Definição de prioris e resumo da posteriori

Fernando de Pol Mayer (LEG/DEST/UFPR)

2019-09-05 (última atualização 2022-02-13)

UNIVERSIDADE FEDERAL DO PARANÁ

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)
        - BUGS (Bayesian interface Using Gibbs Sampler)
        - JAGS (Just Another Gibbs Sampler)
    - Re-amostragem por importância (SIR - *Sampling Importance Resampling*)
    - Monte Carlo Hamiltoniano (HMC) - Stan

- 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 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 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 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 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:

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

- O intervalo **contém** θ com probabilidade γ

- Uma vez que o parâmetro é fixo, o intervalo é aleatório

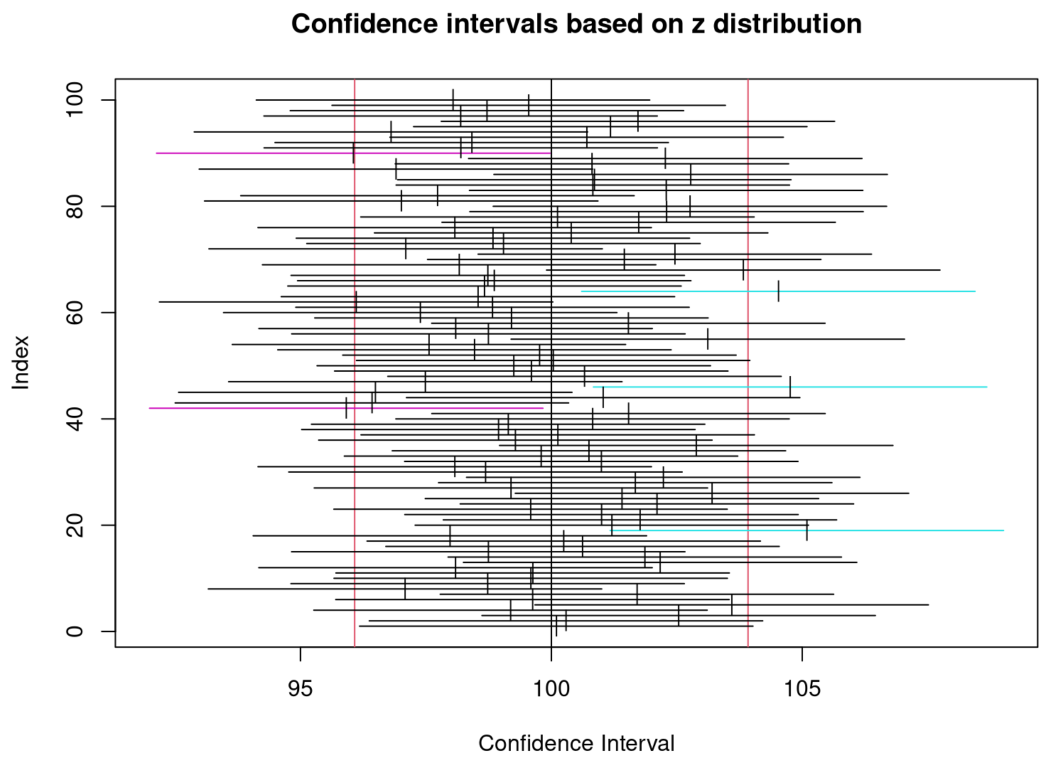
- $\theta \in (T_1, T_2)$ com probabilidade 0 ou 1

Estimação por intervalo

Abordagem clássica

Visualização de 100 intervalos de confiança

```
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.

Exemplo (Kinas e Andrade, 2010)

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 theta (numero médio de avistagens)
theta <- seq(0, 2, length = 200)
```

Exemplo (Kinas e Andrade, 2010)

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(theta, 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
```

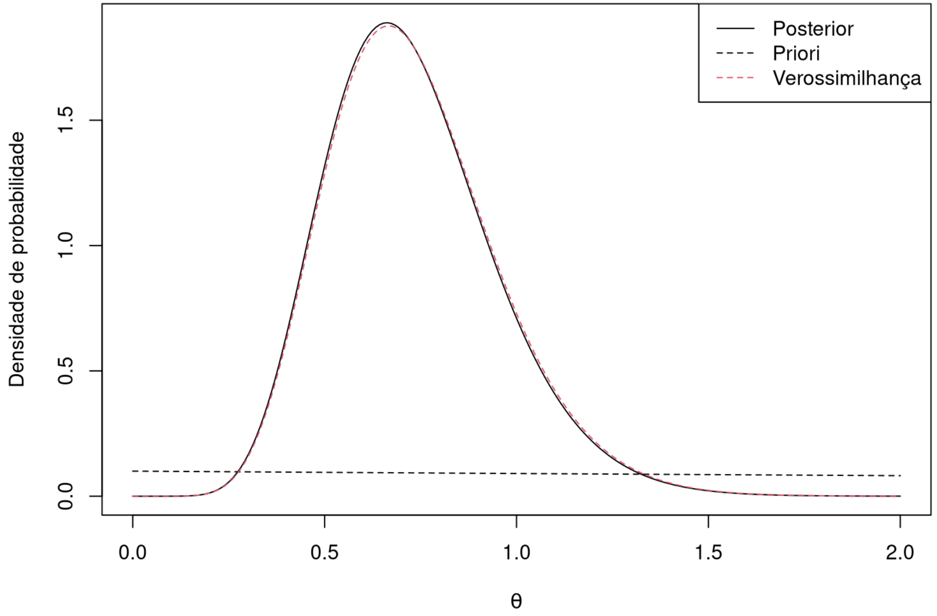
```
(beta.star <- beta + n)
```

```
# [1] 15.1
```

Estimação por intervalo

Cálculo da densidade da posterior com os novos parâmetros

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



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

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

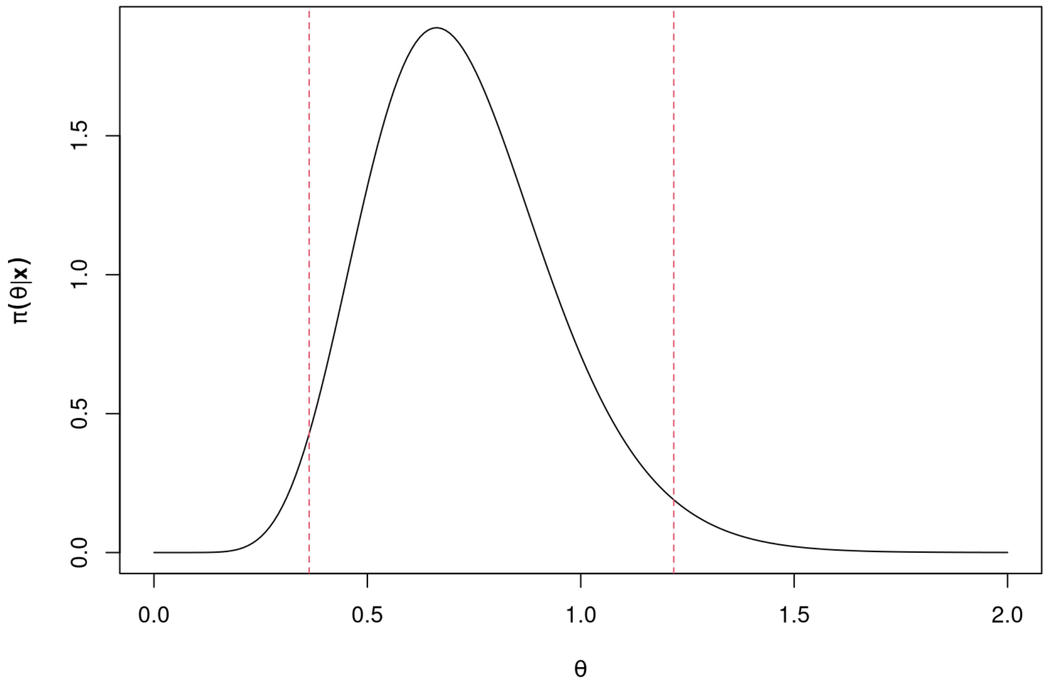
da posterior $\pi(\theta|\mathbf{x})$

```
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(theta, 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 $\pi(\theta) \sim Gama(\alpha = 4.5, \beta = 10)$

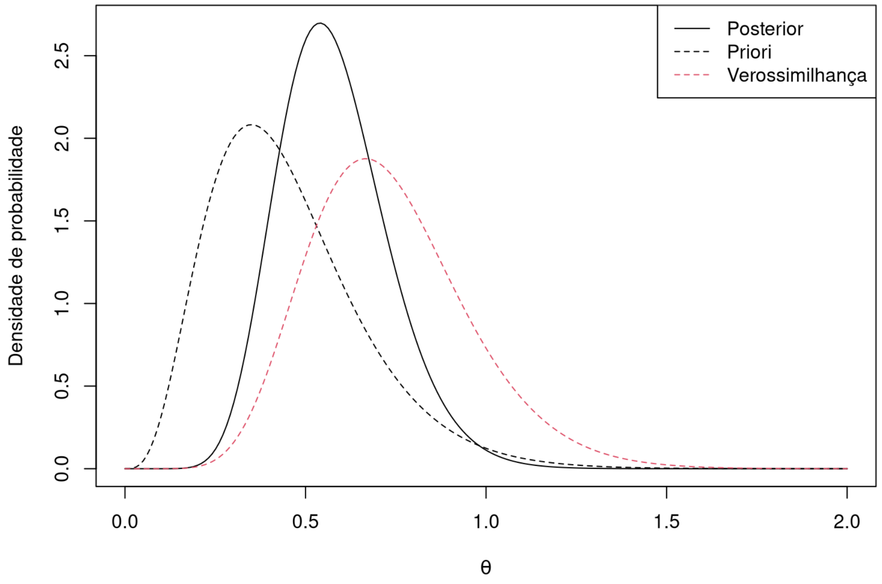
```
alfa.i <- 4.5
beta.i <- 10
## Calcula a densidade da priori
priori.i <- dgamma(theta, alfa.i, beta.i)
```

Estimação por intervalo

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

```r
alfa.star.i <- alfa.i + x
beta.star.i <- beta.i + n
## Cálculo da densidade da posterior com a priori informativa
post.i <- dgamma(theta, alfa.star.i, beta.star.i)
```

```r
## Visualização
plot(theta, post.i, type = "l", xlab = expression(theta),
     ylab = "Densidade de probabilidade")
lines(theta, priori.i, lty = 2)
lines(theta, dgamma(theta, 1 + x, n), col = 2, lty = 2)
legend("topright",  lty = c(1, 2, 2), col = c(1, 1, 2),
       legend = c("Posterior", "Priori", "Verossimilhança"))
```



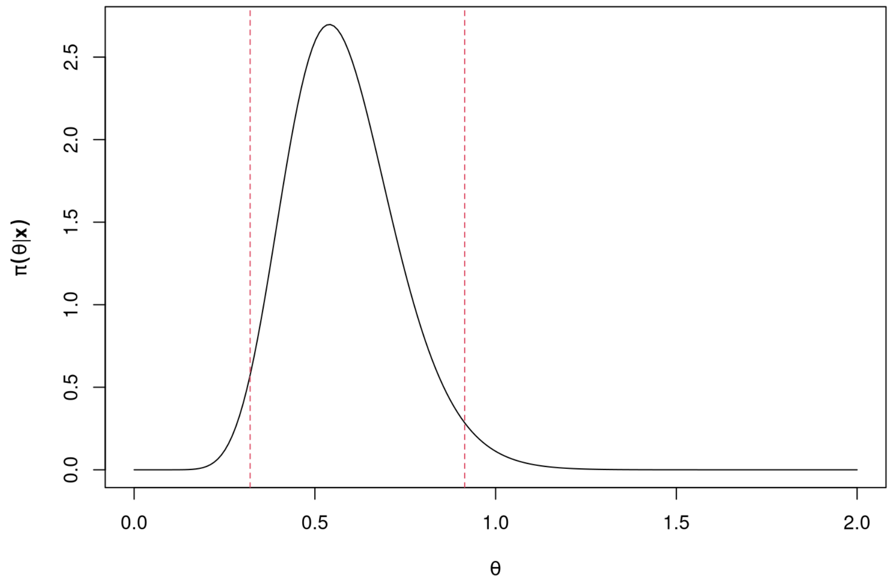
Estimação por intervalo

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(theta, 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

```r
## Inferência com a priori não informativa
qgamma(c(.025, .5, .975), alfa.star, beta.star)
```

```
# [1] 0.3636530 0.7065247 1.2179044
```

```r
## Inferência com a priori informativa
qgamma(c(.025, .5, .975), alfa.star.i, beta.star.i)
```

```
# [1] 0.3209414 0.5667225 0.9144457
```

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