Inferência bayesiana

Fernando Mayer

2019-09-05

Análise Bayesiana de dados

A análise de dados sob a perspectiva Bayesiana resume-se à 3 etapas:

1. Definir a distribuição *à priori*
    - Informativa
    - Não informativa
2. Definir a função de verossimilhança
3. Encontrar a distribuição posterior
    - Famílias conjugadas
    - Simulação estocástica

Todo processo de inferência será baseado na distribuição posterior

Distribuições posteriores

Famílias conjugadas

- As **famílias conjugadas** podem ser utilizadas quando o produto $\pi(\theta) \cdot f(\mathbf{x}|\theta)$ possui **forma fechada**, ou seja, resulta em uma distribuição conhecida
- Portanto, a posterior é obtida diretamente pelo teorema de Bayes
  - $\uparrow$ Resultado simples e conveniente
  - $\downarrow$ Restringe a escolha das distribuições *priori* e de verossimilhança

Definição: família conjugada

Seja $\mathcal{F}$ uma família de distribuições para a verossimilhança $f(\mathbf{x}|\theta)$, e $\mathcal{P}$ uma família de distribuição para a *priori* $\pi(\theta)$. Dizemos que $\mathcal{F}$ e $\mathcal{P}$ são **famílias conjugadas** de distribuições se a distribuição posterior $\pi(\theta|\mathbf{x})$ também for um membro de $\mathcal{P}$.

Distribuições posteriores

Famílias conjugadas

Exemplo: Beta-binomial com $\pi(\theta) \sim Beta(\alpha, \beta)$ e $f(\mathbf{x}|\theta) \sim Bin(n, \theta)$

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \pi(\theta) \cdot f(\mathbf{x}|\theta) \\
&\propto \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{x} \theta^{x}(1-\theta)^{n-x} \\
&\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}
\end{aligned}$$

Portanto, $\pi(\theta|\mathbf{x}) \sim Beta(\alpha^* = \alpha + x, \beta^* = \beta + n - x)$

Distribuições posteriores

Simulação estocástica

- A **simulação estocástica** é necessária quando o produto $\pi(\theta) \cdot f(\mathbf{x}|\theta)$ não possui forma fechada, e resulta em uma distribuição desconhecida
- Portanto é necessário realizar um processo de simulação para se construir a posterior
  - $\uparrow$ Aplicação não se restringe apenas às famílias conjugadas
  - $\downarrow$ O processo de simulação pode ser muito "caro" computacionalmente
- Alguns métodos de simulação:
  - Monte Carlo via Cadeias de Markov (MCMC)
  - Re-amostragem por importância
- Outros: *Integrated Nested Laplace Approximation* - INLA

Distribuições posteriores

Independente de como a posterior foi obtida:

- É um compromisso entre a *priori* (que carrega o **conhecimento prévio**), e a verossimilhança (que expressa o **conhecimento atual** adquirido com o experimento realizado)
- Representa todo o conhecimento existente sobre o problema, portanto a inferência Bayesiana é baseada nesta distribuição

> "A posterior de hoje é a *priori* de amanhã"

Inferência

Abordagem clássica

- θ é uma quantidade **desconhecida**, mas **fixa**
- Uma amostra aleatória $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ é obtida a partir de uma população indexada por θ
- Com base nos valores observados na amostra, o conhecimento sobre θ é obtido

Abordagem Bayesiana

- θ é uma quantidade **desconhecida**, e **aleatória**
- A variabilidade em θ é expressa pela *priori* $\pi(\theta)$
- Uma amostra aleatória $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ é obtida a partir de uma população indexada por θ
- A distribuição **a priori** é **atualizada** com essa informação da amostra, representada pela verossimilhança $f(\mathbf{x}|\theta)$

Estimação por intervalo

Abordagem clássica

- Dizemos que o intervalo aleatório (T_1, T_2), $T_1(\mathbf{X}) \leq T_2(\mathbf{X})$, é um **intervalo de confiança** para θ, com coeficiente de confiança $\gamma = 1 - \alpha$, $(0 < \alpha < 1)$, se

$$P[T_1 < \theta < T_2] = \gamma$$

- Notação:

$$\text{IC}(\theta, \gamma\%) = (T_1, T_2)$$

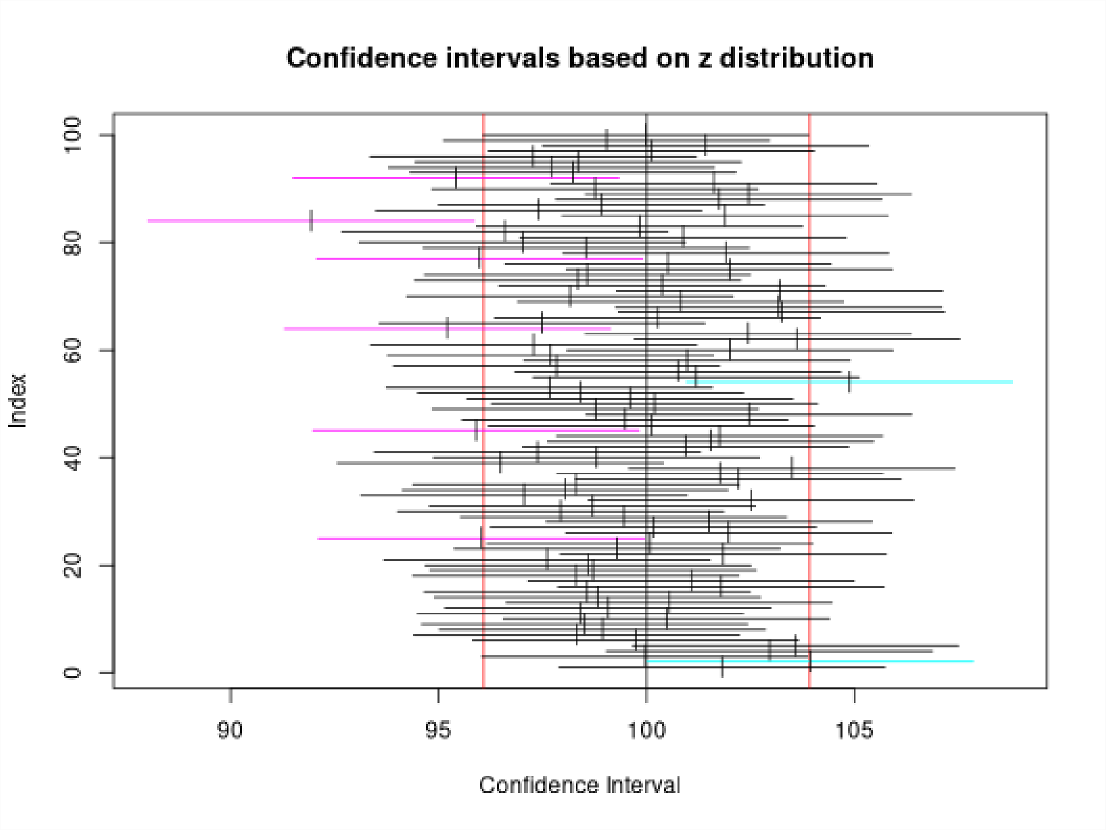
- O intervalo **contém** θ com probabilidade γ
- Uma vez que o parâmetro é fixo, o intervalo é aleatório
- $\theta \in (T_1, T_2)$ com probabilidade 0 ou 1

Estimação por intervalo

Abordagem clássica

Visualização de 100 intervalos de confiança

```r
require(TeachingDemos)
ci.examp(reps = 100)
```



Estimação por intervalo

Abordagem Bayesiana

- Chamamos de **intervalo de credibilidade** de $\gamma = 1 - \alpha$, $(0 < \alpha < 1)$, o intervalo delimitado pelos percentis $[\alpha/2]$ e $[1 - (\alpha/2)]$

$$(\theta_{[\alpha/2]}, \theta_{[1-(\alpha/2)]})$$

  da distribuição posterior $\pi(\theta|\mathbf{x})$ para θ.
- De maneira mais geral, dizemos que (T_1, T_2), $T_1 = \theta_{[\alpha/2]}$ e $T_2 = \theta_{[1-(\alpha/2)]}$, é um intervalo de credibilidade para θ com coeficiente de confiança γ se

$$\int_{T_1}^{T_2} \pi(\theta|\mathbf{x}) d\theta = \gamma$$

- Notação:

$$\text{ICr}(\theta, \gamma\%) = (\theta_{[\alpha/2]}, \theta_{[1-(\alpha/2)]})$$

- Um **intervalo de credibilidade** é então o intervalo de valores mais prováveis de θ, que soma probabilidade γ
- $\theta \in (\theta_{[\alpha/2]}, \theta_{[1-(\alpha/2)]})$ com probabilidade γ

Estimação por intervalo

Interpretação

- **Abordagem clássica**: Temos $\gamma\%$ de confiança de que o intervalo contém θ.
- **Abordagem Bayesiana**: Temos $\gamma\%$ de confiança de que θ pertence a esse intervalo.

Estimação por intervalo

Exemplo

Interesse: média de baleias avistadas em 10 milhas náuticas (MN)

- Em 150 MN navegadas foram realizadas 10 avistagens
- Obter a distribuição posterior $\pi(\theta|\mathbf{x})$ para o número médio de baleias avistadas

Caracterização do problema:

$$\begin{aligned} x &\equiv \text{número de avistagens} = 10 \\ n &\equiv \text{número de unidades amostradas} = 150/10 = 15 \\ \theta &\equiv \text{número médio de avistagens}/10 \text{ MN} \end{aligned}$$

```r
# Número de avistagens
x <- 10
# Unidades amostradas
n <- 15
# Possiveis valores para teta (numero médio de avistagens)
teta <- seq(0, 2, length = 200)
```

Estimação por intervalo

Exemplo

Suposições do problema:

$$\begin{aligned} f(\mathbf{x}|\theta) &\sim Poisson(n\theta) \\ \pi(\theta) &\sim Gama(\alpha, \beta) \\ \pi(\theta|\mathbf{x}) &\sim ? \end{aligned}$$

Posterior conjugada para $\pi(\theta) \sim Gama(\alpha, \beta)$ e $f(\mathbf{x}|\theta) \sim Poisson(n\theta)$:

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) \cdot f(\mathbf{x}|\theta) \\ &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \mathrm{e}^{-\beta\theta} \cdot \frac{n^x}{x!} \theta^x \mathrm{e}^{-n\theta} \\ &\propto \theta^{\alpha+x-1} \mathrm{e}^{-\theta(\beta+n)} \end{aligned}$$

Portanto

$$\pi(\theta|\mathbf{x}) \sim Gama(\alpha^* = \alpha + x, \beta^* = \beta + n)$$

Estimação por intervalo

Considerando que não existe nenhuma informação prévia sobre θ, vamos assumir que $\pi(\theta)$ é uma *priori* não informativa, por exemplo $\theta \sim Gama(\alpha = 1, \beta = 0.1)$

```
alfa <- 1
beta <- 0.1
## Calcula a densidade da priori
priori.ni <- dgamma(teta, alfa, beta)
```

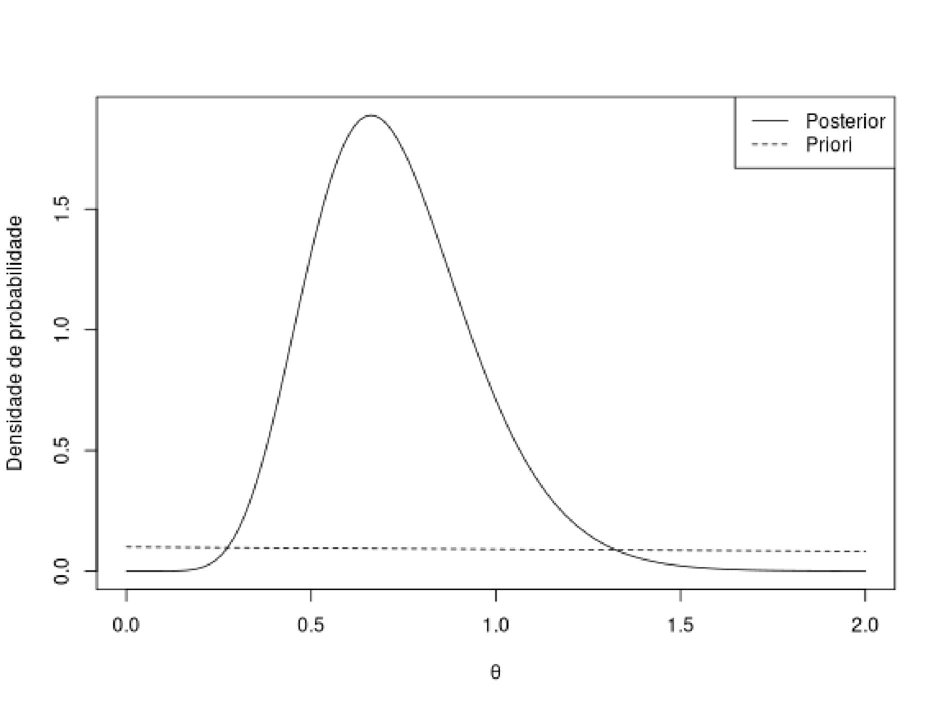
Assim, os parâmetros da posterior $\pi(\theta|\mathbf{x}) \sim Gama(\alpha^* = \alpha + x, \beta^* = \beta + n)$

```
(alfa.star <- alfa + x)
# [1] 11
(beta.star <- beta + n)
# [1] 15.1
```

Estimação por intervalo

Cálculo da densidade da posterior com os novos parâmetros

```
post.ni <- dgamma(teta, alfa.star, beta.star)
plot(teta, post.ni, type = "l", xlab = expression(theta),
     ylab = "Densidade de probabilidade")
lines(teta, priori.ni, lty = 2)
legend("topright", legend = c("Posterior", "Priori"), lty = c(1,2))
```



Estimação por intervalo

A partir da posterior $Gama(\alpha^*, \beta^*)$, podemos encontrar um **intervalo de credibilidade** de $\gamma = 1 - \alpha = 1 - 0.05 = 0.95 = 95\%$, onde os intervalos serão delimitados pelos percentis

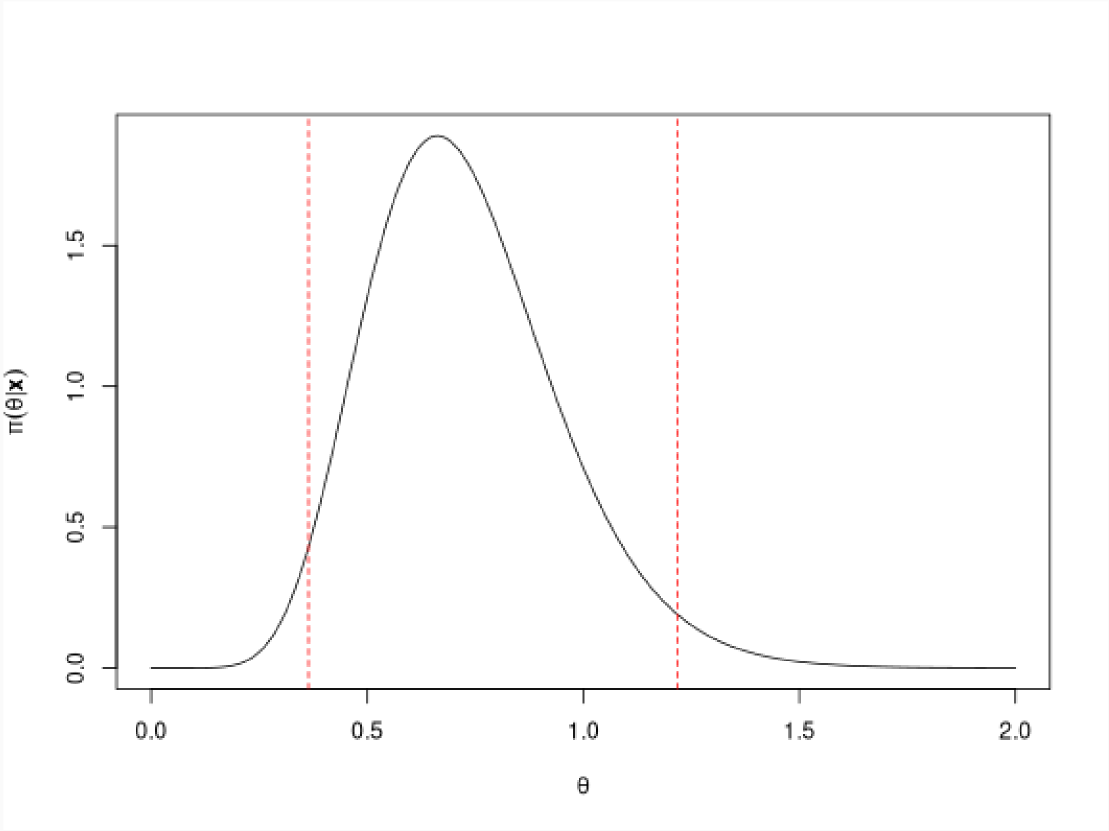
$$(\theta_{[0.025]}, \theta_{[0.975]})$$

da posterior $\pi(\theta|\mathbf{x})$

```
qgamma(c(0.025, 0.975), alfa.star, beta.star)
# [1] 0.363653 1.217904
```

Estimação por intervalo

```
## Gráfico da posterior com intervalo de credibilidade
plot(teta, post.ni, type = "l", xlab = expression(theta),
     ylab = expression(pi(paste(theta, "|", bold(x)))))
abline(v = qgamma(c(0.025, 0.975), alfa.star, beta.star),
       lty = 2, col = 2)
```



Estimação por intervalo

Considere que artigos científicos indicam que na região de estudo deve-se esperar uma média de 0.45 avistagens/10 MN. Podemos utilizar essa informação como uma *priori* **informativa**.

- Igualando essa informação à esperança da Gama:

$$E(X) = \frac{\alpha}{\beta} = 0.45 \quad \Rightarrow \quad \beta = \frac{\alpha}{0.45}$$

- Como a *priori* deve ser informativa, podemos assumir um valor relativamente alto para α, p.ex. $\alpha = 4.5$

$$\beta = \frac{\alpha}{0.45} = \frac{4.5}{0.45} = 10$$

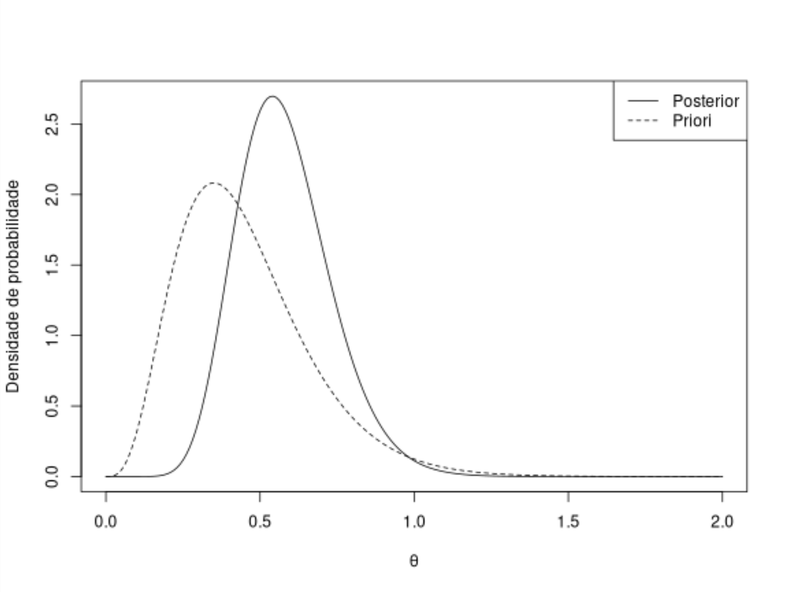
- Portanto ficamos com uma *priori* informativa $\pi(\theta) \sim Gama(\alpha = 4.5, \beta = 10)$

```
alfa.i <- 4.5
beta.i <- 10
## Calcula a densidade da priori
priori.i <- dgamma(teta, alfa.i, beta.i)
```

Estimação por intervalo

A posterior $\pi(\theta|\mathbf{x}) \sim Gama(\alpha^* = \alpha + x, \beta^* = \beta + n)$

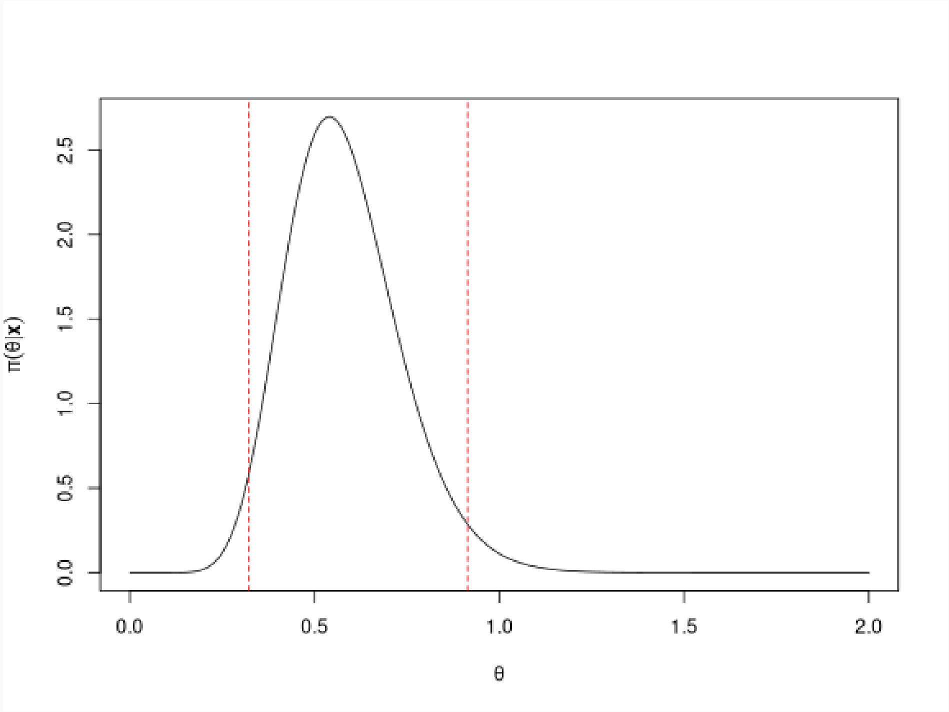
```
alfa.star.i <- alfa.i + x
beta.star.i <- beta.i + n
## Cálculo da densidade da posterior com a priori informativa
post.i <- dgamma(teta, alfa.star.i, beta.star.i)
# Visualização
plot(teta, post.i, type = "l", xlab = expression(theta),
     ylab = "Densidade de probabilidade")
lines(teta, priori.i, lty = 2)
legend("topright", legend = c("Posterior", "Priori"), lty = c(1,2))
```



Estimação por intervalo

Podemos encontrar o **intervalo de credibilidade**

```
qgamma(c(0.025, 0.975), alfa.star.i, beta.star.i)
# [1] 0.3209414 0.9144457
## Gráfico da posterior com intervalo de credibilidade
plot(teta, post.i, type = "l", xlab = expression(theta),
     ylab = expression(pi(paste(theta, "|", bold(x)))))
abline(v = qgamma(c(0.025, 0.975), alfa.star.i, beta.star.i),
       lty = 2, col = 2)
```



Estimação por intervalo

Um resumo simples das duas posteriores obtidas

```
## Inferência com a priori não informativa
qgamma(c(.025, .5, .975), alfa.star, beta.star)
# [1] 0.3636530 0.7065247 1.2179044
## Inferência com a priori informativa
qgamma(c(.025, .5, .975), alfa.star.i, beta.star.i)
# [1] 0.3209414 0.5667225 0.9144457
```

Estimação por intervalo

Comparação das duas posteriores

```
plot(teta, post.i, type = "l", xlab = expression(theta),
     ylab = expression(pi(paste(theta, "|", bold(x)))))
lines(teta, post.ni, lty = 2)
legend("topright",
       legend = c("Posterior (informativa)",
                  "Posterior (não informativa)"), lty = c(1,2))
```

Outro exemplo

Seja $x_1, \ldots, x_n$ uma a.a. de uma distribuição normal de média μ conhecida e variância σ^2 desconhecida. Considere a priori $[\sigma^2] \propto 1/\sigma^2$.

1. Obtenha a expressão da verossimilhança do modelo.
2. Obtenha a expressão da distribuição a posteriori.
3. É possível identificar a posteriori do modelo como alguma distribuição conhecida?
4. Considere que foi tomada a amostra dada pelos valores a seguir, e que $\mu = 10$. Obtenha a expressão da posteriori.

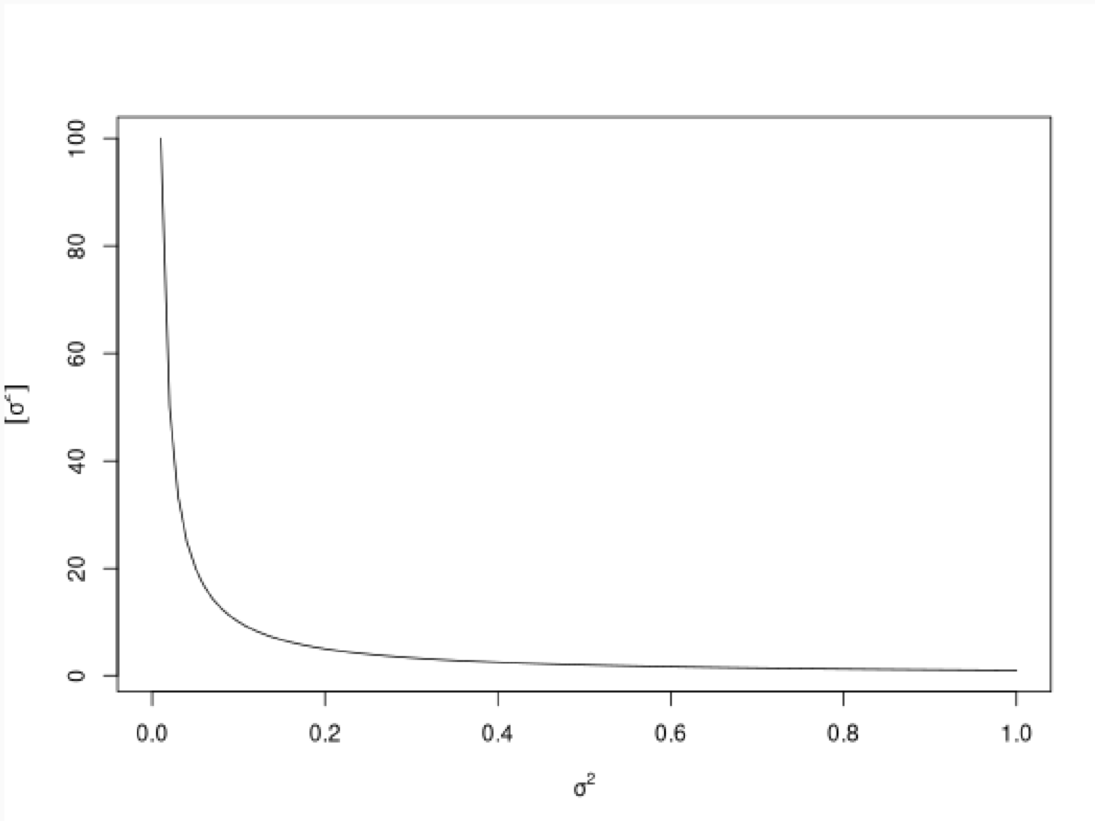
```
# [1] 12.1  8.7 11.3  9.2 10.5  9.7 11.6
```

1. Indique (com comandos do R ou de alguma outra forma) como resumos pontuais e intervalares desta distribuição a posteriori poderiam ser obtidos para fins de inferência

Priori de Jeffreys

O que significa $[\sigma^2] \propto 1/\sigma^2$?

```
x <- seq(0, 1, 0.01)
fx <- 1/x
plot(x, fx, type = "l",
     xlab = expression(sigma^2),
     ylab = expression(group("[", paste(sigma^2), "]")))
```



1. Expressão da verossimilhança

As funções de verossimilhança $L(\sigma^2)$, log-verossimilhança $l(\sigma^2)$, escore $U(\sigma^2)$ e hessiana $H(\sigma^2)$ são dadas por:

$$
\begin{aligned}
L(\sigma^2; y) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\} \\
&= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{1=1}^{n}(y_i - \mu)^2\right\} \\
l(\sigma^2; y) = \log\{L(\sigma^2; y)\} &= (-n/2)\left[\log(2\pi) + \log(\sigma^2) + \frac{\sum_{1=1}^{n}(y_i - \mu)^2}{n}(\sigma^2)^{-1}\right] \\
U(\sigma^2) = \frac{\mathrm{d}l(\sigma^2; y)}{\mathrm{d}\sigma^2} &= (-n/2)[(\sigma^2)^{-1} - \frac{\sum_{1=1}^{n}(y_i - \mu)^2}{n}(\sigma^2)^{-2}] \\
H(\sigma^2) = \frac{\mathrm{d}^2 l(\sigma^2; y)}{\mathrm{d}(\sigma^2)^2} &= (-n/2)[-(\sigma^2)^{-2} + \frac{\sum_{1=1}^{n}(y_i - \mu)^2}{n}(\sigma^2)^{-3}]
\end{aligned}
$$

2. Expressão da distribuição posterior

A expressão da distribuição a posteriori é obtida da forma:

$$
\begin{aligned}
[\sigma^2|y] &\propto [\sigma^2] \cdot L(\sigma^2; y) \\
&= (\sigma^2)^{-1} \cdot (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{1=1}^{n} (y_i - \mu)^2 \right\} \\
&\propto (\sigma^2)^{-(n/2)-1} \exp\left\{ -\frac{\sum_{1=1}^{n} (y_i - \mu)^2}{2\sigma^2} \right\}
\end{aligned}
$$

É possível identificar esta posteriori como alguma distribuição conhecida?

3. Identificar a posteriori como alguma distribuição conhecida

Posterior:

$$[\sigma^2|y] \propto (\sigma^2)^{-(n/2)-1} \exp\left\{-\frac{\sum_{1=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}$$

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

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)}\, x^{\alpha-1}\; \exp -x/\beta\; \mathbb{I}_{x\geq 0}$$

$$E[X] = \alpha\beta \qquad Var[X] = \alpha\beta^2$$

$X \sim \text{IGama}(\alpha, \beta)$

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)}\, x^{-\alpha-1}\; \exp -1/(\beta x)\; \mathbb{I}_{x\geq 0}$$

$$E[X] = \frac{1}{\beta(\alpha-1)} \qquad Var[X] = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)}$$

3. Identificar a posteriori como alguma distribuição conhecida

A expressão

$$[\sigma^2|y] \propto (\sigma^2)^{-(n/2)-1} \exp\left\{-\frac{\sum_{1=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}$$

vista como uma função de σ^2, corresponde ao núcleo de uma densidade *gama inversa* de parâmetros:

$$\alpha = \frac{n}{2} \qquad \text{e} \qquad \beta = \frac{2}{\sum_{1=1}^{n}(y_i-\mu)^2}$$

Esta distribuição também é chamada na literatura de *qui-quadrado inversa escalonada*.

4. Expressão da posteriori para os dados

Para o conjunto de dados temos que

$$\alpha = \frac{n}{2} = \frac{7}{2} = 3,5 \qquad \text{e} \qquad \beta = \frac{2}{\sum_{1=1}^{n}(y_i - \mu)^2} = \frac{2}{11,33} = 0,1765$$

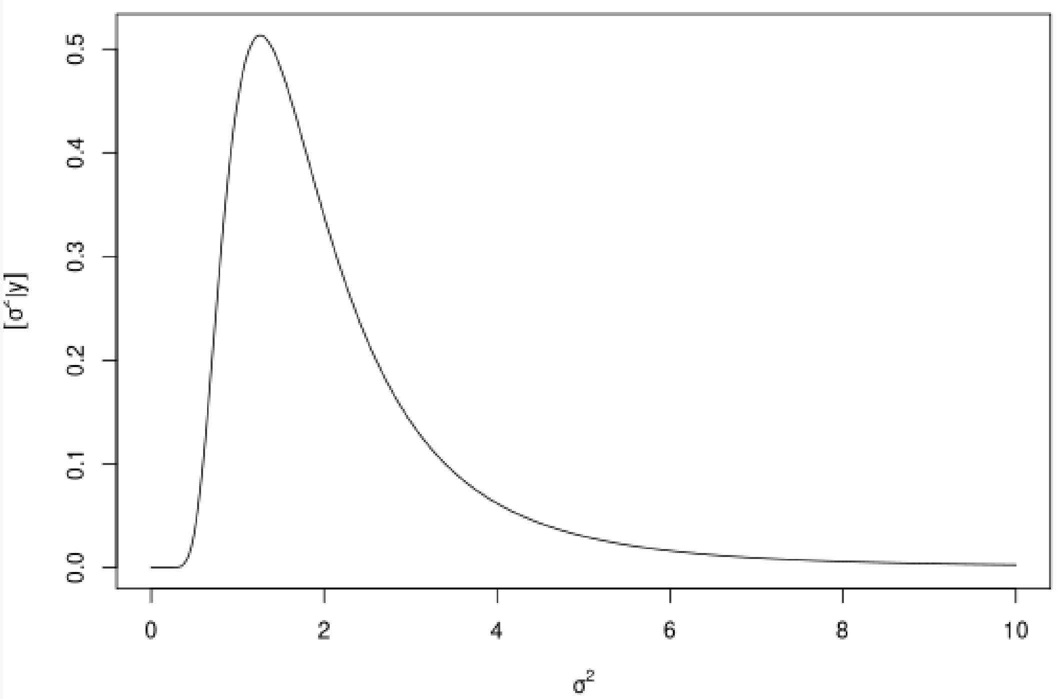
```
y <- c(12.1, 8.7, 11.3, 9.2, 10.5, 9.7, 11.6)
(a.post <- length(y)/2)
# [1] 3.5
(b.post <- 2/sum((y - 10)^2))
# [1] 0.1765225
```

Embora códigos para tal distribuição possam ser obtidos já implementados, vamos definindo uma função da densidade gama-inversa a partir de sua expressão. A seguir usamos a função para obter o gráfico da posteriori obtida aqui.

```
dinvgamma <- function(x, a, b, log = FALSE){
    res <- ifelse(
        x > 0,
        - a * log(b) - log(gamma(a)) - (a+1)*log(x) - 1/(b*x), -Inf)
    if(!log) res <- exp(res)
    return(res)
}
```

4. Expressão da posteriori para os dados

```
curve(dinvgamma(x, a = a.post, b = b.post), from = 0, to = 10, n = 501,
      xlab = expression(sigma^2),
      ylab = expression(group("[", paste(sigma^2,"|",y), "]")))
```



5. Resumos pontuais e intervalares

Resumos pontuais: média da posteior

Neste caso se tem a expressão analítica, que portanto deve ser utilizada.

- Como em muitos casos pode não ser disponível, ilustra-se também a obtenção por integração numérica e por simulação.

- Expressão analítica

$$E[\sigma^2|y] = \frac{1}{\beta(\alpha - 1)} = 2.27$$

```
(e.post <- 1/(b.post * (a.post - 1)))
# [1] 2.266
```

- Integração numérica (a partir da definição de esperança de uma v.a.).

$$E[\sigma^2|y] = \int_0^\infty \sigma^2 f(\sigma^2|y) \mathrm{d}\sigma^2$$

```
Epost <- function(par, ...) par * dinvgamma(par, ..., log = FALSE)
integrate(Epost, lower = 0, upper = 50, a = a.post, b = b.post)
# 2.263282 with absolute error < 6.6e-06
```

5. Resumos pontuais e intervalares

Resumos pontuais: média da posteior

- Por simulação

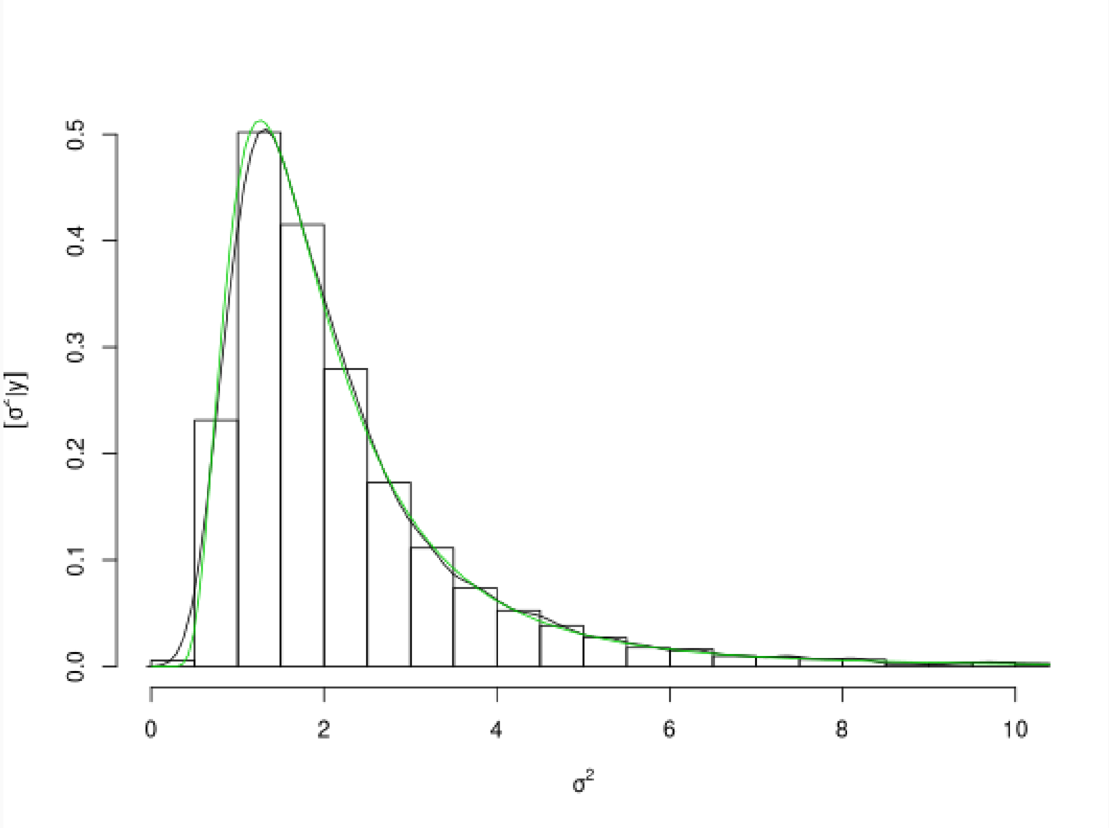
Se $X \sim \text{Gama}(\alpha, \beta)$ então $Y = 1/X \sim \text{IGama}(\alpha, \beta)$.

Portanto, para simular de uma distribuição gama inversa basta tomar o inverso de valores simulados de uma distribuição gama com os os mesmos parâmetros.

```
set.seed(123)
sim <- 1/rgamma(10000, shape = a.post, scale = b.post)
hist(sim, prob = TRUE, ylim = c(0,0.5),
     breaks = c(seq(0, 10, by = 0.5), seq(10, 60, by = 2)),
     xlab = expression(sigma^2),
     ylab = expression(group("[", paste(sigma^2,"|",y), "]")),
     main = "", xlim = c(0, 10))
lines(density(sim))
curve(dinvgamma(x, a = a.post, b = b.post), from = 0.01, to = 30,
      add = TRUE, col = 3, n = 501)
```

5. Resumos pontuais e intervalares

Resumos pontuais: média da posteior




```
mean(sim)
# [1] 2.280075
```

5. Resumos pontuais e intervalares

Resumos pontuais: moda da posteior

Assim como no caso da média, a expressão analítica da moda é conhecida mas ilustra-se também a obtenção por otimização numérica e por simulação.

- Expressão analítica

$$Mo[\sigma^2|y] = \frac{1}{\beta(\alpha+1)} = 1.26$$

```
(mo.post <- 1/(b.post * (a.post + 1)))
# [1] 1.258889
```

- Otimização numérica, maximizando a densidade

```
optimize(dinvgamma, lower = 0, upper = 50, a = a.post, b = b.post,
         log = FALSE, maximum = TRUE)$maximum
# [1] 1.258905
```

- Simulação. Utiliza-se aqui um algorítmo simples tomando-se o ponto de máximo de uma suavização da densidade

```
sim.den <- density(sim, n = 1024)
sim.den$x[which.max(sim.den$y)]
# [1] 1.32702
```

5. Resumos pontuais e intervalares

Resumos pontuais: mediana da posteior

- Expressão analítica: não disponível
- Otimização numérica: usando inversos dos quantis da distribuição gama

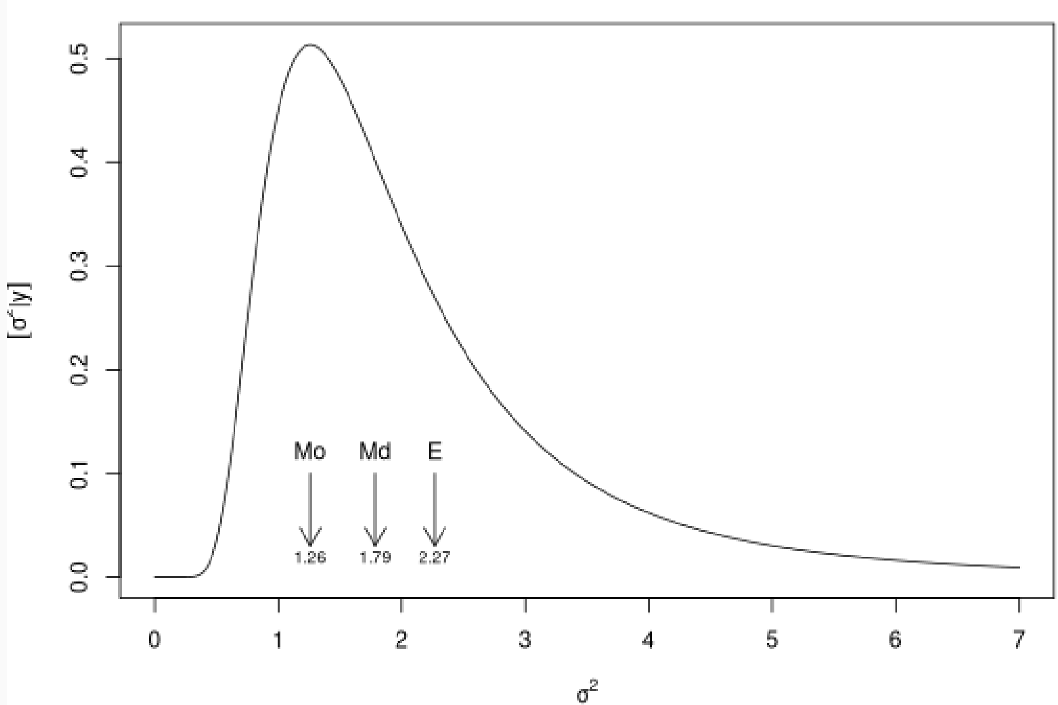
```
(md.post <- 1/qgamma(1 - 0.5, shape = a.post, scale = b.post))
# [1] 1.78543
```

- Simulação

```
median(sim)
# [1] 1.793622
```

5. Resumos pontuais e intervalares

Resumos pontuais: média, moda e mediana



5. Resumos pontuais e intervalares

Resumos intervalares: intervalos de credibilidade

- Expressão analítica: não disponível
- Por otimização numérica. Podemos definir uma função específica para a gama-inversa:

```
qinvgamma <- function(p, a, b, ...) {
    1/qgamma(1 - p, shape = a, scale = b, ...)
}
(ICq <- qinvgamma(c(0.025, 0.975), a = a.post, b = b.post))
# [1] 0.7075605 6.7046610
```

- Por simulação

```
quantile(sim, prob = c(0.025, 0.975))
#      2.5%     97.5%
# 0.7075484 6.8346235
```

5. Resumos pontuais e intervalares

Resumos intervalares: HPD

- Quando a distribuição posterior e assimétrica, o intervalo de credibilidade nem sempre representa a porção central da distribuição adequadamente
- Por isso, usamos o intervalo HPD (*Highest Posterior Density*), que, por definição, sempre será o menor intervalo possível (central)

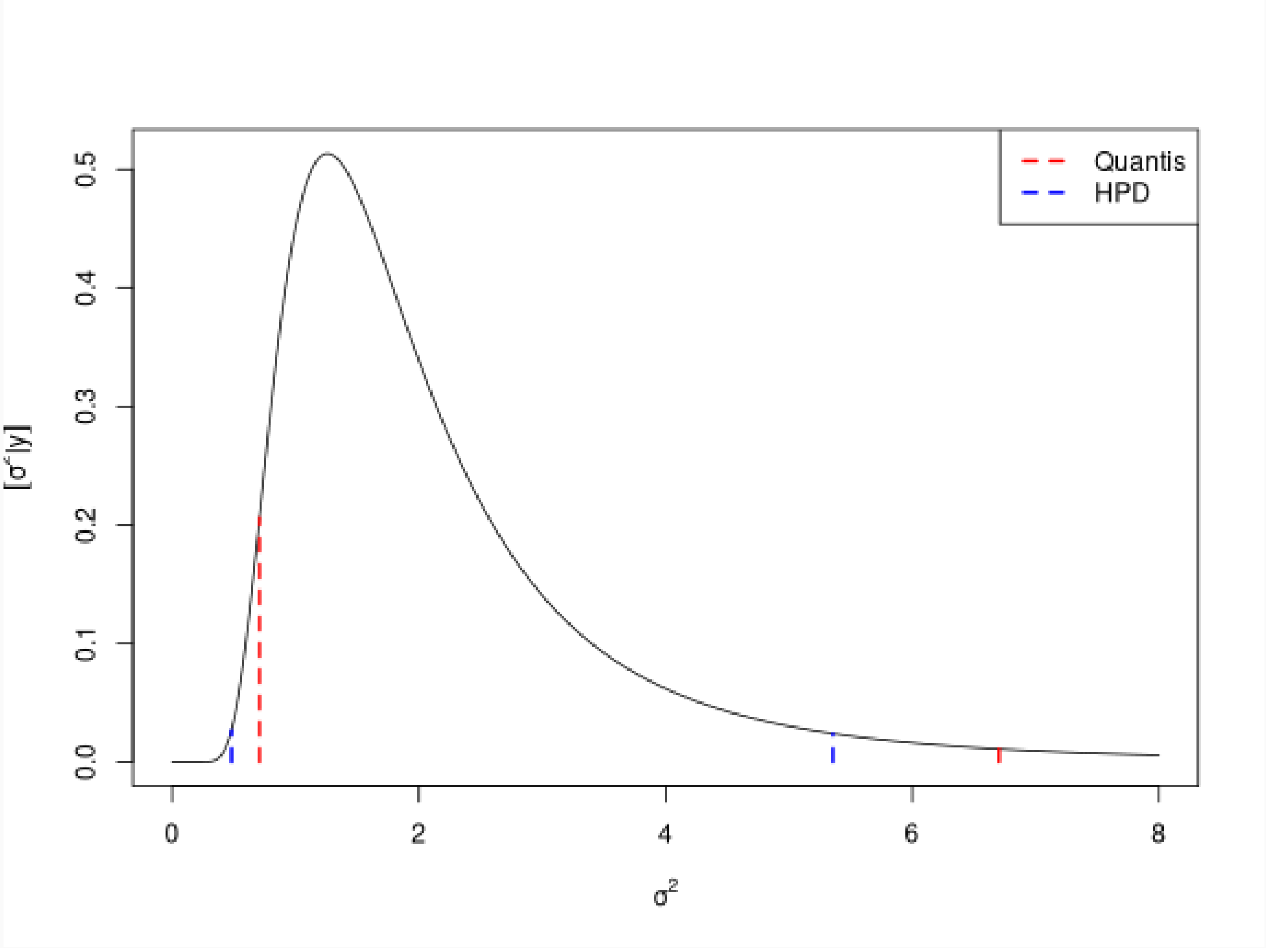
```
library(HDInterval)
(IChpd <- hdi(sim))
#     lower     upper
# 0.4796374 5.3586471
# attr(,"credMass")
# [1] 0.95
```

Veja a diferença:

```
ICq
# [1] 0.7075605 6.7046610
diff(ICq)
# [1] 5.9971
IChpd
# [1] 0.4796374 5.3586471
diff(IChpd)
# [1] 4.87901
```

5. Resumos pontuais e intervalares

Resumos intervalares: quantis x HPD



Gráficos

Priori, posteriori e verossimilhança

Para incluir o gráfico da **verossimilhança** na mesma escala da priori e posteriori é necessário utilizar a função padronizada (de forma a integrar 1).

Há duas formas que serão ilustradas a seguir:

1. reconhecendo o núcleo de alguma distribuição conhecida (quando possível)
2. integrando explicitamente a verossimilhança (analítica ou numericamente).

Neste exemplo a opção (1) é possível pois, para a verossimilhança, tem-se que:

$$L(\sigma^2|y) \propto \text{IGama}(\alpha = (n/2) - 1, \beta = 2/\sum_{i=1}^{n}(y_i - \mu)^2)$$

Gráficos

Priori, posteriori e verossimilhança