Matemática Discreta

Licenciatura em Segurança Informática
em Redes de Computadores
Licenciatura em Engenharia Informática

Teoria dos Números

Probabilidades (Discreta) e Cadeias de Markov

Eliana Costa e Silva – eos@estg.ipp.pt

P.PORTO ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO

Felgueiras, maio de 2022

- Uma **experiência** aleatória é um procedimento cujo resultado é um entre todos os possíveis resultados.
- O **espaço amostral** de uma experiência é o conjunto de todos os resultados possíveis.
- Um **acontecimento** é um subconjunto do espaço amostral.

Lei de Laplace

Seja S um conjunto finito não vazio e A um acontecimento de S.
Temos que a probabilidade de A é

$$p(A) = \frac{\#A}{\#S}.$$

Chamamos **frequência relativa** do acontecimento A, nas n repetições de uma experiência, ao número obtido por:

$$f_A = \frac{n_A}{n},$$

onde n_A é o número de que o acontecimento A ocorre em n experiências.

Propriedades

A frequência relativa f_A apresenta as seguintes propriedades:

- $0 \leq f_A \leq 1$;
- $f_A = 1$ se, e só se, A ocorrer em todas as n repetições
- $f_A = 0$ se, e só se, A nunca ocorrer nas n repetições
- Se A e B forem **acontecimentos mutuamente exclusivos**, e se $f_{A \cup B}$ for a frequência relativa associada ao acontecimento $A \cup B$, então

$$f_{A \cup B} = f_A + f_B$$

Se o número de repetições da experiência for aumentando, o valor da frequência relativa f_A tenderá a "estabilizar" próximo de um determinado valor numérico bem definido:

$$p(A) = \lim_{n \to \infty} f_A.$$

Axiomas

Seja S o espaço amostral associada a uma experiência.
Temos que:

- $p(A) \geq 0$, para todo o acontecimento A de S.
- $p(S) = 1$.
- $p(A \cup B) = p(A) + p(B)$,
  onde A e B são dois acontecimentos mutuamente exclusivos.
- $p\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i)$, onde $A_i \cap A_j = \emptyset$.

Teorema

- $p(\emptyset) = 0$
- $p(\bar{A}) = 1 - p(A)$
- $p(B \setminus A) = p(B) - p(A \cap B)$
- Se $A \subset B$ então $p(B \setminus A) = p(B) - p(A)$
- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- Se A e B forem acontecimentos tais que $A \subset B$, então $p(A) \leq p(B)$.
- $p(A) \leq 1$

Exemplo

Qual a probabilidade de um número inteiro selecionado aleatoriamente de um conjunto de números inteiros não superiores a 100 serem divisíveis por 2 ou por 5?

Exemplo

Qual a probabilidade de um número inteiro selecionado aleatoriamente de um conjunto de números inteiros não superiores a 100 serem divisíveis por 2 ou por 5?

Seja A o acontecimento "o número selecionado ser divisível por 2" e B o acontecimento "o número selecionado ser divisível por 5".

Temos que $A \cap B$ é o acontecimento "o número ser simultaneamente divisível por 2 e 5".

Pretende-se determinar $p(A \cup B)$.

Assim,

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}.$$

Veja o Exemplo 10, página 450 de (Rosen, 2014).

Definição - Probabilidade Condicionada

Sejam A e B dois acontecimentos tais que $p(B) > 0$.
A **probabilidade de A condicionada a B** é dada por

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

Exemplo

Considere um string de bits de comprimento quatro é gerado aleatoriamente tal que cada um dos 16 strings de bits de comprimento quatro sejam igualmente prováveis.
Qual a probabilidade do string de bits conter pelo menos dois 0s consecutivos, dado que o seu primeiro bit é 0?

Solução:
$\frac{5}{8}$

Definição

Dois acontecimentos A e B são **independentes**

se e só se

$$p(A \cap B) = p(A) \times p(B).$$

Observações

- $0 \leq p(A|B) \leq 1$
- $p(S|B) = 1)$
- $p(A_1 \cap A_2|B) = p(A_1|B) + p(A_2|B)$
- $p(A \cap B) = p(A) \times p(B|A) = p(B) \times p(A|B)$

Exemplo

Seja A o acontecimento um string de bits de comprimento quatro é gerado aleatoriamente começa com um 1 e B o acontecimento o string de bits contem um número par de 1s.
Verifique se A e B são independentes.

Existem oito strings de comprimento quatro que começam por 1:

1000, 1001, 1010, 1011, 1100, 1101, 1110 e 1111.

Existem oito strings de comprimento quatro com um número par de 1:

000, 0011, 0101, 0110, 1001, 1010, 1100 e 1111.

Como existem 16 strings de comprimento 4 temos que:

$$p(A) = p(B) = \frac{8}{16} = \frac{1}{2}.$$

Por outro lado, $A \cap B = \{1111, 1100, 1010, 1001\}$, donde $p(A \cap B) = \dfrac{4}{16} = \dfrac{1}{4}$.
Assim,

$$p(A \cap B) = p(A) \times p(B),$$

donde A e B são acontecimentos independentes.

Definição

Dizemos que $A_1, A_2, \ldots, A_n$ são **independentes dois a dois** se

$$p(A_i \cap A_j) = p(A_i) \times p(A_j),$$

para todos os pares de inteiros i e j com $1 \leq i < j \leq n$.
Estes acontecimentos dizem-se **mutuamente independentes** se

$$p(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = p(A_{i_1}) \times p(A_{i_2}) \times \cdots \times p(A_{i_m})$$

onde $i_j, j = 1, 2, \ldots, m$, são inteiros com $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ e $m \geq 2$.

Definição

Seja S um conjunto com n elementos.
A *distribuição uniforme* atribuiu a cada elementos de S a probabilidade $1/n$.

Exemplo

Consideremos o lançamento de um dado equilibrado.

Temos que $S = \{1, 2, 3, 4, 5, 6\}$ e $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \dfrac{1}{6}$.

Experiências de Bernoulli e Distribuição Binomial

Consideremos uma experiência com apenas dois possíveis resultados (p.ex., 0 ou 1).
Cada possível resultado é uma **experiência de Bernoulli** e o seu resultado designamos por **sucesso** ou **insucesso**.
Seja p a **probabilidade de sucesso** e $q = 1 - p$ a **probabilidade de insucesso**.
A probabilidade de se obterem exatamente k sucessos é dada por

$$^{n}C_{k} p^{k} q^{n-k}$$

À função $b(k; n, p) = {}^{n}C_{k} p^{k} q^{n-k}$ chamamos **distribuição binomial**.

Exemplo

Suponha que a probabilidade de ser gerado o bit 0 é 0.9.
Suponho que os bits 0 e 1 são gerados independentemente.
Qual a probabilidade de num string de 10 bits termos exatamente oito bits 0?

Solução: ≈ 0.1937

Distribuição Geométrica

Uma variável aleatória X tem uma distribuição geométrica de parâmetro p se
$p(X = k) = (1 - p)^{k-1}$, para $k = 1, 2, 3, \ldots,$
onde p é um número real tal que $0 \leq p \leq 1$.

Teorema

Teorema de Bayes

Sejam A e B acontecimentos de S tais que $p(A) \neq 0$ e $p(B) \neq 0$. Então,

$$p(B|A) = \frac{p(A|B) \times p(B)}{p(A|B) \times p(B) + p(A|\bar{B}) \times p(\bar{B})}.$$

Teorema

Generalização do Teorema de Bayes

Sejam A um acontecimentos de S e $B_1, B_2, \ldots, B_n$ acontecimentos mutuamente exclusivos tais que $\cup_{i=1}^{n} B_i = S$.

Assuma que $p(A) \neq 0$ e $p(B_i) \neq 0$, $i = 1, \ldots, n$. Então,

$$p(B_j|A) = \frac{p(A|B_j) \times p(B_j)}{\sum_{i=1}^{n} p(A|B_i) \times p(B_i)}$$

Uma aplicação dos dois últimos resultados são os filtros Bayesianos de spam de correio eletrónico. Ver (Rosen, 2014) página 472 até 475.

Definição

Seja X uma variável aleatória definida no espaço amostral S.

- O valor esperado ou **valor médio esperado** de X é dado por:

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

- A **variância** de X é:

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

- O **desvio-padrão** de X é:

$$\sigma[X] = \sqrt{V[X]}.$$

Se o espaço amostral é tal que $S = \{x_1, x_2, \ldots, x_n\}$, então

$$E[X] = \sum_{i=1}^{n} p(x_i)X(x_i)$$

Teorema

- O valor esperado do número de sucessos em n experiências de Bernoulli, onde p é a probabilidade de sucesso de cada experiência é:

$$n\,p$$

- O valor esperado de uma variável X com distribuição de geométrica de parâmetro p é:

$$E[X] = 1/p$$

Propriedades

Sejam n um número inteiro positivo, $X_1, X_2, \ldots, X_n$, n variáveis aleatórias em S, $a, b \in \mathbb{R}$. Temos que:

- $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$
- $E[aX + b] = aE[X] + b$
- Se X_1 e X_2 são independentes, então $E[X_1 \; X_2] = E[X_1]E[X_2]$

Propriedades

Sejam n um número inteiro positivo, $X_1, X_2, \ldots, X_n$, n variáveis aleatórias em S, $a, b \in \mathbb{R}$. Temos que:

- **Fórmula de Bienaymé**:
  Se X_1 e X_2 são independentes, então $V[X_1 + X_2] = V[X_1] + E[X_2]$
  Se $X_i, i = 1, \ldots, n$ são mutuamente independentes, então
  $$V[X_1 + X_2 + \cdots + X_n] = V[X_1] + V[X_2] + \cdots + V[X_n]$$
- **Desigualdade de Cheybyshev**:
  $$p(|X(s) - E[X]| \geq r) \leq V[X]/r^2$$
  onde r é um número real positivo.

Cadeias de Markov

- Uma **cadeia de Markov** em tempo discreto (DTMC) é um caso particular de **processo estocástico** com estados discretos[1] com a propriedade de que a *distribuição de probabilidade do próximo estado depende apenas do estado atual e não na sequência de eventos que precederam* – **propriedade Markoviana**.

- **Memória Markoviana** diz-nos que os estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido.

- **Cadeias de Markov** têm enumeras aplicações como modelos estatísticos de processos do mundo real.
  Ver por exemplo
  `https://www.google.com/patents/US6285999?hl=pt-PT`.

[1] O parâmetro, em geral o tempo, pode ser discreto ou contínuo

Definição

Uma **cadeia de Markov** é uma sequência $X_0, X_1, X_2, X_3, \ldots$ de variáveis aleatórias.

O conjunto de valores que as variáveis aleatórias podem assumir, é chamado de **espaço de estados**, onde X_n denota o estado do processo no instante de tempo n.

Se a distribuição de probabilidade condicional de X_{n+1} nos estados passados é uma função apenas de X_n, então:

$$\Pr(X_{n+1} = x | X_0, X_1, X_2, \ldots, X_n) = \Pr(X_{n+1} = x | X_n)$$

onde x é algum estado do processo.
Esta é a **propriedade de Markov**.

As cadeias de Markov são frequentemente descritas por uma **sequência de grafos orientados**, onde as arestas do grafo n são rotulados pelas probabilidades de ir de um estado no instante de tempo n para outros estados no tempo $n+1$:

$$\Pr(X_{n+1} = x \mid X_n = x_n).$$

Matriz de transição

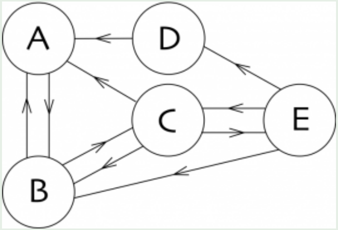
A **matriz de transição**, $T = (t_{ij})$ de uma **cadeia de Markov** no instante de tempo n para o tempo $n+1$ é uma matriz $m \times m$ cujas entradas são a probabilidade de o sistema se mover do estado i para o estado j, com $i, j = 1, \ldots, m$ e m é o número de estados do sistema, i.e.:

$$t_{ij} = \Pr(X_{n+1} = j \mid X_n = i).$$

Temos que $0 \leq t_{ij} \leq 1$ e a soma das entradas de cada coluna da matriz de transição tem de ser igual a 1.

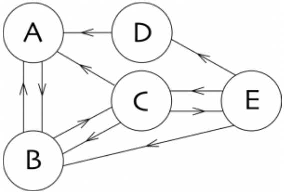
Exemplo Adaptado de http://blog.kleinproject.org/?p=280.

Considere rede constituída por 5 páginas web A, B, C, D, E com os links:



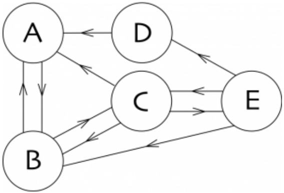
Em cada instante, a probabilidade de começando numa página qualquer terminar numa oura página é dada pela matriz:

$$T = \begin{bmatrix} 0 & 1/2 & 1/3 & 1 & 0 \\ 1 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 0 \end{bmatrix}$$



$$T = \begin{bmatrix} 0 & 1/2 & 1/3 & 1 & 0 \\ 1 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 0 \end{bmatrix}$$

- Uma pessoa que esteja na página A neste instante, estará em B no instante seguinte.
- Se estiver em B neste momento, tem 50% de probabilidade de no instante seguinte estar na página A.
- A probabilidade após dois passos vai ser dada por T^2. Determine esta matriz e interprete.



$$T = \begin{bmatrix} 0 & 1/2 & 1/3 & 1 & 0 \\ 1 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 0 \end{bmatrix}$$

- Consideremos que uma pessoa no instante inicial está na página A, considere $X_0 = [1\ 0\ 0\ 0\ 0]^\top$.
- No instante seguinte a probabilidade da pessoa estar numa página é dada por TX_0, e passados dois instantes é T^2X_0.
- Ou seja, é certo que no instante 1 esteja em B e tem 50% de probabilidade de estar em C (ou em A) no instante 2.

Scilab

```
-->T = [0   1/2   1/3  1  0
-->1  0  1/3  0 1/3
-->0   1/2  0   0  1/3
-->0  0  0  0  1/3
-->0  0  1/3  0  0 ];
```

```
-->X0=[1 0 0 0 0]'
X0  =

  1.
  0.
  0.
  0.
  0.
```

```
-->X1=T*X0
X1  =

  0.
  1.
  0.
  0.
  0.
```

```
-->X2=T^2*X0
X2  =

  0.5
  0.
  0.5
  0.
  0.
```