



Statistics for Engineers

TI0111

HOMEWORK II

NOME COMPLETO: SAVLIO CARVALHO PONTES

NÚMERO DE MATRICULA: 567715

NOME COMPLETO: THAÍS SOUSA BARROS LEAL

NÚMERO DE MATRICULA: 565548

Base Teórica

Esta base teórica apresenta os principais conceitos probabilísticos e estatísticos necessários para o desenvolvimento das análises propostas neste trabalho. São abordadas variáveis aleatórias, funções de distribuição, esperança matemática, variância e as principais distribuições de probabilidade discretas e contínuas utilizadas nas questões, incluindo as distribuições Uniforme, Binomial, Poisson e Normal, além de conceitos de simulação estatística.

Variáveis Aleatórias

Uma variável aleatória é uma função que associa valores numéricos aos resultados de um experimento aleatório. Elas podem ser classificadas em:

- **Discretas:** assumem valores contáveis, como o número de clientes que pedem sobre-mesa ou o número de vencedores em um sorteio.
- **Contínuas:** assumem valores em intervalos reais, como medições de temperatura.

Funções de Distribuição

Para uma variável aleatória discreta X , define-se a **função massa de probabilidade (PMF)** como

$$P(X = x),$$

e a **função distribuição acumulada (CDF)** como

$$F_X(x) = P(X \leq x).$$

Para uma variável aleatória contínua, utiliza-se a **função densidade de probabilidade (PDF)** $f_X(x)$, tal que

$$P(a < X < b) = \int_a^b f_X(x) dx.$$

Expectativa(ou Valor Esperado) e Medidas de Dispersão

A esperança matemática (valor esperado) de uma variável aleatória discreta X é definida por

$$E[X] = \sum_x x P(X = x).$$

A variância é dada por

$$\text{Var}(X) = E[X^2] - E[X]^2,$$

e o desvio-padrão corresponde à raiz quadrada da variância.

Distribuição Uniforme

A distribuição uniforme contínua modela situações em que todos os valores em um intervalo possuem a mesma probabilidade. Se $X \sim \text{Unif}(a, b)$, sua densidade é

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{caso contrário.} \end{cases}$$

Essa distribuição é fundamental na geração de números aleatórios e serve como base para métodos de simulação, como a transformação de Box-Muller.

Distribuição Binomial

A distribuição binomial descreve o número de sucessos em n ensaios independentes, cada um com probabilidade de sucesso p . Se $X \sim \text{Bin}(n, p)$,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Seu valor esperado e variância são

$$E[X] = np, \quad \text{Var}(X) = np(1-p).$$

Distribuição de Poisson

A distribuição de Poisson é utilizada para modelar o número de ocorrências de eventos raros. Se $X \sim \text{Pois}(\lambda)$,

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

O valor esperado e a variância são iguais a λ . Essa distribuição é uma boa aproximação da binomial quando n é grande e p é pequeno, com $\lambda = np$.

Distribuição Normal

A distribuição normal é uma distribuição contínua amplamente utilizada para modelar fenômenos naturais. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, sua densidade é

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Probabilidades são obtidas por meio da função distribuição acumulada da normal.

Simulação e Transformação de Box-Muller

A simulação estatística permite a geração de dados artificiais a partir de distribuições teóricas. A transformação de Box-Muller utiliza variáveis aleatórias uniformes independentes para gerar variáveis com distribuição normal padrão, possibilitando a análise empírica de fenômenos modelados por distribuições normais.

1 Questão 1

Em um restaurante muito frequentado, aproximadamente 70% dos clientes pedem uma sobremesa após o prato principal. Seja X a variável aleatória que representa o número de clientes que pedem sobremesa em uma amostra aleatória de $n = 50$ clientes.

1. Determine a função de distribuição de X .
2. Construa os gráficos da função massa de probabilidade (PMF) e da função distribuição acumulada (CDF) de X .
3. Calcule o valor esperado, a variância e o desvio padrão dev X .
4. Calcule a probabilidade de:
 - (a) $P(X \geq 20)$.
 - (b) $P(30 < X < 43)$.
 - (c) $P(X = 31)$.
5. Suponha que o restaurante estoque sobremesas com base na demanda esperada. Como o uso da distribuição de X poderia ajudar a reduzir desperdício e evitar falta de produtos.
6. Como mudanças em p (por exemplo, sobremesa se torna mais popular, $p = 0.8$) ou em n (número de clientes) afetariam a forma e as probabilidades de X ?

Solução da questão

1.1) Função de distribuição.

Cada cliente pede sobremesa de forma independente com probabilidade $p = 0.7$. Considerando X a variável aleatória que representa o número de clientes que pedem sobremesa numa amostra de 50 pessoas ($n = 50$).

Como o experimento consiste em tentativas independentes com probabilidade constante de sucesso, X segue uma distribuição binomial:

$$X \sim \text{Binomial}(50, 0.7).$$

A função de massa de probabilidade (PMF) é

$$P(X = k) = \binom{50}{k} (0.7)^k (0.3)^{50-k}, \quad k = 0, 1, \dots, 50,$$

e a função de distribuição acumulada (CDF) é

$$F_X(k) = P(X \leq k) = \sum_{i=0}^k \binom{50}{i} (0.7)^i (0.3)^{50-i}.$$

Essas funções permitem descrever completamente o comportamento probabilístico do número de clientes que pedem sobremesa.

1.2) Gráficos da função massa de probabilidade (PMF) e da função distribuição acumulada (CDF) de X .

- o Código em R

Utilizando as funções $dbinom$ e $pbinom$ para o cálculo da PMF e da CDF, respectivamente, os gráficos das Figuras 1 e 2 foram construídos pelo código abaixo.

```
1 #Gráfico da PMF
2 PMF <- dbinom(x, size = n, prob = p)
3 plot(x, PMF,
4       type = "h",
5       lwd = 2,
6       xlab = "Número de clientes que pedem sobremesa",
7       ylab = "P(X = x)",
8       main = "")
9 # Curva Normal de aproximação
10 curve(dnorm(x, mean = valor_esperado, sd = desvio), add = TRUE, col = "red", lwd = 2)
11 legend("topleft", legend = c("Binomial", "Normal"), col = c("black", "red"), lwd = 2)
12
13
14 #Gráfico da CDF
15 CDF = c(0, cumsum(PMF))
16 CDF.plot = stepfun(x,CDF,f=0)
17 par(mar=c(6,5,4,2))
18 plot.stepfun(CDF.plot,col='blue',
19               xlab=TeX('x (valores que X pode assumir)'),ylab=TeX('F_X(x)'),
20               verticals=FALSE,
21               do.points=TRUE,
22               pch=19, lwd=2,
23               cex.lab=1.3,
24               cex.axis=1.2,
25               main="")
26 abline(h=0)
```

Listado 1: Gráficos da PMF e CDF

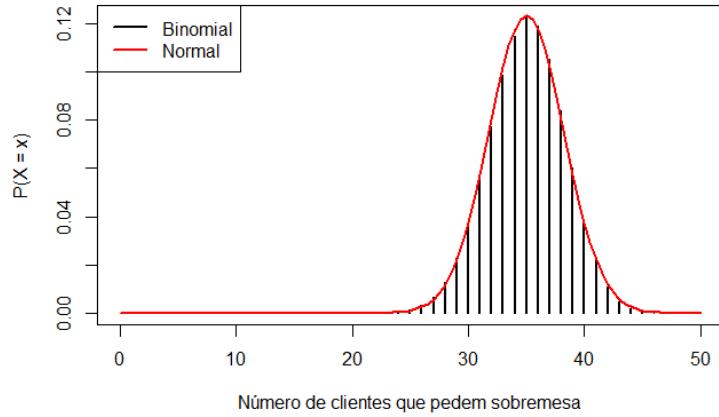


Figura 1: PMF de X com $n = 50$ e $p = 0.7$

- **Análise do gráfico da PMF**

Percebe-se que trata-se de uma distribuição unimodal com um pico em torno do valor esperado ($E(X) = 35$) e que segue a tendência da linha normal vermelha, o que pode ser explicado pelo fato de ser uma amostra grande ($n = 50$) e com probabilidade distante o suficiente tanto de 1 quanto de 0.

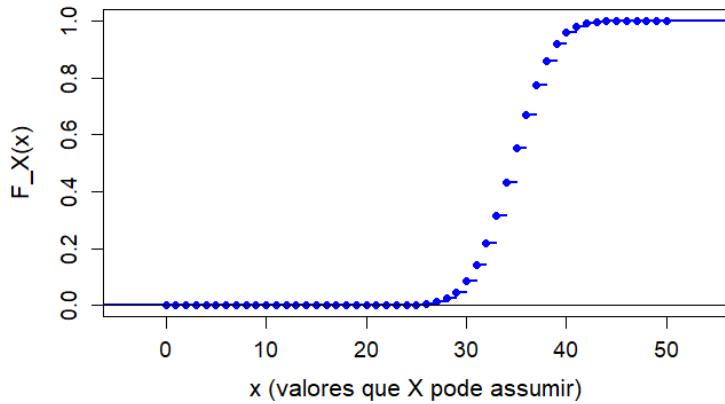


Figura 2: CDF de X com $n = 50$ e $p = 0.7$

- **Análise do gráfico da CDF**

Percebe-se que a área do gráfico para valores pequenos de x é mínima e aumenta rapidamente a partir de $x=30$ até $x=40$, o que pode ser explicado pela maior proximidade com o valor esperado (35), ou seja, nessa faixa, observa-se a concentração da maior parte da probabilidade. Por conta disso, a partir de $x=40$, percebe-se que $F_X(x) \approx 1$, isto é, a probabilidade acumulada já é praticamente total.

1.3) Valor esperado, variância e desvio-padrão.

Para uma variável aleatória binomial, o valor esperado, a variância e o desvio-padrão são dados por:

$$E[X] = np = 50 \cdot 0.7 = 35,$$

$$\text{Var}(X) = np(1 - p) = 50 \cdot 0.7 \cdot 0.3 = 10.5,$$

$$\sigma_X = \sqrt{10.5} \approx 3.24.$$

- Código em R

Para a resolução desse item em código foram aplicadas as fórmulas já definidas para a distribuição binomial, uma vez que não existem funções específicas no R para o cálculo do valor esperado de uma variável aleatória e, consequentemente, da sua variância e seu desvio padrão.

```

1 valor_esperado <- n*p
2 variânci a <- n * p * (1 - p)
3 desvio <- sqrt(variânci a)
4
5 #display dos dados
6 data.frame(valor_esperado, variânci a, desvio)
```

Listado 2: Valor esperado, variância e desvio-padrão.

- Diferenças entre os resultados

Como as fórmulas utilizadas foram as mesmas na solução teórica e nos códigos, uma vez que não existem funções específicas no R para o cálculo do valor esperado de uma variável aleatória discreta, percebe-se que não houve diferença entre os resultados obtidos.

1.4) Cálculo das probabilidades solicitadas.

(a) Probabilidade $P(X \geq 20)$ Sabendo que, pelo complemento da CDF:

$$P(X \geq 20) = 1 - P(X \leq 19).$$

Utilizando a Aproximação da Binomial pela Normal e os valores encontrados no item 3 e aplicando a correção de continuidade, uma vez que se trata de uma variável aleatória discreta:

$$P(X \leq 19) \approx P(Y \leq 19.5), \quad Y \sim \mathcal{N}(35, 3.24^2)$$

Padronizando:

$$P(Y \leq 19.5) = P\left(Z \leq \frac{19.5 - 35}{3.24}\right) = P(Z \leq -4.78)$$

Como $P(Z \leq -4.78) \approx 0$, observa-se que:

$$P(X \geq 20) \approx 1 - 0 = 1$$

Ou seja, esse valor é numericamente muito próximo de 1, pois 20 está distante da média $E[X] = 35$, o que minimiza significativamente a probabilidade de menos de 20 pessoas pedirem sobremesa.

(b) Probabilidade $P(30 < X < 43)$ Como X assume apenas valores inteiros,

$$P(30 < X < 43) = P(31 \leq X \leq 42).$$

Utilizando a Aproximação da Binomial pela Normal:

$$P(31 \leq X \leq 42) \approx P(30.5 \leq Y \leq 42.5), \quad Y \sim \mathcal{N}(35, 3.24^2)$$

Padronizando:

$$\begin{aligned} P(30.5 \leq Y \leq 42.5) &= P\left(\frac{30.5 - 35}{3.24} \leq Z \leq \frac{42.5 - 35}{3.24}\right) \\ &= P(-1.39 \leq Z \leq 2.31) \end{aligned}$$

Usando a tabela da normal padrão:

$$P(Z \leq 2.31) \approx 0.9896, \quad P(Z \leq -1.39) \approx 0.0823$$

Logo:

$$P(31 \leq X \leq 42) \approx 0.9896 - 0.0823 = 0.9073$$

Essa probabilidade concentra a maior parte da massa da distribuição, pois o intervalo está centrado em torno da média.

(c) Probabilidade $P(X = 31)$ Neste caso, utiliza-se diretamente a função de massa de probabilidade:

$$P(X = 31) = \binom{50}{31} (0.7)^{31} (0.3)^{19} \approx 0,05576.$$

Esse valor representa a probabilidade de que exatamente 31 clientes, dentre os 50, peçam sobremesa.

- **Código em R**

Para a resolução desse item em código foram aplicadas as funções próprias do R para calcular as probabilidades, como *pbinom* para a PMF e *dbinom* para calcular a CDF quando necessário.

```

1 # a) P(X >= 20)
2 respostaA <- 1-pbinom(q = 19, size = 50, prob = 0.7)
3 # b) P(30 < X < 43)
4 respostaB <- sum(dbinom(31:42, size = 50, prob = 0.7))
5 # c) P(X = 31)
6 respostaC <- dbinom(31, 50, 0.7)
7
8 cat("a) P(X >= 20) =", respostaA, "\n")
9 cat("b) P(30 < X < 43) =", respostaB, "\n")
10 cat("c) P(X = 31) =", respostaC, "\n")

```

Listado 3: Probabilidades desejadas.

```

1 a) P(X >= 20) = 0.9999972
2 b) P(30 < X < 43) = 0.9079332
3 c) P(X = 31) = 0.05575728

```

Listado 4: Saída do código no console com os dados calculados

- **Diferenças entre os resultados**

Considerando a aproximação de casas decimais nos cálculos da aproximação pela normal e o método utilizado pela função *pbinom* no R:

$$P(X \leq 19) = \sum_{k=0}^{19} \binom{50}{k} (0.7)^k (0.3)^{50-k}$$

percebe-se que as diferenças entre os resultados obtidos não são significativas.

1.5) Demanda esperada.

Do ponto de vista prático, a distribuição de X permite ao restaurante planejar seu estoque de sobremesas com base na demanda esperada e na variabilidade do processo. O valor esperado indica a quantidade média de sobremesas necessárias, enquanto a variância permite avaliar o risco de excesso ou falta de produtos.

1.6) Mudanças em p e n.

Alterações no parâmetro p deslocam a distribuição: por exemplo, se $p = 0.8$, o valor esperado passa a ser $E[X] = 40$, aumentando a demanda média por sobremesas, como pode-se observar na Figura 3, no qual a distribuição está mais à direita, centrada na nova média.

- **Código em R**

```

1 # p = 0.8
2 PMF2 <- dbinom(x, size = n, prob = 0.8)
3 plot(x, PMF2,
4       type = "h",
5       lwd = 2,
6       xlab = "Número de clientes que pedem sobremesa",
7       ylab = "P(X = x)")
8 curve(dnorm(x, mean = 40, sd = sqrt(0.8*0.2*50)), add = TRUE, col = "red", lwd = 2)
9 legend("topleft", legend = c("Binomial", "Normal"), col = c("black", "red"), lwd = 2)

```

Listado 5: PMF com $p=0.8$

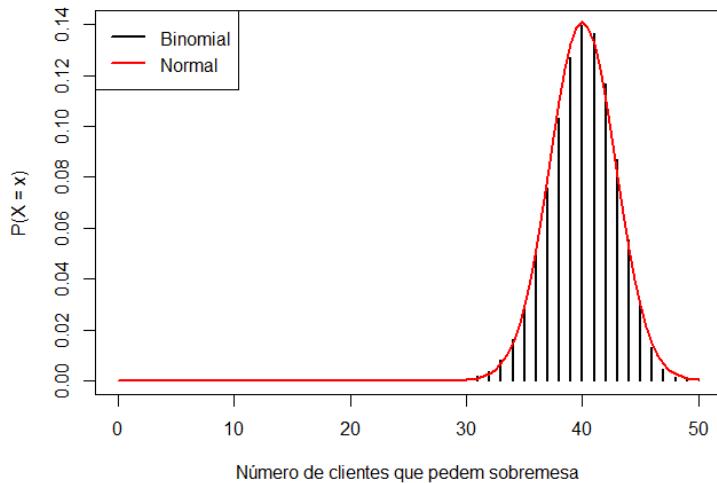


Figura 3: PMF de X com $p = 0.8$

Já um aumento em n torna a distribuição mais concentrada em torno da média, apesar de também resultar num aumento absoluto da variância, uma vez que $\text{Var}(X) = np(1 - p)$. Além disso, observa-se uma redução da variabilidade relativa, o que torna a distribuição de X mais simétrica e próxima da distribuição normal, na qual o gráfico tem o formato de curva de sino, como mostrado na Figura 4 gerada pelo código abaixo.

- o Código em R

```

1 # n=100
2 y <- 0:100
3 PMF3 <- dbinom(y, size = 100, prob = 0.7)
4 plot(y, PMF3,
5       type = "h",
6       lwd = 2,

```

```

7   xlab = "Número de clientes que pedem sobremesa",
8   ylab = "P(X = x)",
9   main = "")
10  curve(dnorm(x, mean = 70, sd = sqrt(0.7*0.3*100)), add = TRUE, col = "red", lwd = 2)
11  legend("topleft", legend = c("Binomial", "Normal"), col = c("black", "red"), lwd = 2)

```

Listado 6: PMF com $p=0.8$

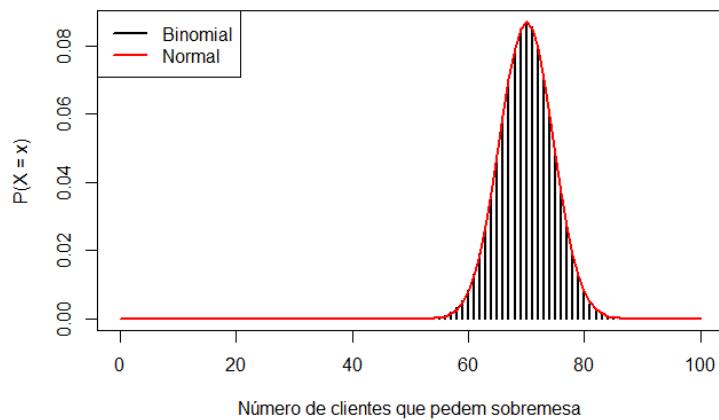


Figura 4: PMF de X com $n = 100$ e $p = 0.7$

2 Questão 2

Um site realiza uma pesquisa online e oferece uma recompensa a um usuário selecionado aleatoriamente que responde a uma série de perguntas. Cada um dos 10 milhões de visitantes diários tem, independentemente, probabilidade $p = 10^{-7}$ de ganhar a recompensa.

1. Encontre uma aproximação simples e adequada para a função de massa de probabilidade (PMF) do número de vencedores em um dia, X . Justifique claramente se essa aproximação é apropriada para os valores dados de n and p .
2. Calcule o valor esperado, $E[X]$, e a variância $\text{Var}(X)$, usando tanto a distribuição exata quanto a aproximada. Comente sobre a semelhança entre os resultados.
3. Suponha que você ganhe a recompensa, mas que possa haver outros vencedores. Seja $W \sim \text{Pois}(1)$ o número de vencedores além de você. Se houver vários vencedores, o prêmio é sorteado aleatoriamente entre todos eles. Encontre a probabilidade de que você realmente receba o prêmio.
4. Gere um grande número de simulações diárias para o número de vencedores. Crie uma comparação visual entre os resultados empíricos e a aproximação considerada no item 1. Descreva brevemente o que a visualização indica sobre a qualidade da aproximação.

Solução da questão

2.1) Aproximação da PMF

O número de vencedores X segue exatamente uma distribuição binomial,

$$X \sim \text{Binomial}(n = 10^7, p = 10^{-7}).$$

Como n (número de experimentos) é muito grande e p (probabilidade de sucesso em cada experimento) é muito pequeno, é apropriado utilizar a aproximação de Poisson. Dessa forma, temos

$$\lambda = np = 1.$$

Assim, consideramos

$$X \approx \text{Poisson}(\lambda = 1).$$

Essa aproximação é adequada pois satisfaz as condições clássicas para a aproximação binomial–Poisson. A função de massa de probabilidade aproximada é dada por

$$P(X = k) = \frac{e^{-1}}{k!}, \quad k = 0, 1, 2, \dots$$

A função de massa de probabilidade da distribuição exata é

$$P(X = k) = \binom{10^7}{k} (10^{-7})^k (1 - 10^{-7})^{10^7 - k}.$$

- o Código em R

```

1 # Por ser uma probabilidade de grande amostra, usaremos a distribuição de poisson
2 # como aproximação
3 taxa <- n * p
4 # Calculamos a taxa com a probabilidade individual e a "população"
5
6 # PMF da distribuição de Poisson = dpois(k,taxa)
7 aproximacao <- function(n,p,k){
8   lambda <- n * p
9   pmf <- (exp(-lambda) * lambda^k)/factorial(k)
10
11  return(pmf) # Retorna a aproximação da probabilidade
12 }
```

Listado 7: Aproximação PMF

2.2) Valor esperado e variância

Como as probabilidades são independentes, a distribuição exata é a Binomial. Logo, tem-se que:

$$E[X] = np = 1$$

$$\text{Var}(X) = np(1 - p) = 1(1 - 10^{-7}) = 0.999\dots$$

Para a distribuição de Poisson com parâmetro $\lambda = 1$, vale que

$$E[X] = \lambda = 1$$

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

Observa-se que os valores do valor esperado e da variância obtidos pela distribuição exata e pela aproximação são praticamente idênticos, indicando que a aproximação de Poisson é adequada neste contexto.

- o Código em R

```

1 #Distribuição Binomial (exata)
2 exp_bin <- n * p
3 var_bin <- n * p * (1-p)
4
5 #Distribuição de Poisson (aproximada)
6 exp_pois <- taxa
7 var_pois <- taxa
8
9 data.frame(
10   Distribuição = c("Binomial (exata)", "Poisson (aproximada)"),
11   Valor_Esperado = c(exp_bin, exp_pois),
12   Variância = c(var_bin, var_pois)
13 )
```

Listado 8: Expectativa e Variância

```

1 Distribuição Valor_Esperado Variância
2   Binomial (exata)           1 0.9999999
3   Poisson (aproximada)       1 1.0000000

```

Listado 9: Saída do código no console com os dados calculados

- **Diferenças entre os resultados**

Como o valor esperado e a variância são dados por expressões conhecidas para as distribuições consideradas, os valores obtidos na solução teórica coincidem com aqueles calculados no código. Assim, não se observam diferenças relevantes entre os resultados, mas indicam a correspondência da aproximação.

2.3) Probabilidade de receber o prêmio

Sabendo que W é o número de vencedores além de você e $W = k$, então existem $k + 1$ vencedores no total (você e mais k pessoas). Como o sorteio é feito de forma uniforme, a probabilidade de você receber o prêmio nesse caso é

$$\frac{1}{k+1}.$$

Como k pode assumir os valores $0, 1, 2, \dots$, a probabilidade total de receber o prêmio é obtida somando todos esses casos, ponderados pela probabilidade de ocorrência de cada k :

$$P(\text{receber o prêmio}) = \sum_{k=0}^{\infty} \frac{1}{k+1} P(W = k).$$

Como $W \sim \text{Poisson}(1)$, temos

$$P(\text{receber o prêmio}) = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{e^{-1}}{k!}.$$

Essa soma resulta em

$$P(\text{receber o prêmio}) = 1 - e^{-1} \approx 0.632120.$$

- **Código em R**

```

1 # Como a probabilidade e a PMF de Poisson é um sorteio entre os ganhadores,
2 # para isso pegaremos cada possível k para calcular a prob total:
3 # P(ganho) = sum(Pw(k) * 1/(k+1))
4
5 k_ganhadores <- 0:30
6 prob_W <- aproximacao(n,p,k_ganhadores)
7 prob_ganho <- sum(prob_W * (1/(k_ganhadores+1)))

```

Listado 10: Cálculo da probabilidade de ganho

2.4) Simulação e comparação com a aproximação

Para avaliar a qualidade da aproximação de Poisson obtida no item 1, foi gerado um grande número de simulações diárias para o número de vencedores. Em cada simulação, o número de vencedores foi modelado por uma variável aleatória com distribuição de Poisson e parâmetro $\lambda = 1$.

A partir das simulações, foi construída a distribuição empírica do número de vencedores, obtida pela frequência relativa de cada valor observado. Em seguida, essa distribuição empírica foi comparada com a função de massa de probabilidade teórica da distribuição de Poisson através da construção do gráfico da Figura 5.

- o Código em R

```
1 set.seed(90) # Setamos uma semente para analisar as mesmas simulações aleatorias.
2 n_simulacoes <- 100000 # Quantidade de simulações a serem realizadas
3 simulacoes <- rpois(n_simulacoes,lambda = taxa) #função que irá simular a distribuição
4
5 #Análise do resultado das simulações;
6 freq_empirica <- table(simulacoes) / n_simulacoes
7
8 #Análise do resultado esperado(Teórico):
9 k_teorico <- as.numeric(names(freq_empirica))
10 freq_teorica <- approximacao(n, p, k_teorico)
11 names(freq_teorica) <- names(freq_empirica)
12 print(round(freq_teorica,4))
13
14 print("--EMPÍRICO--")
15 print(round(freq_empirica,4))
16 print("--TEÓRICO--")
17 print(round(freq_teorica,4))
18
19 # Gráfico de comparação: PMF teórica vs empírica
20 plot(k_teorico, freq_teorica, type = "h",
21       lwd = 3,
22       main = "Simulação vs Teoria (PMF)",
23       xlab = "Número de Vencedores (k)",
24       ylab = "Probabilidade",
25       ylim = c(0, max(freq_teorica, freq_empirica) * 1.1))
26
27 points(k_teorico, freq_empirica,
28         pch = 16, col = "blue")
29
30 legend("topright",
31        legend = c("Teoria (Poisson)", "Simulação"),
32        lwd = c(3, NA),
33        pch = c(NA, 16),
34        col = c("black", "blue"))
```

Listado 11: Simulação das Distribuições

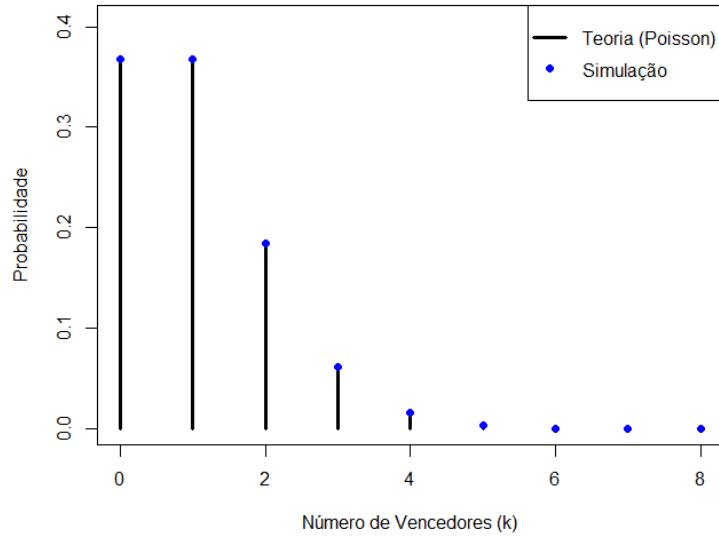


Figura 5: Gráfico de comparação das simulações

- **Análise do gráfico das simulações**

A comparação visual entre as probabilidades empíricas e a distribuição teórica mostra uma forte concordância entre os resultados. Os pontos obtidos pela simulação ficam muito próximos dos valores teóricos da PMF, indicando que a aproximação de Poisson é adequada para modelar o número de vencedores neste problema.

3 Questão 3

Você é responsável por monitorar a temperatura de uma CPU multicore em uma unidade de processamento embarcada. Sob carga normal, a temperatura da CPU apresenta flutuações devido a mudanças na carga de trabalho, nas condições ambientais e na eficiência do sistema de resfriamento. Testes mostram que a temperatura em regime estacionário da CPU segue uma distribuição normal com temperatura média $\mu = 62^\circ\text{C}$ e desvio-padrão $\sigma = 3,5^\circ\text{C}$. Sua tarefa é simular medições de temperatura da CPU e analisar suas propriedades estatísticas.

1. Crie uma função que gere valores com distribuição normal usando a transformação de Box-Muller¹, a partir de entradas aleatórias uniformes. Especificamente:

- (a) Gere duas variáveis aleatórias uniformes independentes: $U_1, U_2 \sim \text{Unif}(0, 1)$.
- (b) Calcule dois valores normais padrão usando as fórmulas de Box-Muller:

$$Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2), \quad Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

Z_1 e Z_2 são variáveis aleatórias independentes com distribuição normal padrão. Concatene-as para formar um vetor Z de valores normais padrão.

- (c) Converta cada valor normal padrão para a distribuição de temperatura da CPU:

$$T = 62 + 3.5 Z$$

2. Use seu gerador de números aleatórios para gerar 1.000 medições de temperatura da CPU.

Gere mais 1.000 valores de temperatura utilizando o gerador de números aleatórios normal embutido do R, com a mesma média e desvio-padrão.

3. Para ambos os conjuntos de dados simulados, calcule:

- (a) Média amostral.
- (b) Desvio-padrão amostral.
- (c) Temperatura mínima e máxima observada.
- (d) Probabilidade empírica e teórica $P(T > 68)$.
- (e) Probabilidade empírica e teórica $P(60 < T < 65)$.
- (f) Probabilidade teórica $P(T > 75)$.

Algum dos conjuntos de dados simulados (1.000 amostras) contém valores acima de 75°C ? Caso não, explique por que eventos raros requerem tamanhos de amostra grandes para serem observados.

4. Visualize os resultados criando:

¹ https://en.wikipedia.org/wiki/Box-Muller_transform

- (a) Um histograma das temperaturas simuladas da CPU (pode plotar os dois conjuntos de dados separadamente ou sobrepostos).
- (b) A função densidade de probabilidade (PDF) normal teórica (média 62 °C, desvio padrão 3,5 °C) sobreposta ao histograma.
5. Discuta seus resultados respondendo às seguintes perguntas: As distribuições empíricas da temperatura da CPU se assemelham à curva normal teórica? Quão próximas estão a média amostral e o desvio-padrão amostral dos valores esperados 62 °C e 3,5 °C? Há diferenças perceptíveis entre o conjunto de dados gerado com seu RNG manual e o produzido pelo RNG embutido do R? Como essa simulação pode ajudar na avaliação de estratégias de resfriamento ou de escalonamento dinâmico de clock? Por que geradores de números aleatórios uniformes são a base dos sistemas de RNG?

Solução da questão

Assume-se que a temperatura da CPU em regime estacionário pode ser modelada por uma distribuição normal com média $\mu = 62$ °C e desvio-padrão $\sigma = 3.5$ °C. Essa modelagem é adequada devido à presença de diversas fontes independentes de variação, como carga de trabalho, condições ambientais e eficiência do sistema de resfriamento.

3.1) Geração de variáveis normais via transformação de Box–Muller

A transformação de Box–Muller permite gerar variáveis aleatórias com distribuição normal padrão a partir de duas variáveis aleatórias independentes com distribuição uniforme no intervalo (0, 1). Sejam ++

$$U_1, U_2 \sim \text{Unif}(0, 1).$$

Os valores

$$Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2), \quad Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

são variáveis aleatórias independentes com distribuição normal padronizada, $Z_1, Z_2 \sim \mathcal{N}(0, 1)$. A concatenação desses valores forma um vetor de observações normais padrão.

Para obter valores compatíveis com a temperatura da CPU, realiza-se a transformação linear

$$T = 62 + 3.5Z,$$

resultando em uma variável aleatória com distribuição $T \sim \mathcal{N}(62, 3.5^2)$.

- Código em R

```

1 gerar_temp <- function(n){ # Função que vai gerar os valores empíricos
2   n_pares <- n/2
3
4   # a) variáveis aleatórias uniformes independentes
5   u1 <- runif(n_pares)

```

```

6   u2 <- runif(n_pares)
7
8   # b) Box-Muller
9   z1 <- sqrt(-2 * log(u1)) * cos(2 * pi * u2)
10  z2 <- sqrt(-2 * log(u1)) * sin(2 * pi * u2)
11
12  z <- c(z1, z2)
13
14  # c) conversão para temperatura
15  temp <- 62 + 3.5 * z
16  return(temp)
17 }
```

Listado 12: Geração de valores

3.2) Simulação das medições de temperatura

- Código em R

Utilizando o gerador implementado via Box–Muller, foram simuladas 1.000 medições de temperatura da CPU. Adicionalmente, foram geradas mais 1.000 medições utilizando o gerador normal embutido do R, com os mesmos parâmetros de média e desvio-padrão, como mostrado abaixo.

```

1 #Dados gerados a partir da Função Box-Muller
2   dados_box <- gerar_temp(1000)
3   head(dados_box) # Vizualização dos primeiros valores
4
5 #Dados gerados a partir da Função rnorm
6   dados_r <- rnorm(n = 1000,mean = 62,sd = 3.5)
7   head(dados_r) # Vizualização dos primeiros valores
```

Listado 13: Simulação dos valores de temperatura

3.3) Estatísticas amostrais e probabilidades

A partir dos dados simulados pelo método de Box–Muller e dos dados gerados pela função `rnorm` do R, foram calculadas estatísticas amostrais básicas e probabilidades de interesse. As estatísticas empíricas foram obtidas diretamente das simulações, enquanto os valores teóricos foram calculados assumindo que a temperatura segue uma distribuição normal $\mathcal{N}(62, 3.5^2)$.

As estatísticas analisadas incluem a média amostral, o desvio-padrão, bem como os valores mínimo e máximo observados. Além disso, foram estimadas probabilidades para eventos específicos, tanto de forma empírica quanto teórica.

A Tabela 1 apresenta a comparação entre os valores empíricos e teóricos obtidos.

Tabela 1: Estatísticas amostrais e probabilidades teóricas e empíricas

Quantidade	Empírico	Teórico
Média amostral	61.92	62.18
Desvio-padrão amostral	3.45	3.577
Temperatura mínima	51.90	49.91
Temperatura máxima	71.99	72.66
$P(T > 68)$	0.043	0.0432
$P(60 < T < 65)$	0.512	0.520
$P(T > 75)$	0	0.0001

Observa-se que a probabilidade de ocorrer uma temperatura acima de 75 °C é extremamente pequena, uma vez que esse valor está vários desvios-padrão acima da média. Assim, nota-se que nenhuma observação acima desse limite foi registrada na amostra simulada, o que evidencia que eventos raros exigem tamanhos amostrais maiores para serem observados empiricamente.

- o Código em R

```

1 # Estatísticas para o seu método (Box-Muller) - EMPÍRICO
2 # Estatísticas para o método do R (rnorm) - TEÓRICO
3
4
5 # a)Mean:
6 media_empirica <- mean(dados_box)
7 media_teorico <- mean(dados_r)
8
9 # b)Desvio Padrão:
10 desvio_empirico <- sd(dados_box)
11 desvio_teorico <- sd(dados_r)
12
13 # c)Max:
14 min_empirico <- min(dados_box)
15 min_teorico <- min(dados_r)
16
17 #c)Min:
18 max_empirico <- max(dados_box)
19 max_teorico <- max(dados_r)
20
21 # d)P(T > 68):
22 pd_empirica <- mean(dados_box > 68) #Média dos valores maiores que 68
23 pd_teorico <- 1 - pnorm(q = 68,mean = 62,sd = 3.5) # P(T > 68) = 1 - P(T <= 68)
24
25
26 # e)P(60 < T < 65):
27 pe_empirica <- mean(dados_box > 60 & dados_box < 65) #Média dos valores no intervalo
28 pe_teorico <- pnorm(q = 65,mean = 62,sd = 3.5) - pnorm(q = 60,mean = 62,sd = 3.5)
29 #P(60 < T < 65) = P(T <= 65) - P(T <= 60)
30

```

```

31 # f)Existe T > 75 ?
32 pf_empirica <- sum(dados_box > 75) #Número de valores maiores que 75 (igual a zero.)
33 # f)P(T > 75):
34 pf_teorico <- 1 - pnorm(q = 75,mean = 62,sd = 3.5) ## P(T > 75) = 1 - P(T <= 75)

```

Listado 14: Cálculo das medidas

3.4) Visualização dos resultados

Foram construídos histogramas das temperaturas simuladas para ambos os conjuntos de dados, sobre os quais foi sobreposta a função densidade de probabilidade da distribuição normal teórica com média 62 °C e desvio-padrão 3.5 °C.

- o Código em R

```

1 # (a) Histograma apenas com os dados simulados
2 hist(dados_box,
3       breaks = 30,
4       col = "lightgray",
5       border = "white",
6       main = "Histograma das Temperaturas Simuladas",
7       xlab = "Temperatura (C)",
8       ylab = "Frequência")
9
10 # Gráfico da normal isolada
11 x_curva <- seq(50, 75, length = 200)
12 y_curva <- dnorm(x_curva, mean = 62, sd = 3.5)
13 plot(x_curva, y_curva, type = "l", col = "red", lwd = 2,
14       main = "Curva Normal Teórica",
15       xlab = "Temperatura (C)",
16       ylab = "Densidade")
17
18 # (b) PDF teórica sobreposta ao histograma
19 hist(dados_box,
20       breaks = 30,
21       probability = TRUE,
22       col = "lightblue",
23       main = "Sobreposição Empírico x Teórico",
24       xlab = "Temperatura (C)",
25       ylab = "Densidade")
26
27 lines(x_curva, y_curva, col = "red", lwd = 3)

```

Listado 15: Gráfico das simulações

3.5) Análise dos resultados

Os resultados obtidos por meio das simulações indicam que a distribuição empírica das temperaturas da CPU se aproxima bem da distribuição normal teórica com média 62 °C

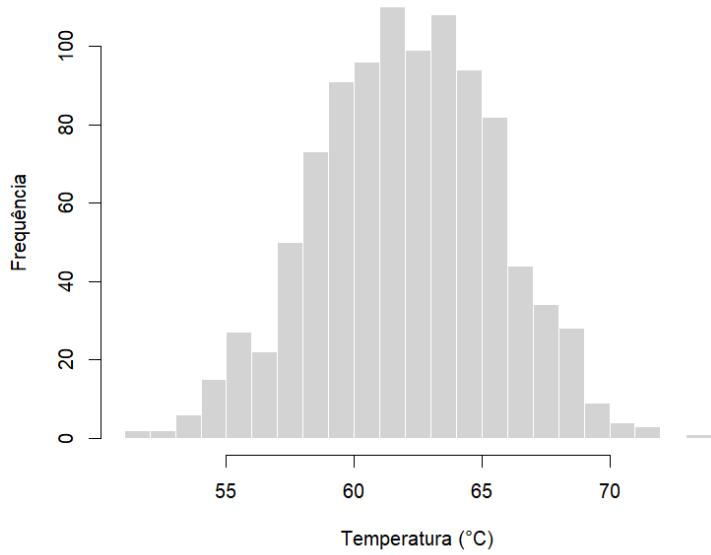


Figura 6: Histograma das temperaturas simuladas

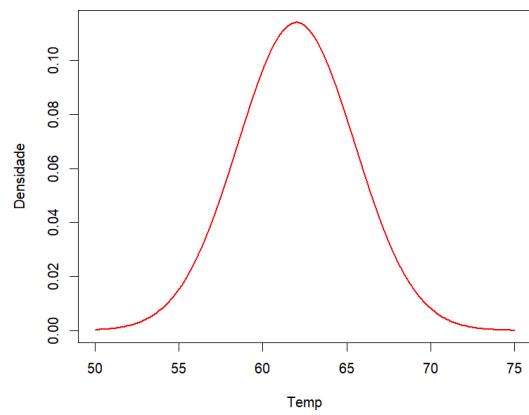


Figura 7: Curva da distribuição normal teórica

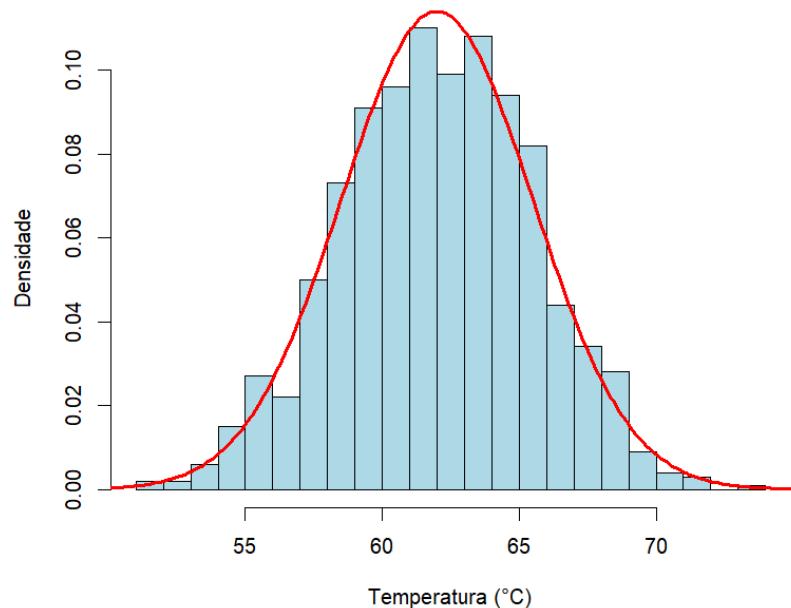


Figura 8: Sobreposição entre distribuição empírica e teórica

e desvio-padrão 3.5°C . O formato do histograma apresenta certa simetria em torno da média e decaimento suave nas caudas, características típicas de uma distribuição normal.

As estatísticas amostrais calculadas a partir dos dados simulados mostram valores coerentes com os parâmetros teóricos, indicando que o modelo adotado representa de forma adequada o comportamento da temperatura da CPU em regime estacionário. Esse tipo de simulação é útil para analisar o comportamento térmico do sistema, permitindo estimar probabilidades associadas a temperaturas críticas e auxiliar na avaliação de estratégias de controle térmico.