Exercícios e Complementos MI678 - 2017

González-López, V. A.*

May 10, 2017

1. Dadas $X_1, X_2 \cdots, X_n$ i.i.d, defina a estatística $T_\mu = \sqrt{n}\frac{(\overline{X}-\mu)}{S_n}$ onde $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ e $S_n^2 = \frac{\sum_{i=1}^n (X_i-\overline{X})^2}{n-1}$. Para rejeitar $H_0 : \mu = \mu_0$ ao nível α é requerido que $P(|T_{\mu_0}| > t_0) \leq \alpha$ onde t_0 é o $\alpha/2$ quantil da distribuição Normal padrão.
Inspecionar o desempenho da regra de decisão em função do aumento do tamanho amostral n comparando os seguintes supostos:
   i. $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, \cdots, n$;
   ii. $X_i \sim \mathrm{Exp}(\alpha)$, $\alpha > 0$, $i = 1, \cdots, n$.

2. Dadas $X_1, X_2 \cdots, X_n$ i.i.d e distribuição acumulada G ($X_i \sim G$). Seja $M_n = \max\{X_1, \cdots, X_n\}$ e F_n a distribuição acumulada de M_n. Mostre as seguintes convergências fracas,
$$F_n(a_n x + b_n) \Rightarrow F(x), \forall x$$
ponto de continuidade de F,
   i. $G(x) = 1 - e^{-\alpha x}, \forall x$, $F(x) = e^{-e^{-\alpha x}}, \forall x$, $a_n = 1$, $b_n = \alpha^{-1}\log(n)$, $\alpha > 0$.
   ii. $G(x) = \begin{cases} 0 & \text{se } x \leq 1 \\ 1 - x^{-\alpha} & \text{se } x > 1 \end{cases}$, $F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{se } x > 0 \end{cases}$, $a_n = n^{1/\alpha}$, $b_n = 0$, $\alpha > 0$.
   iii. $G(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - (1-x)^\alpha & \text{se } 0 < x \leq 1 \\ 1 & \text{se } x \geq 1 \end{cases}$, $F(x) = \begin{cases} e^{-(-x)^\alpha} & \text{se } x \leq 0 \\ 1 & \text{se } x > 0 \end{cases}$, $a_n = n^{-1/\alpha}$, $b_n = 1$, $\alpha > 0$.

3. Dadas $X_1, X_2 \cdots, X_n$ i.i.d tais que $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$. Denote por F_n a distribuição acumulada de $\frac{S_n}{n}$ onde $S_n = \sum_{i=1}^n X_i$, denote ainda por Δ a distribuição degenerada em 0.
É valida a seguinte relação?
$$F_n \Rightarrow \Delta \text{ se, e somente se } P\Big(|\frac{S_n}{n}| > \epsilon\Big) \to 0, \forall \epsilon > 0, \text{ quando } n \to \infty.$$

4. Demonstre os seguintes resultados,
   i. Se $F_n \Rightarrow F$, $\{b_n\}$ é não limitada, então $F_n(a_n x + b_n)$ não converge fracamente, para arbitrária sequencia $\{a_n\}$.
   ii. Se $F_n(x) \Rightarrow F(x)$ e $F_n(a_n x + b_n) \Rightarrow G(x)$ onde F e G são não degeneradas, então $0 < \inf\{a_n\} \leq \sup\{a_n\} < \infty$, $\sup\{|b_n|\} < \infty$.

5. <u>Consistência</u>. Defina (i) consistência em probabilidade e (ii) consistência uniforme, apresente exemplos para tais conceitos. Prove:

*University of Campinas, Brazil. E-mail: veronica@ime.unicamp.br

i. Dada $\{Z_n\}_n$ uma sequência de variáveis e Z_0 um valor constante. Se $Z_n \to^P Z_0$, quando $n \to \infty$, g é uma função contínua em Z_0, então $g(Z) \to^P g(Z_0)$, quando $n \to \infty$. Ou seja a consistência em probabilidade é preservada por continuidade.

ii. Seja X uma variáverl aleatória, não negativa, $f : \mathbb{R} \to \mathbb{R}$, crescente tal que $\mathbb{E}(f(X)) < \infty$. Então, $\forall a \in \mathbb{R}$, $f(a)P(X \geq a) \leq \mathbb{E}(f(X))$.

   a. Identifique a desigualdade em questão se: $f(x) = |x|^k, a > 0$.
   b. Identifique a função f e a variável sobre a qual é considerada a desigualdade de modo a obter a desigualdade de Chebyshev: $P(|X - \mathbb{E}(X)| \geq a) \leq \frac{Var(X)}{a^2}$.
   c. Identifique a função f que aplicada na variável X de média μ e variância σ^2, sobre a qual é considerada a desigualdade, garante a relação: $P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$.

6. Seja $\mathbb{P} = \left\{p = (p_1, \cdots, p_k) : 0 \leq p_j \leq 1, 1 \leq j \leq k, \sum_{j=1}^{k} p_j = 1\right\}$. Suponha que $p_j = Prob(X_1 = x_j)$ onde x_j é a categoria j entre as possíveis k categorias. Considere uma amostra de tamanho n, e seja N_j o total de objetos observados na categoria j, assim $\hat{p}_j = \frac{N_j}{n}, j = 1, \cdots, k$. Dada uma aplicação $q : \mathbb{P} \to R^s$, contínua. Então, $\hat{q}_n = q(\hat{p}_n)$ é uniformemente consistente para $q(p)$, onde $\hat{p}_n = (\hat{p}_1, \cdots, \hat{p}_k)$.

7. Considere $X(\omega) = 0\ \forall \omega \in [0, 1]$, defina $X_n(\omega) = n^2$ se $\omega \in [0, \frac{1}{n}]$ e $X_n(\omega) = 0$ quando $\omega \in (\frac{1}{n}, 1]$. (i) $X_n(\omega) \to X(\omega)$ quando $n \to \infty$?. (ii) $\mathbb{E}(X_n) \to \mathbb{E}(X)$, quando $n \to \infty$?. Comente os achados.

8. Sejam X e $\{X_n\}$ vetores aleatórios no memso espaço, $X_n \Rightarrow X$ (convergência fraca-notação alternativa), então se f é contínua e limitada, $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$.

9. Seja $\{X_n\}$ uma sequência de vetores ateatórios em $\mathbb{R}^k$. (i) Se $X_n \Rightarrow X$, então $\{X_n\}$ é uniformemente tight. (ii) Se $\{X_n\}$ é uniformemente tight, então existe uma subsequência $\{n_j\}$ de $\{n\}$ tal que $X_{n_j} \Rightarrow X$, quando $j \to \infty$, para algum X vetor limite.

10. Dadas as sequências $\{X_n\}$, $\{Y_n\}$ e $\{R_n\}$ dizemos que (i) $X_n = o_p(1)$ se $X_n \to^P 0$; (ii) $X_n = O_p(1)$ se $\{X_n\}$ é uniformemente tight; (iii) $X_n = o_p(R_n)$ se $X_n = Y_n R_n$ e $Y_n \to^P 0$; (iv) $X_n = O_p(R_n)$ se $X_n = Y_n R_n$ e $Y_n = O_p(1)$. Prove:

   i. $o_p(1) + o_p(1) = o_p(1)$;
   ii. $o_p(1) + O_p(1) = O_p(1)$;
   iii. $O_p(1) o_p(1) = o_p(1)$;
   iv. $(1 + o_p(1))^{-1} = O_p(1)$;
   v. $o_p(R_n) = R_n o_p(1)$;
   vi. $O_p(R_n) = R_n O_p(1)$;
   vii. $o_p(O_p(1)) = o_p(1)$.

11. Dada uma sequência de variáveis aleatórias $\{Y_n\}$ e uma constante c.

   i. Demonstre que se $\mathbb{E}(Y_n - c)^2 \to 0$, quando $n \to \infty$, então $Y_n \to^P c$, quando $n \to \infty$.
   ii. Exemplifique o fato seguinte: $Y_n \to^P c$, não implica (i) $\mathbb{E}(Y_n)^2 \to c$ ou (ii) $\mathbb{E}(Y_n - c)^2 \to 0$.
   iii Prove que se existe um valor $M > 0 : P(|Y_n - c| < M) = 1$ para todo n e $Y_n \to^P c$, então: (i) $\mathbb{E}(Y_n)^2 \to c$ e (ii) $\mathbb{E}(Y_n - c)^2 \to 0$.

12. Prove que $n^k = o(e^{an^b})$, para valores a e b positivos.

13. Prove que $\left(1 + \frac{c}{n} + o(\frac{1}{n})\right)^n \to e^c$ quando $n \to \infty$.

14. Prove que se $a_n \asymp b_n$ então $ca_n \asymp b_n\ \forall c \neq 0$.

15. Seja $a_n = \frac{1}{n} + R_n$ onde $R_n = \frac{b}{n\sqrt{n}} + \frac{c}{n^2} + \frac{d}{n^2\sqrt{n}}$. Prove:

i. $R_n \sim \frac{1}{n^2}$ se e somente se $b = 0$ e $c = 1$;

ii. $R_n \asymp \frac{1}{n^2}$ se e somente se $b = 0$ e $c \neq 0$;

iii. $R_n = O(\frac{1}{n^2})$ se e somente se $b = 0$;

iv. $R_n = o(\frac{1}{n^2})$ se e somente se $b = c = 0$.

1 M-estimadores

16. Famílias de Locação- Escala.

    a. Mostre que as seguintes distribuições são famílias de *Locação*: (i)N(μ, σ^2), σ^2 conhecido; (ii) Cauchy(μ,1) onde a densidade é $f(x;\mu) = \frac{1}{\pi}\frac{1}{(1+(x-\mu)^2)}$; (iii) Unif($\theta - \frac{1}{2}$, $\theta + \frac{1}{2}$).

    b. Mostre que as seguintes distribuições são famílias de *Escala*: (i) N(0,σ^2); (ii) Exp(λ); (iii) Unif(θ, 2θ).

    c. Mostre que as seguintes distribuições são famílias de *Locação-Escala*: (i) N(μ, σ^2); (ii) E2(μ, σ) onde a densidade é dada por $f(x;\mu,\sigma) = \frac{1}{2\sigma}e^{-|\frac{x-\mu}{\sigma}|}$; (iii) Cauchy($\mu$,$\sigma^2$) onde a densidade é $f(x;\mu,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\frac{(x-\mu)^2}{\sigma^2}}$.

17. Considere o modelo $x_i = \mu + \sigma u_i$ onde $u_i, i = 1, \cdots, n$ são i.i.d. com distribuição F, e F é dada pela seguinte forma: $F = (1-\epsilon)\Phi + \epsilon F^*$, sendo Φ a distribuição acumulada da N(0,1).

    a. Fixe diferentes valores de ϵ e fixe a F^* na Exponencial (λ), variando a taxa λ. Simule valores de x_i. Se o objetivo é estimar μ compare o desempenho de dois estimadores: $\overline{x}$ e a *mediana*$\{x_1, \cdots, x_n\}$.

    b. Fixe diferentes valores de ϵ e fixe a F^* na Uniforme, variando o intervalo. Simule valores de x_i. Se o objetivo é estimar μ compare o desempenho de dois estimadores: $\overline{x}$ e a *mediana*$\{x_1, \cdots, x_n\}$.

18. Considere as seguintes bases de dados (considere elas marginalmente).

    Table 1: Observações do índice Dow Jones. Período: 16/08/2001-29/09/2001

10.39	10.24	10.32	10.17	10.28	10.23	10.42	10.38	10.22	10.1
9.92	9.95	9.99	10.03	9.84	9.61	8.92	8.90	8.76	8.38

    Assuma o seguinte modelo $x_i = \mu + \sigma u_i, i = 1, \cdots, n$. Compute o M-estimador de locação, de Huber em cada caso apresentado nas tabelas, explore os seguintes valores de k, $k = 1, 1.5, 2, 2.5$. Determine o valor do Median Absolute Deviation MAD para cada situação e inspecione seus resultados com base no boxplot dos dados.

19. No contexto do exercício anterior compare o M-estimador de locação de Huber com o estimador de máxima verossimilhança e com a mediana.

Table 2: Observações da cidade de New York, ano 1973.

Ozônio	Radiação Solar	Vento	Temperatura
91	189	4.6	93
47	95	7.4	87
32	92	15.5	84
20	252	10.9	80
23	220	10.3	78
21	230	10.9	75
24	259	9.7	73
44	236	14.9	81
21	259	15.5	76
28	238	6.3	77
9	24	10.9	71
13	112	11.5	71
46	237	6.9	78
18	224	13.8	67
13	27	10.3	76
24	238	10.3	68
16	201	8.0	82
13	238	12.6	64
23	14	9.2	71
36	139	10.3	81
7	49	10.3	69

Table 3: Medições, em minutos de espera até a próxima erupção do gêiser Old Faithful em Yellowstone National Park, Wyoming, USA.

79	51	20	78	69	74	83	55	76	78	79	73	77	66	80	74	52	48	80	59	90	80

2 Para Entregar

20. Considere cada uma das distribuições listadas a seguir.

   a. Postule uma situação (G_n) tal que $G_n \Rightarrow F$. Onde F é:

      i. Gumbel: $F(x) = e^{-e^{-\alpha x}}, \alpha > 0, \ -\infty < x < \infty$;

      ii. Fréchet: $F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{se } x > 0 \end{cases}, \ \alpha > 0.$

      iii. Weibull: $F(x) = \begin{cases} e^{-(-x)^{\alpha}} & \text{se } x \leq 0 \\ 1 & \text{se } x > 0 \end{cases}, \ \alpha > 0.$

   b. Prove que a classe de distribuições extremais é dada pelos tipos anteriormente listados i, ii e iii.

   c. Apresente as formas não padrão de i, ii e iii.

   d. Existe alguma forma analítica que permite unificar i., ii. e iii. numa única expressão ?

   e. Apresente uma aplicação real usando uma dessas famílias.

21. Determine a função característica das seguintes distribuições:

   i. Distribuição de Laplace $f(x) = \frac{e^{-|x|}}{2}, \ -\infty < x < \infty$.

   ii. Distribuição de Cauchy $f(x) = \frac{1}{\pi(1+x^2)}, \ -\infty < x < \infty$.

iii. Distribuição Normal $f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \ -\infty < x < \infty.$