

Lista Prática

Question 1.

1.1. FUNÇÕES DE VEROSSIMILHANÇA E ESTIMADORES

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

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

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

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

Podemos escrever a função de verossimilhança amostral como

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

Obtendo a log-verossimilhança:

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

O Estimador de Máxima Verossimilhança é dado por

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

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

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

Aplicando transformação logarítmica

$$(7) \quad l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; a, b) = n \log \left(\frac{b^a}{\Gamma(a)} \right) + (a-1) \sum_{i=1}^n \log(y_i - \beta_1 - \beta_2 x_i) - b \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i).$$

O Estimador de Máxima Verossimilhança é dado por

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

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:

$$(9) \quad \begin{aligned} L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; \nu) &= \prod_{i=1}^n \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\nu}\right) \\ &= \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \right]^n \prod_{i=1}^n \left(1 + \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\nu}\right). \end{aligned}$$

A log-verossimilhança, por sua vez,

$$(10) \quad l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}; \nu) = n \left[\log \left(\Gamma\left(\frac{\nu+1}{2}\right) \right) - \log(\sqrt{\nu\pi}) - \log \left(\Gamma\left(\frac{\nu}{2}\right) \right) \right] + \sum_{i=1}^n \log \left(1 + \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\nu} \right).$$

O Estimador de Máxima Verossimilhança é dado por

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

1.2. DISTRIBUIÇÕES ASSINTÓTICAS

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

$$(12) \quad \hat{\boldsymbol{\theta}} \xrightarrow{d} N(\boldsymbol{\theta}_0, \{\mathbf{I}(\boldsymbol{\theta}_0)\}^{-1}),$$

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

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 $\boldsymbol{\theta}$, $|\partial^3 \log(f(y_i|\boldsymbol{\theta}))/\partial \theta_j \partial \theta_k \partial \theta_l|$ é menor que uma função com esperança finita.

Question 2.

2.1. SIMULAÇÕES

O experimento de simulação é conduzido em R [3]¹. Primeiramente, definimos parâmetros gerais para a simulação, como o tamanho da amostra e os valores verdadeiros para β .

```
# 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 $e_1 \sim N(\mu_1, \sigma_1^2)$, $e_2 \sim N(\mu_2, \sigma_2^2)$ 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)])
```

¹Soluções providas por [5] e [4] foram empregadas na análise e reporte deste exercício.


```

"\\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.	Var.
\bar{x}	1.81	4.27	4.97	4.98	5.69	8.46	1.00
\bar{e}_1	-1.95	1.19	2.00	2.01	2.87	5.88	1.53
\bar{e}_2	-5.83	-0.19	1.07	0.99	2.18	6.76	3.03
\bar{e}	-5.83	0.47	1.58	1.47	2.59	5.90	2.50
\bar{y}_1	1.19	6.41	7.63	7.56	8.76	12.92	3.03
\bar{u}	0.03	0.99	1.76	2.18	2.94	13.91	2.82
\bar{y}_2	4.68	7.00	7.96	8.27	9.18	21.09	3.34
\bar{v}	-2163.94	-1.01	-0.01	1.48	0.98	5489.95	17973.79
\bar{y}_3	-2158.61	4.84	6.05	7.57	7.33	5495.47	17972.39

TABELA 1. Estatísticas descritivas

Os momentos teóricos, por sua vez, são os seguintes:

$$(13) \quad \mathbb{E}[e_i] = p\mathbb{E}[e_{1i}] + (1-p)\mathbb{E}[e_{2i}] = 1.5$$

$$(14) \quad \mathbb{V}[e_i] = p(\sigma_1^2 + \mu_1^2) + (1-p)(\sigma_2^2 + \mu_2^2) - [p\mu_1 + (1-p)\mu_2]^2 = 2.5$$

$$(15) \quad \mathbb{E}[y_{1i}] = \beta_0 + \beta_1\mathbb{E}[x_i] + \mathbb{E}[e_i] = 7.6$$

$$(16) \quad \mathbb{V}[y_{1i}] = \mathbb{V}[\beta_1 x_i + e_i] = \beta_1^2 \mathbb{V}[x_i] + \mathbb{V}[e_i] = 3.0625$$

As estimativas observadas para média e variância amostral de y_1 são coerentes com a derivação teórica.

$$(17) \quad \mathbb{E}[u_i] = \frac{a}{b} = 2.1875$$

$$(18) \quad \mathbb{V}[u_i] = \frac{a}{b} = 2.734375$$

$$(19) \quad \mathbb{E}[y_{2i}] = \beta_0 + \beta_1\mathbb{E}[x_i] + \mathbb{E}[u_i] = 8.2875$$

$$(20) \quad \mathbb{V}[y_{2i}] = \mathbb{V}[\beta_1 x_i + u_i] = \beta_1^2 \mathbb{V}[x_i] + \mathbb{V}[u_i] = 3.296875$$

As estimativas observadas para média e variância amostral de y_2 não diferem substancialmente dos momentos teóricos.

A esperança e a variância para v_i não estão definidas, portanto, também não estão para y_{3i} . Como sabemos que a distribuição de v_i é simétrica, bem como a de x_i , sugere-se intuitivamente utilizar a mediana como uma medida de centralidade.

Question 3.

- 1)
- 2) Definindo a log-verossimilhança amostral, que será a função objetivo do problema de otimização, como função de β e σ^2 :

```

# Exercise 3.2
# I'm assuming y - Xb ~ N(0, sigma^2)
# For the remaining exercises, I'll be using OLS estimates as initial conditions

# Objective function

```

```
l_norm0 <- function(theta, y, x){
  -length(y)/2*log(2*pi*theta[3]) + sum(-1/(2*theta[3])*(y - theta[1] - theta[2]*x)^2)
}
```

Iremos utilizar o algoritmo de Nelder-Mead [2] para obter uma solução. Para os três modelos será utilizada como condição inicial a estimativa por Mínimos Quadrados Ordinários.

```
# Model 1

# Initial condition
init_norm = lm(y_1 ~ x) %>%
  {c(coef(.), sigma2 = summary(.)$sigma^2)} # Adding OLS estimate for \sigma^2

mod1_norm <- optim(par = init_norm,
  fn = function(theta){-l_norm0(theta = theta, y = y_1, x=x)})
```

```
print(mod1_norm$par)
## (Intercept)          x      sigma2
##   3.9487238   0.7250645   2.5030414
```

```
# Model 2

# Initial condition

init_gamma = lm(y_2 ~ x) %>%
  {c(coef(.), sigma2 = summary(.)$sigma^2)}

mod2_norm <- optim(par = init_gamma,
  fn = function(theta){-l_norm0(theta = theta, y = y_2, x=x)})
```

```
print(mod2_norm$par)
## (Intercept)          x      sigma2
##   4.644255   0.727031   2.815482
```

```
# Model 3

# Initial condition

init_t = lm(y_3 ~ x) %>%
  {c(coef(.), sigma2 = summary(.)$sigma^2)}

mod3_norm <- optim(par = init_t,
  fn = function(theta){-l_norm0(theta = theta, y = y_3, x=x)})
```

```
print(mod3_norm$par)
## (Intercept)          x      sigma2
##  10.3833465  -0.5662194 17957.6199485
```

Question 4.

```

# Exercise 4

# Model 1

# Objective function
l_normmix <- function(theta, mu1, mu2, sigma2_1, sigma2_2, y, x){
  -length(y)*log(2)+ sum(log(1/sqrt(2*pi*sigma2_1)*
    exp(-1/(2*sigma2_1)*
      (y - theta[1] - theta[2] * x - mu1)^2) +
    1/sqrt(2*pi*sigma2_2)*
    exp(-1/(2*sigma2_2) *
      (y - theta[1] - theta[2] * x - mu2)^2)))
}

# Initial condition previously defined #

mod1_true <- optim(par = init_norm[-3],
  fn = function(theta){-l_normmix(theta = theta, mu1 = 2, mu2 = 1, sigma2_1 = 1.5, s

print(mod1_true$par)
## (Intercept)          x
## 2.4391715 0.7283436

# Model 2

# Objective function
l_gamma <- function(theta, a, b, y, x){
  if(prod(y - theta[1] - theta[2]*x > 0)){ # Checking whether  $u_i > 0$  \forall  $i = 1, \dots$ 
    return(length(y)*log(b^a/gamma(a)) + (a-1)*sum(log(y - theta[1] - theta[2]*x)) - b*s
  }else{
    return(-Inf)
  }
}

# Initial condition previously defined #

# The objective function evaluated at the initial condition returns an error.
# So I'll be subtracting 1

mod2_true <- optim(par = init_gamma[-3]-1,
  fn = function(theta){-l_gamma(theta = theta, a = 1.75, b = .8, y = y_2, x = x)})

print(mod2_true$par)
## (Intercept)          x
## 2.343295 0.751715

# Model 3

# Objective function

```

```

l_t <- function(theta, y, x){
  length(y)*(log(gamma((theta[3]+1)/2)) - log(sqrt(theta[3]*pi)) - log(gamma(theta[3]/2)
    (theta[3]+1)/2*sum(log(1 + ((y - theta[1] - theta[2]*x)^2)/theta[3])))
}

init_t2 = c(coef(
  lm(y_3 ~ x)
), df = 0.0001) # Setting a pretty small degree of freedom as an initial guess

mod3_true <- optim(par = init_t2,
  fn = function(theta){-l_t(theta = theta, y = y_3, x=x)})

```

```

print(mod3_true$par)
## (Intercept)          x          df
##    2.3474134    0.7512420    0.9799201

```

```

plot_curves <- function(...){
  ggplot() +
    geom_point(aes(y =eval(parse(text = paste0('y_', ...))), x=x), alpha = .5, color = p
    geom_abline(slope = beta[2],
      intercept = beta[1],
      linetype = 'F1',
      color = 'grey',
      linewidth = 1) +
    geom_abline(slope = eval(parse(text = paste0('mod', ..., '_true$par[2]'))),
      intercept = eval(parse(text =paste0('mod', ..., '_true$par[1]'))),
      linetype = 'dashed',
      color = purple_rain_colors[1],
      linewidth = 1,
      alpha = .75) +
    geom_abline(slope = eval(parse(text = paste0('mod', ..., '_norm$par[2]'))),
      intercept = eval(parse(text = paste0('mod', ..., '_norm$par[1]'))),
      linetype = 'solid',
      color = purple_rain_colors[1],
      linewidth = 1,
      alpha = .75) +
    labs(x = "x", y = "y")
}

invoke_map(.f = plot_curves, .x = 1:3)

```

REFERÊNCIAS

- [1] WH Greene. Econometric analysis, 7th. saddle river, 2011.
- [2] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
- [4] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, 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.

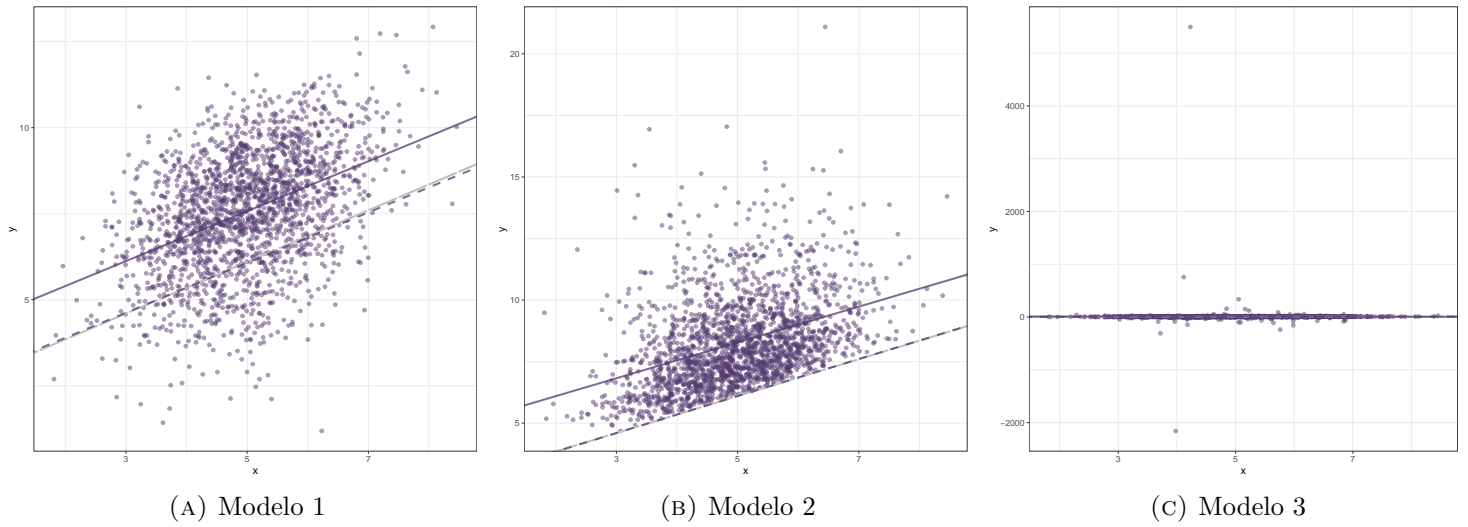


FIGURA 2. Comparativo

- [5] 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.