Inferência Bayesiana

Inferência sobre o parâmetro de forma da Gamma

Eduardo E. R. Junior

Conjunto de dados

O conjunto de dados utilizado neste exemplo é apresentado abaixo:

```
 [1]  0.8  1.5  2.6  1.5  1.4  2.1  3.9  3.3  3.0  3.9  4.2  4.8  6.9  5.5
[15]  7.3  6.7  8.2  8.1 10.2 10.7 12.8
```

São 21 observações que provém de um processo aleatório Gama com parâmetro de locação conhecido (α = 2) e estamos interessados na inferência sobre a dispersão do processo, controlada pelo parâmetro θ.

A distribuição Gama

Uma variável aleatória Y segue o modelo Gama, se sua densidade for dada por

$$f_Y(y \mid \alpha, \theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{\theta^{-1} y} \qquad (1) \qquad\qquad f_Y(y \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta^{y}} \qquad (2)$$

em que $\alpha > 0$, $\theta > 0$ e $\beta > 0$. Aqui são apresentadas duas diferentes parametrizações da distribuição Gama. A primeira apresentada em (1) possui $E[Y] = \alpha\theta$ e a variância por $Var[Y] = \alpha\theta^2$. A segunda (1), por sua vez tem as expressões para média e variância dadas por $E[Y] = \alpha/\beta$ e $V[Y] = \alpha/\beta^2$. As parametrizações adotadas aqui se referem, propositalmente ao parâmetro de escala, sob invertigação. Ainda as duas definições se relacionam sob a transformação $\theta = 1/\beta$.[1]

Inferência via Verossimilhança

Abaixo são definidas analiticamente as funções de *verossimilhança* $L(\theta; y)$; *log-verossimilhança* $l(\theta; y)$; *escore* $U(\theta)$; *Hessiana* $H(\theta)$; e *Informação esperada de Fisher* $I_E(\theta)$. Com elas temos todos os elementos necessários para inferência estatística baseada em verossimilhança, a partir de uma amostra.

[1]Há diversas outras parametrizações que podem ser consideradas para a distribuição Gama. No R, as funções `d`, `p`, `q` e `r` de sufixo `gamma` provém duas parametrizações aqui consideradas.

$$
\begin{aligned}
L(\theta; y) &= \prod_{i=1}^{n} \left(\frac{\theta^{-\alpha}}{\Gamma(\alpha)} y_i^{\alpha-1} e^{\theta^{-1} y_i} \right) \\
l(\theta; y) &= -n\alpha \log(\theta) - n \log(\Gamma(\alpha)) + \\
& \quad (\alpha - 1) \sum_{i=1}^{n} \log(y_i) - \theta^{-1} \sum_{i=1}^{n} y_i \\
U(\theta) &= \frac{-n\alpha}{\theta} + \frac{\sum_{i=1}^{n} y_i}{\theta^2} \\
H(\theta) &= \frac{n\alpha}{\theta^2} - \frac{2\sum_{i=1}^{n} y_i}{\theta^3} \\
I_E(\theta) &= E[-H(\theta; y)] = \frac{n\alpha}{\theta^2}
\end{aligned}
$$

$$
\begin{aligned}
L(\beta; y) &= \prod_{i=1}^{n} \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} y_i^{\alpha-1} e^{\beta y_i} \right) \\
l(\beta; y) &= n\alpha \log(\beta) - n \log(\Gamma(\alpha)) + \\
& \quad (\alpha - 1) \sum_{i=1}^{n} \log(y_i) + \beta \sum_{i=1}^{n} y_i \\
U(\beta) &= \frac{-n\alpha}{\beta} + \sum_{i=1}^{n} y_i \\
H(\beta) &= \frac{-n\alpha}{\beta^2} \\
I_E(\beta) &= E[-H(\beta; y)] = \frac{n\alpha}{\beta^2}
\end{aligned}
$$

Para encontrar $\hat{\theta}$, uma estimativa do parâmetro θ, resolvemos a equação $U(\theta) = 0$, ainda uma medida de incerteza sobre esta estimativa pode ser obtida pelo inverso da informação de Fisher, $Var[\hat{\theta}] = I_E^{-1}(\theta)$

Analiticamente temos:

$$
U(\theta; y) = 0 \Rightarrow \hat{\theta} = (n - \alpha)^{-1} \sum_{i=1}^{n} y_i
$$

Obtendo a estimativa de θ via expressão analítica temos:

```r
##-------------------------------------------
## Obtendo as estimativas analiticamente
alpha.f <- 2
MLEana <- sum(y) / (length(y)*alpha.f); MLEana
```

```
[1] 2.604762
```

```r
sum(y) / (length(y)*alpha.f)
```

```
[1] 2.604762
```

```r
(length(y)*alpha.f) / sum(y)
```

```
[1] 0.3839122
```

E para as demais medidas de incerteza sobre o ajuste definimos as seguintes funções no R

```r
## Escore
Ufun <- function(par, alpha = alpha.f, y,
                 parametrization = c("theta", "alpha")) {
    n <- length(y); soma <- sum(y)
    out <- -n * alpha * par^(-1)
    if (parametrization[1] == "beta") {
        out <- out + soma
    } else {
        out <- out + soma * par^(-2)
    }
    return(out)
}

## Hessiana
Hfun <- function(theta, alpha = alpha.f, y,
                 parametrization = c("theta", "alpha")) {
    n <- length(y)
    out <- -n * alpha * theta^(-2)
    if (parametrization[1] == "theta") {
        out <- -out - 2 * sum(y) * theta^(-3)
    }
    return(out)
}

## Informação esperada
Iefun <- function(theta, alpha = alpha.f, y) {
    ## Invariante a parametrização
    n <- length(y)
    out <- n * alpha * theta^(-2)
    return(out)
}
```

Adicionalmente definimos a função de log-verossimilhança para verificar se as expressões calculadas conferem seu máximo quando aplicadas no exemplo

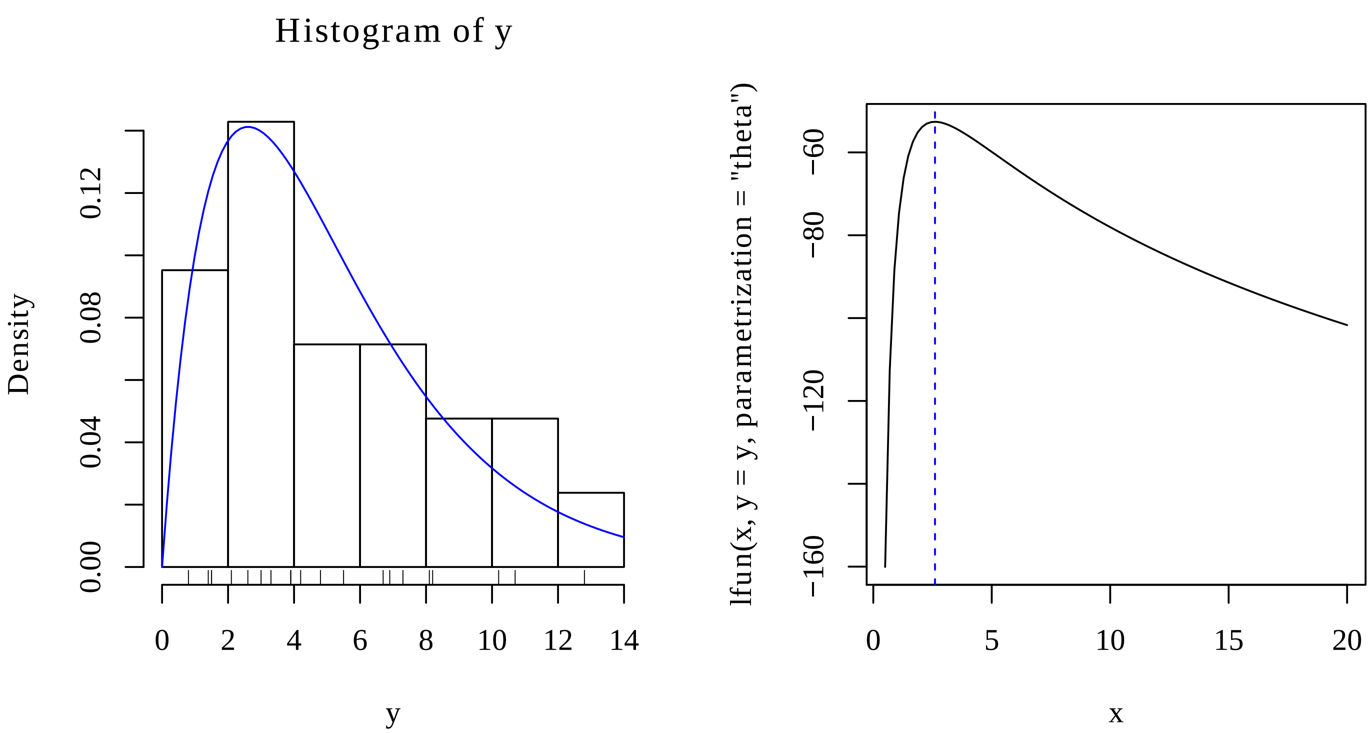
```r
## log-verossimilhança
lfun <- function(par, alpha = alpha.f, y, log = TRUE,
                 parametrization = c("theta", "beta")) {
    if (parametrization[1] == "beta") {
        theta <- 1/par
    } else {
        theta <- par
    }
    n <- length(y)
    ## res <- sum(dgamma(x = y, shape = alpha, scale = theta, log = TRUE))
    out <- -n * alpha * log(theta) - n * lgamma(alpha) +
        (alpha - 1) * sum(log(y)) - theta^(-1) * sum(y)
```

```
    if(!log) out <- exp(out)
    attr(out, "parametrization") <- parametrization
    attr(out, "alpha") <- alpha.f
    return(out)
}
```

Observando o ajuste

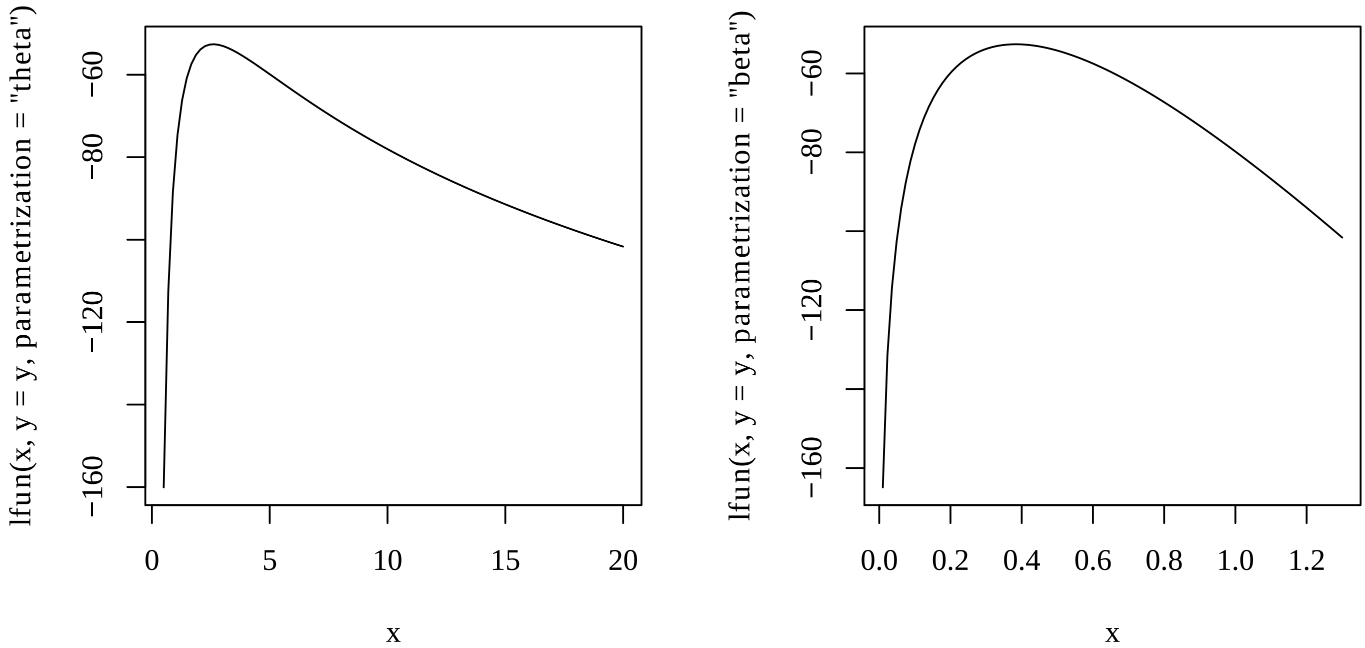
```
## Calculando conforme expressão analitica
alpha.f <- 2
MLEana <- sum(y) / (length(y)*alpha.f)

## Visualizando o ajuste
par(mfrow = c(1, 2))
hist(y, prob = TRUE); rug(y)
curve(dgamma(x, shape = alpha.f, scale = MLEana), col = 4, add = TRUE)
##
curve(lfun(x, y = y, parametrization = "theta"), from = 0.5, to = 20)
abline(v = MLEana, lty = 2, col = 4)
```



Estudando de diferentes parametrizações

```
par(mfrow = c(1, 2))
curve(lfun(x, y = y, parametrization = "theta"), from = 0.5, to = 20)
curve(lfun(x, y = y, parametrization = "beta"), from = 1e-2, to = 1.3)
```



Inferência Bayesiana

Em inferência Bayesiana tratamos o parâmetro de interesse como variável aleatória e, através do teorema de Bayes, definimos sua função de densidade de probabilidade e assim todas as inferências são derivadas dela.

$$f(\theta; y) = f(\theta) L(\theta; y) \tag{3}$$

Priori de Jeffreys

Uma das características da inferência Bayesiana é que podemos (devemos) incluir uma informação a priori a respeito do parâmetro de interesse, ou seja, àquela cujo temos antes de se realizar o experimento e se coletar os dados. Em muitas situações não temos o que acrecentar sobre o parâmetro e desejamos representar nossa ignorância com funções a priori vagas, porém definir funções a priori que sejam vagas não é uma tarefa fácil. Um dos principais defeitos de prioris vagas é que elas são vagas somente em certa escala do parâmetro, ou seja, é como se seu conhecimento sobre o parâmetro variasse conforme a parametrização utilizada.

Uma das saídas para esta situação é adotar a priori de Jeffreys que contém a mesma informação, invariavelmente à transformação do parâmetro. A priori de Jefrreys é definida como $f(\theta) = |I_E(\theta)|^{1/2}$. No exemplo Gama a priori de Jeffreys, tomando a expressão da Informação de Fisher em (6) é

$$\begin{aligned} f(\theta) &= \left| \frac{n\alpha}{\theta^2} \right|^{1/2} \\ &\propto \theta^{-1} \end{aligned} \tag{4}$$

Posteriori

Substituindo as expressões (8) e (2), sem as constantes de proporcionalidades, na expressão (7) temos a função a posteriori da seguinte forma:

$$\begin{aligned} f(\theta; y) &\propto \theta^{-1}\theta^{-n\alpha} \exp\left\{-\theta^{-1}\sum_{i=1}^{n} y_i\right\} \\ &\propto \theta^{-n\alpha-1} \exp\left\{-\sum_{i=1}^{n} y_i(\theta^{-1})\right\} \end{aligned} \tag{5}$$

E neste exemplo conseguimos reconhecer o núcleo da distribuição, pois a distribuição Gama Inversa possui densidade de probabilidade na forma:

$$\begin{aligned} Y &\sim InvGama(a, b) \\ f_Y(y, a, b) &= \frac{b^{-a}}{\Gamma(a)} y^{-a-1} \exp\left\{-b^{-1}y^{-1}\right\} \end{aligned}$$

E notamos que a posteriori segue este modelo *InverseGama* de parâmetros $a = n\alpha$ e $b = 1/\sum_{i=1}^{n} y_i$. Neste exemplo também pôde-se obter as expressões analíticas para as verossimilhanças, encaixando-as em modelos de probabilidade já conhecidos. Abaixo apresentamos os gráficos da posteriori e verossimilhança para as duas parametrizações

```
## Define a função da densidade da InverseGama
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)
}

##-------------------------------------------

## Informações do problema
n <- length(y)
soma <- sum(y)
alpha.f <- 2

par(mfrow = c(1, 2))
## Considerando a parametrização com theta
curve(dinvgamma(x, a = n*alpha.f, b = 1/soma),
      from = 1e-5, to = 5)
curve(dinvgamma(x, a = n*alpha.f + 1, b = 1/soma), add = TRUE,
```

```r
      lty = 2, col = 4)
legend("topleft", c("Verossimilhança", "Posteriori"),
       lty = c(2, 1), col = c(1, 4), bty = "n")
## Considerando a parametrização com beta
curve(dgamma(x, shape = n*alpha.f, rate = soma),
      from = 1e-2, to = 1.3)
curve(dgamma(x, shape = n*alpha.f + 1, rate = soma), add = TRUE,
      lty = 2, col = 4)
legend("topright", c("Verossimilhança", "Posteriori"),
       lty = c(2, 1), col = c(1, 4), bty = "n")
```