

SME0809 - Inferência Bayesiana - Distribuição Normal

Grupo 13 - Francisco Miranda - 4402962 - Heitor Carvalho - 11833351

Outubro 2021

```
library(tidyverse)
library(effectsize)
library(invgamma)
library(dados)

#remotes::install_github("cienciadedatos/dados")

sample <- dados::penguins %>%
  filter(especie == "Pinguim-de-barbicha") %>%
  select(comprimento_bico) %>%
  drop_na()
```

Caso 1: μ desconhecido e σ conhecido

Verossimilhança da distribuição

$$\mathcal{L}(y|\theta) = \prod_{i=1}^n p(y_i|\theta) = \prod_{i=1}^n e^{-1/2\sigma^2(y_i-\theta)^2}$$

A priori da distribuição

$$p(\theta) \propto e^{(-1/2\tau_0^2)(\theta-y_0)^2}$$

A priori não informativa de Jeffreys

A posteriori da distribuição

A posteriori é computada assumindo-se que: 1. Cada observação é independentemente distribuída 2. Cada observação tem a mesma variância

$$p(\theta|y) = p(\theta) p(y|\theta) = p(\theta) \prod_{i=1}^n p(y_i|\theta) = e^{(-1/2\tau_0^2)(\theta-\mu_0)^2} \prod_{i=1}^n e^{(-1/2\sigma^2)(y_i-\theta)^2} = e^{(\frac{-1}{2})(1/\tau_0^2(\theta-\mu_0)^2 + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i-\theta)^2)}$$

Desse modo, a distribuição a posteriori da média θ depende apenas da média amostral $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, sendo assim, \bar{y} é uma estatística suficiente.

Portanto, para n observações, a posteriori apresenta a seguinte distribuição:

$$p(\theta|y_1, \dots, y_n) = p(\theta | \bar{y}) = \mathcal{N}(\theta|y_n, \tau_n^2)$$

Sendo,

$$\mu_n = \frac{\tau_0^{-2}\mu_0 + n\sigma^{-2}\bar{y}}{\tau_0^{-2} + \sigma^{-2}}, \quad \text{e} \quad \tau_n^{-2} = \tau_0^{-2} + n\sigma^{-2}$$

Podemos reescrever $p(\theta|y)$ como:

$$p(\theta|y_n) \propto e^{(-1/2\tau_n^2)(\theta - \mu_n)^2}$$

Logo, para uma distribuição Normal com variância conhecida, a média a posteriori μ_n pode ser interpretada como a média ponderada da média a priori e o valor observado $y = y_1, \dots, y_n$, sendo os pesos proporcionais às precisões de cada um.

Caso 2: μ conhecido e σ desconhecido

Distribuições a priori

Seja Y_i uma amostra aleatória simples de uma distribuição $Y \sim N(\theta, \sigma^2)$, com θ conhecido.

Primeiramente, vamos encontrar a função de verossimilhança de σ^2 .

$$\mathcal{L}(y|\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-(y_i - \theta)^2 / 2\sigma^2} \propto (\sigma^2)^{-\frac{n}{2}} e^{-\left(\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right)}$$

Priori não informativa

Definimos a log-verossimilhança em nosso caso como sendo:

$$\log(\mathcal{L}(y|\sigma^2)) \propto -\frac{n}{2} \log(\sigma^{-2}) - \sigma^{-2} \sum_{i=1}^n (y_i - \theta)^2$$

A distribuição a *priori* de Jeffreys é dada por $\pi(\sigma^2) \propto \sqrt{J(\sigma^2)}$.

$$\begin{aligned} J(\sigma^2) &\propto E \left(-\frac{\partial^2}{\partial \theta^2} \log(L(\theta)) \right) = E \left(-\frac{\partial^2}{\partial \theta^2} \left(-\frac{n}{2} \log(\sigma^{-2}) - (\sigma^2)^{-1} \sum_{i=1}^n (y_i - \theta)^2 \right) \right) \\ &= E \left(-\frac{\partial}{\partial \theta} \left(-\frac{n}{2} (\sigma^2)^{-1} + (\sigma^2)^{-2} \sum_{i=1}^n (y_i - \theta)^2 \right) \right) = E \left(-\frac{n}{2\sigma^2} + 2(\sigma^2)^{-3} \sum_{i=1}^n (y_i - \theta)^2 \right) \\ &= -\frac{n}{2\sigma^2} + 2\sigma^{-4} \sum_{i=1}^n (E(y_i) - \theta)^2 = -\frac{n}{\sigma^2} + 2\sigma^{-4} \sum_{i=1}^n (\theta - \theta)^2 = -\frac{n}{\sigma^2} \propto \sigma^{-2} \end{aligned}$$

Assim, $\pi(\sigma) \propto \sqrt{\sigma^{-2}} = \sigma^{-1}$. Seu parâmetro Φ de escala que faz com que θ mude somente em locação pode ser obtido através do cálculo de

$$\phi \propto \int \pi(\sigma^2) d\sigma^2 = \int \frac{1}{\sigma^2} d\sigma^2 = \log |\sigma^2| + k \propto \log \sigma^2$$

ϕ é uma distribuição imprópria, pois $\int_0^{+\infty} \log(\sigma^2) d\sigma^2$ é divergente. Assim, a *priori* não favorece nenhuma escala em detrimento de outra.

Conjugadas Naturais

O suporte de nosso parâmetro de interesse $\sigma > 0$ permite-nos adotar três distribuições de probabilidade estudadas durante o curso:

1. Gama:

Se $X \sim \text{Gama}(\alpha, \beta)$ então

$$f_X(x|\alpha, \beta) = \frac{\beta^\alpha x^{(\alpha-1)} e^{-\beta x}}{\Gamma(\alpha)} \propto x^{(\alpha-1)} e^{-\beta x}, \quad \alpha > 0, \beta > 0, x > 0$$

2. Gama-Inversa:

Se $X \sim (\alpha, \beta)$ então

$$f_X(x|\alpha, \beta) = \frac{\beta^\alpha x^{-(\alpha+1)} e^{-\beta/x}}{\Gamma(\alpha)} \propto x^{-(\alpha+1)} e^{-\beta/x}, \quad \alpha > 0, \beta > 0, x > 0$$

3. Qui-Quadrado:

Se $X \sim \chi^2(\nu)$ então

$$f_X(x|\nu) = \frac{x^{(\nu/2)-1} e^{-x/2}}{2^{\nu/2} \Gamma(\nu/2)} \propto x^{(\nu/2)-1} e^{-x/2}, \quad \nu > 0, x > 0$$

Note que as duas primeiras estão relacionadas via uma transformação simples e a última é um caso particular delas. Desprezadas as constantes não informativas, as três distribuições são da forma x elevado a uma potência vezes a exponencial de x . Dessa forma, as três distribuições servem como conjugada natural da Normal, em nosso caso. Neste trabalho, optou-se por utilizar a distribuição Gama Inversa.

Fazendo $x = \sigma$, temos uma *priori* da forma:

$$\sigma^{-(\alpha+1)} e^{-\beta/\sigma^2} \Rightarrow \pi(\sigma) \sim \text{Gama-Inv}(\alpha, \beta)$$

Se quisermos torná-la não informativa, basta utilizarmos $\alpha \rightarrow 0, \beta \rightarrow 0$.

Distribuição a *posteriori*

$$\begin{aligned} \pi(\sigma|y) &\propto \mathcal{L}(y|\sigma^2) \pi(\sigma) = (\sigma^2)^{-\frac{n}{2}} e^{-\left(\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right)} \sigma^{-(\alpha-1)} e^{-\beta/\sigma^2} \\ &= (\sigma^2)^{-(\alpha + \frac{n}{2} + 1)} e^{-\frac{1}{\sigma^2} \left(\beta + \frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2\right)} \end{aligned}$$

Dessa forma,

$$\pi(\sigma^2|y) \sim \text{Gama-Inv}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2\right)$$

Exemplo: Comprimento do bico dos pinguins

Cleiton, Eduarda, Larissa e Robertinho estão estudando sobre os Pinguins-de-barbicha. Sabe-se que o comprimento do bico deles tem distribuição Normal com média 48.833 e desvio padrão desconhecido. Os quatro amigos decidem estimar este desvio padrão, cada um define sua *priori* da seguinte forma:

- **Cleiton** nunca viu um pinguim-de-barbicha na vida, nem em fotografia. Dessa forma, ele decide adotar uma *priori* não informativa $\text{Gama-Inv}(\alpha = 0.01, \beta = 0.01)$
- **Eduarda** sabe tudo sobre pinguins, mas nunca viu um pessoalmente. Ela opta por uma $\text{Gama-Inv}(\alpha = 0.5, \beta = 3)$
- **Larissa** adora ir ao zoológico visitar aos pinguins. Ela decide adotar uma $\text{Gama-Inv}(\alpha = 20, \beta = 20)$
- **Robertinho** é um biólogo com muita experiência, que consulta suas anotações sobre pinguins e decide adotar uma *priori* $\text{Gama-Inv}(\alpha = 35, \beta = 186)$

```
sample <- dados::pinguins %>%
  filter(especie == "Pinguim-de-barbicha") %>%
  select(comprimento_bico) %>%
  drop_na() %>% pull()

# gera a priori e a posteriori de uma normal com media conhecida e sigma desconhecido
SigmaNorm <- function(samp, theta = 48.833, alpha = 0.001, beta = 0.001){

  n <- length(samp)
  s <- sum(((samp - theta)/2)^2)

  l_sigma2 <- function(sigma2) sigma2^(-(n/2)) * exp(- 1/sigma2 * s)

  a.post <- alpha + n/2
  b.post <- beta + s

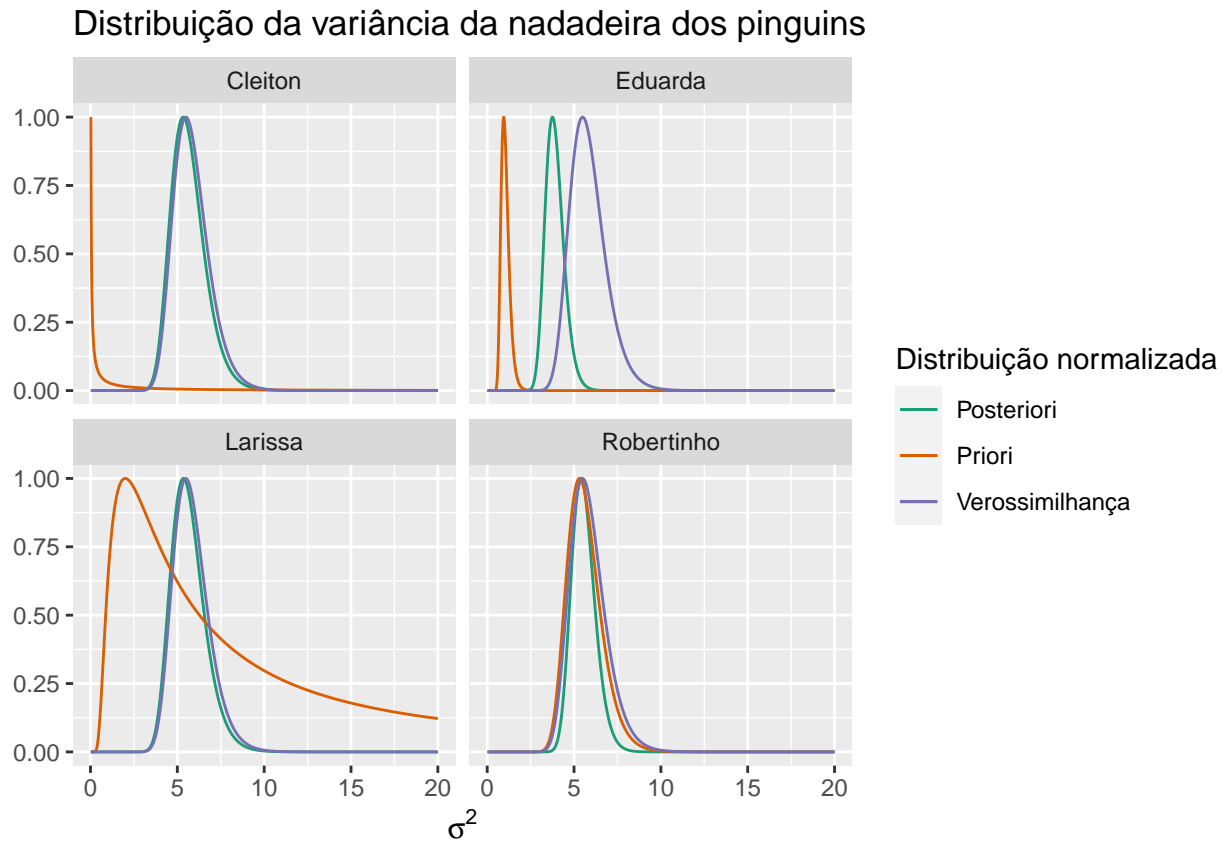
  sigma2 <- seq(0.02, 40, 0.02)

  tibble(sigma2 = sigma2,
    priori = normalize(dinvgamma(sigma2,alpha,beta)),
    post = normalize(dinvgamma(sigma2,a.post,b.post)),
    ver = normalize(l_sigma2(sigma2)),
    alpha1 = a.post,
    beta1 = b.post,
    alpha0 = alpha,
    beta0 = beta)
}
```

```
a <- SigmaNorm(sample, alpha = 0.01, beta = 0.01) %>% mutate(Priori = "Cleiton")
b <- SigmaNorm(sample,alpha = 20, beta = 20) %>% mutate(Priori = "Eduarda")
c <- SigmaNorm(sample,alpha = 0.5, beta = 3) %>% mutate(Priori = "Larissa")
d <- SigmaNorm(sample,alpha = 34, beta = 186) %>% mutate(Priori = "Robertinho")

rbind(a,b,c,d) %>%
  ggplot(aes(x = sigma2)) +
  geom_line(aes(y = post, color = "Posteriori")) +
  geom_line(aes(y = priori, color = "Priori")) +
  geom_line(aes(y = ver, color = "Verossimilhança")) +
  #geom_line(aes(y = 0.03* ver, colour = "Verossimilhança")) +
```

```
scale_colour_brewer(name = "Distribuição normalizada", type = "qual", palette = "Dark2")+
scale_x_continuous(name = expression(sigma^2), limits = c(0, 20))+
theme(axis.title.y=element_blank()) +
ggtitle("Distribuição da variância da nadadeira dos pinguins") +
facet_wrap(~Priori)
```



```
tabDesc <- function(alpha, beta){
  tibble( alpha = alpha,
    beta = beta,
    media = beta/(alpha-1),
    var = beta^2/((alpha-1)^2*(alpha-2)),
    moda = beta/(alpha+1),
    IC2.5 = qinvgamma(0.025, alpha, beta),
    IC97.5 = qinvgamma(0.975, alpha, beta))
}

cbind( Priori = c("Cleiton","Eduarda", "Larissa", "Robertinho"),
  rbind(tabDesc(0.1 , 0.1),
    tabDesc(20,20),
    tabDesc(0.5 , 3),
    tabDesc(34 , 186)
  )) %>%
knitr::kable(digits = 2, caption = "Resumo a priori")
```

Table 1: Resumo a priori

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	0.1	0.1	-0.11	-0.01	0.09	0.10	1.726606e+15
Eduarda	20.0	20.0	1.05	0.06	0.95	0.67	1.640000e+00
Larissa	0.5	3.0	-6.00	-24.00	2.00	1.19	6.109550e+03
Robertinho	34.0	186.0	5.64	0.99	5.31	4.01	7.900000e+00

```
cbind( Priori = c("Cleiton", "Eduarda", "Larissa", "Robertinho"),
      rbind(tabDesc(a$alpha[1], a$beta[1]),
            tabDesc(b$alpha[1], b$beta[1]),
            tabDesc(c$alpha[1], c$beta[1]),
            tabDesc(d$alpha[1], d$beta[1])
      )) %>%
knitr::kable(digits = 2, caption = "Resumo a posteriori")
```

Table 2: Resumo a posteriori

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	34.01	186.78	5.66	1.00	5.34	4.03	7.93
Eduarda	54.00	206.77	3.90	0.29	3.76	2.98	5.10
Larissa	34.50	189.77	5.66	0.99	5.35	4.04	7.92
Robertinho	68.00	372.77	5.56	0.47	5.40	4.38	7.06