# Distribuições não-informativas, estimação, testes de hipósteses

**Prof. Vinícius D. Mayrink**
EST088 - Inferência Bayesiana

Sala: 4073 Email: vdm@est.ufmg.br

1$^o$ semestre de 2025

# 4.1 - Distribuições não infomativas

Algumas vezes a distribuição *a priori* conjugada não é adequada pois:

- ela pode não existir.
- ela pode não representar muito bem as informações iniciais sobre o parâmetro.

Uma alternativa para esta situação é utilizar distribuições *a priori* de referência. Neste caso, a análise estatística é denominada **análise Bayesiana de referência**.

Duas justificativas para optarmos por este caminho:

- não há qualquer informação inicial sobre o parâmetro.
- deseja-se uma análise “objetiva”, ou seja, sem a subjetividade das distribuições de probabilidade expressando nossa informação inicial.

**Princípio da razão insuficiente (postulado de Bayes-Laplace):** Se não há informação para diferenciar entre os distintos valores de $\theta$, deve-se dar a mesma probabilidade a todos os valores.

Aplicar este princípio implica na utilização de uma distribuição *a priori* uniforme para $\theta$. A forma como essa uniforme será empregada dependerá do espaço paramétrico $\Theta$.

Exemplo: Suponha os seguintes espaços paramétricos

1. Se $\Theta = \{\theta_1, \theta_2, \cdots, \theta_q\}$, temos $f_\theta(\theta_i) = 1/q$, para $i = 1, \cdots, q$.
2. Se $\Theta = (a, b)$, temos $f_\theta(\theta) = 1/(b-a)$ para $\theta \in (a, b)$.
3. Se $\Theta = \{\theta_1, \theta_2, \theta_3, \cdots\}$, temos $f_\theta(\theta_i) = C$, para $i = 1, 2, 3, \cdots$ e $C$ constante. Podemos escrever também $f_\theta(\theta_i) \propto 1$.
   Note que $\sum_{i=1}^{\infty} f_\theta(\theta_i) = \sum_{i=1}^{\infty} 1 = \infty$.
4. Se $\Theta = \mathbb{R}$, temos $f_\theta(\theta) = C$, para $\theta \in \mathbb{R}$.
   Novamente, escreve-se $f_\theta(\theta) \propto 1$.
   Note que $\int_{-\infty}^{\infty} f_\theta(\theta)\, d\theta = \int_{-\infty}^{\infty} 1\, d\theta = 1 \times \theta|_{-\infty}^{\infty} = \text{indefinido}$.

Nos casos 3 e 4 (slide anterior) as distribuições *a priori* sugeridas com o uso do princípio da razão insuficiente são **distribuições impróprias**. Isto significa que não são distribuições de probabilidade propriamente ditas.

O uso de distribuições *a priori* impróprias não é um problema se essa escolha gerar uma distribuição *a posteriori* própria.

Uma crítica ao método de Bayes-Laplace é que a distribuição *a priori* <u>não é invariante</u> sob transformações (reparametrizações).

Por exemplo, $\theta \sim U(0,1)$ e deseja-se investigar o novo parâmetro $\phi = \ln(\theta)$. Se não há informação sobre $\theta$, é natural pensar que também não há informação sobre $\phi$. Neste raciocínio, esperava-se que a especificação *a priori* sobre $\phi$ também fosse uniforme. Entretanto, temos

$$f_\phi(\phi) \; = \; f_\theta(\theta = e^\phi) \times \tfrac{d}{d\phi} e^\phi \; = \; 1 \times \exp\{\phi\} \; = \; \exp\{\phi\} \text{ para } \phi \in (-\infty, 0).$$

Claramente a distribuição de $\phi$ não é uniforme.

A crítica detalhada no último slide, motivou o estudioso inglês Harold Jeffreys a propor em 1946 uma outra forma de construir distribuições *a priori* do tipo "não-informativas".

## Distribuição a priori de Jeffreys

Iniciamos com o caso unidimensional, isto é, $\theta$ é um escalar.

Definição: Seja $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$ a função de verossimilhança associada a algum experimento. A **informação de Fisher** correspondente a esta verossimilhança é dada por

$$\mathtt{I}(\theta) = E_{\boldsymbol{Y}|\theta}\left[\left(\frac{d}{d\theta}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right)^2\right].$$

O termo $\frac{d}{d\theta}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$ é comumente chamado de função *score*.

Pode-se mostrar que $E_{\boldsymbol{Y}|\theta}[\frac{d}{d\theta}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})] = 0$, portanto, a $\mathtt{I}(\theta)$ é a variância da função *score*.

Se $\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$ for 2 vezes derivável e se algumas condições de regularidade forem atendidas (fato que ocorrerá para as distribuições da família exponencial), teremos a seguinte forma alternativa da informação de Fisher:

$$\mathtt{I}(\theta) = -E_{\boldsymbol{Y}|\theta}\left[\frac{d^2}{d^2\theta}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right].$$

Se existir independência condicional para $Y_1, Y_2, \cdots, Y_n$ dado $\theta$, podemos ainda escrever a informação de Fisher como segue:

$$\mathtt{I}(\theta) = n\ E_{Y_i|\theta}\left[\left(\frac{d}{d\theta}\ln f_{Y_i|\theta}(y_i)\right)^2\right] \quad \text{ou} \quad \mathtt{I}(\theta) = -n\ E_{Y_i|\theta}\left[\frac{d^2}{d^2\theta}\ln f_{Y_i|\theta}(y_i)\right].$$

Note que nas expressões acima estamos avaliando a f.d.p. ou f.m.p. de $Y_i|\theta$ e não a distribuição conjunta de $\boldsymbol{Y}|\theta$.

$\mathtt{I}(\theta)$ mede quanta informação a variável observável $Y_i$ carrega sobre o desconhecido $\theta$ (indexando a distribuição de $Y_i$).
$\mathtt{I}(\theta)$ grande $\Rightarrow$ muita informação e $\mathtt{I}(\theta)$ pequeno $\Rightarrow$ pouca informação.

Definição: No caso unidimensional, $\mathtt{I}(\theta)$ é um escalar.
A distribuição *a priori* de Jeffreys é dada por $f_\theta(\theta) \propto \sqrt{\mathtt{I}(\theta)}$.

Obs.: A especificação acima é proporcional à raiz da informação de Fisher, portanto, sempre que estivermos trabalhando com uma amostra $\boldsymbol{Y}$ i.i.d. dado $\theta$ poderemos descartar o termo "$n$" que multiplica a esperança, isto é:

$$f_\theta(\theta) \propto \sqrt{n\ E_{Y_i|\theta}\left[\left(\frac{d}{d\theta}\ln f_{Y_i|\theta}(y_i)\right)^2\right]} \quad \text{ou} \quad f_\theta(\theta) \propto \sqrt{-n\ E_{Y_i|\theta}\left[\frac{d^2}{d^2\theta}\ln f_{Y_i|\theta}(y_i)\right]}.$$

Descartando $\sqrt{n}$ teremos

$$f_\theta(\theta) \propto \sqrt{E_{Y_i|\theta}\left[\left(\frac{d}{d\theta}\ln f_{Y_i|\theta}(y_i)\right)^2\right]} \quad \text{ou} \quad f_\theta(\theta) \propto \sqrt{-E_{Y_i|\theta}\left[\frac{d^2}{d^2\theta}\ln f_{Y_i|\theta}(y_i)\right]}.$$

É possível mostrar que este tipo de especificação *a priori* é invariante sob transformações aplicadas ao parâmetro alvo $\theta$.

Exemplo: Adote o modelo $Y_i|\theta \sim \text{Exp}(\theta)$. A amostra $Y_1, Y_2, \cdots, Y_n$ é tratada como condicionalmente independente dado $\theta > 0$. Iremos determinar a distribuição *a priori* de Jeffreys neste contexto.

f.d.p.: $f_{Y_i|\theta}(y_i) \; = \; \theta \; \exp\{-\theta y_i\}$.

Log-densidade: $\ln f_{Y_i|\theta}(y_i) \; = \; \ln(\theta) \; - \; \theta y_i$.

Derivando $1^a$ vez: $\frac{d}{d\theta} \ln f_{Y_i|\theta}(y_i) \; = \; (1/\theta) - y_i$.

Derivando $2^a$ vez: $\frac{d^2}{d^2\theta} \ln f_{Y_i|\theta}(y_i) \; = \; -1/\theta^2$.

$$\mathtt{I}(\theta) \; = \; -E_{Y_i|\theta}\left(-1/\theta^2\right) \; = \; E_{Y_i|\theta}\left(1/\theta^2\right) \; = \; \int_0^\infty (1/\theta^2)\, \theta \exp\{-\theta y_i\} dy_i \; = \; 1/\theta^2.$$

Distribuição *a priori* de Jeffreys para $\theta$: $f_\theta(\theta) \propto \sqrt{1/\theta^2} \; \propto \; 1/\theta$ para $\theta > 0$.

Este resultado tem semelhança com o núcleo da $\text{Ga}(a = 0, b = 0)$:

$$\frac{b^a}{\Gamma(a)} \, \theta^{a-1} \; \exp\{-b\theta\} \; \propto \; \theta^{a-1} \; \exp\{-b\theta\} \; \propto \; \theta^{0-1} \; \exp\{-0\theta\} \; \propto \; 1/\theta.$$

Exemplo (continuação):

A distribuição *a priori* de Jeffreys determinada no último slide é imprópria, pois

$\int_0^\infty f_\theta(\theta)\, d\theta \;=\; \int_0^\infty C/\theta\, d\theta \;=\; C\; \ln(\theta)|_0^\infty =$ indefinido.

O termo $C$ é uma constante positiva.

Entretanto, a distribuição *a posteriori* é própria:

$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;\propto\; f_{\boldsymbol{Y}|\theta}(\boldsymbol{\theta})\; f_\theta(\theta) \;\propto\; \theta^n\; \exp\{-\theta \sum_{i=1}^n y_i\}\; C/\theta.$

$\propto\; \theta^{n-1}\; \exp\{-\theta \sum_{i=1}^n y_i\}.$

A expressão acima é o núcleo da Ga$(a^*, b^*)$ com $a^* = n$ e $b^* = \sum_{i=1}^n y_i$.

Exemplo: Considere $Y_i|\theta \sim \text{Bernoulli}(\theta)$. Admita independência condicional de $Y_1, Y_2, \cdots, Y_n$ dado $\theta \in (0,1)$. A distribuição *a priori* de Jeffreys é como segue.

f.d.p.: $f_{Y_i|\theta}(y_i) \;=\; \theta^{y_i}\,(1-\theta)^{1-y_i}$.

Log-densidade: $\ln f_{Y_i|\theta}(y_i) \;=\; y_i \ln(\theta) \;+\; (1-y_i)\ln(1-\theta)$.

Derivando $1^a$ vez: $\frac{d}{d\theta}\ln f_{Y_i|\theta}(y_i) \;=\; \dfrac{y_i}{\theta} \;+\; \dfrac{1-y_i}{1-\theta}(-1)$.

Derivando $2^a$ vez: $\frac{d^2}{d^2\theta}\ln f_{Y_i|\theta}(y_i) \;=\; -\dfrac{y_i}{\theta^2} \;-\; \dfrac{1-y_i}{(1-\theta)^2}$.

$$\mathtt{I}(\theta) \;=\; -E_{Y_i|\theta}\left[-\tfrac{y_i}{\theta^2} - \tfrac{1-y_i}{(1-\theta)^2}\right] \;=\; E_{Y_i|\theta}\left[\tfrac{y_i}{\theta^2} + \tfrac{1-y_i}{(1-\theta)^2}\right]$$

$$=\; \tfrac{1}{\theta^2}E[y_i] + \tfrac{1}{(1-\theta)^2}E[1-y_i] \;=\; \tfrac{1}{\theta^2}\theta + \tfrac{1}{(1-\theta)^2}(1-\theta) \;=\; \tfrac{1}{\theta} + \tfrac{1}{(1-\theta)} \;=\; \tfrac{1}{\theta(1-\theta)}.$$

Exemplo (continuação):

Distribuição *a priori* de Jeffreys para $\theta$:

$$f_\theta(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}} \propto \theta^{-1/2}(1-\theta)^{-1/2} \quad \text{para} \quad \theta \in (0,1).$$

Perceba que a expressão final acima é o núcleo da Beta(1/2,1/2).

Conclusão: A distribuição *a priori* de Jeffreys no caso Bernoulli é própria.

Exercício: Admitindo independência condicional dado o parâmetro alvo, obtenha a distribuição *a priori* de Jeffreys nos casos abaixo:

- $Y_i|\mu,\sigma^2 \sim N(\mu,\sigma^2)$ com $\sigma^2$ conhecido.
- $Y_i|\mu,\sigma^2 \sim N(\mu,\sigma^2)$ com $\mu$ conhecido.

Verifique também se a distribuição *a posteriori* é própria.

Investigaremos agora a distribuição *a priori* de Jeffreys no caso multidimensional.

Considere que $\theta = \{\theta_1, \cdots, \theta_q\}$ é um vetor de parâmetros.

Neste caso, $\mathtt{I}(\theta)$ é a **matriz de informação de Fisher** de tamanho $q \times q$.

A $i$-ésima linha e $j$-ésima coluna desta matriz contém

$$\mathtt{I}(\theta)_{ij} \;=\; E_{\boldsymbol{Y}|\theta}\left[\left(\frac{d}{d\theta_i}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right)\;\left(\frac{d}{d\theta_j}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right)\right].$$

Se $i = j$, o resultado será $\mathtt{I}(\theta)_{ii} \;=\; E_{\boldsymbol{Y}|\theta}\left[\left(\frac{d}{d\theta_i}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right)^2\right]$.

Diante de algumas condições de regularidade (que são atendidas pelas distribuições da família exponencial), escreve-se

$$\mathtt{I}(\theta)_{ij} \;=\; -E_{\boldsymbol{Y}|\theta}\left[\frac{d^2}{d\theta_i d\theta_j}\ln f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\right].$$

Quando a independência condicional for válida para $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ dado $\theta$, podemos ainda escrever:

- $\mathtt{I}(\theta)_{ij} \; = \; n \; E_{\mathbf{Y}|\theta} \left[ \left( \frac{d}{d\theta_i} \ln f_{Y_k|\theta}(y_k) \right) \; \left( \frac{d}{d\theta_j} \ln f_{Y_k|\theta}(y_k) \right) \right].$
- $\mathtt{I}(\theta)_{ij} \; = \; -n \; E_{\mathbf{Y}|\theta} \left[ \frac{d^2}{d\theta_i d\theta_j} \ln f_{Y_k|\theta}(y_k) \right].$

Nas expressões acima, $k \in \{1, 2, \cdots, n\}$ indica de forma genérica qualquer uma das variáveis em $\boldsymbol{Y}$.

Definição: No caso multidimensional, $\mathtt{I}(\theta)$ é um matriz.

A distribuição *a priori* de Jeffreys é dada por $\quad f_\theta(\theta) \propto \sqrt{|\mathtt{I}(\theta)|}$.

Dentro da raiz, temos o determinante de $\mathtt{I}(\theta)$.

Exemplo: Considere $Y_i|\theta \sim N(\theta_1, \theta_2)$, tal que $\theta = (\theta_1, \theta_2)^\top$ contém a média e a variância da Normal. Admita independência condicional de $Y_1, Y_2, \cdots, Y_n$ dado $\theta$. A distribuição *a priori* de Jeffreys é obtida a seguir para este caso bidimensional.

f.d.p.: $f_{Y_i|\theta}(y_i) \;=\; (2\pi\,\theta_2)^{-1/2}\; \exp\{-\frac{1}{2\theta_2}(y_i^2 - 2\theta_1 y_i + \theta_1^2)\}.$

Log-densidade: $\ln f_{Y_i|\theta}(y_i) \;=\; -\frac{1}{2}[\ln(2\pi) + \ln(\theta_2)] - \frac{1}{2\theta_2}(y_i^2 - 2\theta_1 y_i + \theta_1^2).$

Derivação em $\theta_1$: $\frac{d}{d\theta_1} \ln f_{Y_i|\theta}(y_i) \;=\; -\frac{1}{2\theta_2}(-2y_i + 2\theta_1) \;=\; -\frac{1}{\theta_2}(-y_i + \theta_1).$

Derivação dupla em $\theta_1$: $\frac{d^2}{d\theta_1^2} \ln f_{Y_i|\theta}(y_i) \;=\; -1/\theta_2.$

Derivação em $\theta_1$ e $\theta_2$: $\frac{d^2}{d\theta_1 d\theta_2} \ln f_{Y_i|\theta}(y_i) \;=\; (-y_i + \theta_1)/\theta_2^2.$

Derivação em $\theta_2$: $\frac{d}{d\theta_2} \ln f_{Y_i|\theta}(y_i) = -\frac{1}{2\theta_2} - \frac{1}{2}(y_i^2 - 2\theta_1 y_i + \theta_1^2)\,(-1/\theta_2^2)$.

Derivação dupla em $\theta_2$:
$$\begin{aligned}\frac{d^2}{d\theta_2^2} \ln f_{Y_i|\theta}(y_i) &= \frac{1}{2\theta_2^2} - (y_i^2 - 2\theta_1 y_i + \theta_1^2)/\theta_2^3 \\ &= \frac{1}{2\theta_2^2} - (y_i - \theta_1)^2/\theta_2^3\end{aligned}$$

Derivação em $\theta_2$ e $\theta_1$: $\frac{d^2}{d\theta_2 d\theta_1} \ln f_{Y_i|\theta}(y_i) = \frac{1}{2\theta_2^2}(-2y_i + 2\theta_1) = (-y_i + \theta_1)/\theta_2^2$.

Perceba que $\frac{d^2}{d\theta_1 d\theta_2} = \frac{d^2}{d\theta_2 d\theta_1}$, ou seja, a ordem da derivação não importa.

Matriz de informação de Fisher:

$$\mathtt{I}(\theta) = \begin{bmatrix} -n\ E_{Y_i|\theta}\left[\frac{d^2}{d\theta_1^2} \ln f_{Y_i|\theta}(y_i)\right] & -n\ E_{Y_i|\theta}\left[\frac{d^2}{d\theta_1 d\theta_2} \ln f_{Y_i|\theta}(y_i)\right] \\ -n\ E_{Y_i|\theta}\left[\frac{d^2}{d\theta_2 d\theta_1} \ln f_{Y_i|\theta}(y_i)\right] & -n\ E_{Y_i|\theta}\left[\frac{d^2}{d\theta_2^2} \ln f_{Y_i|\theta}(y_i)\right] \end{bmatrix}.$$

Matriz de informação de Fisher:

$$\mathtt{I}(\theta) \;=\; n \begin{bmatrix} -E_{Y_i|\theta}\left[-1/\theta_2\right] & -E_{Y_i|\theta}\left[(-Y_i+\theta_1)/\theta_2^2\right] \\ -E_{Y_i|\theta}\left[(-Y_i+\theta_1)/\theta_2^2\right] & -E_{Y_i|\theta}\left[\frac{1}{2\theta_2^2} - (Y_i-\theta_1)^2/\theta_2^3\right] \end{bmatrix}.$$

Note que:

- $-E_{Y_i|\theta}\left[(-Y_i+\theta_1)/\theta_2^2\right] \;=\; E_{Y_i|\theta}\left[Y_i/\theta_2^2\right] - \theta_1/\theta_2^2 \;=\; \theta_1/\theta_2^2 - \theta_1/\theta_2^2 \;=\; 0.$
- $-E_{Y_i|\theta}\left[\frac{1}{2\theta_2^2} - (Y_i-\theta_1)^2/\theta_2^3\right] \;=\; -\frac{1}{2\theta_2^2} + E_{Y_i|\theta}\left[(Y_i-\theta_1)^2\right]/\theta_2^3$

  $= \;-1/(2\theta_2^2) + Var_{Y_i|\theta}(Y_i)/\theta_2^3 \;=\; -1/(2\theta_2^2) + \theta_2/\theta_2^3 \;=\; -1/(2\theta_2^2) + 1/\theta_2^2$

  $= \;\dfrac{-1+2}{2\theta_2^2} \;=\; \dfrac{1}{2\theta_2^2}$

Matriz de informação de Fisher: $\mathtt{I}(\theta) \;=\; n \begin{bmatrix} 1/\theta_2 & 0 \\ 0 & 1/(2\theta_2^2) \end{bmatrix}.$

Determinante $|\mathtt{I}(\theta)| \;=\; n^2/(2\theta_2^3)$.

A distribuição *a priori* de Jeffreys conjunta será

$$f_\theta(\theta) \;=\; f_{\theta_1,\theta_2}(\theta_1,\theta_2) \;=\; \sqrt{\frac{n^2}{2\theta_2^3}} \;\propto\; \left(\frac{1}{\theta_2}\right)^{3/2}.$$

Obs.: No caso multidimensional, uma maneira alternativa de construir a distribuição *a priori* de Jeffreys conjunta é:

- assumir $\theta_1, \cdots, \theta_q$ independentes.
- usar o produto das distribuições *a priori* de Jeffreys marginais para determinar a conjunta, ou seja: $f_\theta(\theta) = \prod_{j=1}^q f_{\theta_j}(\theta_j) \propto \prod_{j=1}^q \sqrt{\mathtt{I}(\theta)_{jj}}$.

O contexto do caso normal é mostrado a seguir.

Exemplo: $Y_1, \cdots, Y_n$ dado $\theta = (\theta_1, \theta_2)^\top$ são condicionalmente independentes com $Y_i|\theta \sim N(\theta_1, \theta_2)$. Admita $\theta_1 \perp\!\!\!\perp \theta_2$.

Note pela diagonal principal de $\mathtt{I}(\theta)$ (slide anterior) que:

$f_{\theta_1}(\theta_1) \propto \sqrt{1/\theta_2} \propto 1$ e $f_{\theta_2}(\theta_2) \propto \sqrt{1/(2\theta_2^2)} \propto 1/\theta_2$.

Então: $f_{\theta_1,\theta_2}(\theta_1, \theta_2) \propto 1 \times 1/\theta_2 \propto 1/\theta_2$.

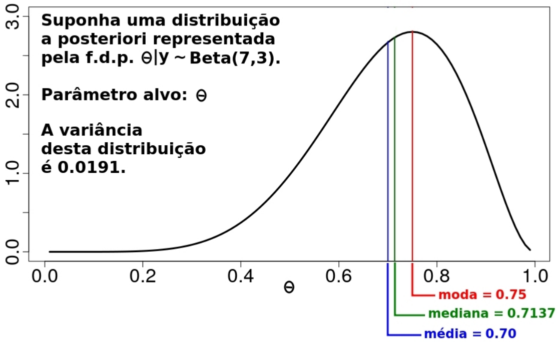
Atenção a 2 problemas relacionados ao uso da distribuição *a priori* de Jeffreys:

1. A análise via distribuição *a priori* Jeffreys viola o princípio da verossimilhança (rever este conceito em slide da Seção 2.2).
2. Se a distribuição *a priori* for imprópria, existe a possibilidade de gerar uma *posteriori* imprópria. Se a *priori* for própria, teremos sempre *posteriori* própria.

# 4.2 - Estimação

Uma vez que a distribuição *a posteriori* foi determinada, o próximo passo da análise Bayesiana é sumarizar a informação a respeito do parâmetro alvo $\theta$ contida nessa distribuição de probabilidade.

O gráfico a seguir exemplifica uma situação envolvendo a distribuição Beta. Destaque é dado para três medidas: média, moda e mediana *a posteriori*.



## Estimação pontual Bayesiana

Qualquer tentativa de sumarizar a informação contida em uma f.d.p. ou f.m.p. deve ser feita com cuidado. O gráfico da f.d.p. ou f.m.p. *a posteriori* é a descrição mais completa que temos no processo de inferência Bayesiana. Entretanto, algumas vezes é útil sumarizar ainda mais a informação *a posteriori*.

A forma mais simples de sumarização é a estimação pontual, onde buscamos determinar um único valor representando a quantidade desconhecida $\theta$. Notação: $\hat{\theta}$ é um estimador pontual (**estimador de Bayes**) para $\theta$.

O gráfico no slide anterior, mostrou 3 possibilidades de estimadores para a probabilidade de sucesso $\theta$:

- média $= E_{\theta|\boldsymbol{Y}}(\theta) = a/(a+b) = 7/(7+3)$;
- moda $= (a-1)/(a+b-2) = (7-1)/(7+3-2)$, o qual é o valor de $\theta$ que maximiza a f.d.p. Beta(7,3);
- mediana obtida por $\int_0^{\hat{\theta}} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) d\theta = 0.5$ `qbeta(0.5,7,3)`.

A escolha de $\hat{\theta}$ pode ser justificada no contexto da teoria da decisão.

Um problema de decisão é completamente especificado por 3 espaços:

- espaço paramétrico $\Theta$;
- espaço amostral $\Omega$ ou $\boldsymbol{U}$ com os possíveis resultados do experimento;
- espaço das possíveis ações $\mathscr{A}$.

Uma **regra de decisão** $\delta$ é uma função definida em $\Omega$ com valores em $\mathscr{A}$, isto é

$$\delta : \Omega \to \mathscr{A}.$$

Uma **função perda** pode ser associada a cada regra de decisão $\delta(\boldsymbol{Y})$ e a cada possível valor de $\theta \in \Theta$.

Ela é interpretada como uma punição por tomarmos a decisão $\delta$ para estimar $\theta$. A função perda é uma função de $\Theta \times \mathscr{A}$ em $\mathbb{R}^+$ e será denotada por $L(\delta, \theta)$.

Definição: O **risco** $R(\delta)$ de uma regra de decisão é a perda esperada *a posteriori* dada por: $R(\delta) = E_{\theta|\boldsymbol{Y}}[L(\delta,\theta)]$. Consequentemente, podemos escrever

$$R(\delta) \; = \; \int_{\Theta} L(\theta,\delta) \; f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \; d\theta.$$

A definição de risco acima introduz uma medida que permite comparar diferentes regras de decisão.

Definição: Uma regra de decisão $\delta^*$ é dita **ótima** se possui risco mínimo, isto é, se $\quad R(\delta^*) < R(\delta) \quad$ para todo $\delta$.

Esta regra é chamada de **regra de Bayes** e seu risco associado é chamado **risco de Bayes**.

Definição: Um **estimador** é uma regra de decisão ótima em relação a uma dada função perda. O valor observado de um estimador é chamado de **estimativa**.

A seguir, a maior parte da apresentação será focada em funções perdas ditas simétricas que são do tipo $L(\delta, \theta) = h(\delta - \theta)$, para alguma função $h$.

Este é o tipo de função perda mais usado. Em vários problemas práticos iremos trabalhar com $\Theta \subset \mathbb{R}$ e funções perda contínuas.

Lema - Função perda quadrática: Seja $L_1(\delta, \theta) = (\delta - \theta)^2$ a função perda associada com a estimação de $\theta$ por $\delta$. Se esta função perda for admitida na análise estatística, o estimador de $\theta$ será $\delta_1 = E_{\theta|\boldsymbol{Y}}(\theta)$, que é a média da distribuição *a posteriori* de $\theta$.

**Prova:** Precisamos mostrar que $\delta_1$ minimiza a função risco. Considere:

$R(\delta) \; = \; E_{\theta|\mathbf{Y}}[(\delta-\theta)^2] = E_{\theta|\mathbf{Y}}\{(\delta-\delta_1 \; + \delta_1 - \theta)^2\}$ em que $\delta_1 = E_{\theta|\mathbf{Y}}(\theta)$.

Então $R(\delta) \; = \; E_{\theta|\mathbf{Y}}[(\delta-\delta_1)^2] \; + \; E_{\theta|\mathbf{Y}}[(\delta_1-\theta)^2] \; + \; 2E_{\theta|\mathbf{Y}}[(\delta-\delta_1)(\delta_1-\theta)]$

$= \; (\delta-\delta_1)^2 \; + \; E_{\theta|\mathbf{Y}}[(\delta_1-\theta)^2] \; + \; 2(\delta-\delta_1)\, E_{\theta|\mathbf{Y}}(\delta_1-\theta)$

$= \; (\delta-\delta_1)^2 \; + \; E_{\theta|\mathbf{Y}}[(\theta-\delta_1)^2] \; + \; 2(\delta-\delta_1)\, [\delta_1 - E_{\theta|\mathbf{Y}}(\theta)]$

$= \; (\delta-\delta_1)^2 \; + \; Var_{\theta|\mathbf{Y}}(\theta) \; + \; 2(\delta-\delta_1)\, [\delta_1-\delta_1]$

$= \; (\delta-\delta_1)^2 + Var_{\theta|\mathbf{Y}}(\theta)$.

Logo, o risco é minimizado quando $\delta = \delta_1$.

Neste caso, o risco de Bayes será $R(\delta_1) = Var_{\theta|\mathbf{Y}}(\theta)$ e $R(\delta_1) \leq R(\delta)$, para todo $\delta$. A igualdade dos riscos ocorre se e só se $\delta = \delta_1 = E_{\theta|\mathbf{Y}}(\theta)$.

A função perda quadrática é algumas vezes criticada por introduzir uma penalidade que aumenta fortemente conforme $\delta - \theta$ cresce.

Em muitas situações, desejamos ter uma função perda que não dá muito destaque a grandes erros de estimação.

O próximo lema apresenta o estimador associado com a função perda absoluta que considera uma punição crescendo linearmente com o erro do estimador.

Lema - Função perda absoluta: Seja $L_2(\delta, \theta) = |\delta - \theta|$ a função perda associada com a estimação de $\theta$ por $\delta$. Em uma análise estatística admitindo esta função perda, o estimador de $\theta$ será $\delta_2 = \text{mediana}(\theta|\boldsymbol{Y})$, ou seja, a mediana da distribuição *a posteriori* de $\theta$.

**Prova:** Primeiramente, alguns resultados importantes precisam ser apontados:

- Regra de Leibniz: $\frac{\partial}{\partial x} \int_{A(x)}^{B(x)} g(x,t)\, dt =$

  $= \int_{A(x)}^{B(x)} \frac{\partial}{\partial x} g(x,t)\, dt + g(x, B(x)) \frac{\partial}{\partial x} B(x) - g(x, A(x)) \frac{\partial}{\partial x} A(x).$

- Se $P(X \geq m) \geq 1/2$ e $P(X \leq m) \geq 1/2$,
  então $m$ será a mediana da variável aleatória $X$.

  Porque "$\geq 1/2$"? Isto se deve ao caso discreto.

  Exemplo: Suponha $x = 1, 2, 3, 4, 5$ com $P(X = x) = 1/5$, para todo $x$.
  Veja que $m = 3$ é a mediana e
  $P(X \geq m) = 3/5 > 1/2$ e $P(X \leq m) = 3/5 > 1/2$.

**Prova (continuação):** $R(\delta) = E_{\theta|\boldsymbol{Y}}(|\delta-\theta|) = \int_{\Theta} |\delta-\theta| \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta$

$$= \int_{\theta\geq\delta} (\theta-\delta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + \int_{\theta<\delta} (\delta-\theta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta$$

$$= \int_{\delta}^{\infty} (\theta-\delta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + \int_{-\infty}^{\delta} (\delta-\theta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta$$

Deseja-se minimizar o risco (derive e iguale a zero).
Aplique a regra de Leibniz para derivar a integral.

$$\frac{\partial}{\partial\delta} R(\delta) = \frac{\partial}{\partial\delta} \int_{\delta}^{\infty} (\theta-\delta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + \frac{\partial}{\partial\delta} \int_{-\infty}^{\delta} (\delta-\theta) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta$$

$$= \int_{\delta}^{\infty} (-1) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + (\infty-\delta) f_{\theta|\boldsymbol{Y}}(\infty|\boldsymbol{y}) \, \frac{\partial}{\partial\delta}\infty - (\delta-\delta) \, f_{\theta|\boldsymbol{Y}}(\delta|\boldsymbol{y}) \, \frac{\partial}{\partial\delta}\delta +$$

$$+ \int_{-\infty}^{\delta} (1) \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + (\delta-\delta) f_{\theta|\boldsymbol{Y}}(\delta|\boldsymbol{y}) \, \frac{\partial}{\partial\delta}\delta - [\delta-(-\infty)] f_{\theta|\boldsymbol{Y}}(-\infty|\boldsymbol{y}) \, \frac{\partial}{\partial\delta}(-\infty)$$

$$= -\int_{\delta}^{\infty} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) d\theta + 0 - 0 + \int_{-\infty}^{\delta} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + 0 - 0$$

$$= -P_{\theta|\boldsymbol{Y}}(\theta \geq \delta|\boldsymbol{y}) + P_{\theta|\boldsymbol{Y}}(\theta \leq \delta|\boldsymbol{y}) = 0.$$

**Prova (continuação):** O valor de $\delta$ que satisfaz

$$P_{\theta|\boldsymbol{Y}}(\theta \geq \delta|\boldsymbol{y}) \;=\; P_{\theta|\boldsymbol{Y}}(\theta \leq \delta|\boldsymbol{y})$$

será $\delta =$ mediana da distribuição *a posteriori*.

É possível mostrar que $\dfrac{\partial^2}{\partial \delta^2} R(\delta) > 0$.

Isto comprova que a mediana minimiza o risco.

Uma outra forma de reduzir o efeito dos erros de estimação é considerar funções perda que permanecem constantes sempre que $|\delta - \theta| > \epsilon$, para algum $\epsilon$ positivo arbitrário. A escolha mais comum de $\epsilon$ é o valor limite $\epsilon \to 0$. Neste caso, para qualquer magnitude do erro, a perda é fixa. Este tipo de perda é conhecido como perda 0-1.

Lema - Função perda 0-1: Seja $L_3(\delta, \theta) = \lim_{\epsilon \to 0} 1_{\{|\delta - \theta| \geq \epsilon\}}$, tal que $1_{\{A\}}$ é função indicadora binária (vale 1 se $A$ ocorre, 0 caso contrário). Quando este tipo de função perda é adotado em uma análise estatística, o estimador de $\theta$ será $\delta_3 = \text{moda}(\theta|\boldsymbol{Y}) =$ moda do distribuição *a posteriori* de $\theta$.

**Prova (caso contínuo):** Note que a indicadora resultará em 1 se $|\delta - \theta| \geq \epsilon$, ou seja $\delta - \theta \geq \epsilon$ ou $\delta - \theta \leq -\epsilon$.
Isto é o mesmo que dizer $\theta \leq \delta - \epsilon$ ou $\theta \geq \delta + \epsilon$.

$$R(\delta) = E_{\theta|\boldsymbol{Y}}[L_3(\delta, \theta)] =$$

$$= \lim_{\epsilon \to 0} \left[ \int_{-\infty}^{\delta - \epsilon} 1 \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + \int_{\delta - \epsilon}^{\delta + \epsilon} 0 \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta + \int_{\delta + \epsilon}^{\infty} 1 \, f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta \right]$$

$$= \lim_{\epsilon \to 0} \left[ 1 - \int_{\delta - \epsilon}^{\delta + \epsilon} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta \right] = 1 - \lim_{\epsilon \to 0} P_{\theta|\boldsymbol{Y}}(\delta - \epsilon < \theta < \delta + \epsilon|\boldsymbol{y}).$$

Veja que $\lim_{\epsilon \to 0} P_{\theta|\boldsymbol{Y}}(\delta - \epsilon < \theta < \delta + \epsilon|\boldsymbol{y}) = f_{\theta|\boldsymbol{Y}}(\delta|\boldsymbol{y})$.

Conclusão: $R(\delta)$ é minimizado quando $f_{\theta|\boldsymbol{Y}}(\delta|\boldsymbol{y})$ é maximizado; portanto, $\delta_3 =$ moda da distribuição *a posteriori*.

Exemplo: Suponha que $Y_1, Y_2, \cdots, Y_n$ dado $\theta$ sejam condicionamlmente independentes com distribuição $Y_i|\theta \sim \text{Bernoulli}(\theta)$. Adote *a priori* que $\theta \sim \text{Beta}(a, b)$. Obtenha os estimadores de Bayes para as funções de perda quadrática, absoluta e 0-1.

Já foi mostrado neste curso que a distribuição *a posteriori* de $\theta$ é

$$\theta|\boldsymbol{Y} \sim \text{Beta}(a^*, b^*) \quad \text{com} \quad a^* = a + \textstyle\sum_{i=1}^n y_i \quad \text{e} \quad b^* = b + n - \textstyle\sum_{i=1}^n y_i.$$

Sob a perda quadrática, estimamos $\theta$ com $\hat{\theta} = E_{\theta|\boldsymbol{Y}}(\theta) = a^*/(a^* + b^*)$.

Sob a perda absoluta, estimamos $\theta$ com $\hat{\theta} = \text{mediana}(\theta|\boldsymbol{Y})$,
que é a solução da igualdade $\int_0^{\hat{\theta}} \frac{\Gamma(a^*+b^*)}{\Gamma(a^*)\Gamma(b^*)} \theta^{a^*-1}(1-\theta)^{b^*-1} \, d\theta = 0.5$.
Se $a^* = b^*$, temos a f.d.p. simétrica. Nesta situação, $\text{mediana}(\theta|\boldsymbol{Y}) = E_{\theta|\boldsymbol{Y}}(\theta)$.

Sob a perda 0-1, estimamos $\theta$ com $\hat{\theta} = \text{moda}(\theta|\boldsymbol{Y}) = (a^* - 1)/(a^* + b^* - 2)$.
Esta moda existirá apenas quando $a^* > 1$ e $b^* > 1$.

Exemplo (continuação): Encontrar a moda é um problema de otimização. Deve-se derivar a f.d.p. e igualar o resultado a 0. Recomenda-se realizar a derivação da log-densidade para facilitar as contas.

$$\ln f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \; = \; \ln\left[\frac{\Gamma(a^*+b^*)}{\Gamma(a^*)\Gamma(b^*)}\right] \; + \; (a^* - 1)\; \ln(\theta) + (b^* - 1)\; \ln(1-\theta)$$

$$\frac{d}{d\theta} \ln f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \; = \; (a^* - 1)\; \frac{1}{\theta} + (b^* - 1)\; \frac{1}{1-\theta}\; (-1) \; = \; 0$$

$$\frac{(a^* - 1)}{\theta} \; - \; \frac{(b^* - 1)}{1-\theta} \; = \; 0$$

$$\frac{(a^* - 1)(1-\theta) \; - \; (b^* - 1)\theta}{\theta(1-\theta)} \; = \; 0$$

$$(a^* - 1) - (a^* - 1)\theta \; - \; (b^* - 1)\theta \; = \; 0$$

$$a^* - 1 \; = \; \theta\,(a^* - 1 + b^* - 1)$$

$\theta \; = \; \dfrac{a^* - 1}{a^* + b^* - 2}$. Para ter maximização, verifique se $\frac{d^2}{d^2\theta} \ln f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) < 0$

## Estimação intervalar Bayesiana

Muitas vezes estamos interessados em avaliar o grau de incerteza presente na distribuição *a posteriori*. Uma maneira de fazer isso é calcular a variância dessa distribuição (variância grande = incerteza grande).

Uma estimativa intervalar do parâmetro alvo também permite esta análise. Determina-se um intervalo dentro do qual há grande chance de encontrar o verdadeiro valor (grande incerteza *a posteriori* = grande tamanho do intervalo).

O significado da estimação intervalar difere nas visões frequentista e Bayesiana.

Na visão frequentista a nomenclatura é **intervalo de confiança** de 95%.

- Significado: para uma coleção grande de replicações da amostra de dados, teremos 95% delas determinando uma estimativa intervalar que incluirá o real valor (fixo) do parâmetro. O intervalo é aleatório (varia perante diferentes amostras).

Na visão Bayesiana a nomenclatura é **intervalo de credibilidade** de 95%.

- Significado: ele define uma região do espaço paramétrico dentro da qual temos probabilidade 0.95 de encontrar o verdadeiro valor do parâmetro. A aleatoriedade aqui está associada ao parâmetro.

Seja $\theta$ uma quantidade desconhecida definida no espaço $\Theta$.

Uma região $\mathscr{C} \subset \Theta$ é uma região de 100(1-$\alpha$)% de credibilidade para $\theta$ se $P_{\theta|\boldsymbol{Y}}(\theta \in \mathscr{C}|\boldsymbol{y}) \geq 1-\alpha$. Neste caso, $1-\alpha$ é o "nível de credibilidade".

Se a distribuição *a posteriori* é contínua, escrevemos a igualdade

$$P_{\theta|\boldsymbol{Y}}(\theta \in \mathscr{C}|\boldsymbol{y}) = 1-\alpha$$

Se $\theta$ é um escalar, a região $\mathscr{C}$ poderá ser um intervalo do tipo $[c_1, c_2]$.

Se $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y})$ é a f.d.p. *a posteriori* do parâmetro escalar $\theta$, dizemos que $[c_1, c_2]$ é um intervalo de credibilidade de $100(1-\alpha)$% de probabilidade se

$$P_{\theta|\boldsymbol{Y}}(c_1 \leq \theta \leq c_2|\boldsymbol{y}) \ = \ \int_{c_1}^{c_2} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \, d\theta \ = \ 1-\alpha.$$

Exemplo: Sejam $Y_1, Y_2, \cdots, Y_n$, dado $\{\mu, \sigma^2\}$, variáveis condicionalmente independentes com $Y_i \sim N(\mu, \sigma^2)$. Admita que $\sigma^2$ é conhecido e assuma a especificação *a priori* $\mu \sim N(m, v)$; $m$ e $v$ são escolhidos pelo analista.

Este cenário já foi estudado anteriormente neste curso.
Recorde que a distribuição *a posteriori* é $\mu|\boldsymbol{Y} \sim N(m^*, v^*)$,

- $v^* = (\frac{n}{\sigma^2} + \frac{1}{v})^{-1}$.
- $m^* = v^*(\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{m}{v})$.

Dado $\boldsymbol{Y}$, podemos escrever: $\dfrac{\mu - m^*}{\sqrt{v^*}} \sim N(0, 1)$.

Muitos intervalos com 100(1-$\alpha$)% de credibilidade podem ser construídos para $\mu$.

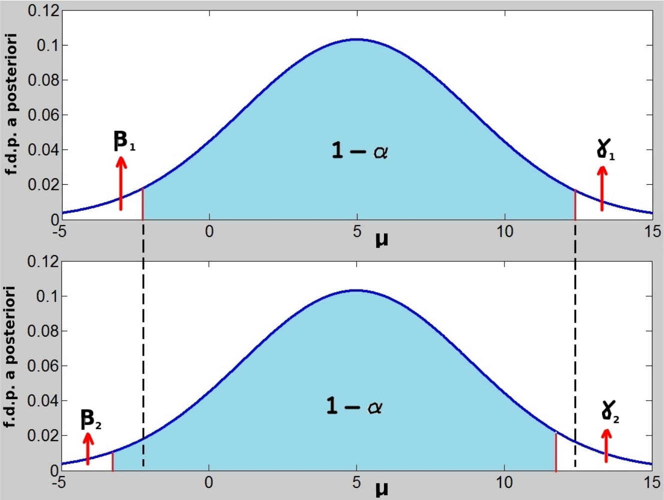
Dois exemplos são apresentados a seguir...

Notação: Se $Z \sim N(0,1)$, então $P(Z > z_\alpha) = \alpha$.

1. $1 - \alpha = P\left(\frac{\mu - m^*}{\sqrt{v^*}} \leq z_\alpha \mid \boldsymbol{Y}\right)$, então $(-\infty,\ m^* + z_\alpha\sqrt{v^*}\ ]$
   é um intervalo com probabilidade $(1-\alpha)$ para $\mu$.
   Entretanto, a amplitude deste intervalo é infinito.

2. Suponha que $\beta + \gamma = \alpha$.
   $1 - \alpha = P\left(z_{1-\beta} \leq \frac{\mu - m^*}{\sqrt{v^*}} \leq z_\gamma \mid \boldsymbol{Y}\right)$, então $[m^* + z_{1-\beta}\sqrt{v^*}, m^* + z_\gamma\sqrt{v^*}\ ]$
   é também um intervalo com probabilidade $(1-\alpha)$ para $\mu$.
   Amplitude $\quad m^* + z_\gamma\sqrt{v^*} - (m^* + z_{1-\beta}\sqrt{v^*}) = \sqrt{v^*}(z_\gamma - z_{1-\beta})$.
   Temos infinitas escolhas para $\beta + \gamma = \alpha$. O próximo slide ilustra duas delas.

Suponha que foi obtido $m^* = 5$ e $v^* = 15$.




Como minimizar a amplitude do intervalo de credibilidade e manter $\beta + \gamma = \alpha$?

Baseado na ilustração do slide anterior, veja que:

1. Se $\beta = 0.01$ e $\gamma = 0.04$, temos $\alpha = 0.05$ e $1 - \alpha = 0.95$.

   Amplitude do intervalo de credibilidade de 95%:

   $\sqrt{v^*}(z_\gamma - z_{1-\beta}) \;=\; \sqrt{15}\,[11.78 - (-4.01)] \;=\; 61.15.$

2. Se $\beta = 0.02$ e $\gamma = 0.03$, temos $\alpha = 0.05$ e $1 - \alpha = 0.95$.

   Amplitude do intervalo de credibilidade de 95%:

   $\sqrt{v^*}(z_\gamma - z_{1-\beta}) \;=\; \sqrt{15}\,[12.28 - (-2.95)] \;=\; 58.98.$

3. Se $\beta = 0.025$ e $\gamma = 0.025$, temos $\alpha = 0.05$ e $1 - \alpha = 0.95$.

   Amplitude do intervalo de credibilidade de 95% (<u>simétrico</u> em relação a $m^*$):

   $\sqrt{v^*}(z_\gamma - z_{1-\beta}) \;=\; \sqrt{15}\,[12.59 - (-2.59)] \;=\; 58.79.$

Temos apenas três casos acima, mas é possível mostrar que a opção 3 (simétrico em relação a $m^*$) será o intervalo mais curto de todos com $\beta + \gamma = 0.05$.

O pesquisador deve preferir o intervalo com a menor amplitude em sua análise. Este intervalo concentra a massa de probabilidade em um espaço menor, por isso ele é dito **intervalo de credibilidade de mais alta densidade a posteriori** (HPD - *Highest Posterior Density*)

Se a distribuição *a posteriori* é simétrica e unimodal (ex.: Normal, t-Student), teremos $E_{\theta|\boldsymbol{Y}}(\theta) = \text{mediana}(\theta|\boldsymbol{Y}) = \text{moda}(\theta|\boldsymbol{Y})$.

No caso de uma f.d.p. simétrica, seja $q_{\theta|\boldsymbol{Y}}(\alpha/2)$ o quantil da distribuição a posteriori deixando uma área de probabilidade $\alpha/2$ abaixo. O intervalo HPD de $100(1-\alpha)\%$ será delimitado pelos valores $q_{\theta|\boldsymbol{Y}}(\alpha/2)$ e $q_{\theta|\boldsymbol{Y}}(1-\alpha/2)$.

Em particular, se $\alpha = 0.05$, teremos o intervalo de credibilidade HPD para $\theta$ denotado por $[q_{\theta|\boldsymbol{Y}}(0.025),\ q_{\theta|\boldsymbol{Y}}(0.975)]$.

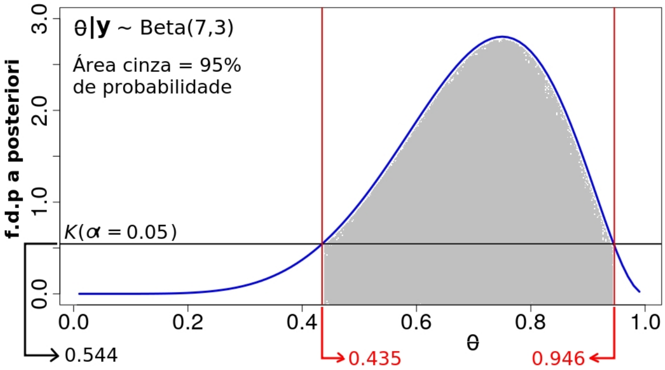
Observação importante: intervalos HPD não são exclusividade de distribuições *a posteriori* simétricas.

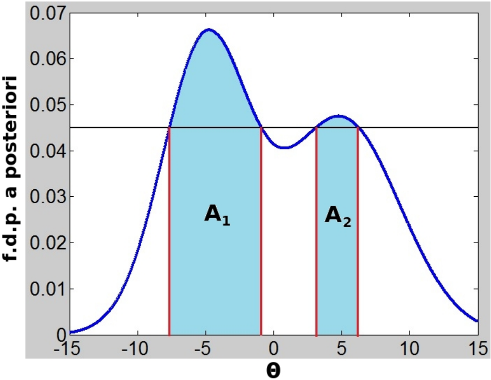
Definição: Um intervalo de credibilidade Bayesiano HPD (*Highest Posterior Density*) de $100(1-\alpha)\%$ para $\theta$ pode ser estabelecido por

$$\mathscr{C} = \{\theta \in \Theta \ : \ f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \geq k(\alpha)\}.$$

Neste caso, $k(\alpha)$ é a maior constante tal que $P(\theta \in \mathscr{C}|\boldsymbol{Y}) \geq 1-\alpha$.

Suponha um estudo em que o cálculo da distribuição *a posteriori* forneceu $\theta|\boldsymbol{Y} \sim \text{Beta}(7, 3)$. A figura a seguir ilustra o intervalo HPD para $\theta$.






Neste exemplo temos uma distribuição *a posteriori* bimodal e a linha horizontal representa uma escolha da constante $\kappa(\alpha)$, a qual estabelece as áreas $A_1$ e $A_2$ tal que $A_1 + A_2 = 1 - \alpha$.

A região HPD para $\theta$ não é um intervalo do tipo $[c_1, c_2]$. Ela é formada por dois intervalos disjuntos $[c_1, c_2] \cup [c_3, c_4]$.

Ideia geral do cálculo numérico de um intervalo de credibilidade HPD. Para diversos valores de $k > 0$ siga os passos:

1. Dado $k$, encontre todas as soluções da equação $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{Y}) = k$;
2. Construa o conjunto $\mathscr{C}(k) = \{\theta \ : \ f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{Y}) \geq k\}$ levando em conta o resultado em (1);
3. Calcule $P_{\theta|\boldsymbol{Y}}[\mathscr{C}(k)] = \int_{\mathscr{C}(k)} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{Y}) \, d\theta$.

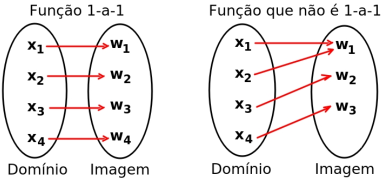
   Se $P_{\theta|\boldsymbol{Y}}[\mathscr{C}(k)] \approx 1 - \alpha$, finalize a busca por $k$ e tome

   os valores obtidos em (1) como limites da região HPD.

   Se esta região for do tipo $[c_1, c_2]$, teremos um intervalo HPD.

Resultado importante: intervalos de credibilidade são invariantes sob transformações 1-a-1 do parâmetro.

Lembrete: em matemática, um função é dita 1-a-1 ou função bijetiva se ela mapeia cada elemento do conjunto domínio para um único elemento do conjunto imagem e cada elemento do conjunto imagem pode ser mapeado de volta para seu correspondente único elemento no domínio. Veja a ilustração a seguir.




Alguns exemplos de funções 1-a-1:

$\log(x)$, $x+1$, $x/2$, $x^3$, $\sqrt{x}$ e muitas outras.

Alguns exemplos de funções que não são 1-a-1:

$x^2$, $|x|$, $x^4$, $\text{sen}(x)$ e muitas outras.

Se $\mathscr{C}$ é um intervalo de $100(1-\alpha)\%$ de credibilidade para $\theta$ e $h(\bullet)$ é uma função que determina uma transformação 1-a-1 no espaço paramétrico de $\theta$, então a imagem de $\mathscr{C}$ sob $h(\bullet)$ dada por $h(\mathscr{C})$ será um intervalo de $100(1-\alpha)\%$ de credibilidade para $\phi = h(\theta)$.

Exemplo: Seja $[c_1, c_2]$ um intervalo de credibilidade de $100(1-\alpha)\%$ para $\theta > 0$. Queremos avaliar a quantidade $\phi = \log(\theta)$. Visto que $\log(\bullet)$ é uma transformação 1-a-1 no espaço paramétrico $\mathbb{R}^+$ de $\theta$, o intervalo de credibilidade de $100(1-\alpha)\%$ para $\phi$ será $[\log(c_1), \log(c_2)]$

Exemplo: Suponha que $\theta|\boldsymbol{Y} \sim \text{Ga}(3, 2)$ é a distribuição *a posteriori* do parâmetro de variância $\theta > 0$.

Um possível intervalo de credibilidade de 95% para $\theta$ pode ser construído usando os seguintes quantis *a posteriori* (comandos `R` em vermelho):

- $q_{\theta|\boldsymbol{Y}}(0.025) = 0.3093$; `qgamma(0.025, 3, rate = 2)`
- $q_{\theta|\boldsymbol{Y}}(0.975) = 3.6123$; `qgamma(0.975, 3, rate = 2)`

Exemplo (continuação): Considere agora a transformação 1-a-1 dada por $\phi = 1/\theta$. O resultado é o parâmetro de precisão $\phi > 0$. Um intervalo de credibilidade de 95% para $\phi$ pode ser obtido a partir do intervalo de 95% para $\theta$:

$0.3093 < \theta < 3.6123$, então $1/3.6123 < 1/\theta < 1/0.3093$.

Conclusão: $0.2768 < \phi < 3.2331$.

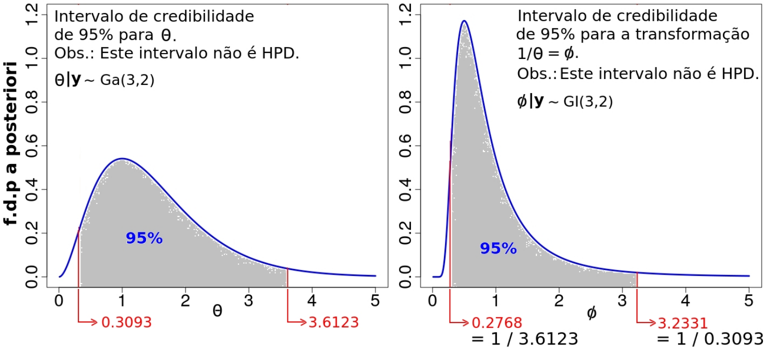
```
n = 1000000;  theta = rgamma(n,3,rate=2);
phi = 1/theta  # transformação 1-a-1

# Limites do intervalo de 95% para theta
theta_c1 = qgamma(0.025,a,rate=b)
theta_c2 = qgamma(0.975,a,rate=b)

# Limites do intervalo de 95% para phi = 1/theta
phi_c1 = 1/theta_c2;   phi_c2 = 1/theta_c1;

# proporção de observações dentro dos intervalos
sum(theta > theta_c1 & theta <= theta_c2) / n
sum(phi > phi_c1 & phi <= phi_c2) / n
```

Resultado das duas proporções acima: `0.950138`




Importante: regiões HPD não são invariantes sob transformações 1-a-1. Após a transformação, o intervalo continuará sendo um intervalo de credibilidade, mas ele pode perder a característica de ser "HPD".

## 4.3 Testes de hipóteses

No contexto Bayesiano, o problema de decidir sobre qual hipótese aceitar é conceitualmente mais simples do que na inferência clássica.

Devemos apenas comparar as hipóteses $H_0$, $H_1$, $H_2$, $\cdots$, $H_k$ através de suas respectivas distribuições *a posteriori* denotadas por $P(H_0|\boldsymbol{Y})$, $P(H_1|\boldsymbol{Y})$, $P(H_2|\boldsymbol{Y})$, $\cdots$, $P(H_k|\boldsymbol{Y})$.

- Escolhe-se a hipótese mais provável *a posteriori*.
- Dentro da teoria de decisão, podemos também selecionar a hipótese que minimiza a perda esperada *a posteriori*.

Considere o caso simples contendo apenas 2 hipóteses no teste

$$H_0 : \theta \in \Theta_0 \quad \text{contra} \quad H_1 : \theta \in \Theta_0^c.$$

Probabilidades *a priori*: $\quad P(H_0) = 1 - P(H_1)$.

Teorema de Bayes fornece: $P(H_0|\boldsymbol{Y}) = \dfrac{P(\boldsymbol{Y}|H_0)\ P(H_0)}{P(\boldsymbol{Y})}$.

Temos também que: $P(H_1|\boldsymbol{Y}) = 1 - P(H_0|\boldsymbol{Y})$

Se $P(H_0|\boldsymbol{Y}) > P(H_1|\boldsymbol{Y})$, existe evidência amostral à favor de $H_0$.

Se $P(H_0|\boldsymbol{Y}) < P(H_1|\boldsymbol{Y})$, devemos rejeitar $H_0$.

Em geral, o que é considerado para a tomada de decisão é a *odds a posteriori* dada pela razão

$$\frac{P(H_0|\boldsymbol{Y})}{P(H_1|\boldsymbol{Y})} = \frac{\frac{P(\boldsymbol{Y}|H_0)\ P(H_0)}{P(\boldsymbol{Y})}}{\frac{P(\boldsymbol{Y}|H_1)\ P(H_1)}{P(\boldsymbol{Y})}} = \frac{P(\boldsymbol{Y}|H_0)}{P(\boldsymbol{Y}|H_1)}\ \frac{P(H_0)}{P(H_1)}$$

O termo vermelho é chamado de **fator de Bayes** (FB).
O termo verde é a *odds a priori*.

Podemos escrever também: $\quad \text{FB} = \dfrac{P(\boldsymbol{Y}|H_0)}{P(\boldsymbol{Y}|H_1)} = \dfrac{P(H_0|\boldsymbol{Y})}{P(H_1|\boldsymbol{Y})} \dfrac{P(H_1)}{P(H_0)}$.

O fator de Bayes mede o quanto os dados observados $\boldsymbol{y}$ são coerentes com as especificações estabelecida em $H_0$ e $H_1$. Isto é avaliado perante a adoção de um modelo probabilístico para os $Y_i$'s.

Se $FB > 1$, temos a indicação de que o modelo sob $H_0$ é mais compatível com as observações em $\boldsymbol{y}$ do que aquele sob $H_1$.

Para obter a *odds a priori*, calcule $P(H_0) = P(\theta \in \Theta_0)$, visto que uma distribuição *a priori* é atribuída para $\theta$. Entretanto, esta conta pode ser problemática dependendo do espaço $\Theta_0$. Veremos isso mais adiante.

Exemplo: Suponha que $Y_1, Y_2, \cdots, Y_n$ dado $\mu$ são i.i.d. com $Y_i|\mu \sim N(\mu, \sigma^2)$. Admita $\sigma^2$ conhecido. Utilize a distribuição *a priori* de Jeffreys e teste, sob a visão Bayesiana, as seguintes hipóteses: $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$.

Visto que $\sigma^2$ é conhecido, a distribuição *a priori* de Jeffreys será
$f_\mu(\mu) \propto 1$ para $\mu \in \mathbb{R}$ (distribuição imprópria).

Verossimilhança: $f_{\boldsymbol{Y}|\mu}(\boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2)\}$.

Distribuição *a posteriori*: $f_{\mu|\boldsymbol{Y}}(\mu|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\mu}(\boldsymbol{y})\, f_\mu(\mu)$.

$\propto \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2)\} \times 1$

$\propto \exp\{-\frac{1}{2\sigma^2} \left[n\mu^2 - 2\mu \sum_{i=1}^n y_i\right]\}$

$\propto \exp\{-\frac{1}{2\sigma^2} n \left[\mu^2 - 2\mu\bar{y}\right]\}$. Este é o núcleo da $N(\bar{y}, \sigma^2/n)$.

Temos *a posteriori*: $\mu|\boldsymbol{Y} \sim N(\bar{y}, \sigma^2/n)$.

$$P(H_0|\boldsymbol{Y}) \;=\; P(\mu \leq \mu_0|\boldsymbol{Y}) \;=\; P(Z \leq \frac{\mu_0 - \bar{y}}{\sigma/\sqrt{n}}|\boldsymbol{Y}), \text{ sendo } Z \sim N(0,1).$$

Lembrete 1: $P(H_1|\boldsymbol{Y}) \;=\; 1 - P(H_0|\boldsymbol{Y})$.

Lembrete 2: adotou-se a distribuição *a priori* de Jeffreys imprópria $f_\mu(\mu) \propto 1$, então $P(H_0) = \int_{-\infty}^{\mu_0} f_\mu(\mu)\, d\mu =$ indefinido. Neste caso, use $P(H_0) = P(H_1) = 0.5$ para simbolizar total incerteza.

Os resultados acima permitem calcular o FB. Rejeite $H_0$, se $0 < FB < 1$.

Adote: $n = 4$, $\bar{y} = 106$, $\sigma^2 = 400$ e $\mu_0 = 100$.

Então: $P(H_0|\boldsymbol{Y}) \;=\; P(Z \leq \frac{100-106}{20/\sqrt{4}}|\boldsymbol{Y}) \;=\; P(Z \leq -0.6) \;=\; 0.2742$.

$$\text{FB} = \frac{P(H_0|\boldsymbol{Y})}{P(H_1|\boldsymbol{Y})} \;=\; \frac{0.2742}{1 - 0.2742} \;=\; 0.3779. \quad \text{Rejeite } H_0, \text{ pois FB} \in (0,1).$$

Obs.: No caso frequentista, o teste do slide anterior é realizado com base na distribuição de probabilidade associada à estatística de teste $\bar{Y} \sim N(\mu, \sigma^2/n)$. Sob $H_0$ e usando a observação $\bar{y} = 106$, calcula-se o valor-p

$$P(\bar{y} > 106|\mu = 100) = 1 - P(Z \leq \tfrac{106-100}{\sigma/\sqrt{n}}) = 1 - P(Z \leq \tfrac{106-100}{20/\sqrt{4}})$$

$$= 1 - P(Z \leq 0.6) = 1 - 0.7258 = 0.2742.$$

Valor-p $> 0.05 \quad \Rightarrow \quad$ não-rejeitar $H_0$.

Note que a conclusão do procedimento Bayesiano foi diferente do caso clássico acima. Eventualmente estas abordagens podem concordar, mas isso nem sempre vai acontecer.

Pode haver dificuldade de uso do fator de Bayes em algumas configurações de hipóteses:

Suponha as hipóteses: $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$.

Veja que o espaço definido em $H_0$ contém apenas o valor $\theta_0$.

Admita que $\theta$ tem espaço paramétrico contínuo e $f_\theta(\theta)$ é uma f.d.p.

Consequência:

- *a priori* temos: $P(H_0) = P(\theta = \theta_0) = \int_{\theta_0}^{\theta_0} f_\theta(\theta)\, d\theta = 0$.
- *a posteriori* temos: $P(H_0|\boldsymbol{Y}) = P(\theta = \theta_0|\boldsymbol{Y}) = \int_{\theta_0}^{\theta_0} f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y})\, d\theta = 0$.

Este resultado indica que <u>nunca</u> vamos escolher $H_0$ na comparação com $H_1$.

A solução para este problema é discutida a seguir.

Admita uma forma especial da f.d.p. *a priori* que é definida de forma a contemplar cada subespaço $\Theta_0$ e $\Theta_0^c$ estabelecido pelas hipóteses $H_0$ e $H_1$.

$$f_\theta(\theta) = \begin{cases} P(H_0)\, f_\theta^{H_0}(\theta), & \text{para } \theta \in \Theta_0. \\ P(H_1)\, f_\theta^{H_1}(\theta), & \text{para } \theta \in \Theta_0^c = \Theta \setminus \Theta_0. \end{cases}$$

Para simplificar a notação, considere $P(H_0) = p$ e $P(H_1) = 1 - p$. Logo

$$f_\theta(\theta) = \begin{cases} p\, f_\theta^{H_0}(\theta), & \text{para } \theta \in \Theta_0. \\ (1-p)\, f_\theta^{H_1}(\theta), & \text{para } \theta \in \Theta_0^c. \end{cases}$$

Um ponto importante é que $\int_\Theta f_\theta(\theta)\, d\theta = 1$, então

$$\int_\Theta f_\theta^{H_0}(\theta)\, d\theta = 1 \quad \text{e} \quad \int_\Theta f_\theta^{H_1}(\theta)\, d\theta = 1.$$

Para simplificar a notação, denote $\Theta_1 = \Theta_0^c$.

A *odds a posteriori* será dada por:

$$\frac{P(H_0|\boldsymbol{Y})}{P(H_1|\boldsymbol{Y})} = \frac{P(\boldsymbol{Y}|H_0)}{P(\boldsymbol{Y}|H_1)} \; \frac{P(H_0)}{P(H_1)} = \frac{p \int_{\Theta_0} f_{\boldsymbol{Y},\theta}(\boldsymbol{y},\theta)\; d\theta}{(1-p) \int_{\Theta_1} f_{\boldsymbol{Y},\theta}(\boldsymbol{y},\theta)\; d\theta}.$$

$$= \frac{p \int_{\Theta_0} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; f_\theta^{H_0}(\theta)\; d\theta}{(1-p) \int_{\Theta_1} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; f_\theta^{H_1}(\theta)\; d\theta}.$$

O fator de Bayes está em vermelho, ou seja $\text{FB} = \dfrac{\int_{\Theta_0} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; f_\theta^{H_0}(\theta)\; d\theta}{\int_{\Theta_1} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; f_\theta^{H_1}(\theta)\; d\theta}$.

Exemplo: $Y_1, Y_2, \cdots, Y_n$ dado $\theta$ são i.i.d. com $Y_i|\theta \sim N(\theta, \sigma^2)$. Assuma $\sigma^2$ conhecido. Deseja-se testar as seguintes hipóteses: $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$.

Exemplo (continuação): Distribuição *a priori*:

$$f_\theta(\theta) = \begin{cases} p \times 1, & \text{para } \theta = \theta_0. \\ (1-p) \times (2\pi v)^{-1/2} \exp\{-\frac{1}{2v}(\theta - m)^2\}, & \text{para } \theta \neq \theta_0. \end{cases}$$

Note que:

- $f_\theta^{H_0}(\theta) = 1$ para $\theta = \theta_0$, ou seja, uma distribuição degenerada em $\theta_0$.
- $f_\theta^{H_1}(\theta)$ é a f.d.p. da $N(m, v)$. Visto que a normal é uma distribuição contínua, não faz diferença a presença ou não de $\theta_0$ no espaço paramétrico de $H_1$. Temos então $\int_{\mathbb{R}\setminus\theta_0} f_\theta^{H_1}(\theta)\, d\theta = \int_{-\infty}^{\infty} f_\theta^{H_1}(\theta)\, d\theta = 1$.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i^2 - 2\theta y_i + \theta^2)\}$.

O cálculo do fator de Bayes é mostrado a seguir.

Exemplo (continuação):

$$P(\boldsymbol{Y}|H_0) \;=\; \int_{\Theta_0} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\; f_\theta^{H_0}(\theta)\; d\theta \;=\; f_{\boldsymbol{Y}|\theta_0}(\boldsymbol{y}) \times 1 \;=\; f_{\boldsymbol{Y}|\theta_0}(\boldsymbol{y})$$

$$= \; (2\pi\sigma^2)^{-n/2}\; \exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (y_i^2 - 2\theta_0 y_i + \theta_0^2)\}.$$

$$= \; (2\pi\sigma^2)^{-n/2}\; \exp\left\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\right\}\; \exp\left\{-\tfrac{1}{2\sigma^2/n}\left[\theta_0^2 - 2\theta_0\bar{y}\right]\right\}.$$

$$P(\boldsymbol{Y}|H_1) \;=\; \int_{\Theta_1} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\; f_\theta^{H_1}(\theta)\; d\theta \;=\; \int_{\mathbb{R}\setminus\theta_0} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\; f_\theta^{H_1}(\theta)\; d\theta$$

$$= \; \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-n/2}\; \exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (y_i^2 - 2\theta y_i + \theta^2)\} \times$$

$$\times\; (2\pi v)^{-1/2} \exp\{-\tfrac{1}{2v}(\theta^2 - 2\theta m + m^2)\}\; d\theta.$$

$$= \; (2\pi\sigma^2)^{-n/2}\, (2\pi v)^{-1/2}\; \exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\}\; \exp\{-\tfrac{m^2}{2v}\} \times$$

$$\times\; \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2\sigma^2/n}\left(\theta^2 - 2\theta\bar{y}\right)\}\; \exp\{-\tfrac{1}{2v}(\theta^2 - 2\theta m)\}\; d\theta.$$

Exemplo (continuação):

$$P(\boldsymbol{Y}|H_1) \;=\; (2\pi\sigma^2)^{-n/2}\,(2\pi v)^{-1/2}\,\exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\}\,\exp\{-\tfrac{m^2}{2v}\}\times$$
$$\times\;\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\left[\theta^2\left(\tfrac{n}{\sigma^2}+\tfrac{1}{v}\right)-2\theta\left(\tfrac{n\bar{y}}{\sigma^2}+\tfrac{m}{v}\right)\right]\}\,d\theta.$$

$$=\;(2\pi\sigma^2)^{-n/2}\,(2\pi v)^{-1/2}\,\exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\}\,\exp\{-\tfrac{m^2}{2v}\}\times$$
$$\times\;\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\left(\tfrac{n}{\sigma^2}+\tfrac{1}{v}\right)\left[\theta^2-2\theta\left(\tfrac{n}{\sigma^2}+\tfrac{1}{v}\right)^{-1}\left(\tfrac{n\bar{y}}{\sigma^2}+\tfrac{m}{v}\right)\right]\}\,d\theta.$$

Dentro da integral temos o núcleo da $N(m^*, v^*)$ com

$$v^* \;=\; \left(\tfrac{n}{\sigma^2}+\tfrac{1}{v}\right)^{-1} \quad \text{e} \quad m^* \;=\; v^*\,\left(\tfrac{n\bar{y}}{\sigma^2}+\tfrac{m}{v}\right)$$

$$P(\boldsymbol{Y}|H_1) \;=\; (2\pi\sigma^2)^{-n/2}\,(2\pi v)^{-1/2}\,\exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\}\,\exp\{-\tfrac{m^2}{2v}\}\times$$
$$\times\;\int_{-\infty}^{\infty}\frac{(2\pi v^*)^{-1/2}}{(2\pi v^*)^{-1/2}}\exp\{-\tfrac{1}{2v^*}\left[\theta^2-2\theta m^*+(m^*)^2-(m^*)^2\right]\}\,d\theta.$$

Exemplo (continuação):

$$P(\boldsymbol{Y}|H_1) \;=\; (2\pi\sigma^2)^{-n/2}\,(2\pi v)^{-1/2}\,\exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\}\,\exp\{-\tfrac{m^2}{2v}\}\times$$
$$\times\,(2\pi v^*)^{+1/2}\exp\{+\tfrac{1}{2v^*}(m^*)^2\}\times$$
$$\times\,\int_{-\infty}^{\infty}(2\pi v^*)^{-1/2}\exp\{-\tfrac{1}{2v^*}\left[\theta^2-2\theta m^*+(m^*)^2\right]\}\,d\theta.$$

A integral resulta em 1.

$$P(\boldsymbol{Y}|H_1) \;=\; (2\pi\sigma^2)^{-n/2}\,\left(\frac{v^*}{v}\right)^{1/2}\,\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^n y_i^2+\frac{m^2}{v}-\frac{(m^*)^2}{v^*}\right]\right\}.$$

Conclusão: use os resultados destacados em vermelho nos últimos slides para determinar o fator de Bayes $\text{FB}=P(\boldsymbol{Y}|H_0)/P(\boldsymbol{Y}|H_1)$.

Para ilustrar de maneira mais concreta o teste de hipóteses que acabou de ser descrito, assuma: $\sigma^2=400$, $n=4$, $\bar{y}=106$, $\sum_{i=1}^n y_i^2=50000$ $\theta_0=100$, $m=105$, $v=100$ $P(H_0)=p=0.5$ e $P(H_1)=1-p=0.5$.

Obs.: assumir $P(H_0)=0.5$, implica em *odds a posteriori* $=$ FB.

Exemplo (continuação):

Note que os valores fornecidos na ilustração são números grandes que serão usados em uma conta envolvendo termos com exponencial. Por questões numéricas, é conveniente trabalhar aqui com a escala logarítmica.

$$P(\boldsymbol{Y}|H_0) \;=\; (2\pi\sigma^2)^{-n/2} \;\exp\left\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2\right\} \;\exp\left\{-\tfrac{1}{2\sigma^2/n}\left[\theta_0^2 - 2\theta_0\bar{y}\right]\right\}.$$

$$\ln P(\boldsymbol{Y}|H_0) \;=\; -\tfrac{n}{2}\ln(2\pi\sigma^2) \;-\; \tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n y_i^2 \;-\; \tfrac{1}{2\sigma^2/n}\left[\theta_0^2 - 2\theta_0\bar{y}\right].$$

$$= \; -\tfrac{4}{2}\ln[2\pi(400)] \;-\; \tfrac{1}{2(400)}50000 \;-\; \tfrac{1}{2(400/4)}\left[100^2 - 2(100)(106)\right]$$

$$= \; -2\ln(2513.27) - 62.5 + 56 \;=\; -22.16.$$

Calcule também: $\quad v^* \;=\; \left(\frac{n}{\sigma^2} + \frac{1}{v}\right)^{-1} \quad$ e $\quad m^* \;=\; v^* \;\left(\frac{n\bar{y}}{\sigma^2} + \frac{m}{v}\right)$

$$v^* \;=\; \left(\tfrac{4}{400} + \tfrac{1}{100}\right)^{-1} \;= 50$$

$$m^* \;=\; v^* \;\left(\frac{4(106)}{400} + \frac{105}{100}\right) \;= 50\;(2.11) \;= 105.5$$

Exemplo (continuação):

$$P(\boldsymbol{Y}|H_1) \;=\; (2\pi\sigma^2)^{-n/2} \;\left(\frac{v^*}{v}\right)^{1/2} \;\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{m^2}{v} - \frac{(m^*)^2}{v^*}\right]\right\}.$$

$$\ln P(\boldsymbol{Y}|H_1) \;=\; -\frac{n}{2}\ln(2\pi\sigma^2) \;+\frac{1}{2}[\ln(v^*) - \ln(v)] \;-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{m^2}{v} - \frac{(m^*)^2}{v^*}\right].$$

$$= \;-\frac{4}{2}\ln[2\pi(400)] \;+\frac{1}{2}[\ln(50) - \ln(100)] \;-\frac{1}{2}\left[\frac{50000}{400} + \frac{105^2}{100} - \frac{(105.5)^2}{50}\right].$$

$$= \;-2\ln[2513.27] \;+\frac{1}{2}[-0.69] \;-\frac{1}{2}\,[125 + 110.25 - 222.61] = -22.32.$$

Conclusão: $\ln(\text{FB}) = \ln P(\boldsymbol{Y}|H_0) - \ln P(\boldsymbol{Y}|H_1) = -22.16 - (-22.32) = 0.16.$

$\text{FB} = \exp\{0.16\} = 1.17.$

A *odds a posteriori* é: $\dfrac{P(H_0|\boldsymbol{Y})}{P(H_1|\boldsymbol{Y})} \;=\; \dfrac{p}{(1-p)}\,\text{FB} \;=\; \dfrac{0.5}{1-0.5}\,1.17 \;=\; 1.17.$

*Odds a posteriori* $> 1 \;\Rightarrow\;$ não rejeitar $H_0 : \theta = \theta_0$.

Lembre-se que na inferência clássica, os intervalos de confiança podem ser usados como ferramenta para testar hipóteses do tipo:

$$H_0 : \theta = \theta_0 \quad \text{contra} \quad H_1 : \theta \neq \theta_0.$$

A lógica é avaliar se o valor $\theta_0$ está contido dentro do intervalo de confiânça construído para $\theta$.

Este mesmo tipo de análise pode ser aplicado na inferência Bayesiana, porém aqui utiliza-se o intervalo de credibilidade de $\theta$.

Seja $[c_1, c_2]$ um intervalo de credibilidade de $100(1-\alpha)\%$ para $\theta$. Deseja-se testar as hipóteses escritas acima. A decisão do pesquisador deve ser:

- rejeitar $H_0$, se $\theta_0 \notin [c_1, c_2]$
- não rejeitar $H_0$, se $\theta_0 \in [c_1, c_2]$

Este tipo de avaliação é empregado, por exemplo, para testar a significância de um coeficiente de regressão. Neste contexto, as hipóteses seriam
$H_0 : \beta_j = 0 \quad \text{contra} \quad H_1 : \beta_j \neq 0$.