Johann Marques Viana Freitas
Econometria I: Prof. Nathalie Gimenes
14 de maio de 2023

Lista Prática

Question 1.

1.1. Funções de Verossimilhança e Estimadores

1) A mistura de normais é caracterizada da seguinte forma:

$$e_i \sim \begin{cases} \mathrm{N}(\mu_1, \sigma_1^2), \text{com probabilidade } 1/2, \\ \mathrm{N}(\mu_2, \sigma_2^2), \text{com probabilidade } 1/2. \end{cases} \tag{1}$$

Para computar a densidade, podemos modelar $f(e_i) \sim \text{Bern}(1/2)$ e calcular $\mathbb{E}_{\text{Bern}(1/2)}\left[f(e_i)\right]$:

$$\mathbb{E}_{\text{Bern}(1/2)}\left[f(e_i)\right] = \frac{1}{2} \times \left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_1\right)^2\right) + \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_2\right)^2\right)\right] \tag{2}$$

Podemos escrever a função de verossimilhança amostral como

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= \prod_{i=1}^{n} \frac{1}{2} \times \left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_1\right)^2\right) + \right. \\ &\left. \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_2\right)^2\right)\right] \\ &= \frac{1}{2^n}\prod_{i=1}^{n} \times \left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_1\right)^2\right) + \right. \\ &\left. \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_2\right)^2\right)\right]. \end{aligned} \tag{3}$$

Obtendo a log-verossimilhança:

$$\begin{aligned} l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= -n\log(2) + \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_1\right)^2\right) + \right. \\ &\left. \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}\left(y_i - \beta_0 - \beta_1 x_i - \mu_2\right)^2\right)\right). \end{aligned} \tag{4}$$

O Estimador de Máxima Verossimilhança é dado por

$$\hat{\boldsymbol{\beta}} = \underset{\beta_0, \beta_1}{\text{argmax}}\ l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2). \tag{5}$$

2) Função de verossimilhança amostral para u, com $u_i \sim \Gamma(a, b)$, definida para $u_i > 0$:

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; a, b) &= \prod_{i=1}^{n} \frac{b^a}{\Gamma(a)}\left(y_i - \beta_1 - \beta_2 x_i\right)^{a-1}\exp\left(-b\left(y_i - \beta_1 - \beta_2 x_i\right)\right) \\ &= \left(\frac{b^a}{\Gamma(a)}\right)^n \prod_{i=1}^{n}\left(y_i - \beta_1 - \beta_2 x_i\right)^{a-1}\exp\left(-b\sum_{i=1}^{n}\left(y_i - \beta_1 - \beta_2 x_i\right)\right). \end{aligned} \tag{6}$$

Aplicando tranformação logarítmica

$$l(\boldsymbol{\beta}|\mathbf{y},\mathbf{x};a,b) = n\text{log}\left(\frac{b^a}{\Gamma(a)}\right) + (a-1)\sum_{i=1}^{n}\log\left(y_i-\beta_1-\beta_2x_i\right) - b\sum_{i=1}^{n}\left(y_i-\beta_1-\beta_2x_i\right). \tag{7}$$

O Estimador de Máxima Verossimilhança é dado por

$$\hat{\boldsymbol{\beta}} = \underset{\beta_0,\beta_1}{\text{argmax}}\ l(\boldsymbol{\beta}|\mathbf{y},\mathbf{x};a,b). \tag{8}$$

3) Sabemos que $v_i \sim \text{Cauchy}(0,1) \sim t(\nu=1)$, em que ν representa os graus de liberdade. Portanto, a função de verossimilhança amostral para $\mathbf{v}$, com $v_i \sim \text{Cauchy}(0,1)$ pode ser escrita como:

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{y},\mathbf{x};\nu) &= \prod_{i=1}^{n}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(y_i-\beta_1-\beta_2x_i)^2}{\nu}\right) \\ &= \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right]^n\prod_{i=1}^{n}\left(1+\frac{(y_i-\beta_1-\beta_2x_i)^2}{\nu}\right). \end{aligned} \tag{9}$$

A log-verossimilhança, por sua vez,

$$l(\boldsymbol{\beta}|\mathbf{y},\mathbf{x};\nu) = n\left[\log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\sqrt{\nu\pi}\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right)\right] + \sum_{i=1}^{n}\log\left(1+\frac{(y_i-\beta_1-\beta_2x_i)^2}{\nu}\right). \tag{10}$$

O Estimador de Máxima Verossimilhança é dado por

$$\hat{\boldsymbol{\beta}} = \underset{\beta_0,\beta_1}{\text{argmax}}\ l(\boldsymbol{\beta}|\mathbf{y},\mathbf{x};\nu). \tag{11}$$

1.2. Distribuições Assintóticas

Vide [1], Teorema 14.1.M2, *se são satisfeitas condições de regularidade*,

$$\hat{\boldsymbol{\theta}} \xrightarrow{d} \text{N}\left(\boldsymbol{\theta}_0, \{\mathbf{I}(\boldsymbol{\theta_0})\}^{-1}\right), \tag{12}$$

em que $\hat{\boldsymbol{\theta}}$ é o estimador de Máxima Verossimilhança de $\boldsymbol{\theta}_0$, e $\mathbf{I}(\boldsymbol{\theta_0}) = -\mathbb{E}\left[\frac{\partial^2\text{log}(L)}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\right]$.

Condições de regularidade

R.1. As primeiras três derivadas de $\log(f(y_i|\boldsymbol{\theta}))$ em relação a $\boldsymbol{\theta}$ são contínuas e finitas para quase todo y_i, e para todo $\boldsymbol{\theta}$.

R.2. As condições necessárias para obter as esperanças da primeira e da segunda derivada de $\log(f(y_i|\boldsymbol{\theta}))$ são satisfeitas.

R.3. Para todos os valores de $\boldsymbol{\theta}$, $|\partial^3\text{log}(f(y_i|\boldsymbol{\theta}))/\partial\theta_j\partial\theta_k\partial_l|$ é menor que uma função com esperança finita.

Question 2.

2.1. Simulações

O experimento de simulação é conduzido em R [2][1]. Primeiramente, definimos parâmetros gerais para a simulação, como o tamanho da amostra e os valores verdadeiros para $\boldsymbol{\beta}$.

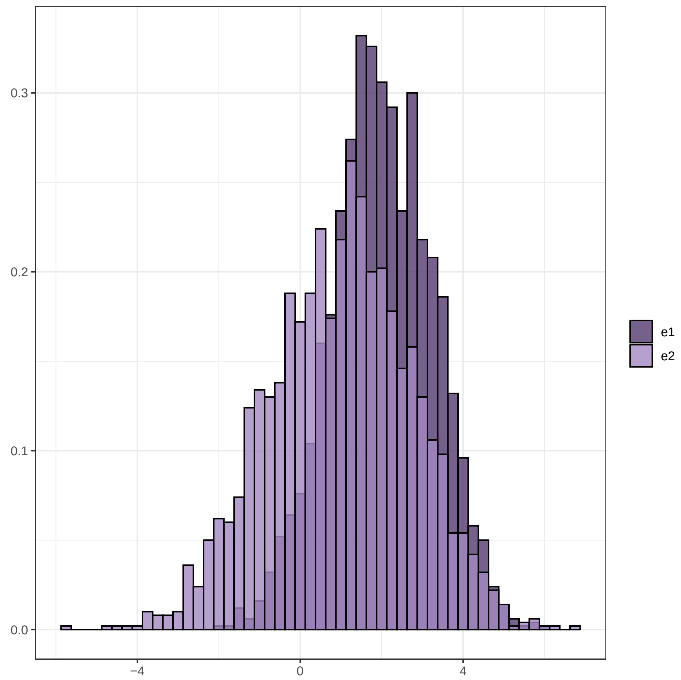
```
# Simulating data generating processes

n = 2000 # Sample size

# Setting random seed
set.seed(1981)

# Exercise 2
# Generating x
x = rnorm(n = n, mean = 5, sd = sqrt(1))

# Defining betas

beta = c(2.35, 0.75)
```

Em seguida, simulamos os três processos de geração de dados. Para obter a mistura de normais, é utilizada uma variável auxiliar $P \sim \text{Bernoulli}(1/2)$ e definimos $e_1 \sim \text{N}(\mu_1, \sigma_1^2)$, $e_2 \sim \text{N}(\mu_2, \sigma_2^2)$ e $e = Pe_1 + (1-P)e_2$.

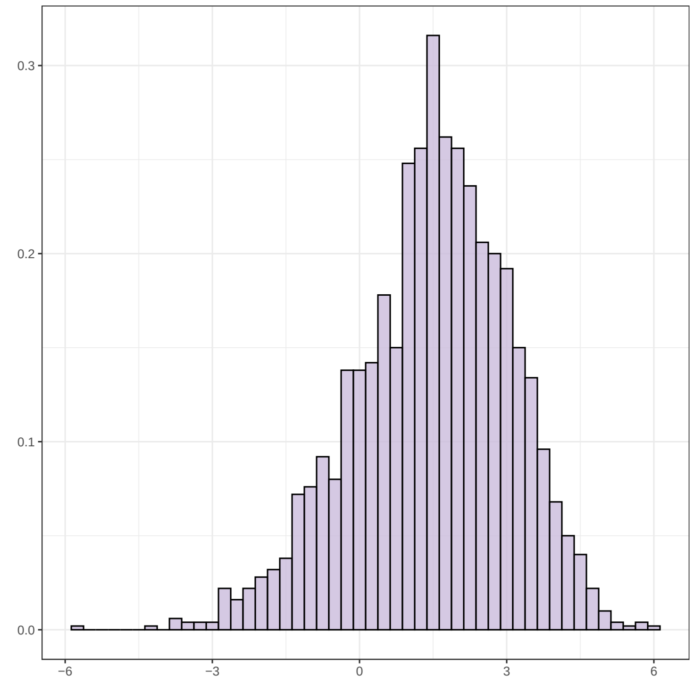
```
# Simulating model 1
# Since \mu_1, \mu_2, \sigma_1 and \sigma_2 were not specified,
# I'm considering the following
mu_1 = 2
mu_2 = 1
sigma_1 = sqrt(1.5)
sigma_2 = sqrt(3)

e1 = rnorm(n = n, mean = mu_1, sd = sigma_1)
e2 = rnorm(n = n, mean = mu_2, sd = sigma_2)

# Defining an auxiliary random variable P ~ Bernoulli(1/2)

P = rbinom(n = n, size = 1, prob = 1/2)

e = P * e1 + (1-P) * e2

y_1 = beta[1] + beta[2] * x + e
```

```
tibble(e1, e2) %>%
  pivot_longer(cols = c('e1', 'e2')) %>%
  ggplot(aes(x = value, group = name)) +
    geom_histogram(aes(y = after_stat(density), fill = name),
                   alpha = .8, position = 'identity', binwidth = 0.25, color = 'black') +
  labs(fill = '', x = '', y = '') + scale_fill_manual(values = purple_rain_colors[c(1,2)])
```

[1]Soluções providas por [4] e [3] foram empregadas na análise e reporte deste exercício.



(A) e_1 e e_2



(B) Mistura

Figura 1. Distribuições de e_1, e_2, e

```
ggplot() +
  geom_histogram(aes(y = after_stat(density), x = e),
                 alpha = .8, position = 'identity', binwidth = 0.25, color = 'black',
                 fill = purple_rain_colors[3]) +
  labs(fill = '', x = '', y = '')
```

```
# Simulating model 2
# Since a and b were not specified, I'm considering the following

a = 1.75
b = .8

u = rgamma(n = n, shape = a, rate = b)

y_2 = beta[1] + beta[2] * x + u
```

```
# Simulating model 3

v = rt(n = n, df = 1)

y_3 = beta[1] + beta[2] * x + v
```

2.2. Estatísticas descritivas

```
print(xtable(map_df(list(x, e1, e2, y_1, u, y_2, v, y_3), ~summary(.x)) %>%
             {cbind(data.frame('Variable' = c("$\\bar{x}$", "$\\bar{e}_1$",
                                              "$\\bar{e}_2$", "$\\bar{y}_1$",
```

```
                                          "$\\bar{u}$", "$\\bar{y}_2$",
                                          "$\\bar{v}$", "$\\bar{y}_3$")),
              .)},
      ,
  caption = 'Estatísticas descritivas'),
  sanitize.text.function=function(x){x}, include.rownames = FALSE)
```

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\bar{x}$	1.81	4.27	4.97	4.98	5.69	8.46
$\bar{e}_1$	-1.95	1.19	2.00	2.01	2.87	5.88
$\bar{e}_2$	-5.83	-0.19	1.07	0.99	2.18	6.76
$\bar{y}_1$	1.19	6.41	7.63	7.56	8.76	12.92
$\bar{u}$	0.03	0.99	1.76	2.18	2.94	13.91
$\bar{y}_2$	4.68	7.00	7.96	8.27	9.18	21.09
$\bar{v}$	-2163.94	-1.01	-0.01	1.48	0.98	5489.95
$\bar{y}_3$	-2158.61	4.84	6.05	7.57	7.33	5495.47

Tabela 1. Estatísticas descritivas

Referências

[1] WH Greene. Econometric analysis, 7th. saddle river, 2011.

[2] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[3] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.

[4] Yihui Xie. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. ISBN 978-1466561595.