Trabalho Final - Probabilidade e Estatística (2025.2)

Equipe 02

Ismael S. Silva, Paulo R. S. Menezes, Ana N. M. de Araujo Kaique D. Sousa, Wanessa R. Santos, Eros R. Simette Thiago C. S. Oliveira, Pedro H. Q. da Silva Marcos G. P. Galdino, Luiz M. R. de Souza, Adrian E. O. Azevedo

Universidade Federal do Ceará (UFC)

21 de julho de 2025

Enunciado da Questão

Considere uma variável aleatória X_i que segue uma distribuição Exponencial com parâmetro $\lambda = 3$.

a) Gere $N = 10.000$ observações de X_i e construa o histograma.

b) Crie uma função ou linha de código que gere uma amostra de tamanho n (ex: $n = 2, 5, 50$) de X_i.

c) Crie uma função ou linha de código que calcule a média de uma amostra de tamanho n.

d) Para cada tamanho de amostra $n \in \{2, 5, 50\}$, gere 10.000 amostras e calcule a média de cada uma. Construa o histograma da distribuição dessas 10.000 médias amostrais $(\bar{X})$ para cada valor de n.

e) Calcule a média e o desvio padrão das 10.000 médias amostrais $(\bar{X})$ para cada n.

f) Com base nos resultados, enuncie o Teorema do Limite Central e explique como ele se aplica a este experimento.

a) Histograma da Distribuição Original

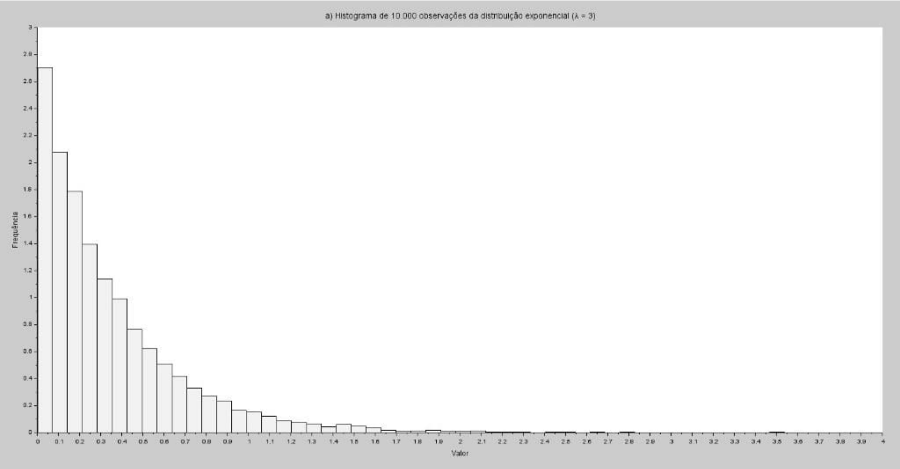


Figura 1: *

b) e c) Geração e Média de Amostras

b) Crie uma função ou linha de código que gere uma amostra de tamanho n de X_i.

```
// O parametro da exponencial no Scilab e a media (1/lambda)
grand(1, n, "exp", 1/3)
```

Essa função cria a amostra com n observações.

c) Crie uma função ou linha de código que calcule a média de uma amostra de tamanho n.

```
mean(grand(1, n, "exp", 1/3))
```

Essa função calcula a média da amostra.

d) Código para Gerar Médias Amostrais

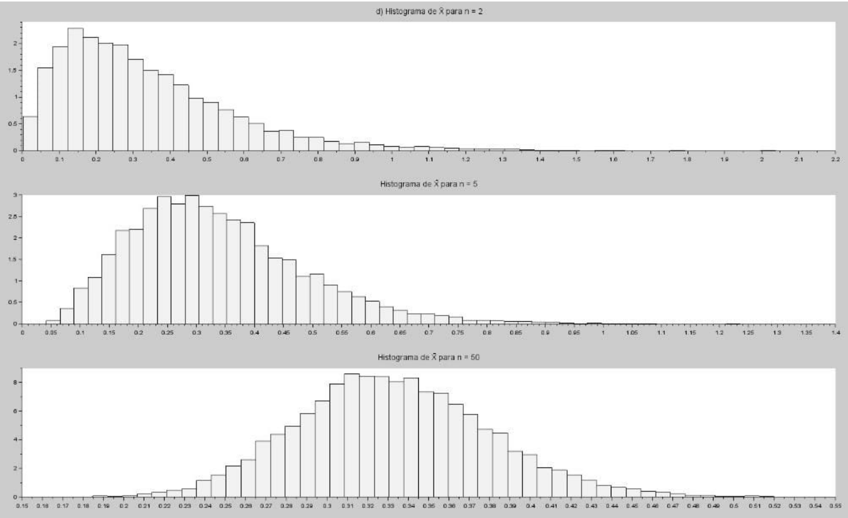
d) Para cada $n \in \{2, 5, 50\}$, gere 10.000 amostras e calcule a média de cada uma. Construa o histograma da distribuição dessas médias.

```
lambda = 3;
media_teorica = 1 / lambda;
N = 10000; // Número de amostras

// Gera 10.000 amostras (linhas) com n elementos (colunas)
Xn2 = grand(N, 2, "exp", media_teorica);
Xn5 = grand(N, 5, "exp", media_teorica);
Xn50 = grand(N, 50, "exp", media_teorica);

// Calcula a média de cada linha (amostra)
Xbar2 = mean(Xn2, 2);
Xbar5 = mean(Xn5, 2);

Xbar50 = mean(Xn50, 2);
```

d) Código para Gerar Médias Amostrais

```
// Código para plotar os histogramas (exemplo para n=50)
histplot(50, Xbar50);
xtitle("Histograma das Medias Amostrais (n=50)");
```

d) Histogramas das Médias Amostrais



e) Estatísticas das Médias Amostrais

e) Calcule a média e o desvio padrão das 10.000 médias amostrais ($\bar{X}$) para cada n.

```
// Média das médias amostrais
media_Xbar2 = mean(Xbar2)   // Resultado esperado: ~0.333
media_Xbar5 = mean(Xbar5)   // Resultado esperado: ~0.333
media_Xbar50 = mean(Xbar50) // Resultado esperado: ~0.333

// Desvio padrão das médias amostrais
dp_Xbar2 = stdev(Xbar2)     // Esperado: (1/3)/sqrt(2) ~ 0.2357
dp_Xbar5 = stdev(Xbar5)     // Esperado: (1/3)/sqrt(5) ~ 0.1490
dp_Xbar50 = stdev(Xbar50)   // Esperado: (1/3)/sqrt(50) ~
    0.0471
```

f) O Teorema do Limite Central (TLC)

f) Com base nos resultados, enuncie o Teorema do Limite Central e explique como ele se aplica a este experimento.

Enunciado do TLC:

> *Dada uma população com média μ e desvio padrão σ finitos, a distribuição das médias amostrais ($\bar{X}$) de amostras de tamanho n retiradas desta população aproxima-se de uma distribuição Normal com média μ e desvio padrão $\sigma/\sqrt{n}$, à medida que n se torna suficientemente grande.*

Aplicação ao experimento:

- A distribuição original é fortemente assimétrica (Gráfico 1).
- No entanto, os histogramas das médias amostrais mostram que, à medida que n aumenta de 2 para 50, a distribuição de $\bar{X}$ torna-se cada vez mais simétrica e com formato de sino, ou seja, Normal.
- As médias das $\bar{X}$ mantiveram-se em torno de $\mu = 1/3$, e os desvios padrão diminuíram visivelmente, aproximando-se de $\sigma/\sqrt{n}$, confirmando as previsões do teorema.

Questão 4

Considere que $X_1, \ldots, X_n$ consiste em n observações de uma amostra aleatória com distribuição estatística contínua uniforme no intervalo $[0, 6]$. Considere a v.a. $\bar{X}$ derivada das v.a.'s $X_1, \ldots, X_n$ da seguinte forma:

$$\bar{X} = \frac{x_1 + \cdots + x_n}{n}.$$

a) Geração das amostras

a) Gere um conjunto de 1200 amostras das v.a.'s $X_1, \ldots, X_n$ com $n = 800$ observações cada. Para cada uma das amostras geradas anteriormente calcule os valores de $\bar{X}$.

Por que? Simular as médias amostrais de $U(0,6)$ para observar seu comportamento e fundamentar os cálculos de intervalo de confiança.

```
// Questao a - Geracao das amostras e calculo das medias
A = grand(1200, 800, "uin", 0, 6);
mediaA = mean(A, "c");
desvioA = stdev(A, "c");
```

b) Intervalos de Confiança (1/3)

b) Para cada amostra das v.a.'s $X_1, \ldots, X_n$ geradas no item anterior, construa o intervalo de confiança para a média populacional $\bar{X}$ com níveis de 90%, 95% e 99%.

Sabemos que cada amostra tem tamanho $n = 800$ e, para $U(0,6)$, $\mu = 3$. Desejamos

$$P\big(L \leq \bar{X} \leq U\big) = \lambda,$$

com $\lambda \in \{0.90, 0.95, 0.99\}$.

Pelo Teorema do Limite Central:

$$\bar{X} \overset{\text{aprox.}}{\sim} \mathcal{N}\big(\mu, \sigma/\sqrt{n}\big).$$

Como n é grande, usamos $\sigma \approx s$. Normalizando:

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \Longrightarrow \quad Z \sim \mathcal{N}(0,1).$$

b) Intervalos de Confiança (2/3)

Após a normalização, temos

$$P(Z_1 \le Z \le Z_2) = \lambda,$$

onde

$$P(Z_1 \le Z) = P\left(L \le \bar{X}\right) = \tfrac{1-\lambda}{2}, \quad P(Z \le Z_2) = P\left(\bar{X} \le U\right) = \tfrac{1-\lambda}{2}.$$

Como $Z_2 = -Z_1 = \rho$, segue

$$P(-\rho \le Z \le \rho) = \lambda \implies P\left(\bar{X} - \rho\,\tfrac{s}{\sqrt{n}} \le \mu \le \bar{X} + \rho\,\tfrac{s}{\sqrt{n}}\right).$$

Assim, o intervalo de confiança é

$$IC = \left(\bar{X} - \rho\,\tfrac{s}{\sqrt{n}},\ \bar{X} + \rho\,\tfrac{s}{\sqrt{n}}\right),$$

com

$$\rho = \begin{cases} 1.645, & \lambda = 0.90, \\ 1.960, & \lambda = 0.95, \\ 2.576, & \lambda = 0.99. \end{cases}$$

b) Intervalos de Confiança (3/3)

```
// Questao b - Calculo dos intervalos de confianca
intervalos90 = zeros(1200, 2);
intervalos95 = zeros(1200, 2);
intervalos99 = zeros(1200, 2);

for i = 1:1200
    intervalos90(i, 1) = mediaA(i, 1) - (1.645 * desvioA(i, 1) / sqrt(800));
    intervalos95(i, 1) = mediaA(i, 1) - (1.960 * desvioA(i, 1) / sqrt(800));
    intervalos99(i, 1) = mediaA(i, 1) - (2.576 * desvioA(i, 1) / sqrt(800));

    intervalos90(i, 2) = mediaA(i, 1) + (1.645 * desvioA(i, 1) / sqrt(800));
    intervalos95(i, 2) = mediaA(i, 1) + (1.960 * desvioA(i, 1) / sqrt(800));
    intervalos99(i, 2) = mediaA(i, 1) + (2.576 * desvioA(i, 1) / sqrt(800));
end
```

c) Cobertura dos Intervalos

c) Verifique o percentual dos intervalos estatísticos com grau de confiança de 90% que realmente contem o valor real da média populacional de $\bar{X}$. E para os graus de confiança de 95% e 99%?

```
acertos90 = 0;
acertos95 = 0;
acertos99 = 0;
for i = 1:1200
    if 3 > intervalos90(i, 1) & 3 < intervalos90(i, 2) then
        acertos90 = acertos90 + 1;
    end
    if 3 > intervalos95(i, 1) & 3 < intervalos95(i, 2) then
        acertos95 = acertos95 + 1;
    end
    if 3 > intervalos99(i, 1) & 3 < intervalos99(i, 2) then
        acertos99 = acertos99 + 1;
    end
end
```

Código: Verificação da Proporção de Cobertura (Parte 2)

```
porcentagem90 = acertos90 / 1200 * 100;
porcentagem95 = acertos95 / 1200 * 100;
porcentagem99 = acertos99 / 1200 * 100;

disp("Porcentagem de intervalos que contem a media real (3):");
disp("90% de confianca: " + string(porcentagem90) + "%");
disp("95% de confianca: " + string(porcentagem95) + "%");
disp("99% de confianca: " + string(porcentagem99) + "%");
```

Enunciado da Questão

Considere a função densidade de probabilidade (fdp) abaixo:

$$f(x) = \begin{cases} 3e^{-3x}, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

(a) Calcule a função de distribuição acumulada (FDA) e sua inversa.

(b) Gere uma amostra de 10.000 observações usando o método da FDA inversa.

(c) Esboce a fdp teórica e o histograma da amostra. Eles são semelhantes?

(d) Calcule a média e a variância populacional.

(e) Calcule a média e a variância amostral.

Item (a): Cálculo da FDA

Dada a fdp da distribuição exponencial com $\lambda = 3$:

$$f(x) = 3e^{-3x} \quad \text{para } x \geq 0$$

A função de distribuição acumulada (FDA), $F(x)$, é a integral da fdp:

$$F(x) = \int_0^x 3e^{-3t}\,dt$$

Resolvendo a integral, temos:

$$\int_0^x 3e^{-3t}\,dt = \left[-e^{-3t}\right]_0^x = -e^{-3x} - (-e^0) = 1 - e^{-3x}$$

Portanto, a FDA é:

$$F(x) = \begin{cases} 1 - e^{-3x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Item (a): Inversa da FDA

Para encontrar a função inversa da FDA, $F^{-1}(u)$, fazemos $u = F(x)$ e isolamos x.

$$u = 1 - e^{-3x}$$

$$e^{-3x} = 1 - u$$

$$\ln(e^{-3x}) = \ln(1-u)$$

$$-3x = \ln(1-u)$$

$$x = -\frac{1}{3}\ln(1-u)$$

A função inversa da FDA, para $0 < u < 1$, é:

$$F^{-1}(u) = -\frac{\ln(1-u)}{3}$$

Item (b): Geração da Amostra (Scilab)

```
// Par metros
lambda = 3;
n = 10000;

// FDA inversa da exponencial
function x = exponencial_inv(u, lambda)
    x = -log(1 - u) / lambda;
endfunction

// 1. Gerar n observaç es  de uma distribuição Uniforme(0,1)
U = rand(n, 1);

// 2. Aplicar a FDA inversa para obter a amostra exponencial
amostra_fda = exponencial_inv(U, lambda);
```

Item (b): Geração da Amostra (Scilab)

```
// Exibindo a média da amostra para verificação
printf("Média amostral: %.4f\n", mean(amostra_fda));
// Valor teórico esperado: 1/3 = 0.3333
```

Item (b): Histograma da Amostra

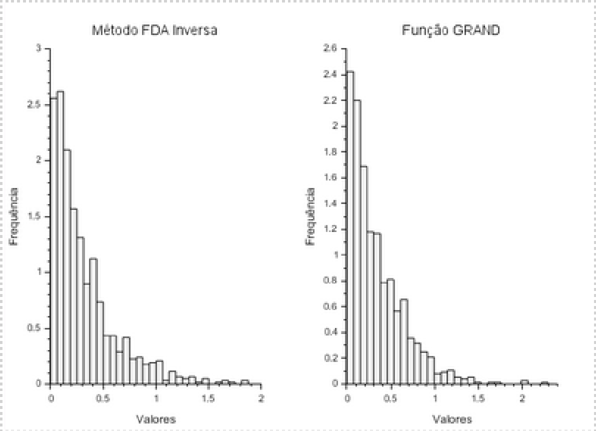


Figura 3: Histograma da amostra de 10.000 observações gerada com o método da FDA inversa.

Item (c): Código da Comparação (Scilab)

```
// Vetor de pontos para a PDF teórica
x_teorico = linspace(0, max(amostra_fda), 200);
pdf_teorica = lambda * exp(-lambda * x_teorico);

// Plotar histograma normalizado e PDF teórica
scf();
histplot(30, amostra_fda, normalization=%t); // Normalizado
plot(x_teorico, pdf_teorica, 'r-', 'LineWidth', 2);

// Configuraç es  do gráfico
title('Comparação: PDF Teórica vs Histograma da Amostra');
xlabel('Valores');
ylabel('Densidade de Probabilidade');
legend(['Histograma (n=10000)'; 'PDF Teórica (  =3)']);
```

Comparação: PDF Teórica vs Histograma

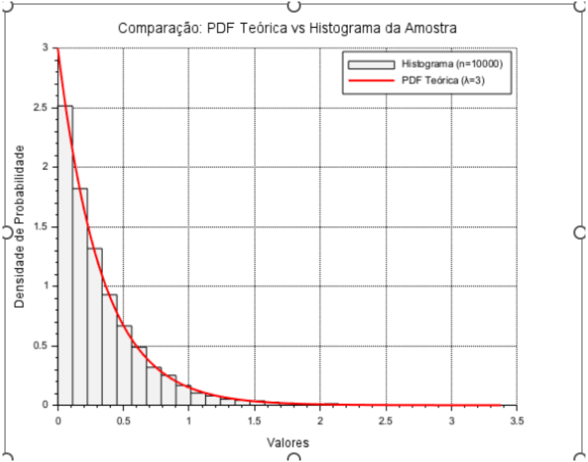


Figura 4: O histograma da amostra se aproxima muito bem da curva da fdp teórica, validando o método de geração.

Item (d): Média e Variância Populacional

A **média** (ou esperança) $\mathbb{E}(X)$ de uma exponencial com parâmetro λ é:

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

Com $\lambda = 3$, temos:

$$\mathbb{E}(X) = \boxed{\frac{1}{3}}$$

A **variância** $\text{Var}(X)$ de uma exponencial com parâmetro λ é:

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Com $\lambda = 3$, temos:

$$\text{Var}(X) = \frac{1}{3^2} = \boxed{\frac{1}{9}}$$

Item (e): Média e Variância Amostral

Usando o Scilab para calcular as estatísticas da amostra gerada:

```
// 'amostra_fda' é o vetor com 10.000 observaç es

media_amostral = mean(amostra_fda);
variancia_amostral = variance(amostra_fda);

printf("Média Teórica: %.4f\n", 1/3);
printf("Média Amostral: %.4f\n\n", media_amostral);

printf("Vari ncia Teórica: %.4f\n", 1/9);
printf("Vari ncia Amostral: %.4f\n", variancia_amostral);
```

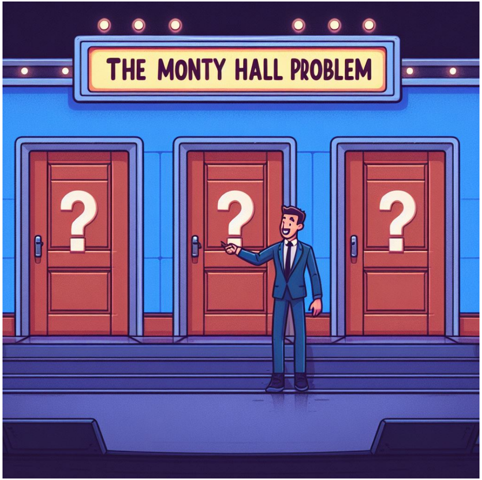
Resultado esperado: Os valores da média e variância amostral devem ser muito próximos dos valores teóricos ($1/3 \approx 0.3333$ e $1/9 \approx 0.1111$).

Questão 17

Considere um programa de TV em que um participante deve escolher uma de três portas. Atrás de uma das portas existe um carro, enquanto nada existe atrás das outras duas.

A produção coloca o prêmio de forma aleatória atrás de uma das portas. Após a escolha de uma porta, o apresentador abre uma das outras portas que não contém o prêmio, e pergunta ao participante se ele deseja trocar sua escolha.

Qual porta escolher?



Simulação - Problema de Monty Hall

```
// --- Simulação do Problema de Monty Hall com Registro de
   Convergência ---

// 1. Definição dos par metros
num_simulacoes = 50000;
nome_arquivo = "monty_hall_convergence.csv";

// 2. Inicialização dos contadores e da matriz para armazenar
   os resultados
vitorias_mantendo = 0;
vitorias_trocando = 0;
resultados_convergencia = zeros(num_simulacoes, 3);
```

Simulação - Problema de Monty Hall

```
// 3. Loop principal da simulação
for i = 1:num_simulacoes
    porta_premiada = grand(1, 1, 'uin', 1, 3);       // Porta
        com o prêmio
    escolha_inicial = grand(1, 1, 'uin', 1, 3);      // Escolha
        do participante

    // Verifica se ganharíamos mantendo ou trocando a escolha
    if escolha_inicial == porta_premiada then
        vitorias_mantendo = vitorias_mantendo + 1;
    else
        vitorias_trocando = vitorias_trocando + 1;
    end
```

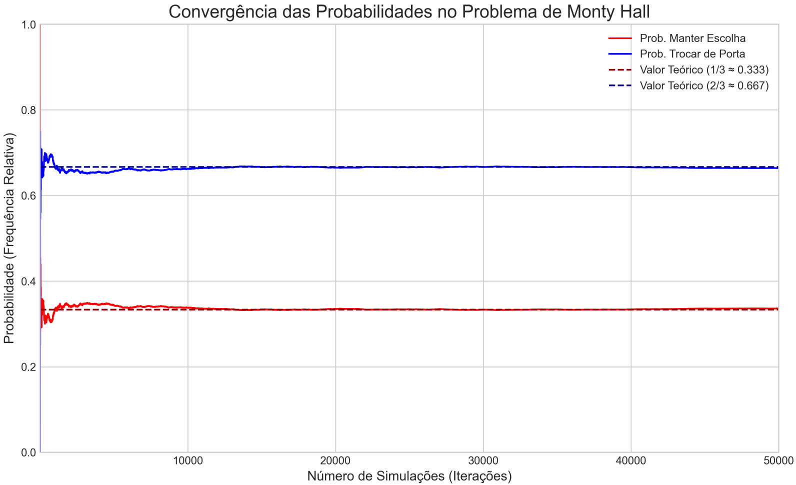
Simulação - Problema de Monty Hall

```
    // Calcula as probabilidades acumuladas até o momento
    prob_manter_atual = vitorias_mantendo / i;
    prob_trocar_atual = vitorias_trocando / i;

    // Armazena os dados de convergência
    resultados_convergencia(i, :) = [i, prob_manter_atual,
        prob_trocar_atual];
end
```

Simulação - Problema de Monty Hall

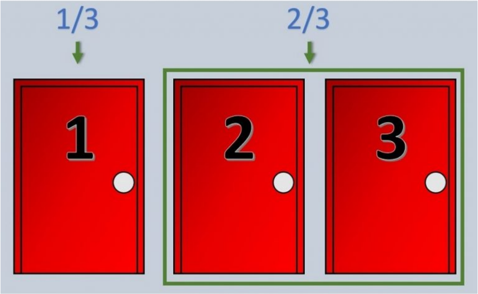
```
// 4. Salvamento dos resultados em arquivo CSV
fid = mopen(nome_arquivo, 'w');
if fid == -1 then
    error("ERRO: Não foi possível criar ou abrir o arquivo para
        escrita: " + nome_arquivo);
end

// Escreve o cabeçalho do CSV
mfprintf(fid, 'Iteracao,Prob_Manter,Prob_Trocar\n');

// Grava os dados linha por linha
for k = 1:size(resultados_convergencia, 1)
    mfprintf(fid, '%d,%.6f,%.6f\n', resultados_convergencia(k,
        1), resultados_convergencia(k, 2),
        resultados_convergencia(k, 3));
end
```

Convergência



Sempre trocando de porta

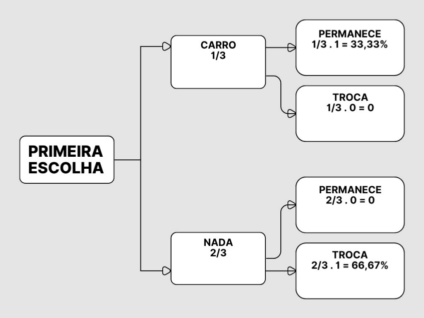
A resposta intuitiva ao problema é a que quando o apresentador revelou uma das portas não premiadas, o participante passaria a ter um novo dilema com duas portas e um prêmio. Portanto, a chance do prêmio estar em uma das portas seria 1/2.
A resposta contra-intuitiva e certa é que é mais vantajoso trocar, pois a chance de ganhar trocando é o dobro da chance de ganhar sem trocar



Solução

Resposta correta: é melhor trocar de porta!

- Probabilidade de ganhar sem trocar: **1/3**
- Probabilidade de ganhar trocando: **2/3**



Referências Bibliográficas I

- MONTGOMERY, D. C.; RUNGER, G. C. *Estatística Aplicada e Probabilidade para Engenheiros*. 6ª ed. LTC, 2016.
- ROSS, S. M. *Probabilidade e Estatística para Engenharia e Ciências*. 9ª ed. Elsevier, 2018.
- WASSERMAN, L. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.
- PAPOULIS, A.; PILLAI, S. U. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.
- GRIMMETT, G.; STIRZAKER, D. *Probability and Random Processes*. Oxford University Press, 3ª ed., 2001.

Obrigado pela atenção!