

Trabalho No.1

CE313 - ESTATÍSTICA NÃO-PARAMÉTRICA

Redigir de maneira individual e entregar na área correspondente no sistema **Microsoft Teams** um relatório eletrônico com as respostas até o dia **14 de abril de 2025**.

1. Seja $X \sim \text{Bernoulli}(1/2)$ e considere todas as possíveis amostras aleatórias de tamanho $n = 4$. Calcule $\bar{X}_n$ e S_n^2 para cada uma das dezesseis amostras. Encontre a função de probabilidade de $\bar{X}_n$ e S_n^2 .
2. Considere a função distância $\Delta(F, G)$,

$$\Delta(F, G) = \int_{-\infty}^{\infty} (F(x) - G(x))^2 \frac{F'(x) + G'(x)}{2} dx,$$

denso F e G distribuições absolutamente contínuas. Sejam $X_1, \dots, X_m$ uma amostra aleatória de F e $Y_1, \dots, Y_n$ uma amostra aleatória de G , independentes. Sabemos que a U -estatística

$$U(X, Y) = \frac{1}{\binom{m}{2} \binom{n}{2}} \sum_{i_1 < i_2} \sum_{k_1 < k_2} \varphi(X_{i_1}, X_{i_2}, Y_{k_1}, Y_{k_2}),$$

onde

$$\varphi(X_1, X_2, Y_1, Y_2) = \begin{cases} 1, & \text{se } \max(X_1, X_2) < \min(Y_1, Y_2) \text{ ou se} \\ & \max(Y_1, Y_2) < \min(X_1, X_2) \\ 0, & \text{caso contrário} \end{cases}$$

permite encontrar o estimador não viciado de mínima variância de $\Delta(F, G)$ como

$$\hat{\Delta}(F, G) = \frac{1}{2} U(X, Y) - \frac{1}{6}.$$

- (a) Escreva uma função R para encontrar estimativas de $\hat{\Delta}(F, G)$, considerando valores diversos de tamanhos de cada amostra.
- (b) Gere amostras de tamanho 50, 100 e 150 das distribuições (i) Normal padrão, (ii) Cauchy padrão e (iii) t -Student(4). Considere estas como possíveis distribuições de referência F .

- (c) Calcule as estimativas de $\hat{\Delta}(F, G)$ para cada amostra das distribuições de referência F obtidas no item (b) e a amostra da distribuição desconhecida G , obtida no arquivo de dados anexo.
 - (d) Segundo seus cálculos, qual é a distribuição desconhecida G ?
3. Sejam $X_1, X_2, \dots, X_m$ e $Y_1, Y_2, \dots, Y_n$ amostras independentes de duas distribuições absolutamente contínuas. Encontre o estimador não viado de mínima variância de: (a) $E(XY)$ e (b) $\text{Var}(X + Y)$.
 4. Podemos considerar a densidade kernel estimada como a função de densidade de uma amostra, como obter então novas amostras?

Como o estimador kernel da função de densidade

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right),$$

é uma mistura de n componentes de kernel, cada um pode ser amostrado independentemente. A única parte que pode exigir tratamento especial é a amostragem da densidade K , embora para a maioria dos kernels implementados R podem ser encontradas funções de mostragem específicas.

Consideremos os dados em:

```
phipsi=read.csv("http://leg.ufpr.br/~lucambio/Nonparam/phipsi.csv",
sep = ",", header = T)
```

Execute o seguinte algoritmo para gerar amostras de tamanho N , a partir da densidade kernel estimada dos dados em `phipsi`:

- (i) Escolha $i \in \{1, \dots, n\}$ aleatoriamente.
- (ii) Obtenha um amostra de tamanho 1 da densidade K , caso utilize-se a densidade gaussiana, a média é X_i e o desvio padrão é h_n .
- (iii) Repita os passos anteriores N vezes

Obtenha uma amostra de tamanho $N = 600$ da variável ψ (`phipsi$psi`) utilizando o algoritmo acima e compare a densidade estimada da nova amostra com a densidade estimada de ψ . Esta comparação pode ser graficamente.