Departamento de Economia/PUC-Rio **2023**
Estatistica

Lista 5

1. Dada uma amostra aleatória $X_1, \dots, X_n$, definimos as estatísticas de ordem como sendo $X_{(j:n)}$ igual a j-ésimo valor na amostra, em ordem ascendente, para $j = 1, \dots, n$. Se X_i tem distribuição F a.c. com densidade f, então

$$f_{X_{(j:n)}}(x) = \frac{n!}{(j-1)!(n-j)!} f(x)(F(x))^{j-1}(1-F(x))^{n-j}.$$

Considere o caso de uma amostra aleatória com distribuição uniforme em $[0, a]$.

(a) Para n ímpar, fixando $j = n + 1/2$ temos que $M = X_{((n+1)/2:n)}$ é a mediana amostral. Mostre que $E[2M] = a$. (Dica: Você pode fazer um argumento que depende apenas de $f(x)$ ser simétrica em torno de $a/2$.)

(b) Calcule a esperança e a variância de $X_{(n:n)}$. Argumente que $X_{(n:n)}$ converge em probabilidade para a.

(c) Que estatística você usaria para estimar a: $2M$, $2\bar{X}$, $X_{(n:n)}$ ou $\frac{n+1}{n} X_{(n:n)}$? Justifique sua resposta.

2. Na questão anterior, estudamos estatísticas de ordem de uma amostra aleatória com distribuição uniforme em $[0, a]$.

(a) Mostre que $X_{(j:n)}$ é uma estatística suficiente para a se e somente se $j = n$.

(b) $R = X_{(n:n)} - X_{(1:n)}$ é ancilar? É suficiente?

(c) Mostre que $X_{(j:n)}/\bar{X}$ é uma estatística ancilar para a (mesmo com $j = n$). Como sempre, $\bar{X}$ é a média amostral.

3. Você observa uma amostra X com distribuição parametrizada por θ, um número real. Você computa o valor de $T(X)$, uma estatística suficiente de θ e de $S(X)$, uma estatística mínima suficiente de θ. Seu orientador observa o valor realizado de T, e o membro externo da banca o valor

realizado de S. Para fazer previsão sobre θ, você utiliza uma estatística $U(X)$ que tem a propriedade de ser não viesada, ou seja, $E_\theta[U] = \theta$, para todo θ. Como o orientador e o membro externo não conhecem X, eles não podem usar U. Eles utilizam, respectivamente, $U_T = E_\theta[U|T]$ e $U_S = E_\theta[U|S]$.

(a) Explique por que U_T e U_S nesse caso são estatísticas, isso é, podem ser calculadas mesmo por alguém que não conhece o valor de θ.

(b) Mostre que U_T e U_S são não viesados: $E_\theta[U_T] = E_\theta[U_S] = \theta$, para todo θ.

(c) Mostre que o membro externo acredita que em média o orientador vai concordar com a sua previsão: $E_\theta[U_T|S] = U_S$.

(d) Mostre que $\mathrm{Var}(U) \geq \mathrm{Var}(U_T) \geq \mathrm{Var}(U_S)$.

4. Sejam $Y_1, \ldots, Y_n$ independentes, com $Y_i \sim N(\beta x_i, 1)$, onde $x_1, \ldots, x_n$ são números fixos conhecidos.

   (a) Com base em $\bar{Y}$, proponha um estimador de método dos momentos para β.

   (b) Qual é o MLE para β? Qual a sua distribuição?

   (c) Se definimos $U_i = Y_i - \beta x_i$, temos que $E(x_i U_i) = 0$. Proponha um estimador de método dos momentos baseado nessa relação.

   (d) Examine a densidade da amostra e verifique que $\sum x_i y_i$ é uma estatística suficiente para β, desde que $\sum_i x_i^2$ seja conhecido. Com base nesse fato, qual dos três estimadores para β acima parece ser preferível?

5. Sejam $X_1, X_2, \ldots, X_n$ i.i.d. com $EX_i = \mu$, $\mathrm{Var}X_i = \sigma^2$. Seja $S^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2$.

   (a) Mostre que

   $$\mathrm{Var}(S^2) = \frac{1}{n}\left(\kappa - \frac{n-3}{n-1}\right)\sigma^4$$

   onde $\kappa = E[(X-\mu)^4]/\sigma^4$ é a curtose da distribuição de X_i.

   (b) Usando o resultado acima, prove que o valor de a que faz com que o erro quadrático médio de aS^2 seja mínimo é

   $$a = \frac{n-1}{n+1+(\kappa-3)(n-1)/n}$$

(c) Sabendo que $\kappa = 3$ na caso da distribuição normal, você agora pode verificar o resultado mencionado em classe, de que o estimador da forma aS^2 que minimiza o MSE é $\frac{1}{n+1}\sum_i (X_i - \bar{X})^2$.

6. Vimos em aula que o número que minimiza o MSE para a esperança é a média amostral: isto é, $a^* = \arg\min_a \sum(x_i - a)^2 = \bar{x}$. Nesse exercício vamos procurar estatística que minimizam o MSE de transformações de x_i.

   (a) Para h estritamente monótona e contínua, considere o problema $\min_a \sum_i (h(x_i) - h(a))^2$. Mostre que a solução desse problema é $a = h^{-1}(\overline{h(x)}) = h^{-1}(\sum_i h(x_i)/n)$. Dica: trabalhe com $b = h(a)$.

   (b) Descubra a transformação h que faça com que média harmônica minimize o MSE de $h(x_i)$.

   (c) Descubra a transformação h que faça com que média geométrica minimize o MSE de $h(x_i)$.

   (d) Mostre que se temos uma amostra aleatória lognormal (μ, σ^2), o MLE de μ é baseado na média geométrica. O estimador obtido depende do fato de σ^2 ser conhecido ou não?

   (e) Qual é o estimador de método dos momentos de μ? Sua resposta depende do fato de σ^2 ser conhecido ou não?