Primeira avaliação (A1)

Disciplina: Inferência Estatística
Professor: Luiz Max de Carvalho

20 de Setembro de 2021

- Por favor, entregue um único arquivo PDF;
- O tempo para realização da prova é de 4 (quatro) horas, mais vinte minutos para upload do documento para o e-class;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo, ou figura geométrica de sua preferência;
- A prova vale 80 pontos; a pontuação restante é contada como bônus.
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto.

Dicas

- Se X tem distribuição exponencial com parâmetro $\lambda > 0$, então, para $x > 0$ as funções de densidade de probabilidade e densidade acumulada são, respectivamente,
$$f_X(x) = \lambda \exp(-\lambda x),$$
$$F_X(x) = 1 - \exp(-\lambda x).$$
- Se X tem distribuição Gama com parâmetros $\alpha > 0$ e $\beta > 0$ e f.d.p.,
$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(\beta x),$$
para $x > 0$, então $W = 1/X$ tem distribuição Gama-inversa, com f.d.p.
$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-(\alpha+1)} \exp(\beta/w),$$
para $w > 0$. Ademais, $E[W] = \beta/(\alpha-1)$ e $\text{Var}(W) = \beta^2/[(\alpha-1)^2(\alpha-2)]$.
- Se $X_i \sim \text{Gama}(\alpha, \beta_i)$, com $\alpha_i > 0$ para todo i e $\beta > 0$, então $Y = \sum_{i=1}^n X_i$ tem distribuição Gama com parâmetros $\alpha_y = \sum_{i=1}^n \alpha_i$ e $\beta_y = \beta$.
- Se $X \sim \text{Gama}(\alpha, \beta)$, $Y = cX$ tem distribuição $\text{Gama}(\alpha, \beta/c)$ para $c > 0$.

1. Circling the square.

Um círculo C_r de raio r é inscrito em uma folha de papel quadrada com lado b. Suponha que desejamos estimar a área A deste círculo. Para tanto, vamos amostrar vetores aleatórios de uma distribuição uniforme definida sobre a folha de papel e, para estimar a área da circunferência, contar a proporção de vetores caindo dentro e fora de C_r e multiplicar esta proporção pela área total da folha de papel.

a) (2,5 pontos) Mostre que se X e Y são variáveis aleatórias i.i.d. com distribuição Uniforme$(0, b)$, então (X, Y) possui função de densidade de probabilidade constante sobre $(0, b) \times (0, b)$;

b) (7,5 pontos) Você deixa cair grãos de milho sobre a folha e conta quantos deles cairam dentro do círculo e fora do círculo (porém na folha). Vamos supor que este mecanismo gera observações i.i.d. uniforme sobre $(0, b)^2$. Represente os grãos que caíram sobre a folha através de $(X_1, Y_1), ..., (X_n, Y_n)$ e defina $Z_i = \mathbb{I}((X_i, Y_i) \in C_r)$, $i = 1, ..., n$ como uma variável indicadora que recebe valor 1 se o grão está dentro da circunferência. Suponha que depois de medir $\boldsymbol{Z}$ você joga fora $\boldsymbol{X}$ e $\boldsymbol{Y}$, isto é, guarda o milho no pote de novo para fazer pipoca mais tarde. Construa um modelo estatístico parametrizado pela área, A, da circunferência que reflete este experimento. Encontre uma estatística suficiente mínima para o parâmetro deste modelo.

   Dica: desenhe um diagrama e considere as áreas envolvidas (evite <u>avaliar</u> integrais!);

c) (5 pontos) Considere $\delta_1(\boldsymbol{Z}) = b^2 \bar{Z}_n$. Este é um estimador não enviesado da área do círculo?

d) (5 pontos) Calcule o erro quadrático médio $R(A, \delta_1)$ de δ_1 e discuta como ele se comporta em relação à quantidade de interesse. O que acontece com $R(A, \delta_1)$ quando A cresce?

2. The shinning.

Suponha que você é a pessoa responsável pelo controle estatístico de qualidade na fábrica de lâmpadas *LuminaEu*. Seu chefe, Astolfo, lhe envia uma planilha com os valores $X_1, X_2, \ldots, X_n$ dos tempos de falha de n lâmpadas (em dias). Você lê no manual da empresa que um modelo exponencial i.i.d. com parâmetro θ é apropriado para análise.

a) (5 pontos) Mostre que o estimador de momentos para θ coincide com o EMV neste caso;

b) (10 pontos) Discuta se o estimador do item anterior é eficiente para amostras finitas. O que acontece assintoticamente?

c) (5 pontos) Conhecendo Astolfo, no entanto, você sabe que ele não saberá interpretar quaisquer estimativas diretas da taxa θ, então decide considerar a probabilidade de excedência[1] $\alpha := \Pr(X_1 > c)$ para um certo $c > 0$. Encontre um estimador de máxima verossimilhança para α;

3. Cool and normal!

Suponha que você é a pessoa responsável por analisar a concentração de ácido em pedaços de queijo vindos da famosa fábrica de frios francesa *J'skeci.* Assumindo uma distribuição normal para as concentrações em n medições independentes de n pedaços distintos, você precisa descobrir a média μ e a variância v desta distribuição.

a) (5 pontos) Considere a priori imprópria

$$\xi(\mu, v) \propto 1/v \tag{1}$$

Mostre que a posteriori $\xi(\mu, v \mid \boldsymbol{x})$ é própria;

Dica: Procure com atenção o núcleo de distribuições conhecidas.

b) (7,5 pontos) Exiba o estimador de Bayes sob perda quadrática para v e o estimador de Bayes sob perda absoluta para μ e discuta se esses estimadores são viesados;

c) (5 pontos) Encontre uma priori conjugada para (μ, v);

d) (2,5 pontos) Mostre que a priori em (1) pode ser vista como um limite particular (dos hiperparâmetros) da priori conjugada do item anterior.

[1] Em inglês, *exceedance probability.*

4. Get your ducks in a row.

Pato Donald, Huguinho, Zezinho e Luisinho estão estudando Inferência Estatística para trabalhar no *hedge fund* do Tio Patinhas. O problema em questão é a estimação do parâmetro θ de uma distribuição uniforme em $(\theta/2, 3\theta/2)$ a partir de uma amostra aleatória $X_1, X_2, \ldots, X_n$. Cada um propôs um estimador diferente para θ e seu trabalho é ajudar o Tio Patinhas a ordenar esses estimadores em ordem de qualidade.

Sejam $M := \max(X_1, X_2, \ldots, X_n)$ e $m := \min(X_1, X_2, \ldots, X_n)$.

Os estimadores escolhidos foram

1. $\delta_{\mathrm{D}}(\boldsymbol{X}) = X_1$, para o Pato Donald;
2. $\delta_{\mathrm{H}}(\boldsymbol{X}) = m$, para Huguinho;
3. $\delta_{\mathrm{Z}}(\boldsymbol{X}) = M$, para Zezinho;
4. $\delta_{\mathrm{L}}(\boldsymbol{X}) = (M + m)/2$, para Luisinho;

Para lhe ajudar na tarefa de julgar estes estimadores, Tio Patinhas enviou o seguinte conjunto de fatos úteis: para $X_1, X_2, \ldots, X_n \sim \text{Uniforme}(a, b)$, temos

$$
\begin{aligned}
E[X_1] &= \frac{a+b}{2},\\
\operatorname{Var}(X_1) &= \frac{(b-a)^2}{12},\\
E[m] &= a + \frac{1}{n+1}(b-a),\\
E[M] &= b - \frac{1}{n+1}(b-a),\\
\operatorname{Var}(m) = \operatorname{Var}(M) &= \frac{n}{(n+1)^2(n+2)}(b-a)^2,\\
\operatorname{Cov}(m, M) &= \frac{(b-a)^2}{(n+1)^2(n+2)},\\
\operatorname{Corr}(m, M) &= \frac{1}{n}.
\end{aligned}
$$

Os patos ainda não sabem Inferência Estatística muito bem, portanto tenha paciência com eles.

a) (2,5 pontos) Os estimadores de Huguinho e Zezinho são viesados. Mostre aos patinhos como construir versões não-viesadas, $\delta_{\mathrm{UH}}(\boldsymbol{X})$ e $\delta_{\mathrm{UZ}}(\boldsymbol{X})$;

b) (2,5 pontos) Discuta se algum dos estimadores do item anterior é inadmissível;

c) (2,5 pontos) Mostre que $\boldsymbol{T} = (m, M)$ é suficiente conjunta para θ;

d) (7,5 pontos) Mostre que $\delta_{\mathrm{L}}(\boldsymbol{X}) = E[\delta_{\mathrm{D}}(\boldsymbol{X}) \mid \boldsymbol{T}]$, isto é, que o estimador de Luisinho é o melhoramento de Rao-Blackwell do estimador do Pato Donald;

e) (5 pontos) Ordene os estimadores $\delta_{\mathrm{D}}(\boldsymbol{X})$, $\delta_{\mathrm{UH}}(\boldsymbol{X})$, $\delta_{\mathrm{UZ}}(\boldsymbol{X})$ e $\delta_{\mathrm{L}}(\boldsymbol{X})$ em termos de erro quadrático médio. Quem propôs o melhor estimador?[2]

[2] No caso de Huguinho e Zezinho, com a sua ajuda.

5.Questão bônus: Boss is boss, ain't it, dad?

Considere mais uma vez o problema da questão 4. Desta vez, Tio Patinhas resolveu propor o próprio estimador, e quer mostrar que esse estimador pode ser melhor que qualquer um dos propostos anteriormente. Para isso, propõe utilizar um estimador da forma

$$\delta_{\mathrm{P}}(\boldsymbol{X}) = (1-\alpha)\delta_{\mathrm{UH}}(\boldsymbol{X}) + \alpha\delta_{\mathrm{UZ}}(\boldsymbol{X}),$$

com $\alpha \in (0, 1)$.

a) (10 pontos) Mostre que δ_{P} é não-viesado e compute seu erro quadrático médio;

   Dica: Lembre-se de que para $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y).$$

b) (10 pontos) Encontre α_{op} que faz com que δ_{P} tenha variância mínima. O estimador $\delta_{\mathrm{P}}^{\mathrm{op}}(\boldsymbol{X}) = (1-\alpha_{\mathrm{op}})\delta_{\mathrm{UH}}(\boldsymbol{X}) + \alpha_{\mathrm{op}}\delta_{\mathrm{UZ}}(\boldsymbol{X})$ domina todos aqueles derivados na questão 4? Justifique.