Avaliação A1 - Inferência Estatística
Soluções

Professor: Philip Thompson
Monitores: Eduardo Adame & Ezequiel Braga

6 de outubro de 2023

Instruções

- A prova vale **10 pontos**. Verifique a pontuação de cada questão e trace sua estratégia para resolvê-las.
- Respostas sem justificativas serão **desconsideradas**;
- Demarque com clareza sua **resposta final** para cada questão. Sugerimos que circule ou desenhe um retângulo em volta desses resultados;
- A questão bônus é um **desafio**. Deixe-a para o final;
- Apenas **uma folha de "cola"** de tamanho A4 frente e verso poderá ser trazida e utilizada como consulta durante a avaliação. A mesma deverá ser **entregue** junto de suas soluções.

Dados úteis

- Se $X \sim \text{Beta}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, então
  - $f_X(x \mid \alpha, \beta) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$, $x \in (0,1)$, $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
  - $\mathbb{E}[X] = \dfrac{\alpha}{\alpha+\beta}$
- Se $X \sim \text{Pareto}(\theta_0, \alpha)$ com parâmetros $\theta_0 > 0$, $\alpha > 0$, mostre que

$$\mathbb{E}[X] = \begin{cases} \dfrac{\alpha\theta_0}{\alpha-1}, & \text{se } \alpha > 1 \\ \infty, & \text{se } \alpha \leq 1 \end{cases}$$

- Se necessário, você pode usar o resultado de que a composição de função bijetora com uma estatística suficiente também é uma estatística suficiente.

Questões

Questão 1 (2 pontos). Uma variável aleatória θ seguindo a distribuição de Pareto com parâmetros θ_0, α positivos tem densidade

$$f(\theta \mid \theta_0, \alpha) := \begin{cases} \dfrac{\alpha\theta_0^\alpha}{\theta^{\alpha+1}}, & \theta \geq \theta_0, \\ 0, & \theta < \theta_0. \end{cases}$$

a) Mostre que a distribuição de Pareto é a *priori* conjugada para amostras seguindo a distribuição Uniforme no intervalo $[0, \theta]$, onde o parâmetro θ é desconhecido.

Solução

Para $X \sim \mathcal{U}[0, \theta]$, temos que:

$$f(x \mid \theta) = \frac{1}{\theta}\mathbb{1}\{0 \leq x \leq \theta\}$$

Logo, para $X_1, \ldots, X_n \sim \mathcal{U}[0, \theta]$ iid, temos:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{1}\{0 \leq x_i \leq \theta\}, \\ &= \frac{1}{\theta^n}\mathbb{1}\{0 \leq \min\{x_1, \ldots, x_n\}\}\mathbb{1}\{\max\{x_1, \ldots, x_n\} \leq \theta\}. \end{aligned}$$

Prosseguindo, sabemos que:

$$\begin{aligned} \xi(\theta \mid \boldsymbol{x}) &\propto f_n(\boldsymbol{x} \mid \theta)\xi(\theta), \\ &\propto \frac{1}{\theta^{n+\alpha+1}}\mathbb{1}\{\max\{x_1, \ldots, x_n\} \leq \theta\}\mathbb{1}\{\theta_0 \leq \theta\}, \\ &\propto \frac{1}{\theta^{n+\alpha+1}}\mathbb{1}\{\max\{x_1, \ldots, x_n, \theta_0\} \leq \theta\}. \end{aligned}$$

Logo, $(\theta \mid \boldsymbol{x}) \sim \text{Pareto}(\max\{x_1, \ldots, x_n, \theta_0\}, n + \alpha)$; e, portanto, a distribuição Pareto é *priori* conjugada para θ de amostra iid Uniforme$[0, \theta]$.

b) Seja $X_1, \ldots, X_n$ uma amostra iid de uma distribuição uniforme no intervalo $[0, \theta]$, onde o parâmetro θ é desconhecido. Tomando como distribuição *a priori* $\xi(\theta)$ a distribuição de Pareto com parâmetros θ_0, α e a função custo quadrática, determine o estimador de Bayes para o parâmetro θ.

Solução

Para função de custo quadrática, temos que o estimador de Bayes será o valor esperado *a posteriori*. Além disso, pelo item anterior sabemos que $(\theta \mid \boldsymbol{x}) \sim \text{Pareto}(\max\{x_1, \ldots, x_n, \theta_0\}, n + \alpha)$ e, pelos dados úteis, temos:

$$\hat{\theta} = \mathbb{E}[\theta \mid \boldsymbol{x}] = \frac{n + \alpha}{n + \alpha - 1}\max\{x_1, \ldots, x_n, \theta_0\}.$$

Como $n > 1, n + \alpha - 1 > 0$.

Questão 2 (2 pontos). Seja $X_1, \dots, X_n$ uma amostra iid de uma distribuição Bernoulli com parâmetro p desconhecido. Tome uma distribuição *a priori* para p sendo a distribuição Beta(α, β) com parâmetros α, β positivos.

a) Determine o estimador de Bayes $\hat{p}$ usando o custo quadrático.

Solução

Antes de tudo, sabemos que o estimador de Bayes sob custo quadrático para p será seu valor esperado *a posteriori*.
Portanto, encontremos a distribuição *a posteriori* para p:

$$\begin{aligned} \xi(p \mid \boldsymbol{x}) &\propto f_n(\boldsymbol{x} \mid p)\xi(p), \\ &\propto p^{\left[\sum_{i=1}^n x_i\right]}(1-p)^{\left[n-\sum_{i=1}^n x_i\right]} \cdot p^{\alpha-1}(1-p)^{\beta-1}, \\ &\propto p^{\left[\alpha+\sum_{i=1}^n x_i - 1\right]}(1-p)^{\left[\beta+n-\sum_{i=1}^n x_i - 1\right]}. \end{aligned}$$

Logo, $(p \mid \boldsymbol{x}) \sim$ Beta$(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$. Pelos dados úteis, sabemos que:

$$\hat{p} = \mathbb{E}[p \mid \boldsymbol{x}] = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \cancel{\sum_{i=1}^n x_i} + \beta + n - \cancel{\sum_{i=1}^n x_i}} = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n}$$

b) Justifique afirmativamente ou negativamente: $\hat{p}$ é uma estatística suficiente mínima?

Solução

Primeiramente, encontremos uma estatística suficiente através do Teorema da Fatorização:

$$\begin{aligned} f_n(\boldsymbol{x} \mid p) &= p^{\left[\sum_{i=1}^n x_i\right]}(1-p)^{\left[n-\sum_{i=1}^n x_i\right]}, \\ &= h(\boldsymbol{x})g(T(\boldsymbol{x}) \mid p), \end{aligned}$$

para $h(\boldsymbol{x}) = 1$, $g(t \mid p) = p^t(1-p)^{n-t}$ e $T(\boldsymbol{x}) = \sum_{i=1}^n x_i$.
Como nosso $\hat{p}$ é função bijetora deste $T(\boldsymbol{x})$ suficiente, ele também é suficiente.
Pelo Teorema 7.8.3 do DeGroot, sabemos que se o estimador de Bayes é suficiente, ele é suficiente mínimo.
Deste modo, nosso $\hat{p}$ é uma estatística suficiente mínima.

Questão 3 (1 ponto). Uma distribuição pertence a família exponencial com parâmetro unidimensional $\theta \in \mathbb{R}$ se a densidade tem forma

$$f(x \mid \theta) = b(\theta)h(x)\exp\{\eta(\theta)T(x)\},$$

para funções b, h, η e T. Seja $X_1, \dots, X_n$ uma amostra desta distribuição e considere a estatística $T(\boldsymbol{X}) = \sum_{i=1}^n T(X_i)$.

a) Mostre que $T(\boldsymbol{X})$ é suficiente.

Solução

Para atender ao pedido da questão, utilizaremos o Teorema da Fatorização, veja:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^n b(\theta)h(x_i)\exp\{\eta(\theta)T(x_i)\}, \\ &= b(\theta)^n \left(\prod_{i=1}^n h(x_i)\right)\exp\{\eta(\theta)\sum_{i=1}^n T(x_i)\} = H(\boldsymbol{x})g(T(\boldsymbol{x}) \mid \theta), \end{aligned}$$

para $H(\boldsymbol{x}) = \prod_{i=1}^n h(x_i)$, $g(t \mid \theta) = b(\theta)^n \exp\{\eta(\theta)\sum_{i=1}^n T(x_i)\}$ com $T(\boldsymbol{x}) = \sum_{i=1}^n T(x_i)$. Logo, $T(\boldsymbol{X})$ é suficiente.

b) Mostre que $T(\boldsymbol{X})$ é eficiente. *Dica*: Lembre-se que um estimador $T(\boldsymbol{X})$ é eficiente se, e somente se, existem funções de θ que permitem escrever $T(\boldsymbol{X})$ satisfazendo uma certa relação, especificada no livro texto.

Solução

Sabemos que $T(\boldsymbol{X})$ será eficiente se, e somente se, existem funções $u(\theta)$ e $v(\theta)$ tais que:

$$T(\boldsymbol{X}) = u(\theta)\frac{\partial}{\partial\theta}\log f_n(\boldsymbol{X} \mid \theta) + v(\theta).$$

O que pode ser confirmado em nosso caso, pois:

$$\begin{aligned} \log f_n(\boldsymbol{X} \mid \theta) &= \log H(\boldsymbol{X}) + n\log b(\theta) + \eta(\theta)T(\boldsymbol{X}), \\ \frac{\partial}{\partial\theta}\log f_n(\boldsymbol{X} \mid \theta) &= n\frac{b'(\theta)}{b(\theta)} + \eta'(\theta)T(\boldsymbol{X}). \end{aligned}$$

Desse modo, percebemos que:

$$T(\boldsymbol{X}) = \left(\frac{\partial}{\partial\theta}\log f_n(\boldsymbol{X} \mid \theta) - n\frac{b'(\theta)}{b(\theta)}\right)\frac{1}{\eta'(\theta)}.$$

O que nos permite encontrar $u(\theta) = \dfrac{1}{\eta'(\theta)}$ e $v(\theta) = -n\dfrac{b'(\theta)}{b(\theta)\eta'(\theta)}$ que satisfazem a relação. E, portanto, $T(\boldsymbol{X})$ é eficiente.

Questão 4 (2.5 pontos). Seja $X_1, \dots, X_n$ uma amostra de uma distribuição parametrizada por $\theta > 0$ com densidade

$$f(x \mid \theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

a) Determine o MLE para θ.

Solução

Calculemos a verossimilhança:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^n \theta x_i^{\theta-1} \mathbb{1}\{0 < x_i < 1\}, \\ &= \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1} \mathbb{1}\{0 < \min\{x_1, \dots, x_n\}\} \mathbb{1}\{\max\{x_1, \dots, x_n\} < 1\}. \end{aligned}$$

Evidentemente $\mathbb{1}\{0 < \min\{x_1, \dots, x_n\}\} \mathbb{1}\{\max\{x_1, \dots, x_n\} < 1\} = 1$ onde estamos buscando um valor máximo. Seguindo, temos:

$$\log f_n(\boldsymbol{x} \mid \theta) = n \log \theta + (\theta - 1) z,$$

para $z = \log(\prod_{i=1}^n x_i)$. Concluímos ao encontrar $\hat{\theta}$ tal que $\frac{\partial}{\partial \theta} \log f_n(\boldsymbol{x} \mid \theta) = \frac{n}{\theta} + z = 0$:

$$\frac{n}{\hat{\theta}} + z = 0 \implies \hat{\theta} = -\frac{n}{\log \prod_{i=1}^n x_i}.$$

b) Justifique afirmativamente ou negativamente se o MLE para θ é suficiente.

Solução

Pelo Teorema da Fatorização, podemos encontrar h e g tais que:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \theta) &= \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1} \mathbb{1}\{0 < \min\{x_1, \dots, x_n\}\} \mathbb{1}\{\max\{x_1, \dots, x_n\} < 1\}, \\ &= h(\boldsymbol{x}) g(T(\boldsymbol{x}) \mid \theta). \end{aligned}$$

ao definir $h(\boldsymbol{x}) = \mathbb{1}\{0 < \min\{x_1, \dots, x_n\}\} \mathbb{1}\{\max\{x_1, \dots, x_n\} < 1\}$ e $g(t \mid \theta) = \theta^n t^{\theta-1}$, com $T(\boldsymbol{x}) = \prod_{i=1}^n x_i$. Como $\hat{\theta}$ é função bijetora desse $T(\boldsymbol{x})$, ele é suficiente.

c) Justifique afirmativamente ou negativamente se o MLE para θ é suficiente mínimo.

Solução

Como o $\hat{\theta}$ é MLE e suficiente, ele é suficiente mínimo (Teorema 7.8.3 do DeGroot).

Questão 5 (2.5 pontos). Seja $X_1, \ldots, X_n$ uma amostra de uma distribuição Poisson(λ).

a) Encontre o estimador MLE para λ.

Solução

Calculemos, primeiramente, a verossimilhança:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \lambda) &= \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}, \\ &= \frac{1}{\prod_{i=1}^{n} x_i!} \left(e^{-n\lambda} \lambda^{\left[\sum_{i=1}^{n} x_i\right]} \right). \end{aligned}$$

Desse modo, calcular a log-verossimilhança e sua derivada em relação a λ:

$$\begin{aligned} \log f_n(\boldsymbol{x} \mid \lambda) &= -\log \left(\prod_{i=1}^{n} x_i! \right) - n\lambda + \left(\sum_{i=1}^{n} x_i \right) \log \lambda, \\ \frac{\partial}{\partial \lambda} \log f_n(\boldsymbol{x} \mid \lambda) &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}. \end{aligned}$$

Para concluir, encontramos $\hat{\lambda}$ tal que $\frac{\partial}{\partial \lambda} \log f_n \left(\boldsymbol{x} \mid \lambda \right) |_{\hat{\lambda}} = 0$:

$$-n + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} = 0 \implies \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

b) Compute a informação de Fisher $I(\lambda)$.

Solução

Sabemos que $I(\lambda) = -\mathbb{E}_\lambda \left[\frac{\partial^2}{\partial \lambda^2} \log f(x) \right]$, portanto, primeiramente basta encontrar:

$$\frac{\partial^2}{\partial \lambda^2} \log f(x) = \frac{\partial}{\partial \lambda} \left(-n + \frac{x}{\lambda} \right) = -\frac{x}{\lambda^2}.$$

Desse modo:

$$I(\lambda) = -\mathbb{E}_\lambda \left[-\frac{x}{\lambda^2} \right] = \frac{1}{\lambda}$$

Note que daí $I_n(\lambda) = n \cdot I(\lambda) = \dfrac{n}{\lambda}$

Questão 6 (**Questão bônus** - 1 ponto). O modelo "bag-of-words" é utilizado na área de processamento de linguagem natural. Neste modelo, há um alfabeto de k palavras ($k \geq 2$) e a i-ézima palavra tem probabilidade θ_i de ocorrer. Aqui, $0 \leq \theta_i \leq 1$ e $\theta_1 + \ldots + \theta_k = 1$.

Numa amostra de n palavras, denotamos por x_i o números de ocorrências da i-ézima palavra; portanto $x_1 + x_2 + \ldots + x_n = n$. Determine o MLE desta amostra para estimar o vetor $\theta = (\theta_1, \ldots, \theta_k)$.

Dica: a distribuição em questão é chamada *distribuição multinomial*. Ela generaliza a distribuição binomial ($k = 2$) onde há apenas duas palavras (por exemplo, 'cara' e 'coroa').

Solução

Duas soluções são convenientes para este problema. Vamos começar pela *brute force*. Temos:

$$\begin{aligned} f_i(x_i \mid \theta) &= \theta_1^{\mathbb{1}\{x_i=1\}} \cdots \theta_k^{\mathbb{1}\{x_i=k\}}, \\ f_n(\boldsymbol{x} \mid \theta) &\propto \theta_1^{x_1} \cdots \theta_k^{x_k}, \\ \log f_n(\boldsymbol{x} \mid \theta) &= \log C_{\boldsymbol{x}} + x_1 \log \theta_1 + \cdots + x_k \log \theta_k. \end{aligned}$$

Podemos reparametrizar este modelo em termos somente dos parâmetros $(\theta_1, \ldots, \theta_{k-1}$ utilizando o fato que $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$:

$$\log f_n(\boldsymbol{x} \mid \theta) = \log C_{\boldsymbol{x}} + x_1 \log \theta_1 + \cdots + x_k \log \left(1 - \sum_{j=1}^{k-1} \theta_j\right).$$

Deste modo, basta encontrar $\hat{\boldsymbol{\theta}}$ que satisfaça o seguinte sistema:

$$\frac{\partial}{\partial \theta_j} \log f_n(\boldsymbol{x} \mid \theta) = \frac{x_j}{\theta_j} - \frac{x_k}{\sum_{j=1}^{k-1} \theta_j} = \frac{x_j}{\theta_j} - \frac{x_k}{\theta_k} = 0, \ \forall\, j = 1, \ \ldots, \ k-1$$

Ou seja,

$$\frac{x_1}{\hat{\theta}_1} = \frac{x_2}{\hat{\theta}_2} = \cdots = \frac{x_k}{\hat{\theta}_k} = \alpha.$$

Concluindo, basta encontrar α através de:

$$1 = \sum_{j=1}^{n} \hat{\theta}_j = \frac{\sum_{j=1}^{k} x_j}{\alpha} = \frac{n}{\alpha} \implies \alpha = n.$$

O que nos leva a:

$$\hat{\theta}_j = \frac{x_j}{n}, \ (j = 1, \ \ldots, \ n)$$

A outra solução é resolver o problema de otimização:

$$\max_{\theta \in \mathbb{R}} \log f_n(\boldsymbol{x} \mid \theta), \ \text{sujeito a} \sum_{j=1}^{k} \theta_j = 1.$$

Que é simples de resolver através dos multiplicadores de Lagrange (ou Condições de Karush-Kuhn-Tucker).