de um espaço amostral S. A probabilidade do evento $E_1 \cup E_2$ é dada por

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Prova. Para calcular $p(E_1 \cup E_2)$ devemos somar $p(s)$ para todo $s \in E_1 \cup E_2$.

Mas sabemos que cada elemento $s \in E_1 \cup E_2$ satisfaz exatamente uma das seguintes propriedades:

1. $s \in E_1$ e $s \notin E_2$,
2. $s \notin E_1$ e $s \in E_2$, ou
3. $s \in E_1$ e $s \in E_2$.

Logo, para calcular $p(E_1 \cup E_2)$ podemos somar $p(E_1) + p(E_2)$, e depois subtrair do resultado $p(E_1 \cap E_2)$, uma vez que os elemento da interseção de E_1 e E_2 foram contados duas vezes. $\square$

- **Teorema:** Sejam $E_1, E_2, \ldots$ uma sequência de eventos mutuamente disjuntos em um espaço amostral S. Então

$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i).$$

(Note que o teorema é válido quando $E_1, E_2, \ldots$ consiste em um número contável de eventos, mesmo que infinito.)

Prova. Exercício para o aluno!

Probabilidade condicional

- Muitas vezes estamos interessados na probabilidade de um evento E ocorrer dado que um outro evento F já ocorreu.

  Por exemplo, podemos estar interessados na:

  1. probabilidade de uma criança nascer daltônica dado que ela é do sexo feminino (o que faz sentido visto que daltonismo ocorre com frequências diferentes em cada sexo), ou
  2. probabilidade de um aluno ser aprovado em Teoria da Informação dado que ele não fez as listas de exercícios (o que é com certeza uma probabilidade bem baixa!).

  A isto chamamos de **probabilidade condicional** do evento E dado o evento F.

- **A probabilidade condicional de E dado F**, denotada por $p(E \mid F)$, é definida como

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

Probabilidade condicional

- $\boxed{\text{Exemplo 12:}}$ Um dado não-viciado de 20 faces (*RPG, alguém?*) é rolado.

  Qual a probabilidade de obtermos um resultado divisível por 3 dado que obtivemos um número par?

  Solução: O espaço amostral do experimento é $S = \{1, 2, 3, \ldots, 20\}$.

  Sejam E o evento de o resultado ser um número divisível por 3 e F o evento de o resultado ser um número par.

  Devemos calcular

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

Probabilidade condicional

- Exemplo 12: (Continuação)

  Note que

  - $F = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ (os resultados pares), e
  - $E \cap F = \{6, 12, 18\}$ (os resultados ao mesmo tempo pares e divisíveis por 3).

  Como o dado é não-viciado, a distribuição sobre resultados é uniforme e portanto

$$\begin{aligned} p(F) =& 10/20 = 1/2 \qquad \text{e} \\ (E \cap F) =& 3/20. \end{aligned}$$

  Logo

$$p(E \mid F) = \frac{3/20}{1/2} = \frac{3}{10}.$$

- $\boxed{\text{Exemplo 13:}}$ Qual a probabilidade de um casal com dois filhos ter dois garotos, dado que o casal tem ao menos um garoto?

  Assuma que a probabilidade de se ter um garoto ou uma garota é uniforme, ou seja, que $p(BB) = p(BG) = p(GB) = p(GG) = 1/4$, onde B representa um garoto (*boy*) e G representa uma garota (*girl*), e o espaço amostral do experimento é dado por $S = \{BB, BG, GB, GG\}$.

  Solução: Seja E o evento de o casal ter dois garotos, e seja F o evento de o casal ter ao menos um garoto.

  Devemos calcular

  $$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

  Note que $F = \{BB, BG, GB\}$ e $E \cap F = \{BB\}$.

Probabilidade condicional

- $\boxed{\text{Exemplo 13:}}$ (Continuação)

  Uma vez que a probabilidade é uniforme, temos que $p(F) = 3/4$ e $p(E \cap F) = 3/4$.

  Logo

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

$\triangleleft$

Independência

- Intuitivamente, dois eventos são independentes se a ocorrência de um não altera a probabilidade da ocorrência do outro.

- Formalmente, E **e** F **são eventos independentes** se a probabilidade de E dado F for igual à probabilidade de E:

$$p(E \mid F) = p(E).$$

- Note que pela definição de probabilidade condicional, $p(E \mid F) = p(E \cap F)/p(F)$, e podemos escrever $p(E \cap F) = p(F)p(E \mid F)$.

  Mas quando E e F são independentes, $p(E \mid F) = p(E)$ e podemos escrever $p(E \cap F) = p(E)p(F)$.

  Esta observação leva à seguinte definição alternativa de independência.

  E **e** F **são eventos independentes** se, e somente se,

$$p(E \cap F) = p(E)p(F).$$

Independência

- $\boxed{\text{Exemplo 14:}}$ Um dado não-viciado de 20 faces é rolado.

  Seja E o evento em que obtemos um resultado divisível por 3, e seja F o evento em que obtemos um número par.

  Os eventos E e F são independentes?

  Solução: E e F são independentes se, e somente se, $p(E \cap F) = p(E)p(F)$. Sabemos que:

  - $E = \{3, 6, 9, 12, 15, 18\}$ é o evento de obtermos um número divisível por 3, e como o dado é justo $p(E) = 6/20 = 3/10$,
  - $F = \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$ é o evento de obtermos um número par, e como o dado é justo $p(F) = 10/20 = 1/2$, e
  - $E \cap F = \{6, 12, 18\}$ é o evento de obtermos um número par e divisível por 3, e como o dado é justo $p(E \cap F) = 3/20$.

  Agora verificamos que $p(E \cap F) = 3/20$ e $p(E)p(F) = 3/10 \cdot 1/2 = 3/20$ são iguais, logo os eventos são independentes. $\triangleleft$

Independência

- $\boxed{\text{Exemplo 15:}}$ Um casal tem dois filhos.

  Assuma que a probabilidade de se ter um garoto ou uma garota é uniforme, ou seja, que $p(BB) = p(BG) = p(GB) = p(GG) = 1/4$ onde B representa um garoto (*boy*) e G representa uma garota (*girl*), e o espaço amostral do experimento é dado por $S = \{BB, BG, GB, GG\}$.

  Seja E o evento de o casal ter dois garotos, e seja F o evento de o casal ter ao menos um garoto.

  Os eventos E e F são independentes?

Independência

- $\boxed{\text{Exemplo 15:}}$ (Continuação)

  Solução: E e F são independentes se, e somente se, $p(E \cap F) = p(E)p(F)$. Sabemos que:

  - $E = \{BB\}$ é o evento de o casal ter dois garotos, e $p(E) = 1/4$,
  - $F = \{BB, BG, GB\}$ é o evento de o casal ter ao menos um garoto, e $p(F) = 3/4$, e
  - $E \cap F = \{BB\}$ é o evento de o casal ter pelo menos um garoto e, ao mesmo tempo, ter dois garotos, e $p(E \cap F) = 1/4$.

  Agora verificamos que $p(E \cap F) = 1/4$ e $p(E)p(F) = 1/4 \cdot 3/4 = 3/16$ não são iguais, logo os eventos não são independentes.

  $\triangleleft$

Ensaios de Bernoulli e a distribuição binomial

- Suponha que um experimento tenha apenas dois resultados possíveis.

  Por exemplo:

  1. um bit gerado aleatoriamente,
  2. o lançamento de uma moeda.

- Cada realização do experimento é chamada de um **ensaio de Bernoulli**.

  Os resultados possíveis do experimento são normalmente chamados de **sucesso** e **fracasso**.

  Se a probabilidade de sucesso em um ensaio de Bernoulli é p, a probabilidade de fracasso é $(1 - p)$.

- Muitos problemas podem ser podem ser resolvidos determinando-se a probabilidade de ocorrerem k sucessos em um experimento consistindo em n ensaios de Bernoulli mutuamente independentes.

  (Ensaios de Bernoulli são mutuamente independentes se a probabilidade de sucesso em cada ensaio mantém-se constante e igual a p.)

Ensaios de Bernoulli e a distribuição binomial

- Ensaios de Bernoulli dão origem a um tipo de distribuição muito particular, chamado de distribuição binomial.

  O próximo exemplo motiva tal distribuição.

- $\boxed{\text{Exemplo 16:}}$ Uma moeda viciada produz como resultado cara com probabilidade $3/4$ e coroa com probabilidade $1/4$.

  Se lançarmos a moeda 10 vezes, qual a probabilidade de obtermos exatamente 3 caras?

  Solução: O espaço amostral do experimento é o conjunto S de todos os resultados possíveis para 10 lançamentos das moedas, ou seja, $|S| = 2^{10} = 1\,024$.

  Estamos interessados no evento E em que exatamente 3 destes lançamentos resultem em caras.

  É fácil ver que $|E| = C(10, 3)$.

Ensaios de Bernoulli e a distribuição binomial

- $\boxed{\text{Exemplo 16:}}$ (Continuação)

  Cada lançamento é independente dos demais, logo a probabilidade de cada resultado em E é

$$\left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^7,$$

  pois caras ocorrem 3 vezes com probabilidade $3/4$ cada, e coroas ocorrem $10 - 3 = 7$ vezes com probabilidade $1/4$ cada.

  A probabilidade do evento E é a soma das probabilidades de cada um de seus resultados, e como há $C(10, 3)$ resultados em E:

$$p(E) = C(10, 3) \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^7 = 120 \cdot \frac{27}{16\,384} = \frac{3\,240}{16\,384}.$$

$\triangleleft$

Ensaios de Bernoulli e a distribuição binomial

- **Teorema:** A probabilidade de k sucessos em n ensaios de Bernoulli independentes, cada ensaio com probabilidade de sucesso p e de fracasso $(1-p)$, é

$$C(n,k)p^k(1-p)^{n-k}.$$

Prova. O resultado de n enasaios de Bernoulli é uma n-tupla $(t_1, t_2, \ldots, t_n)$, onde $t_i = S$ (sucesso) ou $t_i = F$ (fracasso) para cada $i = 1, 2, \ldots n$.

Uma vez que os n ensaios são independentes, a probabilidade de um resultado consistir em k sucessos e $n-k$ falhas (em qualquer ordem) é $p^k(1-p)^{n-k}$.

Como há $C(n,k)$ n-tuplas contendo exatamente k sucessos (e, consequentemente, $n-k$ fracassos), a probabilidade de k sucessos é

$$C(n,k)p^k(1-p)^{n-k}.$$

$\square$

- Chamamos a distribuição gerada por ensaios de Bernoulli de **distribuição binomial**.

Variáveis aleatórias

- Muitos problemas estão interessados em um valor numérico associado ao resultado de um experimento.

  Porém, nem todo espaço amostral é constituído por números.

- Uma **variável aleatória** é uma função do espaço amostral de um experimento para o conjunto dos números reais.

  Em outras palavras, uma variável aleatória associa um número real a cada resultado possível do experimento.

- $\boxed{\text{Exemplo 17:}}$ Suponha que uma moeda seja lançada três vezes.

  Seja $X(s)$ a variável aleatória que mapeia o resultado s do experimento para o número de caras em s.

  Chamando H de cara (*heads*) e T de coroa (*tails*), o espaço amostral do experimento é $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

  A variável aleatória $X(s)$ então assume os valores

$$
\begin{aligned}
&X(HHH) = 3, \\
&X(HHT) = X(HTH) = X(THH) = 2, \\
&X(TTH) = X(THT) = X(HTT) = 1, \qquad \text{e} \\
&X(TTT) = 0.
\end{aligned}
$$

$\lhd$

Variáveis aleatórias

- $\boxed{\text{Exemplo 18:}}$ Suponha que uma moeda seja lançada três vezes.

  Seja $Y(s)$ a variável aleatória que mapeia o resultado s do experimento para 1 se o último resultado obtido em s foi cara, e para 0 em caso contrário.

  Seja $Z(s)$ a variável aleatória que mapeia o resultado s do experimento para o número de coroas obtidas nos dois primeiros lançamentos em s.

  Chamando H de cara (*heads*) e T de coroa (*tails*), o espaço amostral do experimento é $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

  A variável aleatória $Y(s)$ então assume os valores

  $$\begin{aligned} &Y(HHH) = Y(HTH) = Y(THH) = Y(TTH) = 1, && \text{e} \\ &Y(HHT) = Y(HTT) = Y(THT) = Y(TTT) = 0, \end{aligned}$$

  e a variável aleatória $Z(s)$ então assume os valores

  $$\begin{aligned} &Z(HHH) = Z(HHT) = 0, \\ &Z(HTH) = Z(HTT) = Z(THH) = Z(THT) = 1, && \text{e} \\ &Z(TTH) = Z(TTT) = 2. \end{aligned}$$

Variáveis aleatórias

- A importância da nomenclatura:
  - variáveis aleatórias não são variáveis, são funções,
  - variáveis aleatórias não são aleatórias!
- O que muitas vezes queremos dizer quando nos referimos a "variável aleatória" é a distribuição de probabilidade sobre os valores que a variável aleatória pode assumir.
- A **distribuição de uma variável aleatória** X em um espaço amostral S é o conjunto de pares

$$(r, p(X = r)),$$

para todo $r \in X(S)$, onde $p(X = r)$ é a probabilidade de X assumir o valor r.

Uma distribuição é usualmente descrita especificando-se $p(X = r)$ para cada $r \in X(S)$.

- $\boxed{\text{Exemplo 19:}}$ Suponha que uma moeda não-viciada seja lançada três vezes.

  Seja $X(s)$ a variável aleatória que mapeia o resultado s do experimento para o número de caras em s.

  Qual a distribuição da variável aleatória X?

  Solução: A variável aleatória pode assumir os valores 0, 1, 2, ou 3.

  Se a moeda é justa, cada um dos 8 resultados do espaço amostral tem a mesma probabilidade, ou seja, $1/8$.

  Assim, podemos calcular

$$\begin{aligned}
p(X = 3) &= p(HHH) = 1/8, \\
p(X = 2) &= p(HHT) + p(HTH) + p(THH) = 3/8, \\
p(X = 1) &= p(TTH) + p(THT) + p(HTT) = 3/8, \qquad \text{e} \\
p(X = 0) &= p(TTT) = 1/8,
\end{aligned}$$

  e a distribuição de X é $\{(0, 1/8), (1, 3/8), (2, 3/8), (3, 1/8)\}$.

O Teorema de Bayes

O Teorema de Bayes: Introdução

- Em muitas situações temos que determinar a probabilidade de um certo evento ocorrer levando em conta evidências parciais para este evento.
- Algumas evidências contam a favor da ocorrência do evento, outras contam contra.
- À medida que novas evidências são coletadas, é possível obter uma estimativa mais precisa da probabilidade do evento de interesse ocorrer.

O Teorema de Bayes: Introdução

- Por exemplo, um júri deve decidir se um réu é culpado ou inocente.
  - Assuma que, a princípio, a probabilidade de o réu ser culpado é de 50%.
  - Ao longo do julgamento, a acusação aponta evidências da culpa do réu: a arma utilizada no crime pertence ao réu, havia sangue da vítima nas roupas do réu, etc.

    Dadas estas evidências, a probabilidade do réu ser culpado aumenta.
  - Já a defesa aponta evidências da inocência do réu: um álibi indica que o réu não estava na cena do crime, o réu não tinha motivo aparente para cometer o crime, etc.

    Dadas estas evidências, a probabilidade do réu ser culpado diminui.
  - O júri deve considerar cada evidência e atualizar apropriadamente a probabilidade de o réu ser culpado.
  - Ao final do julgamento, o júri considera o réu culpado se probabilidade de culpa dadas as evidências é alta o suficiente (por exemplo, acima de 95%).

O Teorema de Bayes: Introdução

- O **Teorema de Bayes** provê um método preciso para atualizar a probabilidade de um evento levando em conta evidências conhecidas.

- O Teorema de Bayes é a base de **métodos Bayesianos** de inferência, utilizados em áreas como:

  1. inteligência artificial,
  2. aprendizado de máquina,
  3. engenharia,
  4. medicina,
  5. direito,
  6. ...

O Teorema de Bayes

- Antes de apresentar o Teorema de Bayes, vamos apresentar o seguinte lema.
- **Lema:** Sejam E e F dois eventos em um espaço amostral S. Então

$$p(E) = p(E \cap F) + p(E \cap \overline{F}).$$

Prova. Primeiro mostramos que o evento E pode ser escrito como $(E \cap F) \cup (E \cap \overline{F})$:

$$\begin{aligned}
E &= E \cap S && \text{(pois } E \subseteq S\text{)} \\
&= E \cap (F \cup \overline{F}) && \text{(pois } F \cup \overline{F} = S\text{)} \\
&= (E \cap F) \cup (E \cap \overline{F}) && \text{(pela distritibutividade).}
\end{aligned}$$

Notando que $(E \cap F)$ e $(E \cap \overline{F})$ são disjuntos (ou seja, não possuem nenhum elemento em comum), temos que

$$|E| = |(E \cap F) \cup (E \cap \overline{F})| = |(E \cap F)| + |(E \cap \overline{F})|.$$

O Teorema de Bayes

- **Prova.** (Continuação)

  Assim, podemos escrever

$$\begin{aligned} p(E) &= \frac{|E|}{|S|} \\ &= \frac{|(E \cap F) \cup (E \cap \overline{F})|}{|S|} \\ &= \frac{|(E \cap F)| + |(E \cap \overline{F})|}{|S|} \\ &= \frac{|(E \cap F)|}{|S|} + \frac{|(E \cap \overline{F})|}{|S|} \\ &= p(E \cap F) + p(E \cap \overline{F}). \end{aligned}$$

□

O Teorema de Bayes

- **Teorema:** **(Teorema de Bayes)** Sejam E e F eventos de um espaço amostral S tais que $p(E) \neq 0$ e $p(F) \neq 0$. Então

$$p(F \mid E) = \frac{p(F)p(E \mid F)}{p(E)} = \frac{p(F)p(E \mid F)}{p(F)p(E \mid F) + p(\overline{F})p(E \mid \overline{F})}.$$

Prova. Podemos escrever:

$$\begin{aligned}
p(F \mid E) &= \frac{p(E \cap F)}{p(E)} && \text{(pela def. de prob. condicional)} \\
&= \frac{p(F)p(E \mid F)}{p(E)} && \text{(pois } p(E \mid F) = p(E \cap F)/p(F)\text{)} \\
&= \frac{p(F)p(E \mid F)}{p(E \cap F) + p(E \cap \overline{F})} && \text{(pelo lema anterior)} \\
&= \frac{p(F)p(E \mid F)}{p(F)p(E \mid F) + p(\overline{F})p(E \mid \overline{F})} && \text{(pela def. de prob. condicional).}
\end{aligned}$$

$\square$

O Teorema de Bayes

- $\boxed{\text{Exemplo 20:}}$ A população de uma certa cidade é composta de 52% de homens e 48% de mulheres.

  Um entrevistador para uma pessoa aleatoriamente em uma rua desta cidade para fazer uma pesquisa.

  Qual a probabilidade da pessoa selecionada ser homem?

  Solução: A probabilidade de uma pessoa aleatoriamente selecionada ser homem é 0.52 (e a probabilidade da pessoa ser mulher é 0.48).

  $\triangleleft$

O Teorema de Bayes

- $\boxed{\text{Exemplo 21:}}$ Considere a mesma cidade do exemplo anterior, em que a população é composta de 52% de homens e 48% de mulheres.

  Um entrevistador para uma pessoa aleatoriamente em uma rua desta cidade para fazer uma pesquisa.

  O entrevistador nota que a pessoa selecionada está fumando. Nesta cidade, 9% dos homens e 2% das mulheres são fumantes.

  Qual a probabilidade da pessoa selecionada ser homem?

  Solução: Para resolver esta questão, vamos formalizá-la com mais cuidado.

  Vamos usar M para representar que a pessoa é homem (*male*), e $\overline{M}$ para representar que a pessoa é mulher (*not male*).

  Vamos usar C para representar que a pessoa fuma cigarros, e $\overline{C}$ para representar que a pessoa não fuma.

- $\boxed{\text{Exemplo 21:}}$ (Continuação)

  Queremos calcular a probabilidade da pessoa selecionada ser homem, dado que ela fuma, ou seja, queremos calcular $p(M \mid C)$.

  Para isto, podemos usar o Teorema de Bayes, que diz que

$$p(M \mid C) = \frac{p(M)p(C \mid M)}{p(C)} = \frac{p(M)p(C \mid M)}{p(M)p(C \mid M) + p(\overline{M})p(C \mid \overline{M})}.$$

  Note que

  - $p(M) = 0.52$ e $p(\overline{M}) = 0.48$, como calculamos no exemplo anterior,
  - $p(C \mid M) = 0.09$, já que 9% dos homens fumam, e
  - $p(C \mid \overline{M}) = 0.02$, já que 2% das mulheres fumam.

O Teorema de Bayes

- $\boxed{\text{Exemplo 21:}}$ (Continuação)

  Substituindo os valores encontrados na fórmula de Bayes, obtemos

$$
\begin{aligned}
p(M \mid C) &= \frac{p(M)p(C \mid M)}{p(M)p(C \mid M) + p(\overline{M})p(C \mid \overline{M})} \\
&= \frac{0.52 \cdot 0.09}{0.52 \cdot 0.09 + 0.48 \cdot 0.02} \\
&= \frac{0.0468}{0.0564} \\
&\approx 0.83,
\end{aligned}
$$

  de onde concluímos que se a pessoa selecionada fuma, com 83% de chance ela é homem.

$\triangleleft$

O Teorema de Bayes: Interpretação

- O Teorema de Bayes pode ser descrito como uma maneira de atualizar a probabilidade de um evento F quando uma nova evidência E é apresentada.
- A fórmula

$$p(F \mid E) = \frac{p(F)p(E \mid F)}{p(E)}$$

indica que a probabilidade $p(F \mid E)$ do evento F dado o evento E:

1. aumenta proporcionalmente à probabilidade inicial $p(F)$ do evento F;
2. aumenta proporcionalmente à probabilidade $p(E \mid F)$ da evidência E ocorrer dado que o evento F ocorreu;
3. diminui proporcionalmente à probabilidade $p(E)$ da evidência E ocorrer independentemente de F.

O Teorema de Bayes: Interpretação

- Podemos reescrever o Teorema de Bayes da seguinte forma.

  Se a probabilidade inicial de F é $p(F)$, quando uma evidência E é apresentada calculamos a probabilidade atualizada $p(F \mid E)$ como

$$\underbrace{p(F \mid E)}_{\text{prob. atualizada}} = \underbrace{p(F)}_{\text{prob. inicial}} \cdot \underbrace{\frac{p(E \mid F)}{p(E)}}_{\text{fator de correção}} .$$

  Note que podemos ver que a probabilidade atualizada $p(F \mid E)$ é proporcional à probabilidade inicial $p(F)$ e a um fator de correção $p(E \mid F)/p(E)$.

- Estas observações têm implicações relevantes em exemplos práticos.

  É a relação entre a probabilidade inicial de F e o fator de correção que determinam o quanto a probabilidade de F deve ser atualizada dada a evidência E.

  O exemplo seguinte demonstra consequências contra-intuitivas, porém importantes, destas observações.

O Teorema de Bayes: Interpretação

- $\boxed{\text{Exemplo 22:}}$ Suponha que uma certa doença rara se manifeste em apenas 1 em cada 10 000 indivíduos de uma população.

  Suponha ainda que exita para esta doença um exame que, apesar de bastante acurado, não é perfeito:

  1. o exame resulta em positivo em 95% dos casos em que a pessoa testada tem a doença, e
  2. o exame resulta em negativo em 99% dos casos em que a pessoa testada não tem a doença.

  Neste caso, qual a probabilidade de uma pessoa que testa negativo para o exame não ter a doença?

  E qual a probabilidade de uma pessoa que testa positivo para o exame realmente ter a doença?

O Teorema de Bayes: Interpretação

- $\boxed{\text{Exemplo 22:}}$ (Continuação)

  Solução: Vamos chamar de D o evento em que a pessoa tem a doença e de $\overline{D}$ o evento em que a pessoa não tem a doença.

  Vamos chamar de E o evento em que a pessoa testa positivo para o exame, e $\overline{E}$ o evento em que a pessoa testa negativo para o exame.

  Queremos calcular

  - $p(\overline{D} \mid \overline{E})$: a probabilidade de a pessoa não ter a doença dado que o exame deu negativo, e
  - $p(D \mid E)$: a probabilidade de a pessoa ter a doença dado que o exame deu positivo.

- $\boxed{\text{Exemplo 22:}}$ (Continuação)

Vamos começar calculando $p(\overline{D} \mid \overline{E})$. Pelo Teorema de Bayes,

$$p(\overline{D} \mid \overline{E}) = \frac{p(\overline{D})p(\overline{E} \mid \overline{D})}{p(\overline{E})} = \frac{p(\overline{D})p(\overline{E} \mid \overline{D})}{p(D)p(\overline{E} \mid D) + p(\overline{D})p(\overline{E} \mid \overline{D})}.$$

Para fazer a substituição de valores na fórmula, note que:

- a probabilidade de uma pessoa na população em geral ter a doença é $p(D) = 1/10\,000 = 0.0001$,
- a probabilidade de uma pessoa na população em geral não ter a doença é $p(\overline{D}) = 9\,999/10\,000 = 0.9999$,
- a probabilidade do exame resultar negativo para quem tem a doença é $p(\overline{E} \mid D) = 0.05$.
- a probabilidade do exame resultar negativo para quem não tem a doença é $p(\overline{E} \mid \overline{D}) = 0.99$, e

- $\boxed{\text{Exemplo 22:}}$ (Continuação)

Substituindo os valores na fórmula, obtemos

$$
\begin{aligned}
p(\overline{D} \mid \overline{E}) &= \frac{p(\overline{D})p(\overline{E} \mid \overline{D})}{p(\overline{E})} \\
&= \frac{p(\overline{D})p(\overline{E} \mid \overline{D})}{p(D)p(\overline{E} \mid D) + p(\overline{D})p(\overline{E} \mid \overline{D})} \\
&= \frac{0.9999 \cdot 0.99}{0.0001 \cdot 0.05 + 0.9999 \cdot 0.99} \\
&\approx 0.9999,
\end{aligned}
$$

de onde concluímos que uma pessoa que testa negativo para o exame tem 99.99% de chance de não ter a doença.

O Teorema de Bayes: Interpretação

- $\boxed{\text{Exemplo 22:}}$ (Continuação)

  Vamos agora calcular $p(D \mid E)$. Pelo Teorema de Bayes,

  $$p(D \mid E) = \frac{p(D)p(E \mid D)}{p(E)} = \frac{p(D)p(E \mid D)}{p(D)p(E \mid D) + p(\overline{D})p(E \mid \overline{D})}.$$

  Para fazer a substituição de valores na fórmula, note que:

  - como já calculamos, na população em geral a probabilidade de uma pessoa ter a doença é $p(D) = 0.0001$, e a probabilidade da pessoa não ter a doença é $p(\overline{D}) = 0.9999$,
  - a probabilidade do exame resultar positivo para quem tem a doença é $p(E \mid D) = 0.95$, e
  - a probabilidade do exame resultar positivo para quem não tem a doença é $p(E \mid \overline{D}) = 0.01$.

O Teorema de Bayes: Interpretação

- $\boxed{\text{Exemplo 22:}}$ (Continuação)

Substituindo os valores na fórmula, obtemos

$$\begin{aligned} p(D \mid E) &= \frac{p(D)p(E \mid D)}{p(E)} \\ &= \frac{p(D)p(E \mid D)}{p(D)p(E \mid D) + p(\overline{D})p(E \mid \overline{D})} \\ &= \frac{0.0001 \cdot 0.95}{0.0001 \cdot 0.95 + 0.9999 \cdot 0.01} \\ &\approx 0.0094, \end{aligned}$$

de onde concluímos que uma pessoa que testa positivo para o exame tem apenas 0.94% de chance de ter realmente a doença.

Isto acontece porque o fator de correção $\frac{p(E|D)}{p(E)} \approx 94$ não é suficiente para alterar drasticamente a probabilidade inicial $p(D) = 0.0001$ da pessoa estar doente, que era muito baixa.

$\triangleleft$

O Teorema de Bayes: Generalização

- O Teorema de Bayes pode ser generalizado para o caso em que as evidências não são apenas do tipo presente/ausente como a seguir.

- **Teorema: (Teorema de Bayes Generalizado)** Seja E um evento de um espaço amostral S e sejam $F_1, F_2, \ldots, F_n$ eventos mutuamente exclusivos tais que $\bigcup_{i=1}^{n} = S$.

  Se $p(E) \neq 0$ e $p(F_i \neq 0)$ para todo $i = 1, 2, \ldots n$, então

$$p(F_j \mid E) = \frac{p(F_j)p(E \mid F_j)}{\sum_{i=1}^{n} p(F_i)p(E \mid F_i)}.$$

  Prova. Para o aluno!

Valor esperado e Variância

Valor esperado e variância: Introdução

- O **valor esperado** de uma variável aleatória pode ser visto como uma média ponderada dos valores que esta variável assume quando o experimento é reexecutado várias vezes.

  O valor esperado de uma variável aleatória representa uma espécie de ponto médio ao redor do qual os outros valores da variável estão distribuídos.

- A **variância** de uma variável aleatória é uma medida do quão esparsamente os valores da variável aleatória estão distribuídos em torno de seu valor esperado.

- O *valor esperado* (ou *esperança*) de uma variável aleatória $X(s)$ no espaço amostral S é dado por

$$E(X) = \sum_{s \in S} p(s)X(s).$$

  Se o espaço amostral é infinito, o valor esperado só existe se o somatório for absolutamente convergente.

- De forma equivalente, podemos escrever

$$E(X) = \sum_{r \in X(S)} p(X = r) \cdot r,$$

  pois simplesmente no somatório agrupamos todos os valores que a variável aleatória pode assumir vezes a probabilidade dela assumir tal valor.

Valor esperado

- $\boxed{\text{Exemplo 23:}}$ Seja X o número que obtemos ao rolar um dado não-viciado de 6 faces.

  Qual o valor esperado de X?

  Solução: A variável aleatória X assume os valores $\{1, 2, 3, 4, 5, 6\}$ com probabilidade uniforme $1/6$.

  Logo

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$$

$\triangleleft$

- $\boxed{\text{Exemplo 24:}}$ Suponha que uma moeda seja lançada três vezes.

  Seja $X(s)$ a variável aleatória que mapeia o resultado s do experimento para o número de caras em s.

  Qual o valor esperado de X?

  Solução: Chamando H de cara (*heads*) e T de coroa (*tails*), o espaço amostral do experimento é $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

  Como a moeda é não-viciada, qualquer um dos 8 resultados possui probabilidade $1/8$.

- $\boxed{\text{Exemplo 24:}}$ (Continuação)

Assim, podemos calcular:

$$
\begin{aligned}
E(X) = &\sum_{s \in S} p(s)X(s) \\
= &\frac{1}{8} \cdot X(HHH) + \frac{1}{8} \cdot X(HHT) + \frac{1}{8} \cdot X(HTH) + \frac{1}{8} \cdot X(HTT) + \\
&\frac{1}{8} \cdot X(THH) + \frac{1}{8} \cdot X(THT) + \frac{1}{8} \cdot X(TTH) + \frac{1}{8} \cdot X(TTT) \\
= &\frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 2 + \frac{1}{8} \cdot 2 + \frac{1}{8} \cdot 1 + \frac{1}{8} \cdot 2 + \frac{1}{8} \cdot 1 + \frac{1}{8} \cdot 1 + \frac{1}{8} \cdot 0 \\
= &\frac{3}{2}.
\end{aligned}
$$

$\triangleleft$

Valor esperado da distribuição binomial

- Como vimos, a distribuição binomial é a distribuição que surge da execução de n ensaios de Bernoulli independentes.

  Recorde que na distribuição binomial, a probabilidade de k sucessos durante a execução de n ensaios de Bernoulli independentes (onde cada ensaio tem probabilidade de sucesso p) é dada por $C(n, k)p^k(1 - p)^{n-k}$.

- **Teorema: (Valor esperado da distribuição binomial)** O número esperado de sucessos quando n ensaios de Bernoulli independentes são executados, onde p é a probabilidade de sucesso em cada ensaio, é

$$np.$$

  Prova. Para o aluno!

Valor esperado da distribuição binomial

- $\boxed{\text{Exemplo 25:}}$ Uma moeda viciada ao ser lançada resulta em cara $3/4$ das vezes, e em coroa $1/4$ das vezes.

  Qual o número esperado de caras se a moeda for lançada 100 vezes?

  Solução: O experimento é um processo de Bernoulli com $n = 100$ ensaios, em que a probabilidade de sucesso (cara) é $p = 3/4$.

  Portanto, o número esperado de caras no experimento é

  $$np = 100 \cdot \frac{3}{4} = 75.$$

  $\triangleleft$

Linearidade do valor esperado

- O seguinte teorema mostra que o valor esperado é uma função linear.

  Em particular, o valor esperado da soma de variáveis aleatórias é a soma dos valores esperados das variáveis aleatórias.

- **Teorema: (Linearidade do valor esperado)** Se X_i, onde $i = 1, 2, \ldots, n$ é um inteiro positivo, são variáveis aleatórias em um espaço amostral S, e a e b são números reais, então

  1. $E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$, e
  2. $E(aX + b) = aE(X) + b$.

Prova. Vamos provar cada item separadamente.

- **Prova.** (Continuação)

  Para o primeiro item:

$$
\begin{aligned}
E(X_1 + X_2 + \ldots + X_n) &= \sum_{s \in S} p(s)(X_1(s) + X_2(s) + \ldots + X_n) \\
&= \sum_{s \in S} [p(s)X_1(s) + p(s)X_2(s) + \ldots + p(s)X_n] \\
&= \sum_{s \in S} p(s)X_1(s) + \sum_{s \in S} p(s)X_2(s) + \ldots + \sum_{s \in S} p(s)X_n \\
&= E(X_1) + E(X_2) + \ldots + E(X_n).
\end{aligned}
$$

Linearidade do valor esperado

- **Prova.** (Continuação)

  Para o segundo item:

$$
\begin{aligned}
E(aX+b) &= \sum_{s \in S} p(s)(aX(s)+b) && \text{(pela def.)} \\
&= \sum_{s \in S} [p(s)aX(s) + p(s)b] && \text{(distribuindo a multiplicação)} \\
&= \sum_{s \in S} p(s)aX(s) + \sum_{s \in S} p(s)b && \text{(separando o somatório)} \\
&= a\sum_{s \in S} p(s)X(s) + b\sum_{s \in S} p(s) && \text{(constantes em evidência)} \\
&= aE(X) + b \cdot 1 && \text{(pois } \textstyle\sum_{s \in S} p(s) = 1\text{)} \\
&= aE(X) + b
\end{aligned}
$$

$\square$

- $\boxed{\text{Exemplo 26:}}$ Cinquenta clientes passam por um restaurante durante um dia. Cada um dos clientes tem uma probabilidade uniforme de deixar uma gorjeta entre 0 e 10 reais, em valores inteiros apenas.

  Qual o valor esperado do total de gorjetas deixadas pelos clientes ao final de um dia?

  Solução: O espaço amostral do experimento são todas as 50-tuplas em que cada elemento na tupla representa a gorjeta deixada por um cliente específico.

  Se formos calcular o valor esperado diretamente neste espaço amostral, teremos um trabalho muito grande: há 11^{50} resultados possíveis no espaço amostral!

  O problema pode ser simplificado usando a linearidade do valor esperado.

- $\boxed{\text{Exemplo 26:}}$ (Continuação)

  Seja X_i o valor de gorjeta deixado pelo cliente i, sendo $1 \leq i \leq 50$.

  Sabemos que o cliente deixa uma gorjeta entre 0 e 10 reais uniformemente, ou seja, cada valor inteiro entre 0 e 10 é deixado como gorjeta com probabilidade $1/11$, logo

  $$E(X_i) = \frac{1}{11}\left(0+1+2+3+4+5+6+7+8+9+10\right) = 5.$$

  Estamos interessados em calcular $E\left(\sum_{i=1}^{50} X_i\right)$, e pela linearidade do valor esperado podemos fazer

  $$E\left(\sum_{i=1}^{50} X_i\right) = \sum_{i=1}^{50} E(X_i) = \sum_{i=1}^{50} 5 = 50 \cdot 5 = 250,$$

  e concluir que a gorjeta esperada ao final do dia é de 250 reais.

$\triangleleft$

Linearidade do valor esperado

- $\boxed{\text{Exemplo 27:}}$ Um par de dados de 6 faces, ambos não-viciados, é lançado.

  Qual o valor esperado da soma dos números obtidos?

  Solução: O espaço amostral do experimento são pares (i, j), com $1 \leq i \leq 6$ representando o resultado do lançamento do primeiro dado, e $1 \leq j \leq 6$ representando o resultado do lançamento do segundo dado.

  Seja X_1 a variável aleatória que mapeia cada resultado para o número do primeiro dado, ou seja, $X_1(i, j) = i$.

  Seja X_2 a variável aleatória que mapeia cada resultado para o número do segundo dado, ou seja, $X_2(i, j) = j$.

  Queremos calcular $E(X_1 + X_2)$, e pela linearidade do valor esperado sabemos que $E(X_1 + X_2) = E(X_1) + E(X_2)$.

  Já vimos em um exemplo anterior que o valor esperado do lançamento de um dado de 6 faces não-viciado é 3.5, concluímos:

$$E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7.$$

$\triangleleft$

Variáveis aleatórias independentes

- Já vimos que dois eventos E_1 e E_2 são independentes se, e somente se,

$$p(E_1 \cap E_2) = p(E_1)p(E_2).$$

Vamos agora estender o conceito de independência para variáveis aleatórias.

- Duas **variáveis aleatórias** X e Y num espaço amostral S são **independentes** se para todo valor de r_1 e r_2:

$$p(X(s) = r_1 \wedge Y(s) = r_2) = p(X(s) = r_1) \cdot p(Y(s) = r_2).$$

Ou seja, X e Y são independentes se a probabilidade de X assumir o valor r_1 e Y assumir o valor r_2 simultaneamente for igual ao produto da probabilidade de X assumir o valor r_1 pela probabilidade de Y assumir o valor r_2.

Variáveis aleatórias independentes

- $\boxed{\text{Exemplo 28:}}$ Dois dados não-viciados de 6 faces são lançados.

  Seja X_1 a variável aleatória correspondendo ao resultado obtido no primeiro dado, e X_2, a variável aleatória correspondendo ao resultado obtido no segundo dado.

  X_1 e X_2 são independentes?

  Solução: As variáveis são independentes se $p(X_1 = r_1 \wedge X_2 = r_2) = p(X_1 = r_1)p(X_2 = r_2)$.

  O espaço amostral de X_1 e X_2 é o mesmo: $S = \{1, 2, 3, 4, 5, 6\}$. Além disso, a distribuição de X_1 e X_2 é uniforme, ou seja, $1/6$ para qualquer resultado.

  Logo, deduzimos que para quaisquer valores de r_1 e r_2

$$p(X_1 = r_1)p(X_2 = r_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

- $\boxed{\text{Exemplo 28:}}$ (Continuação)

  Agora vamos considerar a probabilidade $p(X_1 = r_1 \wedge X_2 = r_2)$ de os dois dados lançados assumirem valores ao mesmo tempo.

  Nesse caso o espaço amostral tem 36 resultados possíveis (um para cada combinação dos dois dados). Como todos os resultados são igualmente prováveis, a probabilidade de cada um é $1/36$.

  Assim, verificamos que para quaisquer $r_1, r_2 \in S$

$$p(X_1 = r_1 \wedge X_2 = r_2) = p(X_1 = r_1)p(X_2 = r_2) = \frac{1}{36},$$

  e as variáveis X_1 e X_2 são independentes.

$\triangleleft$

Variáveis aleatórias independentes

- $\boxed{\text{Exemplo 29:}}$ Dois dados não-viciados de 6 faces são lançados.

  Seja X_1 a variável aleatória correspondendo ao resultado obtido no primeiro dado, e X_2, a variável aleatória correspondendo à soma dos resultados de cada dado.

  X_1 e X_2 são independentes?

  Solução: As variáveis são independentes se para todo r_1, r_2 $p(X_1 = r_1 \land X_2 = r_2) = p(X_1 = r_1)p(X_2 = r_2)$.

  Vamos mostrar um contra-exemplo que prova que as variáveis não são independentes.

  Tomando $r_1 = 1$ e $r_2 = 12$, temos que

  $$p(X_1 = 1 \land X_2 = 12) = 0,$$

  pois não há como o primeiro dado resultar no valor 1 e a soma dos dados resultar no valor 12 ao mesmo tempo.

- $\boxed{\text{Exemplo 29:}}$ (Continuação)

Por outro lado:

- $p(X_1 = 1) = 1/6$, pois o dado é não-viciado, e
- $p(X_2 = 12) = 1/36$, pois apenas em 1 dos 36 resultados possíveis para o lançamento de dois dados a soma dos resultados dá 12 (quando ambos os dados resultam em 6).

Logo,

$$p(X_1 = 1)p(X_2 = 12) = \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{216},$$

o que difere de $p(X_1 = 1 \wedge X_2 = 12) = 0$ e, portanto, X_1 e X_2 não são independentes.

$\triangleleft$

Variáveis aleatórias independentes

- Se duas variáveis aleatórias são independentes, o valor esperado de seu produto pode ser calculado como o produto de seus valores esperados.

- **Teorema:** Se X e Y são variáveis independentes em um espaço amostral S, então

$$E(XY) = E(X)E(Y).$$

Prova. Para o aluno!

Variância

- O valor esperado de uma distribuição sozinho pode levar a comparações enganosas entre distribuições de probabilidade.

  Distribuições de probabilidade muito distintas podem ter valores esperados parecidos.

- A variância de uma variável aleatória é uma medida do quão dispersos são os valores desta variável em torno de seu valor esperado.

- Seja X uma variável aleatória num espaço amostral S. A **variância** de X, denotada por $V(X)$, é dada por

$$V(X) = \sum_{s \in S} p(s)[X(s) - E(X)]^2.$$

  O desvio-padrão de X, denotado por $\sigma(X)$, é definido como $\sigma(X) = \sqrt{V(X)}$.

Variância

- **Teorema:** Se X é uma variável aleatória num espaço amostral S, então
$$V(X) = E(X^2) - [E(X)]^2.$$

Prova.

$$\begin{aligned}
V(X) &= \sum_{s \in S} p(s)[X(s) - E(X)]^2 \\
&= \sum_{s \in S} p(s)[(X(s))^2 - 2X(s)E(X) + (E(X))^2] \\
&= \sum_{s \in S} [p(s)(X(s))^2 - p(s)2X(s)E(X) + p(s)(E(X))^2] \\
&= \sum_{s \in S} p(s)(X(s))^2 - \sum_{s \in S} p(s)2X(s)E(X) + \sum_{s \in S} p(s)(E(X))^2 \\
&= E(X^2) - 2E(X)\sum_{s \in S} p(s)X(s) + (E(X))^2 \sum_{s \in S} p(s) \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \cdot 1 \\
&= E(X^2) - [E(X)]^2
\end{aligned}$$

$\square$

Variância

- **Teorema:** Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes dois-a-dois num espaço amostral S, então

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n).$$

Prova.

Exercício para o aluno!

Apêndice

Falácias probabilísticas

- Uma **falácia** é um raciocínio errado com aparência de verdadeiro.

  Uma falácia é um argumento logicamente inconsistente, sem fundamento, inválido, ou falho na tentativa de provar o que alega.

- O termo "falácia" deriva do verbo latino "fallere", que significa "enganar".

  Argumentos que se destinam à persuasão podem parecer convincentes para grande parte do público apesar de conterem falácias, mas não deixam de ser falsos por causa disso.

- Reconhecer as falácias é por vezes difícil.

  Os argumentos falaciosos podem ter validade emocional, íntima, psicológica, mas não validade lógica.

- Exemplos de **falácias lógicas** são:

  1. *ad hominem*,
  2. *ad populum*,
  3. espantalho,
  4. etc.

- **Falácias probabilísticas** são por vezes ainda mais difíceis de se identificar, porque vão contra nossa intuição.

Falácias probabilísticas: Ignorar a frequência de base

- Falácia de **ignorar a frequência de base**:

  considerar apenas a probabilidade condicional de um evento dadas informações específicas, ignorando a frequência de base do evento.

- $\boxed{\text{Exemplo 30:}}$ Pedro é introvertido, inteligente e gosta de biologia.

  Qual a profissão mais provável de Pedro: vendedor de uma loja ou neurocirurgião?

  A resposta é vendedor de uma loja.

  Muitas pessoas ficam tentadas a responder neurocirurgião, porque a descrição de Pedro se assemelha ao esteriótipo de um neurocirurgião.

  Entretanto, a proporção-base na população é de centenas (ou milhares) de vendedores para cada neurocirurgião, logo é mais provável que ele seja apenas um vendedor atípico do que um cirurgião típico.

$\triangleleft$

Falácias probabilísticas: Ignorar a frequência de base

- $\boxed{\text{Exemplo 31:}}$ Relembre o exemplo que vimos sobre um exame quem tem 95% de probabilidade de resultar positivo se o indivíduo testado possui uma dada doença.

  Se um indivíduo testa positivo para a doença neste exame, a princípio não se pode determinar qual a probabilidade deste indivíduo realmente ter a doença.

  A frequência de base da ocorrência da doença na população, assim como a probabilidade o teste resultar positivo para alguém que não tem a doença, tem que ser levada em consideração.

  Como vimos no exemplo citado, se a frequência de base na população é de 1 indivíduo contaminado para cada 10 000 indivíduos, e a probabilidade do exame resultar em um falso positivo é de 1%, a probabilidade de um indivíduo que testa positivo realmente ter a doença é de apenas 0.94%.

$\triangleleft$

Falácias probabilísticas: Falácia do apostador

- Falácia do **apostador**:

  assumir que a distribuição de probabilidade em um conjunto pequeno de experimentos deve necessariamente refletir a distribuição de um conjunto grande de experimentos.

- Exemplo 32: *"As últimas 5 rodadas da roleta deram um número vermelho: isto é sinal de que muito provavelmente o próximo número será preto. Vou apostar todas as minhas fichas no preto!"* ◁

- Outra versão da falácia do apostador, oposta à primeira, é assumir que a probabilidade de um evento necessariamente segue a distribuição dos experimentos mais recentes.

- Exemplo 33: *"As últimas 5 rodadas da roleta deram um número vermelho: isto é sinal de que estamos em uma "onda vermelha", e muito provavelmente o próximo número será também vermelho. Vou apostar todas as minhas fichas no vermelho!"* ◁

- **Paradoxo de Simpson**:

  dados agregados podem sugerir uma correlação que pode ser completamente revertida quando os dados são analisados de forma segmentada.

- $\boxed{\text{Exemplo 34:}}$ Uma pesquisa nos EUA descobriu que:

  1. o salário médio dos membros do grupo A é de $1\,590$ US$ por mês, e
  2. o salário médio dos membros do grupo B é de $2\,500$ US$ por mês.

  - Um jornal computa a razão $^{1\,590}/_{2\,500} = 0.64$ e divulga:

    "Desigualdade de salários nos EUA: pessoas do grupo A recebem, em média, 64 centavos para cada dólar recebido por pessoas do grupo B!"

  - Outro jornal separa as pessoas em profissões e divulga:

    "Desigualdade de salários nos EUA: em todas as profissões, pessoas do grupo A sempre recebem 10% a mais que pessoas do grupo B!"

<u>As duas manchetes estão corretas</u>. Como isso é possível?

Falácias probabilísticas: Paradoxo de Simpson

- Exemplo 34: (Continuação)

Quando consultamos os dados da pesquisa, observamos o seguinte:

Grupo	**Profissão 1**	**Profissão 2**	**Profissão 3**
Salário do grupo A	1 100 USD	2 200 USD	3 300 USD
Salário do grupo B	1 000 USD	2 000 USD	3 000 USD
Pessoas do grupo A	600	300	100
Pessoas do grupo B	100	300	600

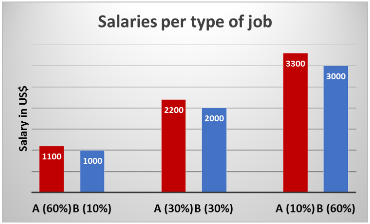
- Média salarial A: $^{600 \cdot 1100 \text{ USD} + 300 \cdot 2200 \text{ USD} + 100 \cdot 3300 \text{ USD}} /_{1000} = 1590$ USD
- Média salarial B: $^{100 \cdot 1000 \text{ USD} + 300 \cdot 2000 \text{ USD} + 600 \cdot 3000 \text{ USD}} /_{1000} = 2500$ USD
- Mas, controlando por profissão, cada membro A sempre ganha 10% a mais que cada membro do grupo B na mesma profissão!

Ou seja, há desigualdade, mas quem seria o grupo "beneficiado"?

Falácias probabilísticas: Paradoxo de Simpson

- Exemplo 34: (Continuação)

Isto ocorre porque, apesar de em cada profissão cada membro do grupo A sempre ganhar 10% a mais que membros do grupo B, existem mais membros do grupo B nas faixas salariais bem pagas, o que puxa a média para cima!



Se há desigualdade, não é a de que membros do grupo A são mais bem pagos que membros do grupo B na mesma profissão.

A desigualdade está na escolha das profissões!

◁

Fórmulas para permutações e combinações

Nº **arranjamentos de** r **elementos; conjunto de** n	Ordem dos elementos **não importa**	Ordem dos elementos **importa**
Elementos **não podem** se repetir	r-combinação: $C(n,r) = \frac{n!}{r!(n-r)!}$	r-permutação: $P(n,r) = \frac{n!}{(n-r)!}$
Elementos **podem** se repetir	r-combinação com repetição: $C(n+r-1,r)$	r-permutação com repetição: n^r