PL 1

Probabilidade: conceitos base

Palavras chave: probabilidade, estimação de probabilidade, experiência aleatória, espaço de amostragem, eventos, casos favoráveis, simulação, Matlab.

Responda às seguintes questões utilizando sempre o Matlab para efetuar os cálculos necessários:

1. Muito breve introdução ao Matlab baseada no documento Matlab num instante do Professor José Vieira.

   Criação de Matrizes:

   Para criar, por exemplo, a seguinte matriz com duas linhas e três colunas

   $$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

   basta introduzir o comando:

   ```
   >> A= [1 2 3;4 5 6]
   ```

   ou em alternativa

   ```
   >> A= [1,2,3;4,5,6]
   ```

   Os vectores coluna e linha são casos particulares de matrizes e são criados utilizando uma notação similar, por exemplo o vetor linha $v = [123]$ poderá ser obtido com

   ```
   >> v= [ 1 2 3]
   ```

   e o correspondente vetor coluna com

   ```
   >> v= [ 1; 2; 3]
   ```

   ou transpondo o vetor linha

   ```
   >> v= [1 2 3]'
   ```

   Índices

   O elemento da linha i e da coluna j de uma matriz A é designado por A (i, j). Por exemplo o elemento da linha 1 e coluna 3 da matriz A é designado por $A(1,3)$. Em notação Matlab, para obter o elemento $A(1,3)$ da matriz A definida anteriormente, pode-se escrever

   ```
   >> A(1,3)
   ```

   e obtém-se

   ```
   ans= 3
   ```

   Para alterar o valor do elemento $A(1,3)$ para 7 basta fazer

   ```
   >> A(1,3)= 7
   ```

   Os índices das matrizes são listas de números inteiros positivos que podem ser armazenadas em vectores declarados previamente. Se pretendermos por exemplo, extrair a segunda linha da matriz A podemos fazer

```
>> v= A(2,[1 2 3])
```

ou declarando primeiro um vector para os índices das colunas

```
>> k= [1 2 3]
>> v= A(2,k)
```

O operador ":"

O Matlab permite gerar sequências de números de forma rápida se fizermos uso do operador ":". Se quisermos gerar o vector $a = [1, 2, 3, ..., 10]$ podemos fazer

```
>> a= 1:10
```

A notação geral para o operador ":" é a seguinte

```
Número_inicial : incremento : Número_final
```

e permite a geração de sequências de números inteiros como no exemplo anterior ou mesmo de números reais. Eis alguns exemplos:

```
>> e= 0:pi/20:2*pi
>> f= 10:-1:-10
```

O operador ":" pode ser utilizado na geração de vectores de índices obtendo-se uma notação muito compacta. Por exemplo, se quisermos obter as colunas ímpares da matriz

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 6 & 1 & 2 \end{bmatrix}$$

podemos fazer

```
>> B=A(1:3,1:2:3)
```

No caso anterior são indexadas todas as linhas da matriz. Para simplificar a notação, quando se pretende todas as linhas e não se conhece exactamente o número de linhas de uma matriz, pode-se utilizar a notação

```
>> B=A(:,1:2:3)
```

Se quiséssemos obter a primeira linha da matriz A podíamos fazer

```
>> A(1,:)
```

(a) Crie um vetor linha com uma sequência de números pares com início em 4 e a terminar no número 100.

(b) Crie um vetor linha com uma sequência decrescente de números inteiros com início em 5 e a terminar em -5.

(c) Crie um vetor linha com uma sequência de números reais igualmente espaçados com 100 elementos pertencentes ao intervalo [0 ... 1].

(d) Crie uma matriz aleatória usando o comando `>> B= rand(20,30)` (20 linhas e 30 colunas). Construa um comando que permita extrair para uma matriz C uma sub-matriz de B constituída pelas linhas de 10 a 15 e as colunas de 9 a 12.

(e) Gere uma sequência, x, a começar em $-\pi$ e a acabar em π com um passo de $\pi/15$.

(f) Corra o comando `>> plot(x, sin(4*pi*x)`. A que corresponde o gráfico obtido?

2. Considere a experiência aleatória de lançar **3 vezes** uma moeda equilibrada. Pretende-se estimar por simulação a probabilidade de se obter **2 caras** no fim dos **3 lançamentos** .

   Para estimar a probabilidade por simulação, é necessário executar várias vezes a experiência aleatória de lançar 3 vezes uma moeda equilibrada e calcular a percentagem de vezes em que o resultado deu 2 caras. Em Matlab, uma forma possível de implementar este simulador é a seguinte (assumindo que a experiência é executada 10000 vezes):

```
%% Código 1

% Gerar uma matriz com 3 linhas e 10000 colunas de números aleatórios
% entre 0.0 e 1.0 (ou seja, cada coluna representa uma experiência):
  experiencias = rand(3,10000);
% Gerar uma matriz com 3 linhas e 10000 colunas com o valor 1 se o valor
% da matriz anterior for superior a 0.5 (ou seja, se saiu cara) ou com o
% valor 0 caso contrário (ou seja, saiu coroa):
  lancamentos = experiencias > 0.5; % 0.5 corresponde a 1 - prob. de caras
% Gerar um vetor linha com 10000 elementos com a soma dos valores de cada
% coluna da matriz anterior (ou seja, o número de caras de cada experiência):
  resultados= sum(lancamentos);
% Gerar um vetor linha com 10000 elementos com o valor 1 quando o valor do
% vetor anterior é 2 (ou seja, se a experiência deu 2 caras) ou 0 quando é
% diferente de 2:
  sucessos= resultados==2;
% Determinar o resultado final dividindo o número de experiências com 2
% caras pelo número total de experiências:
 probSimulacao= sum(sucessos)/10000
```

Note-se que o código proposto está desenvolvido passo a passo para mais fácil compreensão. Muitas das operações podem ser combinadas por forma a evitar o uso de matrizes intermédias e tornar a execução do código mais eficiente. Além disso, é útil usar variáveis iniciais para os parâmetros do problema para ser mais fácil alterar o código e adaptá-lo a outros casos de interesse. Assim, uma outra forma de implementar o mesmo simulador é, por exemplo, a seguinte:

```
%% Código 1 - segunda versão

N= 1e5;   %número de experiências
p = 0.5; %probabilidade de cara
k = 2;    %número de caras
n = 3;    %número de lançamentos
lancamentos = rand(n,N) > p;
sucessos= sum(lancamentos)==k;
probSimulacao= sum(sucessos)/N
```

(a) Implemente no Matlab as 2 versões de código fornecidas.

(b) Estime, usando as 2 versões do código, a probabilidade de obter 2 caras em 3 lançamentos de uma moeda equilibrada (execute várias vezes a simulação).

3. Estime a probabilidade de obter 6 caras em 15 lançamentos de uma moeda equilibrada? [1]

4. Qual é a probabilidade de obter pelo menos 6 caras em 15 lançamentos de uma moeda equilibrada?

5. Para facilitar o cálculo de outras situações similares às que tratou nos pontos anteriores, crie uma função em Matlab que permita estimar a probabilidade por simulação. A função deve ter os seguintes parâmetros de entrada: p, número de lançamentos, número de caras pretendidas e número de experiências a realizar. Deve utilizar nomes adequados para a função e para os parâmetros de entrada.

   (a) Aplique a função para voltar a estimar as probabilidades das questões anteriores assim como estimar as probabilidades para todo o espaço de amostragem[2] para os seguintes números de lançamentos: 20, 40 e 100.

   (b) Faça 3 gráficos, usando a função `stem`, das probabilidades estimadas para 20, 40 e 100 lançamentos.

[1] Adapte o Código 1 e/ou a segunda versão do código 1 de forma apropriada para resolver as questões de 3 a 5

[2] O espaço de amostragem, S, é o conjunto de todos os resultados possíveis para a experiência aleatória. Por exemplo, no problema 1 $S = \{0, 1, 2, 3\}$.

6. Pretende-se agora calcular de forma analítica as probabilidades estimadas nos exercícios anteriores.

   Considere para isso a seguinte expressão

$$P(k) = C_k^n p^k (1-p)^{n-k}$$

   em que p é a probabilidade de ocorrer o acontecimento em que estamos interessados (por exemplo: se o acontecimento for "sair cara" em cada lançamento, $p = 0.5$), k é o número de acontecimentos que ocorreram em n repetições da experiência aleatória.

   Em Matlab, esta expressão é determinada da seguinte forma:

```
%% Codigo 2-  cálculo analítico de probabilidade em séries experiências de Bernoulli
% Dados relativos ao problema 1
p = 0.5;
k = 2;
n = 3;
prob= nchoosek(n,k)*p^k*(1-p)^(n-k); % nchoosek(n,k)= n!/(n-k)!/k!
```

   Calcule o valor analítico para cada um dos problemas anteriores e compare os resultados obtidos com os valores estimados.

7. Considere um processo de produção fabril que produz torneiras em que a probabilidade de cada torneira ser produzida com defeito é de 30%. No processo de controlo de qualidade, é selecionada uma amostra de 5 peças.

   (a) Calcule analiticamente e por simulação a probabilidade de 3 peças da amostra serem defeituosas.

   (b) Calcule analiticamente e por simulação a probabilidade de, no máximo, 2 das peças da amostra serem defeituosas.

   (c) Baseado em simulação, construa no Matlab o histograma representativo da distribuição de probabilidades do número de peças defeituosas da amostra.

Para revisão, responda, sozinho, às seguintes questões. Sempre que apropriado deve utilizar simulação ou análise combinatória para chegar à(s) resposta(s):

R1 Quantas sequências diferentes de 10 bits há? E de n bits?

R2 Quantas sequências diferentes de 10 símbolos do alfabeto (A,C,G,T) há? E de n símbolos do mesmo alfabeto?

R3 Um teste tem n perguntas com respostas possíveis Verdadeiro ou Falso. Forneça uma expressão para calcular o número de maneiras diferentes de responder ao teste. Qual a probabilidade de acertar todas as respostas, escolhendo-as à sorte com igual probabilidade?

R4 Quantas chaves distintas pode ter o Totoloto antigo (5 números em 49)? E o Euromilhões (5 números em 50 e duas estrelas em 11)?

R5 Considere um baralho com 20 cartas. Dessas cartas, 10 são vermelhas e numeradas de 1 a 10. As restantes 10 são pretas e também numeradas de 1 a 10.

   (a) De quantas maneiras diferentes se podem dispor as 20 cartas numa fila?

   (b) Calcule a probabilidade de se obter uma sequência constituída alternadamente por cartas pretas e vermelhas.

R6 Lançam-se dois dados e toma-se nota da soma de pontos obtida.

   (a) Indique o espaço de amostragem (conjunto de valores possíveis) da soma.

   (b) Calcule a probabilidade de se obter a soma 9.

R7 Um conjunto de 50 peças contém 8 peças defeituosas. Escolhem-se aleatoriamente 10 peças. Qual a probabilidade de encontrar 3 defeituosas?

R8 Quantas *passwords* diferentes se podem obter nas seguintes situações:

(a) comprimento 5 e cada posição contendo um dígito entre 0 e 9;

(b) comprimento 5 e cada posição contendo uma letra mínuscula sem acentos.

(c) Qual a probabilidade de acertar em cada um dos dois casos anteriores escolhendo uma password aleatoriamente ?

(d) Qual a alteração ao valor destas probabilidades de fizermos 3 tentativas completamente independentes?

R1:
(a) Para calcular o número de sequências diferentes de 10 bits, podemos usar a análise combinatória. Cada bit pode ser 0 ou 1, portanto, temos 2 opções para cada bit. Como há 10 bits, o número total de sequências é 2^10 = 1024.

(b) Para n bits, o número de sequências diferentes será 2^n.

R2:
(a) Para calcular o número de sequências diferentes de 10 símbolos do alfabeto (A, C, G, T), temos 4 opções para cada símbolo. Como há 10 símbolos, o número total de sequências é 4^10 = 1,048,576.

(b) Para n símbolos do mesmo alfabeto, o número de sequências diferentes será 4^n.

R3:
Para um teste com n perguntas de resposta Verdadeiro ou Falso, cada pergunta tem 2 opções de resposta. Portanto, o número total de maneiras diferentes de responder ao teste é 2^n. A probabilidade de acertar todas as respostas, escolhendo-as ao acaso com igual probabilidade, é de 1/2^n.

R4:
(a) No Totoloto antigo, há 5 números a serem escolhidos de um conjunto de 49. Isso pode ser calculado usando a combinação, C(49, 5), que é igual a 2,118,760 chaves distintas.

(b) No Euromilhões, há 5 números a serem escolhidos de um conjunto de 50 e 2 estrelas a serem escolhidas de um conjunto de 11. O número total de chaves distintas é C(50, 5) * C(11, 2), que é igual a 2,118,760 * 55 = 116,531,800 chaves distintas.

R5:
(a) Para dispor as 20 cartas numa fila, considerando que as cartas de cada cor são idênticas entre si, podemos usar a permutação. Teríamos 20 cartas para a primeira posição, 19 para a segunda, 18 para a terceira, e assim por diante. Portanto, o número de maneiras diferentes de dispor as cartas é 20!.

(b) Para calcular a probabilidade de obter uma sequência alternada de cartas pretas e vermelhas, primeiro, precisamos contar o número de maneiras de organizar as cartas dessa forma. Temos 10 cartas pretas e 10 vermelhas, e podemos começar com qualquer uma delas. Portanto, o número de maneiras é 2 * 10! (multiplicado por 2 para contar as duas possibilidades de início).

A probabilidade é então (2 * 10!) / 20!.

R6:
(a) O espaço de amostragem da soma de pontos obtida ao lançar dois dados consiste em todos os possíveis valores de soma de 2 a 12.

(b) A probabilidade de obter a soma 9 pode ser calculada contando quantas combinações de resultados dos dados somam 9 (como 3+6, 4+5, 5+4, 6+3). Existem 4 dessas combinações. Como há um total de 36 resultados possíveis ao lançar dois dados, a probabilidade é 4/36 = 1/9.

R7:
Para calcular a probabilidade de encontrar 3 defeituosas ao escolher aleatoriamente 10 peças de um conjunto de 50, podemos usar a fórmula da combinação. O número de maneiras de escolher 3 defeituosas em um grupo de 8 defeituosas é C(8, 3), e o número de maneiras de escolher 7 não defeituosas em um grupo de 42 não defeituosas é C(42, 7).

A probabilidade é dada por (C(8, 3) * C(42, 7)) / C(50, 10).