MPEI 2023-2024

Resolução de exercícios
(consolidação)

Exercício - Inquérito futebolístico

- f: fracção da população que gosta de futebol
- Queremos fazer uma sondagem/inquérito a *n* pessoas
- Quantas pessoas devemos inquirir para ter uma confiança (probabilidade) de 95% de que não cometemos um erro superior a 1 %

- Considere:
  - Resultado de um inquérito à pessoa i:

$$X_i = \begin{cases} 1, & se\ gosta \\ 0, & se\ não\ gosta \end{cases}$$

  - $M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$ fracção de "gosta" na amostras

Resolução

- Sugestões ?
- Uma das formas (veremos outra) é usando a Desigualdade de Chebyshev …

- O que diz a desigualdade ?
- $P(|M_n - E[M_n]| \geq \epsilon) \leq \frac{Var(M_n)}{\epsilon^2}$

O que sabemos ?

- ϵ = ?
- ϵ = 0.01

- $Var(M_n)$ = ?
- $Var(M_n) = \frac{Var(X_i)}{n}$

$Var(X_i) = ?$

- Todas as X_i são v. a. de Bernoulli
  - Mas não sabemos p (o inquérito é para estimar isso)

- Para o nosso caso **é útil o valor máximo** de $Var(X_i)$. Qual esse valor ?

- $Var(X_i) = p(1-p) \leq \frac{1}{4}$

Voltando à desigualdade

- Substituindo temos:

- $P(|M_n - E[M_n]| \geq 0{,}01) \ \leq \ \frac{\frac{\frac{1}{4}}{n}}{0{,}01^2} = \frac{1}{4\ n\ 10^{-4}}$

- Como queremos $P(\quad) \leq 0{,}05$
- $\frac{1}{4\ n\ 10^{-4}} \leq 0{,}05$
- $n = ?$
- $n \geq 50\ 000$ (valor conservador)

E se erro aceitável $(\epsilon) = 0,05$?

- $P(|M_n - E[M_n]| \geq 0,05) \ \leq \ \frac{1}{4\,n\,(0,05)^2}$

- Obtendo-se *n* de:
- $\frac{1}{4\,n\,(0,05)^2} \leq 0,05$

- $n \geq 2000$

Discussão

- Problemas com os valores de *n* que obtivemos:

1. São muito grandes

2. Baseiam-se numa desigualdade que apenas pode dar um majorante/minorante
   – E não um valor "exato"

- Veremos de seguida que se pode fazer melhores estimativas de *n*
  – Mas para isso precisamos saber mais sobre a distribuição de M_n

Resolução usando TLC

- Pretendemos $P(|M_n - f| \leq 0{,}05) \geq 0{,}95$

- O evento que nos interessa calcular a probabilidade é $|M_n - f| \leq 0{,}05$
- Pretendemos portanto $\mathrm{P}\left(\left|\frac{S_n - \boldsymbol{nf}}{n}\right| \leq 0{,}05\right)$
- Como $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ manipulamos para obter $\sqrt{n}\sigma$ no denominador, obtendo

$$P\left(\left|\frac{S_n - nf}{\sqrt{n}\sigma}\right| \leq \frac{0{,}05\sqrt{n}}{\sigma}\right)$$

Resolução usando TLC (cont.)

- Como Z_n tende para $N(0,1)$
- Teremos:

$$P(|M_n - f| \leq 0{,}05) \approx P(|Z| \leq 0{,}05\frac{\sqrt{n}}{\sigma})$$

- E usando majorante para a variância

$$p(1-p) \leq 1/4 \qquad (\Rightarrow \sigma = 1/2)$$

- $P(|M_n - f| \leq 0{,}05) \quad \leq P(|Z| \leq 0{,}1\sqrt{n})$

$P(|Z| \leq 0.1\sqrt{n})$?

- $P(|Z| \leq 0{,}1\sqrt{n})$
- $= P(-0{,}1\sqrt{n} \leq Z \leq 0{,}1\sqrt{n}\ )$
- $= F_{N(0,1)}(0{,}1\sqrt{n}) - F_{N(0,1)}(-0{,}1\sqrt{n})$

- Para permitir usar tabelas, coloquemos em função de $Q(z) = 1 - F_{N(0,1)}(z)$
  - Sabe-se também que $F_{N(0,1)}(-z) = Q(z)$
- $= 1 - \ Q(0{,}1\sqrt{n}) - Q(0{,}1\sqrt{n})$
- $= 1 - 2\, Q(0{,}1\sqrt{n})$

Terminando...

- $1 - 2\,Q(0{,}1\sqrt{n})$ terá de ser $\geq 0{,}95$

- $1 - 2\,Q(0{,}1\sqrt{n}) \geq 0{,}95$
- $\Rightarrow Q(0{,}1\sqrt{n}) \geq \mathbf{0,025}$
- $\Rightarrow 0{,}1\sqrt{n} \geq \mathbf{1,96}$ por consulta a tabela

- Resolvendo em ordem a n temos, finalmente,
- $\sqrt{n} \geq (19{,}6) \quad \Rightarrow n \geq 384{,}16$

- $n = 385$ é o número mínimo que procurávamos

Para mais informação

- Capítulo 7, “Somas de variáveis aleatórias e situações limite”, do livro de F. Vaz e A. Teixeira

Problema – Ganho médio num jogo

- Considere o seguinte jogo de cartas entre si e um amigo:
  - Apenas se usam as figuras (Rei, Dama e Valete), o Ás e o Joker de um dos naipes
  - As cartas são baralhadas honestamente e extraídas uma a uma
  - Sempre que se extrai uma figura ganha 1 Euro, cada Ás obriga-o a pagar 1 Euro ao seu amigo, cada Joker implica pagar 2 Euros ao seu amigo.
  - Volta a colocar-se a carta no baralho e baralha-se

- Questão: **Qual o seu ganho médio ao fim de uma longa sequência de jogadas ?**

Fonte: "O ACASO", J. M. Sá, Gradiva

Exemplo

- Para perceber melhor, analisemos um possível jogo:

Saiu	**A**	**R**	**A**	**R**	**A**	**J**	**V**	**D**	**V**	**…**
Seu ganho	-1	1	-1	1	-1	-2	1	1	1	
Ganho acumulado	-1	0	-1	0	-1	-3	-2	-1	0	…

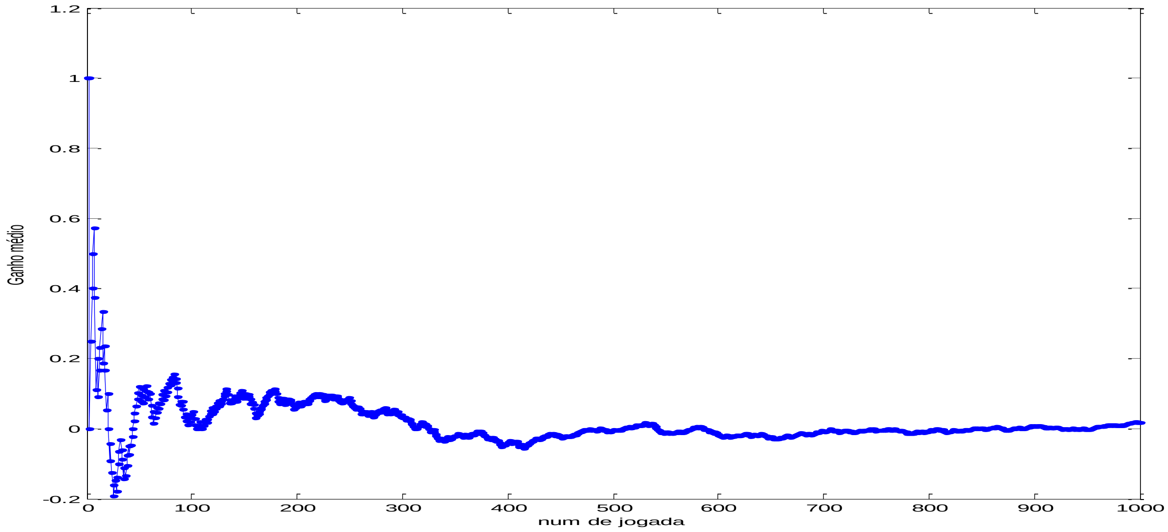
Simulemos ...

- Para simplificar, consideremos a seguinte codificação:
  Rei=1, Dama=2, Valete=3, Ás=4, Joker=5

```
% simular N extracções
e= floor (rand(1,N)*5)+1;   % inteiros de 1 a 5 (equiprováveis)

%  calcular ganho acumulado ao longo do jogo
g=e;  g(e==1 | e==2 | e=3)=1;      g(e==4 )=-1;      g(e==5 )=-2;
gacum= cumsum(g)

%  ganho médio ao longo do tempo
n=1:N;
gmed= gacum ./ n;

Plot(n,gmed, '.')
```

Exemplo de resultado

- 1000 jogadas



- É apenas uma de um número muito grande de sequências possíveis (são possíveis 5^{1000})

Resolução …

- Tentemos agora resolver sem simulação e usando o que já sabemos …

- Sugestões ?

Variável aleatória Ganho

- O Ganho em cada extracção de uma carta (experiência aleatória) pode ser considerada uma variável aleatória (G)
- Pode assumir os valores {-2,-1,1}

- Função de probabilidade ?

$$p_G(G = -2) = ?$$
$$p_G(G = -1) = ?$$
$$p_G(G = 1) = ?$$

Função de probabilidade de G

- $p_G(G = -2) = 1/5$
  - Equivalente a P("sair um Joker")

- $p_G(G = -1) = 1/5$

- $p_G(G = 1) = 3/5$

Ganho médio …

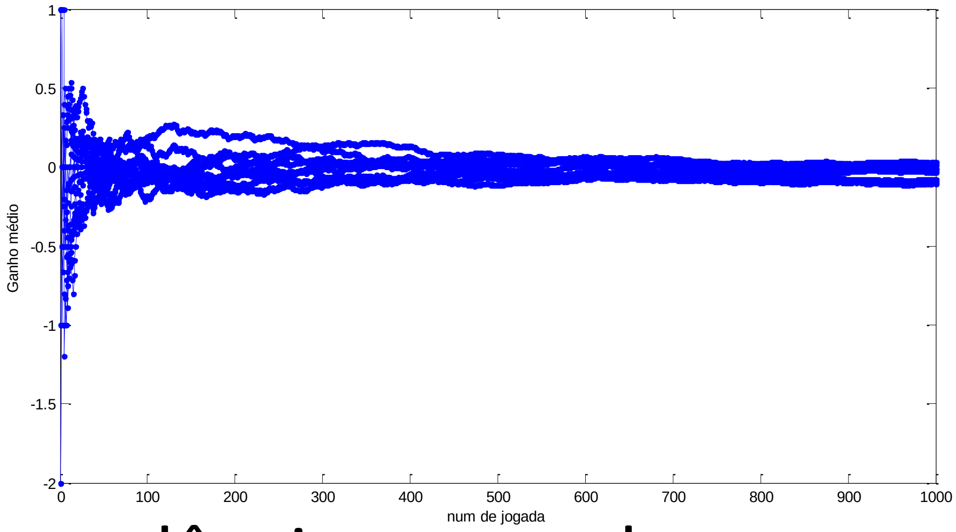
- O "Ganho médio esperado" ao fim de um grande número de jogadas é designado por **esperança matemática da variável Ganho**

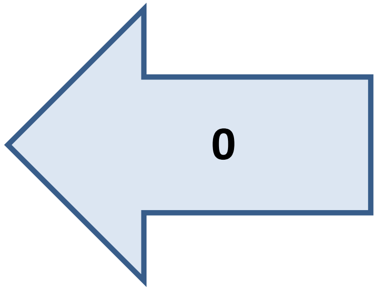
- $\mathrm{E}[G] \ = \sum x_i p(X = x_i)$

- Aplicando ao nosso caso…
- $\mathrm{E}[G] \ = 1 \times \frac{3}{5} + (-1) \times \frac{1}{5} + (-2) \times \frac{1}{5} = 0$

Longa sequência

- É altura de esclarecer um pouco mais o que se entende por “longa sequência”…
- Repetindo o jogo (ou melhor simulando..), temos





- Existe uma tendência para todas as curvas estabilizarem em torno da esperança (0)

Problemas com mais do que uma v.a

3. (2 valores) Considere duas variáveis X e Y com a função massa de probabilidade conjunta seguinte:

$X \setminus Y$	0	1
0	$\frac{1}{9}$	$\frac{2}{9}$
1	$\frac{2}{9}$	$\frac{4}{9}$

a) Calcule as probabilidades marginais de X e de Y.

b) Calcule a média e a variância de X.

Exercícios de mini-testes

Exercício 1 (2015-2016)

8.0 **2)** Considere que tem um pequeno conjunto de páginas web identificadas pelas letras A a F com as seguintes ligações entre si no dia 1 de janeiro de 2016: a página A tem links para as páginas B a E; a página B tem links para as páginas B, D e F; pode chegar-se às páginas A e B através da página C; D apenas tem links para E e B; E tem links para A e B; F possui links para todas as outras páginas, excepto para ela própria.

Considerando que se pretende obter o pagerank das páginas e que se inicializa esse valor com um valor igual para todas as páginas e igual a 1/6:

3.0 **2.a)** Qual o valor da estimativa do pagerank de cada página ao fim de três iterações do processo de cálculo?

Resposta: A ____________________ B ____________________

B ____________________ D ____________________

E ____________________ F ____________________

2.0 $\boxed{\textbf{2.b)}}$ Represente num gráfico a evolução do valor do pagerank de cada uma das páginas em função da iteração? Deve utilizar um número de iterações suficiente para que os valores estabilizem.

Código matlab/octave:

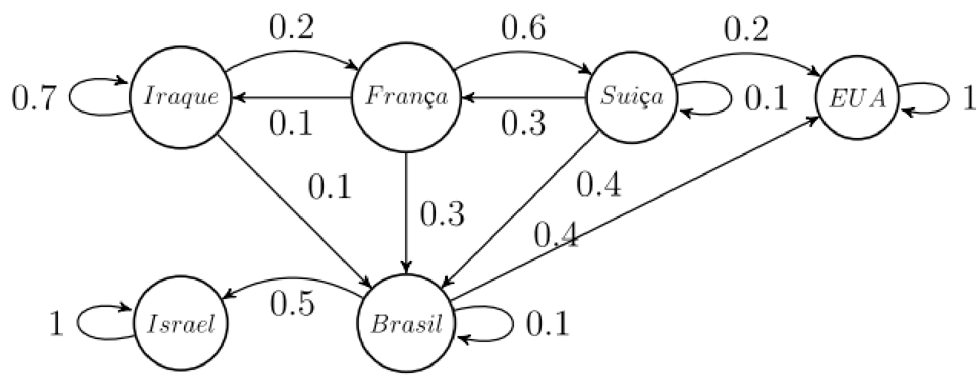
3.0 $\boxed{\textbf{2.c)}}$ Qual a página com o maior valor do pagerank e qual o seu valor?

Resposta: Página? __________
pagerank? __________

Código matlab/octave:

Exercício de 2018-2019

$\boxed{\textbf{2)}}$ Considere o conjunto de países C={Iraque,França,Brasil,Suiça,EUA,Israel} e a informação da figura seguinte relativa à probabilidade de, ao fim de um mês, um terrorista ter viajado de um país para outro ou ter permanecido nesse país:



2.0 $\boxed{\textbf{2.a)}}$ Represente em Matlab a matriz de transição T na sua forma canónica, sendo T_{ji} a probabilidade de viagem de i para j. Considere permanecer num país como equivalente a uma viagem de i para i e que cada transição corresponde a 1 mês. Represente também o vector estado v correspondente à seguinte situação: terrorista está na Europa e com igual probabilidade de estar num dos países possíveis.

Código Matlab:

2.0 $\boxed{\textbf{2.b)}}$ Qual a média (valor esperado) do número de meses necessários para um terrorista inicialmente no Iraque vir a terminar os seus dias em Israel ou nos EUA?

Resposta: ____________

Código Matlab:

2.0 $\boxed{\textbf{2.c)}}$ Qual a probabilidade de um terrorista que esteja inicialmente no Iraque se encontrar no EUA passados 5 meses? Qual a probabilidade de estar em Israel 50 meses depois de se encontrar no Brasil?

Resposta: P[Iraque $\rightarrow$ EUA, em 5 meses] = ____________
P[Brasil $\rightarrow$ Israel, em 50 meses] = ____________

Código Matlab: