UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

Disciplina: INFERÊNCIA ESTATÍSTICA
Curso: Graduação em Estatística
Código: EST0035
Semestre: 2025.1
Professor: Frederico Machado Almeida

LISTA DE EXERCÍCIOS #01

Observações:

- Questões para entregar: 2, 4, 7, 8 e 9
- **A lista deve ser feita preferencialmente em duplas**.
- Demais questões são apenas para estudar.
- Prazo de entrega: **22/04/2025**

Q1. Nos itens (i)-(iii) escolha a alternativa que melhor responde a pergunta de interesse:

(i) Qual das seguintes alternativas descreve a propriedade de um estimador não-viesado (ou não-tendencioso) para um parâmetro θ?
  - (a) A forma da distribuição amostral à aproximadamente normal.
  - (b) O centro da distribuição amostral coincide com o valor de θ.
  - (c) A distribuição a amostral em questão apresenta menor variância entre todas as possíveis distribuições amostrais do estimador.
  - (d) O centro da distribuição amostral coincide com o valor do desvio-padrão populacional.

(ii) Suponha que a idade dos estudantes do curso de Estatística na UnB segue uma distribuição assimétrica com média de 25 anos, e desvio-padrão igual a 4 anos. Se uma amostra de 200 estudantes for selecionada repetidamente, qual das seguintes afirmações a cerca da distribuição amostral da média está **incorreta?**
  - (a) A média da distribuição amostral é aproximadamente igual a 25 anos.
  - (b) O desvio-padrão da distribuição amostral é igual a 4 anos.
  - (c) A forma da distribuição amostral é aproximadamente normal.
  - (d) Todas as alternativas anteriores estão corretas.

(iii) O Teorema Central de Limite é muito importante na estatística porque afirma que:
  - (a) Para qualquer tamanho da população, a distribuição da média amostral $\bar{X}_n$ é aproximadamente normal.
  - (b) Para n suficientemente grande, a população é não-viesada.
  - (c) Para qualquer população, a distribuição da média amostral $\bar{X}_n$ é aproximadamente normal, independentemente do seu tamanho de amostra.
  - (d) Para n suficientemente grande, a distribuição da média amostral $\bar{X}_n$ é aproximadamente normal, independentemente da população onde a amostra foi selecionada.

Q2. Considere a amostra aleatória de dimensão $n = 2$, digamos (X_1, X_2) retirada de uma população da variável aleatória X, que denota o número de animais de estimação por família, cuja distribuição é a seguinte:

x	0	1	2	3
$P(X = x)$	0,60	0,25	0,10	0,05

(a) Qual a probabilidade de obter a amostra (3,1), ou seja, qual a probabilidade de a primeira família selecionada ter trás animais de estimação e a segunda família selecionada ter 1 animal de estimação?

Resposta: Vamos considerar uma amostra de dimensão $n = 2$ (duas famílias), onde a seleção de cada família é feita de forma independente. Assim, a probabilidade conjunta de obter o par (x_1, x_2) é dada por:

$$\mathbb{P}((x_1, x_2)) = \mathbb{P}(X = x_1) \times \mathbb{P}(X = x_2).$$

Queremos que a primeira família tenha 3 animais de estimação e a segunda, 1. Portanto:

$$\mathbb{P}((3, 1)) = \mathbb{P}(X = 3) \times \mathbb{P}(X = 1) = 0.05 \times 0.25 = 0.0125.$$

Assim, a probabilidade de obter a amostra $(3, 1)$ é 0,0125.

(b) Liste todas as possíveis amostras daquela dimensão que pode obter.

Resposta: Como cada família pode ter 0, 1, 2 ou 3 animais, o espaço amostral S é dado por:

$$S = \{(x_1, x_2) : x_1, x_2 \in \{0, 1, 2, 3\}\}.$$

Isso resulta nas 16 amostras ordenadas:

$$\begin{aligned} S = \{\, & (0,0),\ (0,1),\ (0,2),\ (0,3), \\ & (1,0),\ (1,1),\ (1,2),\ (1,3), \\ & (2,0),\ (2,1),\ (2,2),\ (2,3), \\ & (3,0),\ (3,1),\ (3,2),\ (3,3)\,\}. \end{aligned}$$

(c) Qual das amostras possíveis de dimensão 2 é a mais provável?

Resposta: Calculando todas as probabilidades possíveis, de acordo com:

$$\mathbb{P}((x_1, x_2)) = \mathbb{P}(X = x_1) \times \mathbb{P}(X = x_2).$$

Temos:

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$
$X_1 = 0$	$0,360$	$0,1500$	$0,060$	$0,0300$
$X_1 = 1$	$0,150$	$0,0625$	$0,025$	$0,0125$
$X_1 = 2$	$0,060$	$0,0250$	$0,010$	$0,0050$
$X_1 = 3$	$0,030$	$0,0125$	$0,005$	$0,00125$

Desse modo, o par que maximiza o produto é:

$$(0, 0) : \quad \mathbb{P}((0, 0)) = 0,60 \times 0,60 = 0,36.$$

Portanto, a amostra $(0, 0)$ é a mais provável.

(d) Qual a probabilidade de uma amostra selecionada ao acaso fornecer uma média amostral $\bar{X}_n$ igual a 2,5?

Resposta: Probabilidade de a média amostral ser $\bar{X}_n = 2,5$. A média da amostra de duas observações é definida por:

$$\bar{X}_n = \frac{X_1 + X_2}{2}.$$

Para que $\bar{X}_n = 2,5$, devemos ter:

$$\frac{X_1 + X_2}{2} = 2,5 \quad \Longrightarrow \quad X_1 + X_2 = 5.$$

Como $x \in \{0, 1, 2, 3\}$, os únicos pares que satisfazem $X_1 + X_2 = 5$ são (2,3) e (3,2).
Somando as probabilidades obtidas no item anterior, temos:

$$\mathbb{P}\left(\bar{X}_n = 2,5\right) = 0,005 + 0,005 = 0,01.$$

Portanto, a probabilidade de a média amostral ser 2,5 é 0,01.

Q3. O conteúdo em litros de garrafas de azeite de certa marca segue uma distribuição normal com média $\mu = 0,99$ litros e desvio-padrão $\sigma = 0,02$ litros. Qual a probabilidade de o conteúdo médio numa amostra de 16 garrafas selecionadas aleatoriamente ser superior a um litro?

Q4. Assuma que $Y_1, \cdots, Y_n$ denotam variáveis aleatórias independentes e identicamente distribuídas, com fdp dada por,

$$f(x|\mu) = \begin{cases} e^{\mu - y} & y > \mu,\ -\infty < \mu < \infty \\ 0 & \text{caso contrÃ¡rio,} \end{cases}$$

que denota a distribuição exponencial de dois parâmetros. Se $Z_1 = \text{Min}\{Y_1, \cdots, Y_n\}$ denota a menor estatística de ordem. Assim,

(a) Especifique o espaço paramétrico e o suporte associado a distribuição de Y.

Resposta: o espaço paramétrico é $\Theta = \{\mu : \mu \in \mathbb{R}\}$, e o suporte de Y é, $\mathcal{Y} = \{y : y > \mu\}$.

(b) Verifique se $T_{1n} = \bar{Y}_n$ e $T_{2n} = Y_{(1)}$ são estimadores não viciados para μ.

Resposta: observe que, $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Assim, antes de mais nada vamos computar a $\mathbb{E}_\mu(Y)$, tal que,

$$\mathbb{E}_\mu(Y) = \int_\mu^\infty y f(y|\mu)\, dy = \int_\mu^\infty y e^{(\mu - y)} dy = -\int_\mu^\infty y d\left[e^{(\mu - y)}\right] = \mu + 1.$$

(i) Assim, $\mathbb{E}_\mu(T_{1n}) = \mathbb{E}_\mu\left(\bar{Y}_n\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\mu(Y_i) = \frac{1}{n}\sum_{i=1}^n(\mu + 1) = \mu + 1 \neq \mu$.
Portanto, $\bar{Y}_n$ é um estimador viesado para μ.

(ii) como $T_{2n} = Y_{(1)}$, então, vamos obter a fda de $Y_{(1)}$, tal que,

$$F_Y(y|\mu) = \int_\mu^y f(z|\mu)\,dz = \begin{cases} 0, & \text{se } y \le \mu \\ 1 - e^{(\mu-y)}, & \text{se } y > \mu. \end{cases}$$

Desta forma, a fda de $Y_{(1)}$ é dada por,

$$F_{T_{2n}}(t) = \mathbb{P}(T_{2n} \le t) = 1 - \underbrace{\mathbb{P}\left(Y_{(1)} > t\right)}_{(i)}, \text{ segue da def. apresentada em aula.}$$

Observe que, a expressão em (i) é tal que,

$$\begin{aligned} \mathbb{P}\left(Y_{(1)} > t\right) &= \mathbb{P}(min\{Y_1, Y_2, \cdots, Y_n\} > t) = \mathbb{P}(Y_1 > t, Y_2 > t, \cdots, Y_n > t) \\ &\overset{iid}{=} \prod_{i=1}^n \mathbb{P}(Y_i > t) = \prod_{i=1}^n \mathbb{P}(Y_1 > t) = \{\mathbb{P}(Y_1 > t)\}^n \\ &= \{1 - \mathbb{P}(Y_1 \le t)\}^n = \{1 - F_{Y_1}(t)\}^n = e^{n(\mu-t)}. \end{aligned}$$

Observe que, a FDA de Y_1 é a mesma de Y. Assim, substituindo quantidade em (i) na FDA de $Y_{(1)}$ obtemos:

$$\begin{aligned} F_{Y_{(1)}}(t) &= \mathbb{P}\left(Y_{(1)} \le t\right) = 1 - \underbrace{\mathbb{P}\left(Y_{(1)} > t\right)}_{(i)} \\ &= \mathbb{P}\left(Y_{(1)} \le t\right) = 1 - e^{n(\mu-t)}, \end{aligned}$$

com fdp dada por $f_{Y_{(1)}}(t) = \frac{d}{dt} F_{Y_{(1)}}(t) = ne^{n(\mu-t)}$, $\forall t > \mu$. Então, para verificar se $T_{1n} = Y_{(1)}$ é ou não um estimador viesado para μ, vamos calcular sua esperança matemática. Isto é,

$$\mathbb{E}_\mu(T_{2n}) = \mathbb{E}_\mu\left(Y_{(1)}\right) = \int_\mu^\infty t f_{Y_{(1)}}(t|\mu)\,dt = \int_\mu^\infty t n e^{n(\mu-t)} dt = -\int_\mu^\infty t d\left[e^{n(\mu-t)}\right] = \mu + 1/n,$$

o resultado anterior é uma consequência direta da aplicação do método de integração por partes. Portanto, podemos concluir que T_2 é um estimador viesado para μ, pois, $\mathbb{E}_\mu(T_{2n}) = \mu + 1/n \neq \mu$.

(c) Encontre e compare os EQMs dos dois estimadores apresentados no item (b). Faça um gráfico dos EQMs como função de μ.

Resposta: observe que, $Var_\mu(T_{2n}) = Var_\mu\left(Y_{(1)}\right)$. Portanto, precisamos computar o segundo momento $\mathbb{E}_\mu(T_{2n}^2)$. Para tal,

$$\begin{aligned} \mathbb{E}_\mu\left(T_{2n}^2\right) &= \int_\mu^\infty t^2 f_{T_{2n}}(t|\mu)\,dt = \int_\mu^\infty t^2 n e^{n(\mu-t)} dt = -\int_\mu^\infty t^2 d\left[e^{n(\mu-t)}\right] \\ &= -\left\{ \lim_{a\to\infty} t^2 e^{n(\mu-t)}\Big|_\mu^a - \frac{2}{n} \underbrace{\int_\mu^\infty t n e^{n(\mu-t)} dt}_{\mathbb{E}_\mu(T_{2n})} \right\} = \mu^2 + \frac{2}{n}\left(\mu + \frac{1}{n}\right). \end{aligned}$$

Desta forma, $Var\left(T_{2n}\right) = \mathbb{E}_{\mu}\left(T_{2n}^2\right) - \mathbb{E}_{\mu}^2\left(T_{2n}\right) = \frac{1}{n^2}$. Consequentemente, o EQM de T_{2n} será dado por, $EQM\left(T_{2n}\right) = Var\left(T_{2n}\right) + \left[b\left(T_{2n}\right)\right]^2 = Var\left(T_{2n}\right) = 1/n^2 + 1/n^2 = 2/n^2$, que não depende de μ. Então, o gráfico do $EQM_{\mu}\left(T_{2n}\right)$ será uma linha horizontal estabilizada em $2/n^2$.

Ademais, para o estimador $T_{1n} = \bar{Y}_n$ segue que, $Var\left(\bar{Y}_n\right) = 1/n^2 \sum_{i=1}^{n} Var\left(Y_i\right) \overset{iid's}{=} \frac{n}{n^2} Var\left(Y_1\right) = \frac{1}{n} Var\left(Y_1\right)$. Como $\mathbb{E}_{\mu}\left(Y\right) = \mathbb{E}_{\mu}\left(Y_1\right) = \mu + 1$ e

$$\begin{aligned}
\mathbb{E}_{\mu}\left(Y^2\right) &= \int_{\mu}^{\infty} y^2 f\left(y|\mu\right) dy = \int_{\mu}^{\infty} y^2 e^{(\mu-y)} dy = -\int_{\mu}^{\infty} y^2 d\left[e^{(\mu-y)}\right] \\
&= \mu^2 + 2\underbrace{\int_{\mu}^{\infty} y e^{(\mu-y)} dy}_{\mathbb{E}_{\mu}(Y)} = \mu^2 + 2\left(\mu+1\right).
\end{aligned}$$

Assim, $Var_{\mu}\left(Y_1\right) = \mathbb{E}_{\mu}\left(Y^2\right) - \mathbb{E}_{\mu}^2\left(Y_1\right) = \mu^2 + 2\mu + 2 - \left(\mu+1\right)^2 = \mu^2 + 2\mu + 2 - \mu^2 - 2\mu^2 - 1 = 1$. Consequentemente, $Var_{\mu}\left(T_{1n}\right) = n/n^2 = 1/n$. Como $b\left(T_{1n}\right) = \mathbb{E}_{\mu}\left(T_{1n}\right) - \mu = 1$, resultado em:

$$EQM_{\mu}\left(T_{1n}\right) = Var_{\mu}\left(T_{1n}\right) + b^2\left(T_{1n}\right) = 1/n + 1 = \frac{n+1}{n},$$

que não depende de μ. Ou seja, o gráfico do EQM para T_{1n} será uma linha horizontal no ponto $\frac{n+1}{n}$, para todo $n \in \mathbb{N}$.

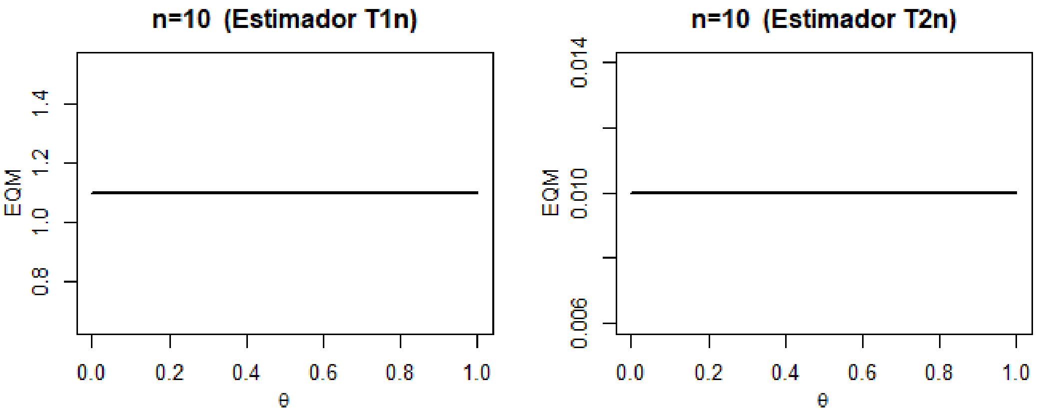

Figure 1: Gráfico dos EQM's para os estimadores T_{1n} e T_{2n}, fixando $n = 10$.

Q5. Seja $X_1, \cdots, X_n$ uma amostra aleatória proveniente de uma distribuição de Bernoulli com probabilidade de sucesso θ, e seja $W_n = \sum_{i=1}^{n} X_i$ uma estatística qualquer. Compare os erros quadráticos médios de dois estimadores de θ, respectivamente, $T_{1n} = W_n/n$ e $T_{2n} = \left(W_n + 1\right)/\left(n+2\right)$.

Q6. Assuma que $X_1, \cdots, X_n$ denota uma amostra aleatória de tamanho n da variável aleatória X extraída de uma distribuição da $\mathcal{N}\left(\mu, 1\right)$. Considere os estimadores $T_{1n} = \bar{X}_n$ e $T_{2n} = T_{1n} + X_1$. Encontre o EQM de T_{1n} e T_{2n} como função de μ. Faça um gráfico do EQM para $n = 15$.

Q7. Considere as variáveis $X_1, X_2, \cdots, X_9$ com distribuição binomial em que, $X_i \sim Binom\,(n_i, p = 0,5)$, com $i = 1, 2, \cdots, 9$ e as variáveis, $Y_j \sim \mathcal{N}\,(2, 1)$ para $j = 1, 2$ (Y_j's são independentes entre si).

(a) Deduza a distribuição amostral de $T = \sum_{i=1}^{5} X_i$

Resposta: Cada X_i segue uma binomial com número de ensaios igual a i e probabilidade de sucesso 0,5. Assim, temos:

$$\begin{aligned} X_1 &\sim \text{Bin}(1,\ 0.5), \\ X_2 &\sim \text{Bin}(2,\ 0.5), \\ X_3 &\sim \text{Bin}(3,\ 0.5), \\ X_4 &\sim \text{Bin}(4,\ 0.5), \\ X_5 &\sim \text{Bin}(5,\ 0.5). \end{aligned}$$

Utilizando a propriedade de que a soma de variáveis binomiais independentes com o mesmo valor de p resulta em uma binomial com número total de ensaios igual à soma dos ensaios individuais, definimos

$$T = \sum_{i=1}^{5} X_i.$$

O total de ensaios é:

$$n_T = 1 + 2 + 3 + 4 + 5 = 15.$$

Portanto,

$$T \sim Binom(15; 0, 5).$$

Uma forma de provar que a soma de binomiais independentes com o mesmo valor de p resulta em uma binomial com número total de ensaios igual à soma dos ensaios individuais é utilizando a função geratriz de momentos:

$$\begin{aligned} M_X(t) = E[e^{tX}] &= \sum_{k=0}^{n} e^{tk} P(X = k) \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \end{aligned}$$

Sendo que, pelo Binômio de Newton:

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-x} y^k$$

Portanto:

$$M_X(t) = (1 - p + pe^t)^n$$

Logo:

$$\begin{aligned} M_{X_i}(t) &= (1 - 0.5 + 0.5e^t)^i \\ &= (0.5 + 0.5e^t)^i \end{aligned}$$

Sendo assim, pelas propriedades das funções geratrizes de momentos:

$$\begin{aligned} M_T(t) = \prod_{i=1}^{5} M_{X_i}(t) &= \prod_{i=1}^{5} (0.5 + 0.5e^t)^i \\ &= (0.5 + 0.5e^t)^{1+2+3+4+5} \\ &= (0.5 + 0.5e^t)^{15} \end{aligned}$$

Portanto, vão-se que $T \sim Binon(15; 0, 5)$

(b) Calcule o valor esperado e a variância da estatística T.
Resposta: Para uma variável $X \sim Binom(n, p)$, vale:

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

Aplicando à variável $T \sim Binom(15; 0, 5)$:

$$E[T] = 15 \times 0,5 = 7,5$$

$$Var(T) = 15 \times 0,5 \times (1 - 0,5) = 15 \times 0,5 \times 0,5 = 3,75.$$

(c) Deduza a distribuição amostral de $R = \sum\limits_{i=1}^{9} X_i - \sum\limits_{j=1}^{2} Y_j$.

Resposta: Similarmente ao item a, temos que:

$$M_{\sum_{i=1}^{9} X_i}(t) = (0,5 + 0,5e^t)^{\sum_{i=1}^{9} i} = (0,5 + 0,5e^t)^{45}$$

Portanto, $\sum_{i=1}^{9} X_i \sim Bin(45; 0, 5)$.
Sabendo que se $X_1, X_2, ..., X_n$ são tais que, X_i é $N(\mu, \sigma_2)$, então, $\sum_{i=1}^{n}(X_i)$ é $N(n\mu, n\sigma_2)$ e como $n = 45$ é suficientemente grande $(n > 30)$, logo, pelo TLC:

$$\begin{aligned} \sum_{i=1}^{9} X_i &\sim N(np, np(1 - p)) \\ &\Rightarrow N(45 \times 0,5; 45 \times 0,5 \times 0,5) \\ &\Rightarrow N(22,5; 11,25) \end{aligned}$$

Sendo assim, como $Y_i \sim N(2, 1)$:

$$\sum_{j=1}^{2} Y_j \sim N(4, 2)$$

Portanto:

$$R = \sum_{i=1}^{9} X_i - \sum_{j=1}^{2} Y_j \sim N(18,5; 13,25)$$

Observação: Nesse exercício tem-se que a normalidade de $\sum_{j=1}^{2} Y_j$ é exata, enquanto a de $\sum_{i=1}^{9} X_i$ é assintótica. Considerou-se que ambas são do mesmo tipo para que valesse a propriedade da soma de normais.

Q8. Seja $X_1, \cdots, X_n$ uma amostra aleatória da variável aleatória $X \sim \mathbb{U}(0, \theta)$. Se $X_{(n)} = \text{Max}\{X_1, \cdots, X_n\}$ denota a maior observação do conjunto de dados, obtenha:

(a) A função de densidade de probabilidade da estatística $T_n = X(n)$.

Resposta: Como $X \sim \mathcal{U}(0, \theta)$, segue que, $F_X(x) = \int_0^x f(y)dy = \int_0^x \theta^{-1}dy = \frac{x}{\theta}$, para todo $0 \leq x \leq \theta$. Assim, sendo $T_1 = X_{(n)}$ um estimador para θ, temos que,

$$\begin{aligned} F_{T_1}(t) &= \mathbb{P}\left(X_{(n)} \leq t\right) = \mathbb{P}\left(max\{X_1, X_2, \cdots, X_n\} \leq t\right) \\ &= \mathbb{P}\left(X_1 \leq x, X_2 \leq x, \cdots, X_n \leq t, \right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq t\right) \overset{iid's}{=} \prod_{i=1}^{n} \mathbb{P}\left(X_1 \leq t\right) \\ &= \left[\mathbb{P}\left(X_1 \leq t\right)\right]^n \\ &= \left(\frac{t}{\theta}\right)^n, \ \forall\, 0 \leq t \leq \theta. \end{aligned} \tag{1}$$

Assim, a fda de T_1 pode ser apresentada da seguinte forma,

$$F_{T_1}(t) = \begin{cases} 0 & t < 0 \\ \left(\frac{t}{\theta}\right)^n & 0 \leq t < \theta \\ 1 & t \geq \theta. \end{cases}$$

Como consequência, a fdp de T_1 é dada por: $f_{T_1}(t) = \frac{d}{dt}F_{T_1}(t) = \frac{n}{\theta^n}t^{n-1}$, para todo $0 \leq t \leq \theta$.

(b) Mostre T_n é um estimador viesado para θ, e obtenha a função viés, $b(T_n)$.

Resposta: com base na fdp de T_n dada a esperança matemática de T_1 é dada por:

$$\mathbb{E}_\theta(T_1) = \int_0^\theta t f_{T_1}(t) = \frac{n}{\theta^n}\int_0^\theta t^n dt = \frac{n\theta}{n+1} \neq \theta, \tag{2}$$

então, concluímos que $T_1 = X_{(n)}$ é um estimador viesado para θ.

(c) Do item (b), obtenha o valor da constante $c > 0$, tal que cT_n seja um estimador não-viesado para θ.

Resposta: observe que, $T_n^* = cT_n = cX_{(n)}$. Segue do item (b) que $\mathbb{E}_\theta(T_n) = \frac{n\theta}{n+1}$, então, $\mathbb{E}_\theta(T_n^*) = \mathbb{E}_\theta(cT_n) = \theta$, consequentemente, $c\frac{n\theta}{n+1} = \theta \longleftrightarrow c = \frac{n+1}{n}$. Ou seja, $T_n^* = \left(1 + \frac{1}{n}\right)X_{(n)}$ é um estimador não-viesado para θ.

(*d*) Considere o seguinte estimador $Y_n = 2\bar{X}_n$. É correto afirmar que Y_n é um estimador não-viesado para θ? Justifique a sua resposta.

Resposta: observe que $Y_n = 2\bar{X}_n$. Então, temos que lembrar que $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ é um estimador não-viesado para a média μ, sendo $\mu = \mathbb{E}(X_i)$. Assim, como $X_i \overset{iid}{\sim} \mathcal{U}(0,\theta)$, segue que, $\mathbb{E}_\theta(X_i) = \frac{\theta}{2}$. Consequentemente, $\mathbb{E}_\theta(Y_n) = 2\mathbb{E}_\theta\left(\bar{X}_n\right) = \frac{2\theta}{2} = \theta$. Assim, podemos concluir que Y_n é um estimador não-viesado para θ.

Q9. Definimos a variável $\epsilon = \bar{X}_n - \mu$ como sendo o erro amostral da média, onde $\bar{X}_n$ é a média de uma amostra aleatória simples de tamanho n de uma população com média μ e desvio-padrão σ:

(a) Determine a média, $E(\epsilon)$ e variância $Var(\epsilon)$.

Resposta: note que, $\epsilon = \bar{X}_n - \mu$. Se $\mathbb{E}_\mu(X_i) = \mu$, então, $\mathbb{E}_\mu\left(\bar{X}_n - \mu\right) = 0$. De igual forma, $Var(\epsilon) = Var\left(\bar{X}_n - \mu\right) = Var\left(\bar{X}_n\right) = \sigma^2/n$.

(b) Se a população é normal com $\sigma = 20$; que proporção das amostras de tamanho 100 terá erro amostral absoluto maior do que 2 unidades?

Resposta: segue do enunciado que $n = 100$ e $\sigma = 20$, supondo que, $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Então, queremos calcular a seguinte probabilidade,

$$
\begin{aligned}
\mathbb{P}(|\epsilon| > 2) &= \mathbb{P}\left(|\bar{X}_n - \mu| > 2\right) = \mathbb{P}\left(\bar{X}_n > 2 + \mu\right) + \mathbb{P}\left(\bar{X}_n < \mu - 2\right) \\
&= \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{2}{\sigma/\sqrt{n}}\right) + \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{-2}{\sigma/\sqrt{n}}\right) \quad \text{padronizaÃ§Ã£o} \\
&= 1 - \mathbb{P}\left(\frac{-2}{\sigma/\sqrt{n}} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{2}{\sigma/\sqrt{n}}\right) \\
&= 1 - \mathbb{P}\left(\frac{-2}{20/\sqrt{100}} \leq \frac{\bar{X}_n - \mu}{20/\sqrt{100}} \leq \frac{2}{20/\sqrt{100}}\right) \\
&= 1 - \mathbb{P}(-1 \leq Z \leq 1) \\
&= 1 - \{\Phi(1) - \Phi(-1)\} \\
&= 1 - (0,8413 - 0,1587) \\
&= 0,3174.
\end{aligned}
$$

A probabilidade apresentada anteriormente pode ser obtida de uma outra forma, levando em consideração a simetria da distribuição normal. Isto Ã©,

$$
\mathbb{P}(|\epsilon| > 2) = \underbrace{2\mathbb{P}(\epsilon > 2)}_{(i)} = \underbrace{2\mathbb{P}(\epsilon < -2)}_{(ii)}.
$$

Observe que, de (i) temos o seguinte resultado,

$$
\begin{aligned}
(i)\ 2\mathbb{P}(\epsilon > 2) &= 2\mathbb{P}\left(\bar{X}_n - \mu > 2 - \mathbb{E}(\epsilon)\right) = 2\mathbb{P}\left(\bar{X}_n - \mu > 2 - \mathbb{E}(\epsilon)\right) \\
&= 2\mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{2 - \mathbb{E}(\epsilon)}{\sigma/\sqrt{n}}\right) = 2\mathbb{P}\left(Z > \frac{2 - 0}{20/\sqrt{100}}\right) \\
&= 2\mathbb{P}(Z > 1) = 2(1 - \Phi(1)) = 2(1 - 0,8413) = 0,3174.
\end{aligned}
$$

$$\begin{aligned}(ii)\ 2\mathbb{P}\left(\epsilon < -2\right) &= 2\mathbb{P}\left(\bar{X}_n - \mu < -2 - \mathbb{E}\left(\epsilon\right)\right) = 2\mathbb{P}\left(\bar{X}_n - \mu < -2 - \mathbb{E}\left(\epsilon\right)\right)\\ &= 2\mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{-2 - \mathbb{E}\left(\epsilon\right)}{\sigma/\sqrt{n}}\right) = 2\mathbb{P}\left(Z < \frac{-2-0}{20/\sqrt{100}}\right)\\ &= 2\mathbb{P}\left(Z < -1\right) = 2\Phi\left(-1\right) = 2 \times 0,1587 = 0,3174.\end{aligned}$$

Ou seja, a proporção de amostras de tamanho 100 com erro amostral absoluto maior do que 2 é de 31,74%. onde $\Phi\left(\cdot\right)$ denota a acumulada da distribuição normal padrão. Isto é, $\Phi\left(a\right) = \mathbb{P}\left(Z \leq a\right)$.

(c) Com base na informação do item (b), qual deve ser o valor de δ para que $P(|\epsilon| > \delta) = 0,01$?

Resposta: queremos encontrar δ, tal que, $\mathbb{P}\left(|\epsilon| > \delta\right) = \mathbb{P}\left(|\bar{X}_n - \mu| > \delta\right) = 0,01$. Ou seja,

$$\begin{aligned}\mathbb{P}\left(|\epsilon| > \delta\right) &= \mathbb{P}\left(|\bar{X}_n - \mu| > \delta\right) = \mathbb{P}\left(\bar{X}_n > \delta + \mu\right) + \mathbb{P}\left(\bar{X}_n < \delta - \mu\right)\\ &= \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{\delta}{\sigma/\sqrt{n}}\right) + \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < -\frac{\delta}{\sigma/\sqrt{n}}\right)\\ &= \mathbb{P}\left(Z > \frac{\delta}{\sigma/\sqrt{n}}\right) + \mathbb{P}\left(Z < -\frac{\delta}{\sigma/\sqrt{n}}\right) = 2\mathbb{P}\left(Z > \frac{\delta}{\sigma/\sqrt{n}}\right).\end{aligned}$$

Segue do item (c) e da simetria da distribuição normal que $\mathbb{P}\left(Z > \frac{\delta}{\sigma/\sqrt{n}}\right) = \mathbb{P}\left(Z < -\frac{\delta}{\sigma/\sqrt{n}}\right)$. Desta forma, $\mathbb{P}\left(|\epsilon| > \delta\right) = 2\mathbb{P}\left(Z > \frac{\delta}{\sigma/\sqrt{n}}\right) = 2\left[1 - \Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right)\right] = 0,01 \implies \Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right) = 1 - 0,01/2 \implies \Phi\left(\frac{\delta}{20/\sqrt{100}}\right) = 0,995 \implies \frac{\delta}{\sigma/\sqrt{n}} = \Phi^{-1}\left(0,995\right) \implies \delta = \left(\frac{\sigma}{\sqrt{n}}\right)\Phi^{-1}\left(0,995\right) = \left(\frac{20}{\sqrt{100}}\right) \times 2,576 = 5,152.$

(d) Qual deve ser o tamanho da amostra para que 95% dos erros amostrais absolutos sejam inferiores a 1 unidade?

Resposta: queremos o valor n tal que, $\mathbb{P}\left(|\bar{X}_n - \mu| < 1\right) = 0,95$. Ou seja,

$$\begin{aligned}\mathbb{P}\left(|\bar{X}_n - \mu| < 1\right) &= \mathbb{P}\left(-1 < \bar{X}_n - \mu < 1\right)\\ &= \mathbb{P}\left(\frac{-1}{\sigma/\sqrt{n}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{1}{\sigma/\sqrt{n}}\right)\\ &= \mathbb{P}\left(\frac{-1}{20/\sqrt{n}} < Z < \frac{1}{20/\sqrt{n}}\right)\\ &= 1 - 2\mathbb{P}\left(Z < -\frac{1}{20/\sqrt{n}}\right) \quad \text{simetria da distribuiÃ§Ã£o normal.}\end{aligned}$$

Assim, $1 - 2\mathbb{P}\left(Z < -\frac{1}{20/\sqrt{n}}\right) = 0,95 \iff 1 - 2\Phi\left(-\frac{\sqrt{n}}{20}\right) = 0,95 \iff \Phi\left(-\frac{\sqrt{n}}{20}\right) = \frac{1-0,95}{2} \iff \Phi\left(-\frac{\sqrt{n}}{20}\right) = 0,025 \iff -\frac{\sqrt{n}}{20} = \Phi^{-1}\left(0,025\right) = -1.960 \iff \sqrt{n} = 20 \times 1,960 \iff \left(\sqrt{n}\right)^2 = \left(1,960 \times 20\right)^2 \approx 1537$. Ou seja, o tamanho de amostra suficiente para que 95% dos erros amostrais absolutos sejam inferiores a 1 un. é $n = 1537$.

Q10. Em uma sondagem, perguntou-se a 1002 membros de determinado sindicato se eles haviam votado na última eleição para a direção do sindicato e 701 responderam afirmativamente. Os registros oficiais obtidos depois da eleição mostram que 61% dos membros aptos a votar de fato votaram. Calcule a probabilidade de que, dentre 1002 membros selecionados aleatoriamente, no mínimo 701 tenham votado, considerando que a verdadeira taxa de votantes seja de 61%. O que o resultado sugere?

Q11. Uma caixa contem duas bolas pretas e uma bola branca. Seja X o número de bolas pretas retiradas da caixa numa única extração.

(a) Qual a distribuição da variável aleatória X?

(b) Considere uma amostra aleatória de tamanho $n = 9$ e seja $T = \sum_{i=1}^{9} X_i$. Deduza a distribuição amostral de T.

(c) Calcule a $\mathbb{P}(T \leq 3)$.

Q12. Sejam $X_1, \cdots, X_n$ um amostra aleatória de tamanho n da distribuição da variável aleatória X com fdp dada por

$$f(x|\theta) = \frac{2x}{\theta^2},\ 0 < x < \theta,\ \theta > 0.$$

(a) Especifique o espaço paramétrico e o suporte associado à distribuição de X.

(b) Verifique se $\hat{\theta}_1 = \bar{X}_n$ e $\hat{\theta}_2 = X_{(n)}$ são estimadores viciados para θ. Item Encontre e compare os $EQM's$ dos dois estimadores. Faça um gráfico dos $EQM's$ como função de θ.