Estatística Computacional

Inferência Bayesiana - Parte 1

Prof. Paulo Cerqueira Jr
Faculdade de Estatística - FAEST
Programa de Pós-graduação em Matemática e Estatística - PPGME

https://github.com/paulocerqueirajr

Introdução

Introdução

- Impulsionado por Jeffreys, Good, Savage, de Finetti e Lindley.
- Adota-se a interpretação subjetiva de probabilidade.
- O valor verdadeiro de θ não é conhecido, mas a intensidade da incerteza sobre θ pode assumir diferentes graus de incerteza, representados por modelos de probabilidade sobre θ (quantidade aleatória não observável).
- Há duas fontes de informação na realização da inferência: a amostral (dados observados) e o seu conhecimento prévio (experiência pessoal).
- O processo inferencial consiste em calibrar (atualizar) o grau de incerteza sobre θ usando os dados observados, através do Teorema de Bayes.

Introdução

- Thomas Bayes (1702-1761) foi um matemático e pastor presbiteriano inglês.
- Em 1719 ingressou na Universidade de Edinburgh para estudar lógica e teologia.
- Mudou-se para Tunbridge Wells, Kent, por volta de 1734, onde permaneceu como ministro da Capela do Monte Sião até 1752.
- Thomas Bayes é conhecido por ter formulado o caso especial do teorema de Bayes (artigo publicado por um de seus pupilos após sua morte).
- Bayes foi eleito membro da Royal Society em 1742.

Introdução

Thomas Bayes (1702-1761)

Teorema de Bayes

Considere um espaço de probabilidade $(\Omega, \mathcal{A}, P)$ com eventos $A \subseteq \Omega, A \in \mathcal{A}$, cuja probabilidade de ocorrência é $P(A)$. Seja $A_1, A_2, \ldots, A_n$ partição finita de Ω, com

$$P(A_i) > 0, \quad A_i \cap A_j = \emptyset, \forall i \neq j, \quad \bigcup_i A_i = \Omega$$

B um outro evento qualquer com $P(B) > 0$, que pode ser escrito como

$$B = \bigcup_i (A_i \cap B).$$

Logo, $P(B) = \sum_i P(A_i \cap B) = \sum_i P(B \mid A_i)P(A_i)$, e pelas propriedades de probabilidades condicionais, o Teorema de Bayes é dado por

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_i P(B \mid A_i)P(A_i)}, \ i = 1, \ldots, n.$$

Processo inferencial

No cenário inferencial Bayesiano usa-se o Teorema de Bayes.

`Distribuição a posteriori:` representada pela função (f.p. ou f.d.p.) $h(\theta \mid \mathbf{x})$.

`Distribuição a priori:` substituímos $P(A_i)$ pela distribuição `a priori`, representada pela função (f.p. ou f.d.p.) $h(\theta)$

`Função de verossimilhança:` substitui-se $P(B \mid A_i)$ por $L(\theta \mid \mathbf{x})$.

Logo,

$$h(\theta \mid \mathbf{x}) = \frac{L(\theta \mid \mathbf{x}) \times h(\theta)}{\sum_{\theta} L(\theta \mid \mathbf{x}) \times h(\theta)}, \quad \text{se } \theta \text{ for discreto.}$$

$$h(\theta \mid \mathbf{x}) = \frac{L(\theta \mid \mathbf{x}) \times h(\theta)}{\int_{\theta} L(\theta \mid \mathbf{x}) \times h(\theta) d\theta}, \quad \text{se } \theta \text{ for contínuo.}$$

Processo inferencial

A distribuição `a posteriori` é comumente apresentada,

$$h(\theta \mid \mathbf{x}) = \frac{L(\theta \mid \mathbf{x}) \times h(\theta)}{f(\mathbf{x})} \propto L(\theta \mid \mathbf{x}) \times h(\theta),$$

em que dizemos que a distribuição `a posteiori` é proporcional a $L(\theta \mid \mathbf{x})$ e $h(\theta)$ a menos de uma constante $f(\mathbf{x})$.

A constante $f(\mathbf{x})$ é chamada de `distribuição marginal dos dados` ou `preditiva`.

Vantagens

- Permite a inclusão de informação subjetiva relevante durante o processo inferencial.
- Fornece maior flexibilidade na modelagem dos dados, especialmente para conjuntos de dados com estruturas mais complexas.
- O fato da inferência ser baseada em uma distribuição de probabilidade contribui para a robustez e interpretabilidade dos resultados.
- Permite a incorporação sequencial de informação de maneira bem natural

Inferência Bayesiana no modelo Bernoulli

Seja $X_1, X_2, \ldots, X_n$ uma a.a. em que $X \sim Ber(\theta)$. Dessa forma, temos que a função de verossimilhança é dada por

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum_i X_i}(1-\theta)^{n-\sum_i X_i}.$$

Neste caso, nosso interesse consiste em inferir sobre θ através da distribuição `a posteriori`.

Note que θ é uma probabilidade, ou seja, $\theta \in [0, 1]$.

Logo, devemos pensar em uma distribuição `a priori` que esteja nesse suporte.

Inferência Bayesiana no modelo Bernoulli

No caso, Bernoulli, usamos a distribuição Beta θ, ou seja, $\theta \sim Beta(\alpha, \beta)$, $\alpha > 0, \beta > 0$, em que

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \theta \in [0,1].$$

Determine a distribuição `a posteriori`!

Como θ é contínuo, temos

$$h(\theta \mid \mathbf{x}) = \frac{\theta^{\sum_i X_i}(1-\theta)^{n-\sum_i X_i} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \theta^{\sum_i X_i}(1-\theta)^{n-\sum_i X_i} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta}$$

Inferência Bayesiana no modelo Bernoulli

$$h(\theta \mid \mathbf{x}) = \frac{\theta^{\sum_i X_i}(1-\theta)^{n-\sum_i X_i} \times \theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \theta^{\sum_i X_i}(1-\theta)^{n-\sum_i X_i} \times \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta}$$

- Note que no denominador temos o núcleo de uma distribuição $Beta(\alpha + \sum_i X_i, \beta + n - \sum_i X_i)$, logo,

$$h(\theta \mid \mathbf{x}) = \frac{\theta^{\sum_i X_i + \alpha - 1}(1-\theta)^{n-\sum_i X_i + \beta - 1}}{\int_0^1 \theta^{\sum_i X_i + \alpha - 1}(1-\theta)^{n-\sum_i X_i + \beta - 1} d\theta}$$

$$Beta\,(\alpha + \sum_i X_i, \beta + n - \sum_i X_i)$$

Inferência Bayesiana no modelo Bernoulli

$$h(\theta \mid \mathbf{x}) = \frac{\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\sum_i X_i + \alpha)\Gamma(\beta + n - \sum_i X_i)}\theta^{\sum_i X_i + \alpha - 1}(1-\theta)^{n - \sum_i X_i + \beta - 1}}{\underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\sum_i X_i + \alpha)\Gamma(\beta + n - \sum_i X_i)}\theta^{\sum_i X_i + \alpha - 1}(1-\theta)^{n - \sum_i X_i + \beta - 1} d\theta}_{=1}}$$

Portanto, a distribuição `a posteriori` é dada por

$$h(\theta \mid \mathbf{x}) = \frac{\Gamma(\alpha+\beta+n)}{\left(\Gamma(\sum_i X_i + \alpha)\Gamma(\beta + n - \sum_i X_i)\right)}\theta^{\sum_i X_i + \alpha - 1}(1-\theta)^{n - \sum_i X_i + \beta - 1}$$

ou seja, $\theta \mid \mathbf{x} \sim Beta(\sum_i X_i + \alpha, \beta + n - \sum_i X_i)$.

Inferência Bayesiana no modelo Bernoulli

Distribuição Beta:

Sabemos da distribuição Beta que

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Logo, como $\theta \mid \mathbf{x} \sim Beta\left(\sum_i X_i + \alpha, \beta + n - \sum_i X_i\right)$, a média `a posteriori`:

$$E(\theta \mid \mathrm{x}) = \frac{\alpha + \sum_{i=1}^{n} X_i}{(\alpha + \sum_{i=1}^{n} X_i) + (\beta + n - \sum_{i=1}^{n} X_i)} = \frac{\alpha}{\alpha + \beta + n} E(\theta) + \frac{n}{\alpha + \beta + n} \bar{X}.$$

Exemplo

> Suponha que em uma amostra de 12 indivíduos tenham 9 fumantes. Suponha também que inicialmente você sabe que 50% das pessoas são fumantes. Como a informação amostral modifica sua informação individual?

- `Solução:` Temos que $E(\theta) = 0,5 \Rightarrow \theta \sim Beta(\alpha = 5, \beta = 5)$. Além disso, os valores amostrais indicam que $n = 12$ e $\sum_{i=1}^{12} X_i = 9$.

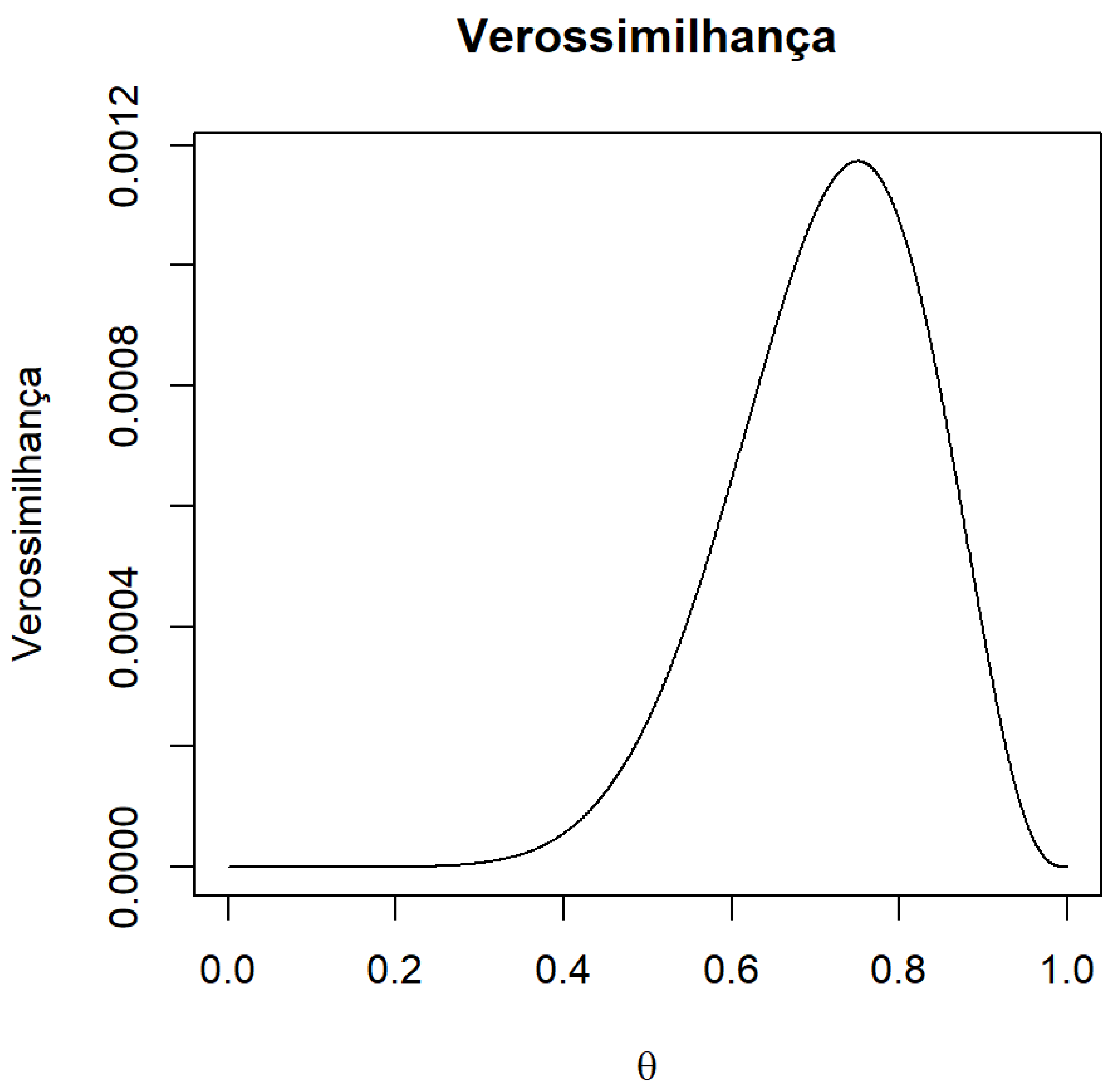
A distribuição `a posteriori` é dada por

$$\theta \mid \mathbf{x} \sim Beta(5 + 9 = 14; 5 + 12 - 9 = 8)$$

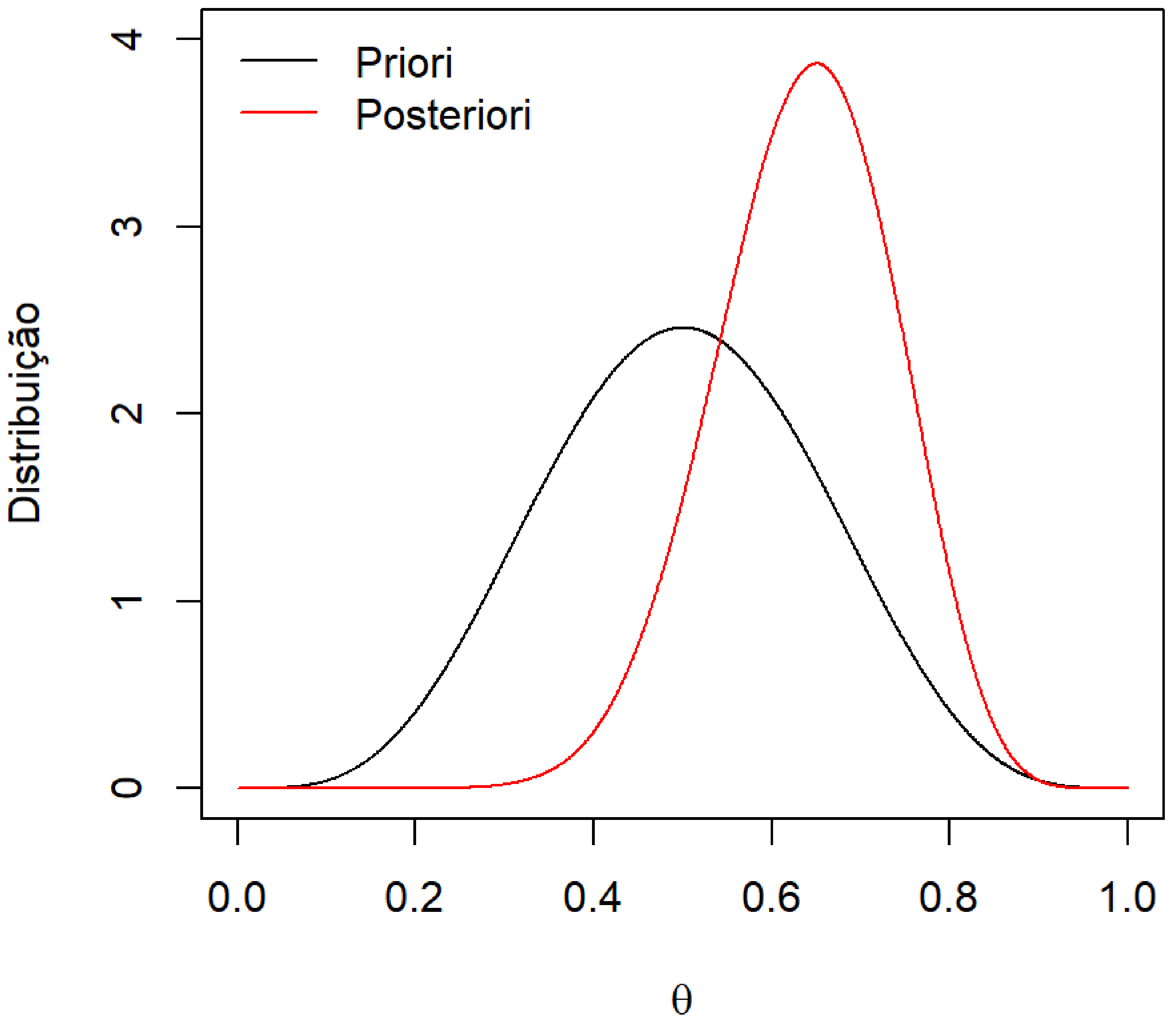
Logo, $E(\theta \mid \mathbf{x}) = 0,6363$ e $Var(\theta \mid \mathbf{x}) = 0,0101$.

`A informação amostral modifica a informação individual`.

Exemplo



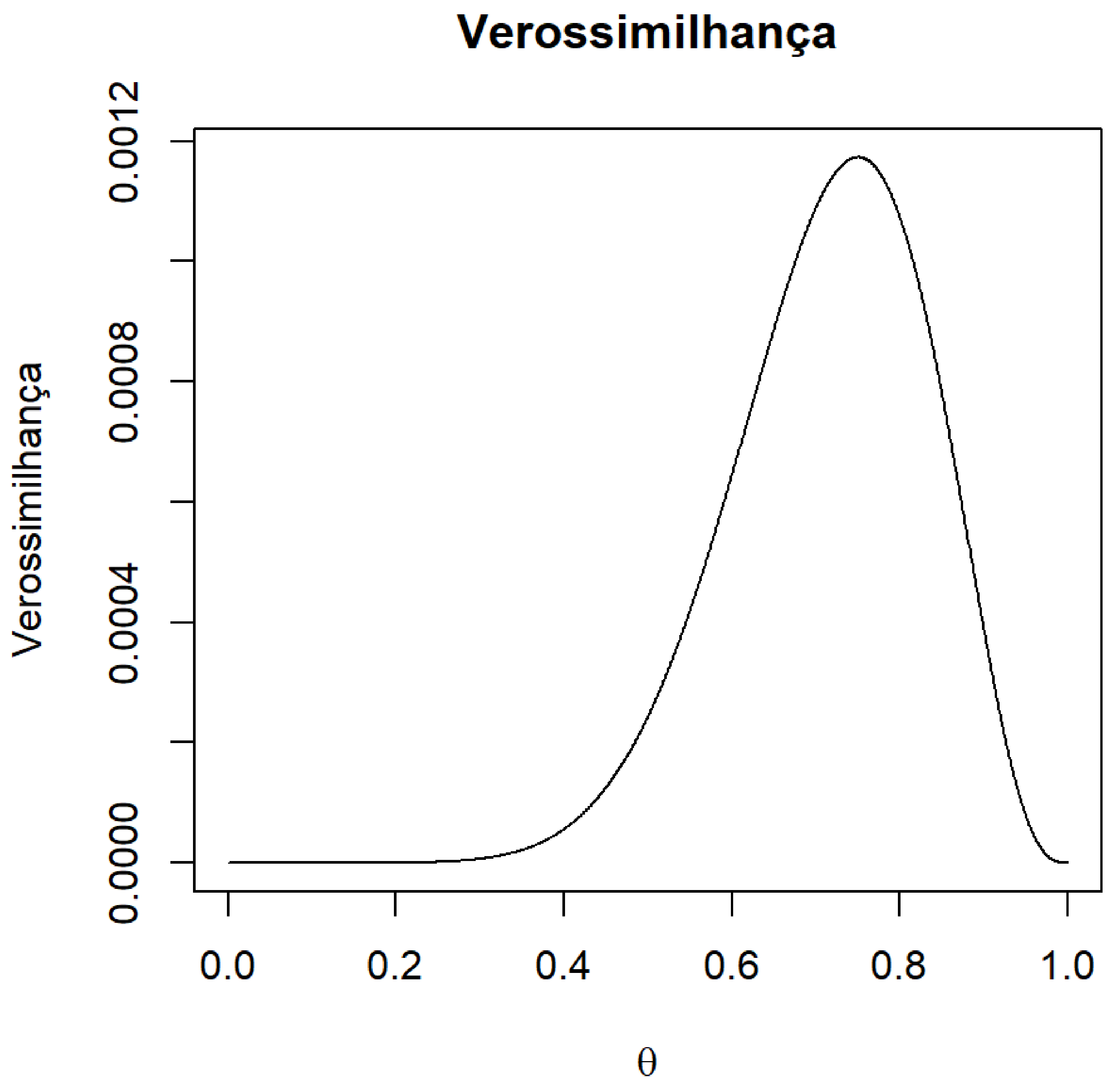
Verossimilhança
Verossimilhança
0.0000
0.0004
0.0008
0.0012
0.0
0.2
0.4
0.6
0.8
1.0
θ

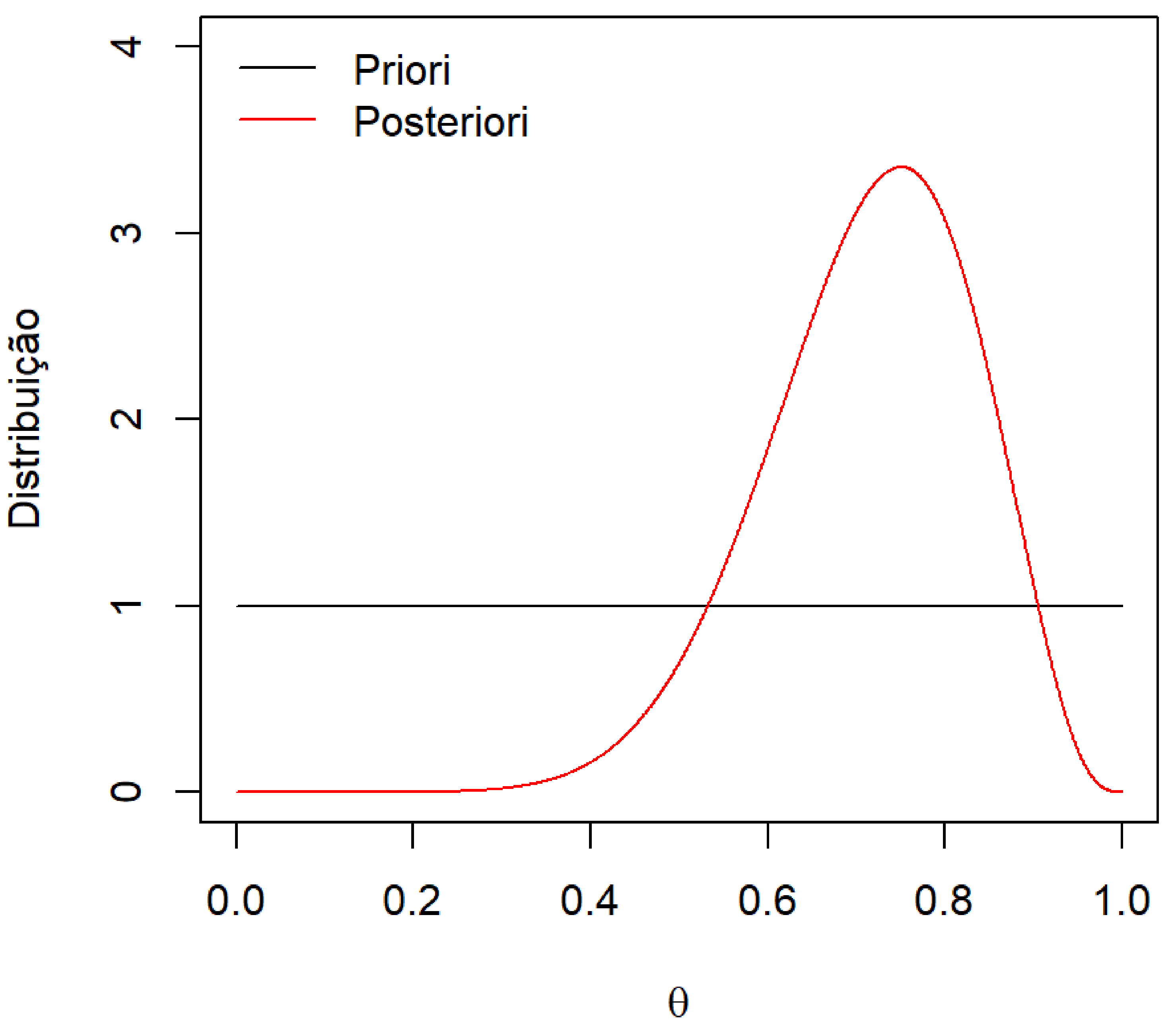


Priori
Posteriori
Distribuição
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
θ

Exemplo

- Distribuição `a priori` $\theta \sim Beta(1, 1)$.





Inferência Bayesiana no modelo Normal

- Seja $X_1, X_2, \ldots, X_n \mid \mu \sim N(\mu, \sigma^2)$ uma a.a. em que $\mu \in \mathrm{R}$ é desconhecido e $\sigma^2 \in \mathrm{R}^+$ é conhecido. Objetivo de determinar a distribuição `a posteriori` para μ.
- Dessa forma, temos que a função de verossimilhança é dada por

$$L(\mu \mid \mathbf{x}) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2} \right)^{0,5} \exp\left\{ -\frac{1}{2\sigma^2}(X_i - \mu)^2 \right\}.$$

- Assuma, a distribuição `a priori` para $\mu \sim N(a, b^2)$, da seguinte forma,

$$h(\mu) = \left(\frac{1}{2\pi b^2} \right)^{0,5} \exp\left\{ -\frac{1}{2b^2}(\mu - a)^2 \right\}.$$

Inferência Bayesiana no modelo Normal

Dessa forma, usando a relação do caso contínuo, temos que a distribuição `a posteriori` para μ é expressa da seguinte forma

$$h(\mu \mid \mathbf{x}) = \frac{L(\mu \mid \mathbf{x}) \times h(\mu)}{\int_\mu L(\mu \mid \mathbf{x}) \times h(\mu) d\mu} = \left(\frac{1}{2\pi B^2}\right)^{0,5} \exp\left\{-\frac{1}{2B^2}(\mu - A)^2\right\}, \quad -\infty < \mu < \infty.$$

com

$$A = E(\mu \mid \mathbf{x}) = \hat{\mu} = \frac{\frac{1}{b^2}a + \frac{n}{\sigma^2}\bar{X}}{\frac{1}{b^2} + \frac{n}{\sigma^2}} \quad \text{e} \quad B^2 = \frac{1}{\frac{1}{b^2} + \frac{n}{\sigma^2}}.$$

Logo, $\mu \mid \mathbf{x} \sim N(A, B^2)$.

Exemplo

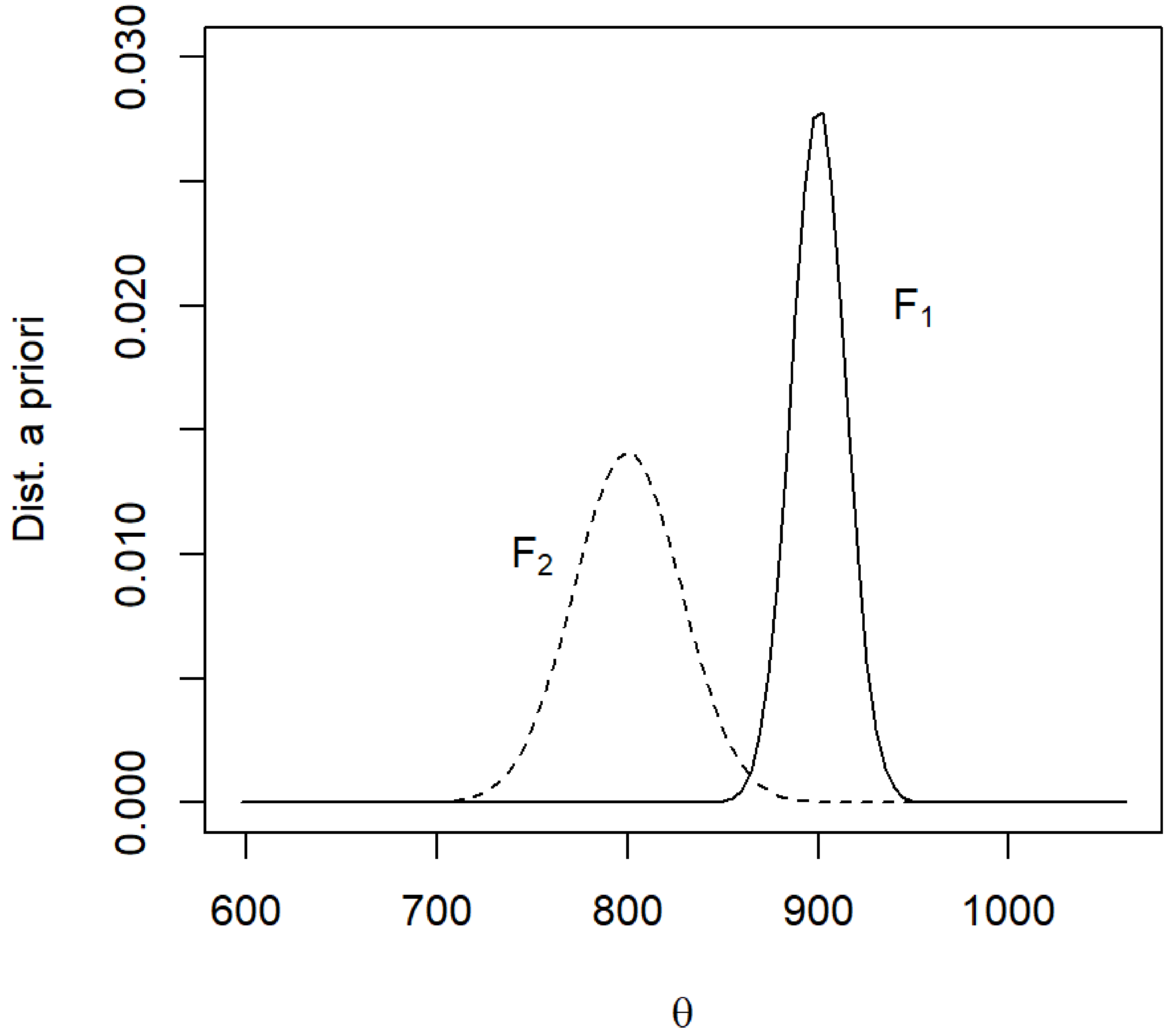
> Suponha no exemplo anterior que μ é uma constante física, de uma característica $X \sim N(\mu, 40^2)$ e dois físicos F_1 e F_2 pretendem estimá-la. Suponha que F_1 atribui a priori $N(900, 20^2)$ e F_2 atribui $N(800, 80^2)$. Observa-se que F_1 tem maior precisão a priori. Após observar uma amostra de tamanho $n = 50$ com $\bar{X} = 870$, tem-se:

$$F_1 \Longrightarrow \mu \mid \mathbf{x} \sim N(872.2, 5.44^2)$$

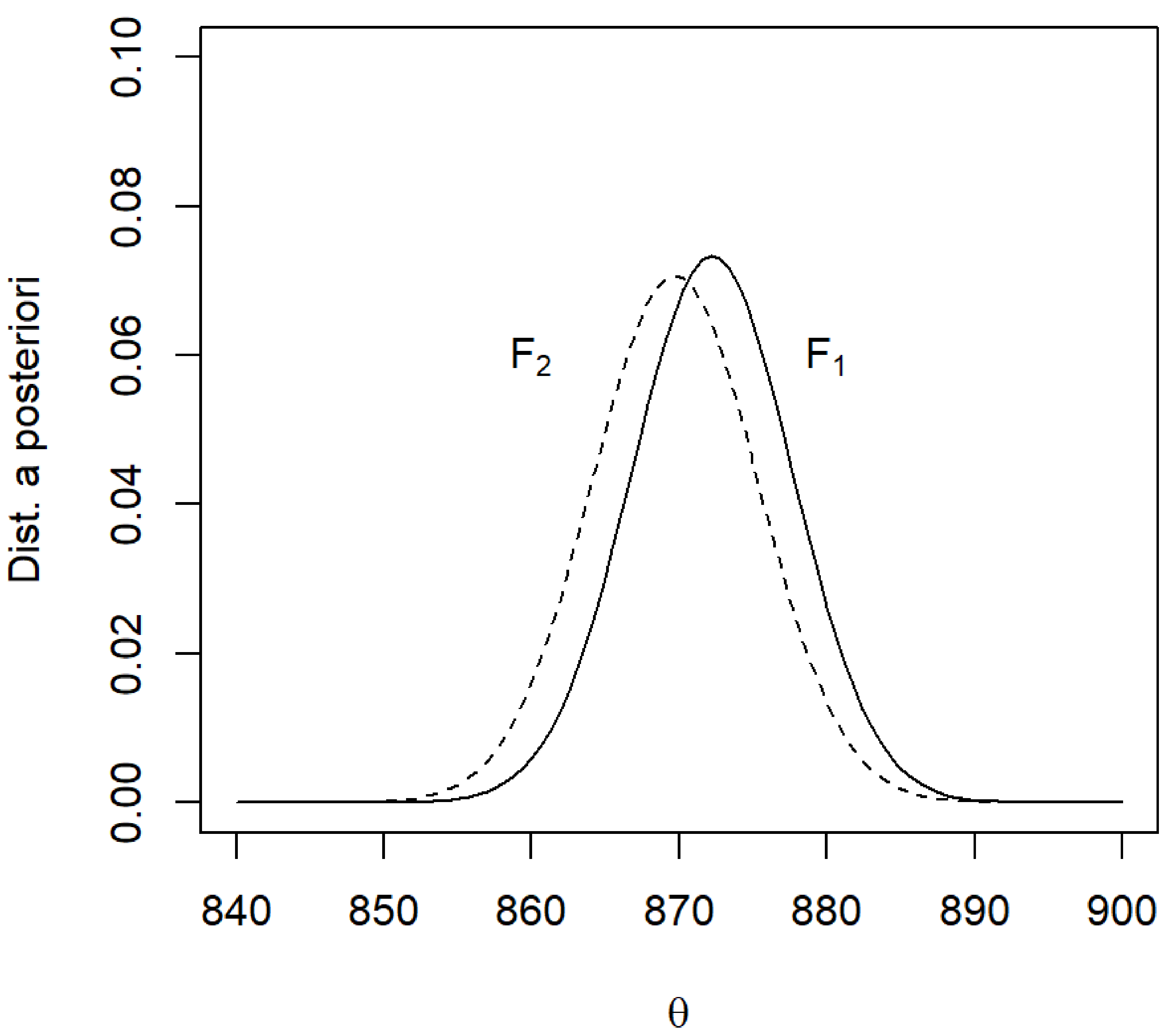
$$F_2 \Longrightarrow \mu \mid \mathbf{x} \sim N(869.7, 5.64^2)$$

- Observa-se que as distribuições a posteriori dos 2 físicos pouco diferem, pois a informação amostral atenuou o afastamento inicial entre as prioris.

Exemplo



Dist. a priori
0.000
0.010
0.020
0.030
600
700
800
900
1000
θ
F_1
F_2



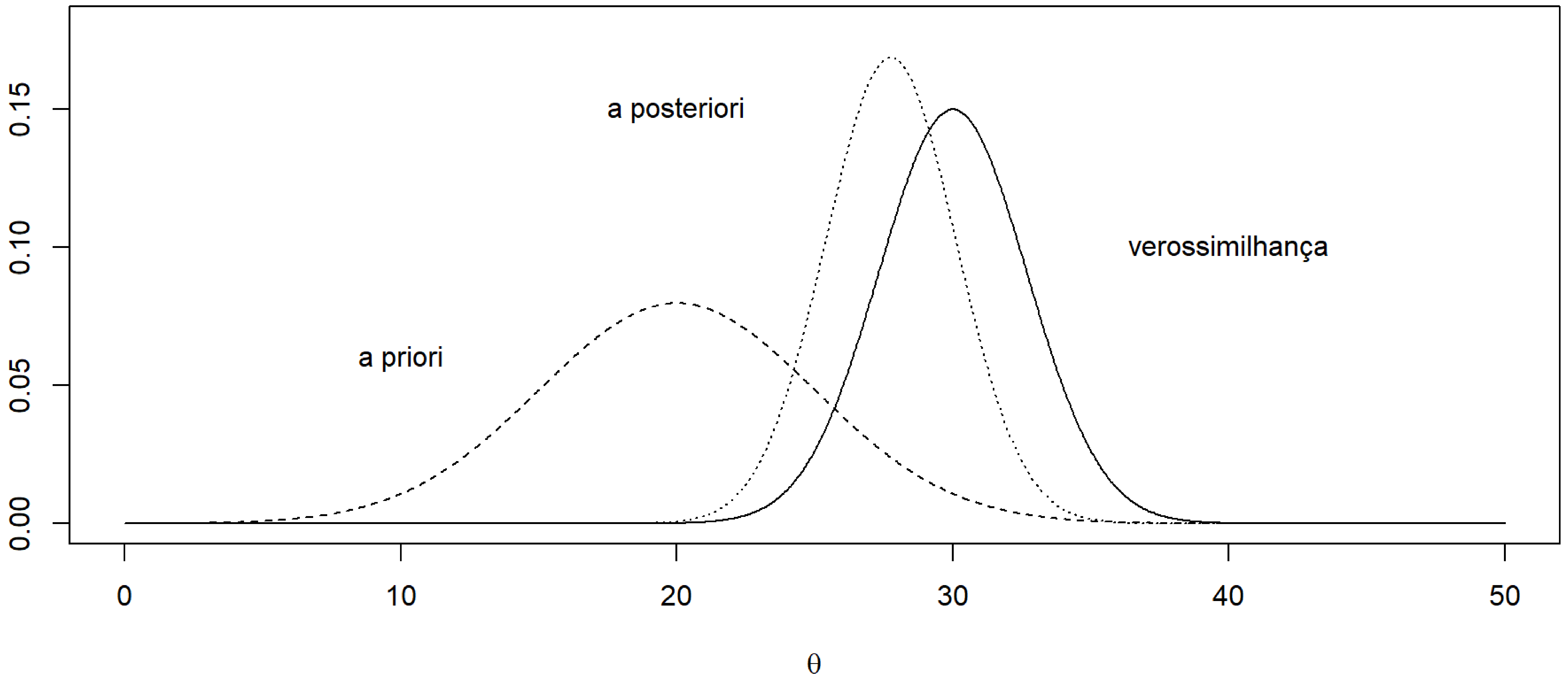
Dist. a posteriori
0.00
0.02
0.04
0.06
0.08
0.10
840
850
860
870
880
890
900
θ
F_1
F_2

Exemplo

- Considere uma amostral causal de dimensão $n = 5, X_1, X_2, \ldots, X_5 \mid \mu \sim N(\mu, \sigma^2)$ com $\sigma = 6$ e adimita-se que $\mu \sim N(a, b^2)$, com $a = 20$ e $b = 5$. Objetivo de determinar a distribuição `a posteriori` para μ.
- Suponha que $\bar{X} = 30$. Logo, a distribuição `a posteriori` é dada por

$$\mu \mid \mathbf{x} \sim N(27.764, 2.364^2). \quad \text{(faça as contas!!)}$$

Exemplo



a posteriori
verossimilhança
a priori
0.00
0.05
0.10
0.15
0
10
20
30
40
50
θ