



**WORKSHOP  
DE BIOINFORMÁTICA  
APLICADA À GENÔMICA E  
MELHORAMENTO ANIMAL**



# **Nivelamento em inferência Bayesiana**



**Fabyano Fonseca e Silva**

**Prof. Adjunto IV - Dep Zootecnia – UFV  
Estatística Genômica e Bioinformática**

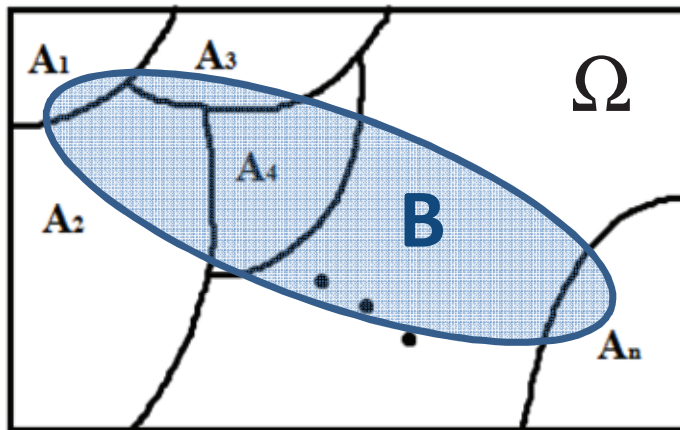
**Campo Grande, 14/07 a 15/07 de 2014**

# 1. Teorema de Bayes (a nível de eventos)

**Origem:** Publicação póstuma em 1763 – Thomas Bayes

“An Essay Towards Solving a Problem in the Doctrine of Chances”

OBS.: Inicialmente aplicado para calcular probabilidades de ocorrência de eventos sujeitos as seguintes condições:



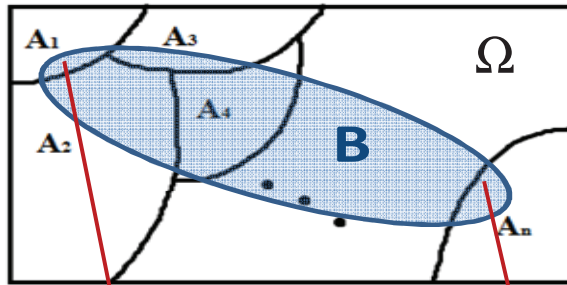
$$P(A_i \cap A_j) = 0 \quad \forall i, j$$

$$P(A_i) > 0 \quad \forall i$$

$$\sum_{i=1}^n P(A_i) = 1$$

$$P(A_i \cap B) > 0 \quad \forall i$$

- A probabilidade condicional  $P(A_i | B)$  representa a contribuição da ocorrência do evento  $B$  para a ocorrência do evento  $A_i$ .
- A probabilidade condicional  $P(B | A_i)$  representa a contribuição da ocorrência do evento  $A_i$  para a ocorrência do evento  $B$ .
- $P(A_i)$  é a probabilidade de ocorrência de  $A_i$  independentemente da ocorrência do evento  $B$ .



**OBS.: Teorema da Prb. Condicional**

$$P(B|A_i) = \frac{P(A_i \cap B)}{P(A_i)} \Rightarrow P(A_i \cap B) = P(B|A_i)P(A_i)$$

$$P(B) = \underbrace{P(A_1 \cap B)} + \dots + \underbrace{P(A_n \cap B)}$$

$$P(B) = P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)$$

**Teorema da Prob. Total**

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

Teorema da Prb. Condicional  
 Teorema da Prob. Total

**Teorema de Bayes**

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)}$$

**Em um quartel existem 2 tipos de fuzis,  $A_1$  (de alta precisão) e  $A_2$  (de baixa precisão), cujas avaliações de performance indicaram, respectivamente, probabilidades de acerto de 95% e 80%. Além disso, o comandante sabe que a maioria dos fuzis são de baixa precisão.**

**Uma vez que um atirador acertou o alvo, é mais provável que ele tenha usado um fuzil do tipo  $A_1$  ou  $A_2$ ?**

**a) Considere que a subjetividade do comandante o levou a assumir que 40% dos fuzis são do tipo  $A_1$  e 60% do tipo  $A_2$ .**

B: acertar o alvo;  $A_1$  : alta precisão;  $A_2$  : baixa precisão

$P(B|A_1)=0,95$ ;  $P(B|A_2)=0,80$ ;  $P(A_1)=0,40$  e  $P(A_2) = 0,60$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{0,95 \times 0,40}{0,95 \times 0,40 + 0,80 \times 0,60} = 0,44$$

$P(A_2|B) = 1 - 0,44 = 0,56 \Rightarrow$  prevaleceu a maior quantidade de  $A_2$

**b) Considere que novas avaliações experimentais (dados) indicaram que as probabilidades de acerto dos fuzis do tipo  $A_1$  e  $A_2$  foram de 95% e 60%, respectivamente.**

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{0,95 \times 0,40}{0,95 \times 0,40 + 0,60 \times 0,60} = 0,51$$

$P(A_2|B) = 1 - 0,51 = 0,49 \Rightarrow$  prevaleceu a maior qualidade de  $A_1$

## 2. Teorema de Bayes (a nível de variáveis aleatórias)

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)}$$

- Substituindo os eventos A e B, respectivamente, pelas v.a. X e Y:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \begin{cases} P(X|Y): \text{fdp condicional de X dado Y} \\ P(Y|X): \text{fdp condicional de Y dado X} \\ P(X): \text{fdp de X e } P(Y): \text{fdp de Y} \end{cases}$$

**OBS.:** O principal objetivo da Estatística é fazer inferência a respeito de parâmetros desconhecidos, e para estimar tais parâmetros é preciso dados (observações amostrais)

- A ideia foi associar a distribuição de probabilidade do parâmetro ( $\theta$ ) a distribuição da v.a. X, e associar a distribuição conjunta dos dados amostrais dado os parâmetros a distribuição condicional de Y dado X

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

## Teorema de Bayes definido em termos de densidades de probabilidade

$$P(X | Y) = \frac{P(Y | X)P(X)}{P(Y)}$$

$\theta$     $y$

$\left\{ \begin{array}{l} \theta : \text{vetor de parâmetros} \\ y : \text{vetor de observações} \end{array} \right.$

$$P(\theta | y) = \frac{P(y, \theta)}{P(y)} = \frac{P(y, \theta)}{\int_{\theta \in \Theta} P(y, \theta) d_{\theta}} = \frac{P(y | \theta)P(\theta)}{\underbrace{\int_{\theta \in \Theta} P(y | \theta)P(\theta) d_{\theta}}_{\text{não depende de } \theta}}$$

$$P(\theta | y) \propto P(y | \theta)P(\theta)$$

Distribuição  
*a posteriori* de  $\theta$

Distribuição conjunta dos dados dado os  
parâmetros: função de verossimilhança

Distribuição  
*a priori* de  $\theta$

## 2.1. Função de verossimilhança: $P(y | \theta)$

É a distribuição conjunta dos dados amostrais  $(y_1, y_2, \dots, y_n)$  condicional aos parâmetros.

Assume-se:  $y_i | \theta \stackrel{iid}{\sim} D(\theta)$ , sendo  $y_i$  uma observação amostral que segue uma distribuição de probabilidade  $D$  com conjunto de parâmetros  $\theta$  e f.d.p. (caso contínuo) ou f.p. (caso discreto) dada por  $P(y_i | \theta)$ .

**OBS.:** Sendo  $x$  e  $y$  duas v.a. independentes, a distribuição conjunta das mesmas é dada por:

$$P(xy) = P(x)P(y)$$

**Dist. Prob. Conjunta das observações amostrais (assumindo-se iid) dado os parâmetros:**

$$P(y_1, y_2, \dots, y_n | \theta) = P(y_1 | \theta) \underbrace{P(y_2 | \theta) \dots P(y_n | \theta)}$$

$$P(y | \theta) = \prod_{i=1}^n P(y_i | \theta) \quad \text{f.d.p. ou f.p. de } D(\theta)$$

## 2.1. Função de verossimilhança: $P(\mathbf{y} \mid \theta)$

➤ **Exemplo 1:** Se  $y_1, y_2, \dots, y_n$  são amostras aleatórias, em que  $y_i \sim \text{Bernoulli}(p)$

$$P(y_i \mid p) = p^{y_i} (1-p)^{1-y_i}$$

$$p(\mathbf{y} \mid p) = \prod_{i=1}^n P(y_i \mid p) = \prod_{i=1}^n [p^{y_i} (1-p)^{1-y_i}] = p^{\sum_{i=1}^n y_i} (1-p)^{n - \sum_{i=1}^n y_i}$$

➤ **Exemplo 2:** Se  $y_1, y_2, \dots, y_n$  são amostras aleatórias, em que  $y_i \sim \text{Poisson}(\lambda)$

$$P(y_i \mid \lambda) = \frac{\exp(-\lambda) \lambda^{y_i}}{y_i!}$$

$$P(\mathbf{y} \mid \lambda) = \prod_{i=1}^n P(y_i \mid \lambda) = \prod_{i=1}^n \left[ \frac{\exp(-\lambda) \lambda^{y_i}}{y_i!} \right] = \frac{\exp(-n\lambda) \lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}$$



## 2.1. Função de verossimilhança: $P(y | \theta)$

➤ **Exemplo 3:** Se  $y_1, y_2, \dots, y_n$  são amostras aleatórias, em que:

$$y_i = a + bx_i + e_i, \quad i = 1, 2, \dots, n,$$

**OBS.:**

$$\theta = [a, b, \sigma^2]$$

$$e_i | a, b, \sigma^2 \sim N(0, \sigma^2) \iff y_i | a, b, \sigma^2 \sim N(a + bx_i, \sigma^2)$$

$$P(y | a, b, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} [y_i - (a + bx_i)]^2 \right\}$$

$$P(y | a, b, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (a + bx_i)]^2 \right\}$$

## 2.2. Distribuição a priori: $P(\theta)$

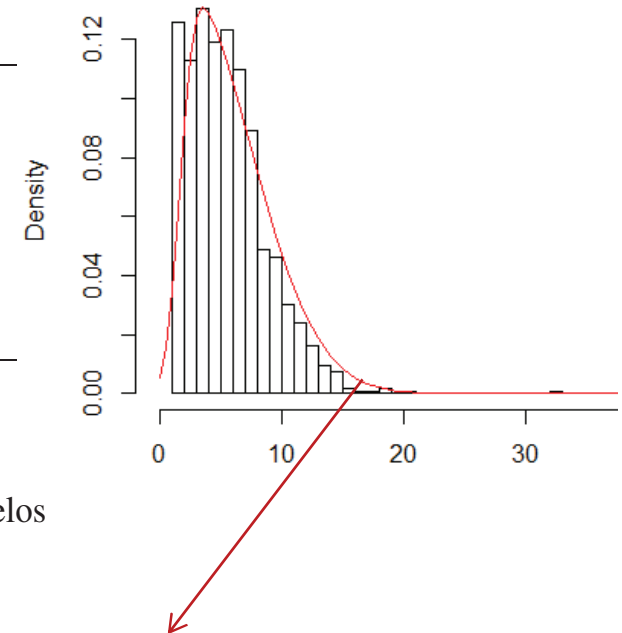
- É a distribuição assumida para o parâmetro, independentemente dos dados.
- É por meio desta distribuição que o pesquisador incorpora o conhecimento prévio a respeito do parâmetro a ser estimado (subjetividade).

### ➤ Ex. Informatividade baseada em estudos prévios

Observação: Uma maneira de se obter a distribuição a priori é por meio de meta-análise (utilizando estudos anteriores). Considerando um parâmetro  $\theta$  qualquer:

Referências *	$\theta$
1	$\theta_1$
2	$\theta_2$
...	...
N	$\theta_N$

\* Estudos semelhantes ao realizado pelo pesquisador.



O objetivo é plotar um histograma verificando qual a distribuição formada pelos  $\theta$ 's e assumi-la como sendo a distribuição a priori de  $\theta$ .

**OBS. Pode-se ajustar uma distribuição ao histograma Ex. Gama(2,3)**

**OBS.** Em estudos prospectivos a distribuição a priori pode ser usada como um mecanismo de atualização de informação (atualização do conhecimento).

1ª análise:

$$P(\theta | \mathbf{y}_1) \propto P(\mathbf{y}_1 | \theta)P(\theta)$$

2ª análise:

$$P(\theta | \mathbf{y}_2) \propto P(\mathbf{y}_2 | \theta)P(\theta | \mathbf{y}_1)$$

Atualização: Distribuições a posteriori tornam-se dist. a priori em novos estudos.

**OBS.** Caso não se tenha nenhuma informação prévia, pode-se utilizar uma distribuição a priori não informativa.

Theor Appl Genet

DOI 10.1007/s00122-013-2089-6

ORIGINAL PAPER

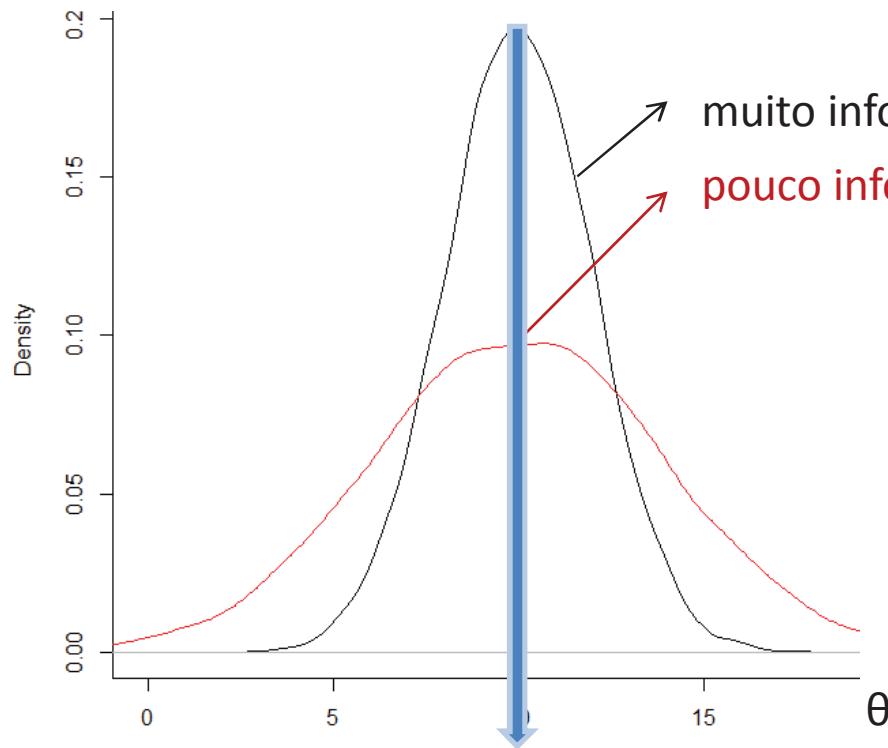
## Bayesian inference of mixed models in quantitative genetics of crop species

Fabyano Fonseca e Silva · José Marcelo Soriano Viana ·  
Vinícius Ribeiro Faria · Marcos Deon Vilela de Resende

## 2.2.1. Classificação de Dist. a priori quanto a informatividade

### 2.2.1.1. Dist. a priori informativas

- É aquela que “afeta” a distribuição a posteriori, fazendo com que as informações nela contida não sejam provenientes apenas dos dados (Função de Verossimilhança). Ou seja, acrescenta uma informação a mais na análise, a qual não está baseada nos dados.
- Quanto menor a variância da distribuição a priori maior é a informatividade presente nessa distribuição.



muito informativa: alta densidade em torno de  $\theta^*$

**pouco informativa:** baixa densidade em torno de  $\theta^*$

**OBS.** A Informatividade é regulada pelos valores assumidos para os parâmetros da distribuição a priori, os quais são denominados Hiperparâmetros.

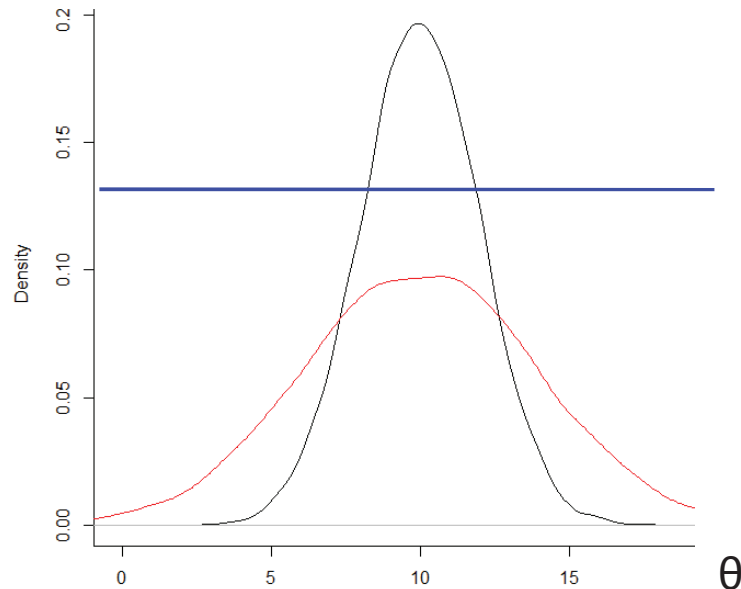
$\theta^*$ : valor esperado a priori (subjetivo) para o parâmetro de interesse

## 2.2.1.2. Dist. a priori não informativas

➤ Essas distribuições fazem com que a distribuição a posteriori dependa apenas dos dados observados (Função de Verossimilhança).

➤ Dist. a priori não informativa “Própria”:

- Distribuição Uniforme:
- Qualquer distribuição que apresente variância extremamente grande

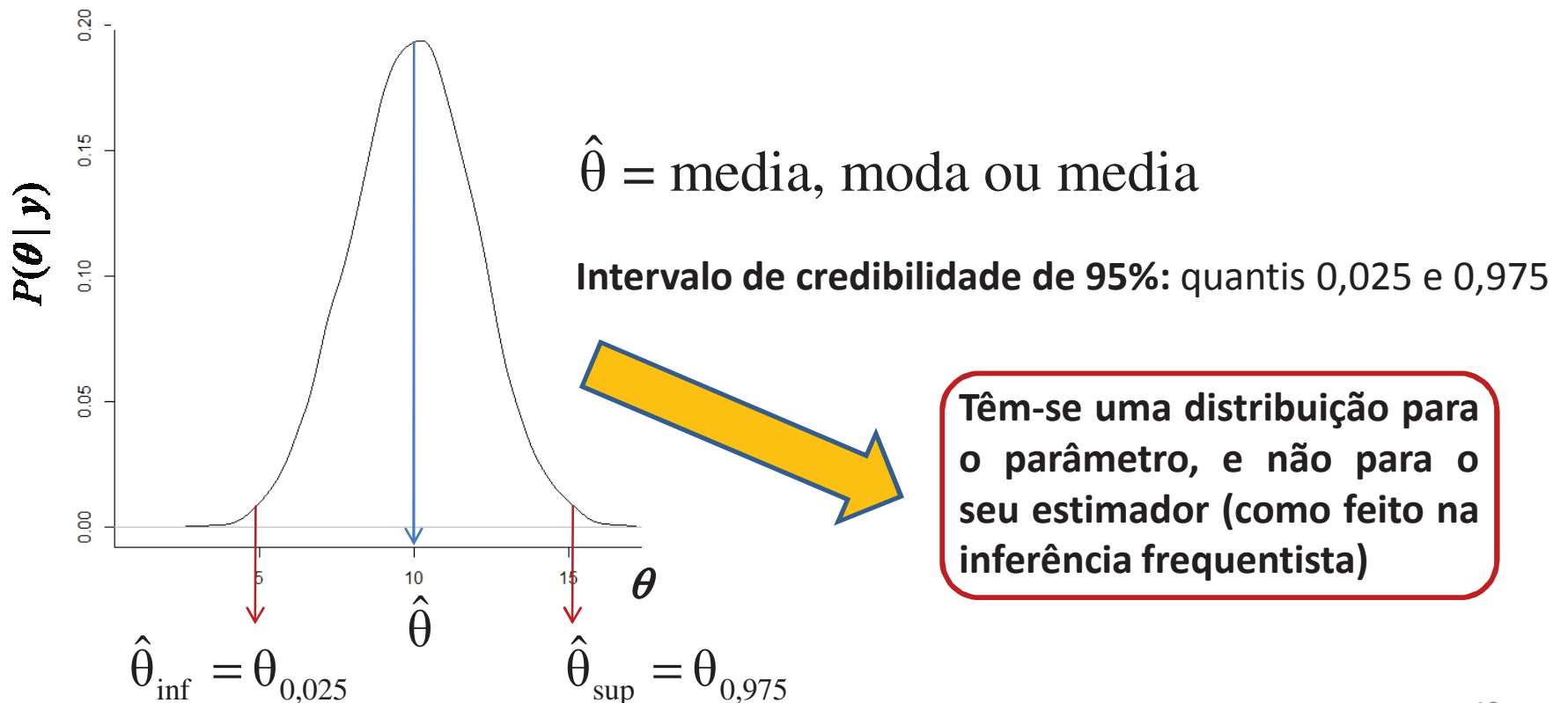


Priori não informativa: todos os valores para o parâmetro são igualmente prováveis

➤ Dist. a priori não informativa “Imprópria”: geralmente uma constante, como por exemplo a Priori de Jeffreys

## 2.3. Distribuição a posteriori: $P(\theta | y)$

- É a distribuição do parâmetro condicionada aos dados, de forma que toda a inferência a respeito do parâmetro é feita por meio desta distribuição.
- Os estimadores bayesianos são caracterizados pela média, moda ou mediana desta distribuição
- A precisão dos estimadores bayesianos está fundamentada na variância desta distribuição
- A estimação intervalar bayesiana está fundamentada nos quantis desta distribuição



### 3. Inferência Bayesiana Uni-paramétrica

**Ex1.** Assumindo:  $y_i \sim \text{Poisson}(\lambda)$  e  $\lambda \sim \text{Gama}(a, b)$ .

Função de verossimilhança:

$$P(y | \lambda) = \prod_{i=1}^n P(y_i | \lambda) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \propto e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}$$

Distribuição a priori:

$$P(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \propto \lambda^{a-1} e^{-b\lambda}.$$

Portanto, a distribuição a posteriori é dada por:  $P(\lambda | y) \propto P(y | \lambda)P(\lambda)$

$$P(\lambda | y) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i} \lambda^{a-1} e^{-b\lambda} \propto \lambda^{\sum_{i=1}^n y_i + a - 1} e^{-(n+b)\lambda} \propto \lambda^{(\sum_{i=1}^n y_i + a) - 1} e^{-(n+b)\lambda}$$

Tome  $a^* = \sum_{i=1}^n y_i + a$  e  $b^* = n + b$ ,

$P(\lambda | y) \propto \lambda^{a^*-1} e^{-b^*\lambda} \Rightarrow$  Núcleo da f.d.p. Gamma:

$$\lambda | y \sim \text{Gama}(a^*, b^*) \quad \text{ou} \quad \lambda | y \sim \text{Gamma}\left(\sum_{i=1}^n y_i + a, n + b\right).$$

- Qual é o estimador bayesiano?

$$\text{Se } X \sim \text{Gama}(a, b) \rightarrow E(X) = \frac{a}{b}.$$

$$\text{Desta forma, } E(\lambda | y) = \frac{a^*}{b^*} = \frac{\sum_{i=1}^n y_i + a}{n + b}.$$

Dist. a posteriori

## 4. Inferência Bayesiana multi-paramétrica

➤ quando dois ou mais parâmetros precisam ser estimados.

$y_i \sim D(\boldsymbol{\theta})$ , sendo  $\boldsymbol{\theta} = [\theta_1, \theta_2]'$  e  $P(y_i)$  a f.d.p. de  $D(\boldsymbol{\theta})$ .

$P(\mathbf{y} | \boldsymbol{\theta}) = \prod_{i=1}^n P(y_i)$  é a função de verossimilhança

$P(\theta_1)$  é dist. a priori do parâmetro  $\theta_1$ .

$P(\theta_2)$  a dist. priori do parâmetro  $\theta_2$ .

Assim, a distribuição a posteriori é dada por:

$$P(\theta_1, \theta_2 | \mathbf{y}) = \frac{P(\mathbf{y} | \theta_1, \theta_2)P(\theta_1, \theta_2)}{P(\mathbf{y})} \propto P(\mathbf{y} | \theta_1, \theta_2)P(\theta_1)P(\theta_2),$$

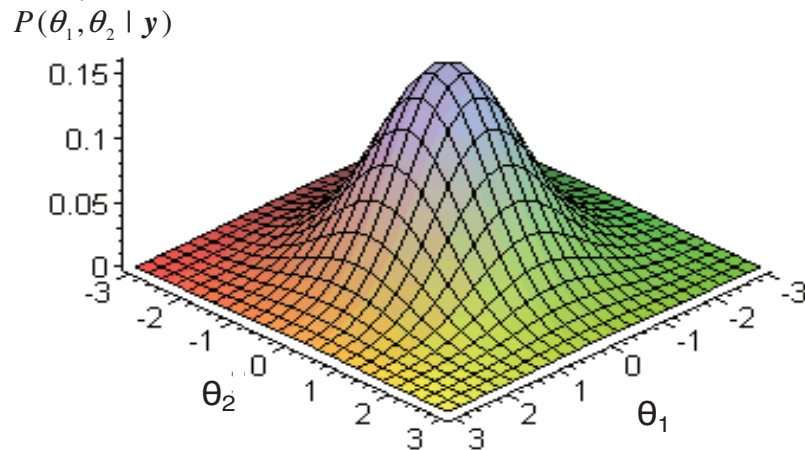
em que  $P(\mathbf{y}) = \int \int P(\mathbf{y} | \theta_1, \theta_2)P(\theta_1)P(\theta_2)d\theta_1d\theta_2$ .

**Distribuição conjunta a posteriori**

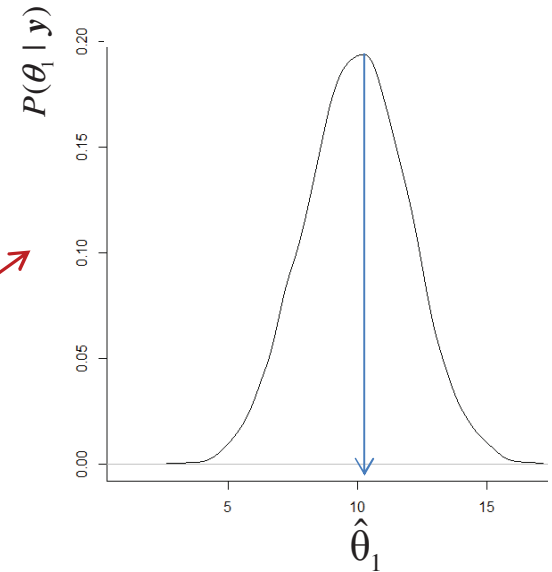


## Distribuição conjunta a posteriori é multi-dimensional

Ex. Situação envolvendo 2 parâmetros ( $\theta_1$  e  $\theta_2$ )

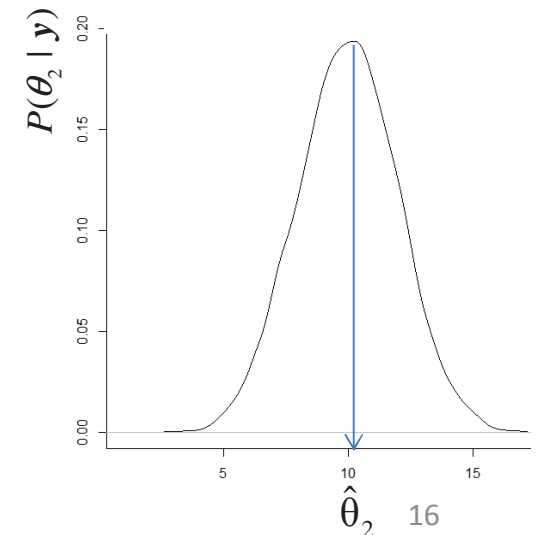


- Como obter o estimador bayesiano de  $\theta_1$ ?



Distribuição marginal a posteriori  $\theta_1$ :  $P(\theta_1 | y) = \int P(\theta_1, \theta_2 | y) d\theta_2$

- Como obter o estimador bayesiano de  $\theta_2$ ?



Distribuição marginal a posteriori  $\theta_2$ :  $P(\theta_2 | y) = \int P(\theta_1, \theta_2 | y) d\theta_1$

## Generalizando para p parâmetros obtém-se:

Função de Verossimilhança:  $P(\mathbf{y} | \theta_1, \theta_2, \dots, \theta_p)$ .

Distribuições a priori:  $P(\theta_1), P(\theta_2), \dots, P(\theta_p)$ .

Distribuição conjunta a posteriori:

$$P(\theta_1, \theta_2, \dots, \theta_p | \mathbf{y}) \propto P(\mathbf{y} | \theta_1, \theta_2, \dots, \theta_p) P(\theta_1) P(\theta_2) \dots P(\theta_p).$$

Inferência Marginal (Distribuições marginais a posteriori):

$$\text{Marginal de } \theta_1: P(\theta_1 | \mathbf{y}) = \int \int \dots \int P(\theta_1, \theta_2, \dots, \theta_p | \mathbf{y}) d\theta_2 \dots d\theta_p.$$

$$\text{Marginal de } \theta_2: P(\theta_2 | \mathbf{y}) = \int \int \dots \int P(\theta_1, \theta_2, \dots, \theta_p | \mathbf{y}) d\theta_1 \dots d\theta_p.$$

$\vdots$

$$\text{Marginal de } \theta_p: P(\theta_p | \mathbf{y}) = \int \int \dots \int P(\theta_1, \theta_2, \dots, \theta_p | \mathbf{y}) d\theta_1 d\theta_2 \dots d\theta_{p-1}$$

**Problema: na maioria das vezes é impossível obter solução analítica**

**Motivação:** exemplo análise bayesiana de uma regressão LASSO, com m=60000 SNPs em GWS

- **Distribuição a posteriori (Teorema de Bayes):  $P(\theta|Y) \propto P(Y|\theta)P(\theta)$**

$$\underbrace{P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y})}_{\text{Posteriori}} \propto \underbrace{P(\mathbf{y} | \mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2)}_{\text{verossimilhança}} \underbrace{\prod_{i=1}^m P(a_i | \tau_i, \sigma_e^2)}_{\text{priori } a_i} \underbrace{\prod_{i=1}^m P(\tau_i^2 | \lambda^2)}_{\text{priori } \tau_i^2} \underbrace{P(\lambda^2)}_{\text{priori } \lambda^2} \underbrace{P(\sigma_e^2)}_{\text{priori } \sigma_e^2}$$

$$P(a_1 | \mathbf{y}) \propto \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_2} \dots d_{a_m} d_{\tau_1} \dots d_{\tau_m} d_{\lambda} d_{\sigma_e^2}$$

$$P(a_2 | \mathbf{y}) \propto \int \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_1} d_{a_3} \dots d_{a_m} d_{\tau_1} \dots d_{\tau_m} d_{\lambda} d_{\sigma_e^2}$$

⋮

$$P(a_m | \mathbf{y}) \propto \int \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_1} d_{a_2} \dots d_{a_{m-1}} d_{\tau_1} \dots d_{\tau_m} d_{\lambda} d_{\sigma_e^2}$$

$$P(\tau_1 | \mathbf{y}) \propto \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_1} \dots d_{a_m} d_{\tau_2} \dots d_{\tau_m} d_{\lambda} d_{\sigma_e^2}$$

$$P(\tau_2 | \mathbf{y}) \propto \int \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_1} \dots d_{a_m} d_{\tau_2} \dots d_{\tau_m} d_{\lambda} d_{\sigma_e^2}$$

⋮

$$P(\tau_m | \mathbf{y}) \propto \int \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_1} \dots d_{a_m} d_{\tau_1} \dots d_{\tau_{m-1}} d_{\lambda} d_{\sigma_e^2}$$

$$P(\lambda | \mathbf{y}) \propto \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_2} \dots d_{a_m} d_{\tau_1} \dots d_{\tau_m} d_{\sigma_e^2}$$

$$P(\sigma_e^2 | \mathbf{y}) \propto \int \dots \int \int \dots \int \int \int P(\mathbf{a}, \boldsymbol{\tau}, \lambda, \sigma_e^2 | \mathbf{y}) d_{a_2} \dots d_{a_m} d_{\tau_1} \dots d_{\tau_m} d_{\lambda}$$

**Como resolver estas  
integrais??**

**Impossível!!!**

**Solução: amostrar valores das  
marginais a posteriori de  
forma indireta por meio de  
amostras das distribuições  
condicionais completas  
(MCMC)**

## 5. Algoritmos MCMC (Cadeias de Markov via Simulação Monte Carlo )

- A técnica de gerar valores de uma distribuição de probabilidade é chamada de simulação Monte Carlo, e quando tais valores são gerados sob um processo (cadeia) de Markov, tem-se o MCMC.
- Em inferência Bayesiana, métodos MCMC visam gerar valores aleatórios das distribuições marginais a posteriori, de forma que após a geração de uma grande quantidade destes valores (se  $N \rightarrow \infty$ ), pode-se assumir que se conhece tais distribuições (evita a resolução das integrais múltiplas relatada anteriormente)
- Estes valores das dist. marginais a posteriori são gerados indiretamente por meio de uma nova classe de distribuição, denominada **distribuição condicional completa a posteriori – D.C.C.P.** (serão vistas detalhadamente)
- Para obter a D.C.C.P. de um certo parâmetro, basta simplesmente assumir que todos os outros parâmetros são constantes na distribuição conjunta a posteriori

## ➤ Distribuição condicional completa a posteriori – D.C.C.P

$$\underbrace{P(\theta_1, \theta_2, \theta_3 | y)}_{\text{Dist conjunta a posteriori}} \propto \underbrace{P(y | \theta_1, \theta_2, \theta_3)}_{\text{Fun. Verossimilhança}} \underbrace{P(\theta_1)}_{\text{Dist a priori de } \theta_1} \underbrace{P(\theta_2)}_{\text{Dist a priori de } \theta_2} \underbrace{P(\theta_3)}_{\text{Dist a priori de } \theta_3}$$

### • D.C.C.P. para $\theta_1$

$$P(\theta_1 | y, \theta_2, \theta_3) \propto \underbrace{P(y | \theta_1, \theta_2, \theta_3)P(\theta_1)}_{\text{apenas termos que contenham } \theta_1}$$



$$P(\theta_2)P(\theta_3) \propto P(y | \theta_1, \theta_2, \theta_3)P(\theta_1)$$

### • D.C.C.P. para $\theta_2$

$$P(\theta_2 | y, \theta_1, \theta_3) \propto \underbrace{P(y | \theta_1, \theta_2, \theta_3)P(\theta_2)}_{\text{apenas termos que contenham } \theta_2}$$



$$P(\theta_1)P(\theta_3) \propto P(y | \theta_1, \theta_2, \theta_3)P(\theta_2)$$

### • D.C.C.P. para $\theta_3$

$$P(\theta_3 | y, \theta_1, \theta_2) \propto \underbrace{P(y | \theta_1, \theta_2, \theta_3)P(\theta_3)}_{\text{apenas termos que contenham } \theta_3}$$



$$P(\theta_1)P(\theta_2) \propto P(y | \theta_1, \theta_2, \theta_3)P(\theta_3)$$

Têm-se o produto da função de veros. pela dist. a priori de cada um dos parâmetros!

Este produto resulta na f.d.p. de uma dist. de prob. conhecida?

Se a D.C.C.P. de um certo parâmetro permitir identificar o núcleo (forma básica) de uma dist conhecida pode-se utilizar o algoritmo **Gibbs sampler**, caso contrário, deve-se optar por um algoritmo mais complexo, como o **Metropolis-Hastings**

**Esta identificação constitui a parte mais trabalhosa de uma análise bayesiana: requer um grande conhecimento de distribuições de probabilidade e de recursos matemáticos (completamento de quadrados, leis de integração e de derivação, propriedades de funções)**

## ➤ 5.1 Algoritmo Gibbs Sampler:

Utilizado quando se tem uma D.C.C.P cujo núcleo pode ser identificado como sendo o núcleo da f.d.p. de uma dist de prob. conhecida

**Notação:**  $\theta_i^{(k)}$  (parâmetro  $i$ ,  $i=1,2,\dots,p$ , na iteração  $k$ ,  $k=1,2,\dots,N$ ). OBS.  $t=0$  representa valores iniciais



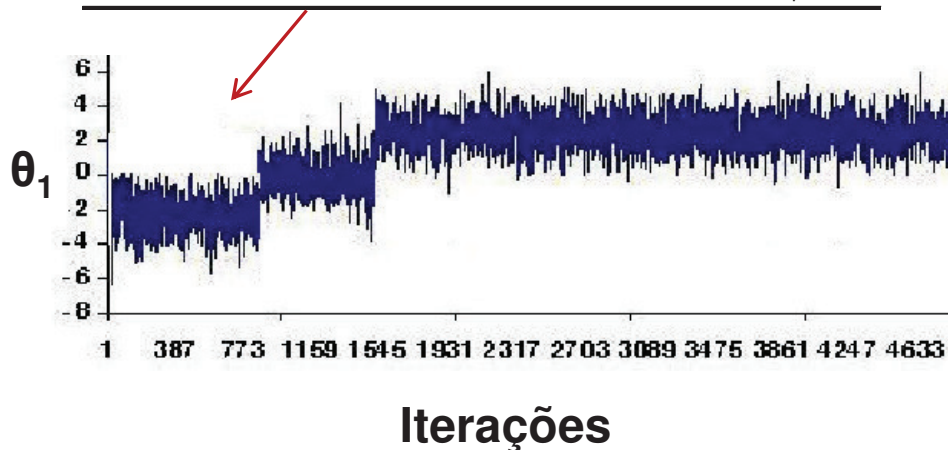
**OBS.:**  
Valores gerados  
diretamente por meio  
de dist. conhecidas  
(simulação Monte  
Carlo)

## ➤ Algoritmo Gibbs Sampler:

Ao final do processo (quando  $k=N$ ), o tem-se a seguinte estrutura de valores gerados:

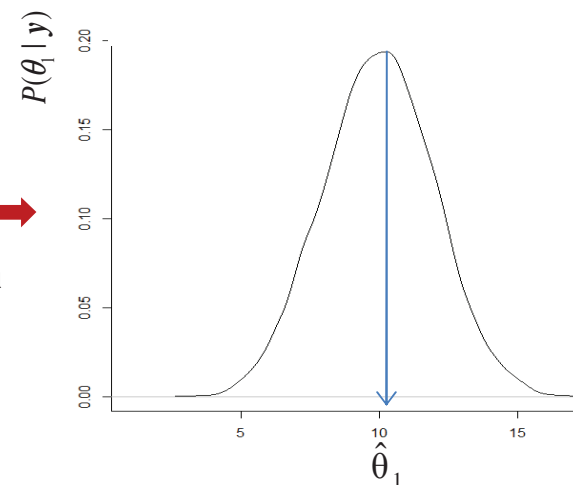
Iteração	$\theta_1$	$\theta_2$	...	$\theta_p$
1	$\theta_1^{(1)}$	$\theta_2^{(1)}$	...	$\theta_p^{(1)}$
2	$\theta_1^{(2)}$	$\theta_2^{(2)}$	...	$\theta_p^{(2)}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
N	$\theta_1^{(N)}$	$\theta_2^{(N)}$	...	$\theta_p^{(N)}$

**OBS.** Em cada coluna, tem-se um conjunto de valores gerados (cadeia) para cada um dos parâmetros ( $\theta_i$ )



Se  $N \longrightarrow \infty$

Dist. marginal a posteriori de  $\theta_1$



**OBS.** Trata-se de um algoritmo iterativo, portanto faz-se necessário verificar a sua convergência (avaliar se realmente  $N$  tende ao infinito, se as amostras são independentes e se a cadeia de markov está em equilíbrio)

## ➤ 5.2 Algoritmo Metropolis-Hastings:

- Utilizado quando se tem uma D.C.C.P cujo núcleo NÃO pode ser identificado como sendo o núcleo da f.d.p. de uma dist de prob. Conhecida
- O algoritmo baseia-se então na utilização de uma distribuição CANDIDATA para gerar valores das D.C.C.P , portanto tem-se um processo de geração de valores das D.C.C.P indiretamente por meio de distribuições candidatas

### Gibbs Sampler

Gera-se  $\theta_1^{(1)}$  de  $P(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y)$  → D.C.C.P. de  $\theta_1$   
Gera-se  $\theta_1^{(2)}$  de  $P(\theta_1|\theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_p^{(1)}, Y)$  → D.C.C.P. de  $\theta_1$   
⋮  
Gera-se  $\theta_1^{(N)}$  de  $P(\theta_1|\theta_2^{(N-1)}, \theta_3^{(N-1)}, \dots, \theta_p^{(N-1)}, Y)$  → D.C.C.P. de  $\theta_1$

**OBS.:**

Valores gerados diretamente por meio de dist. conhecidas (simulação Monte Carlo)

### Metropolis-Hastings

Gera-se  $\theta_1^{(k)}$  de uma dist candidata, testa-se se o valor gerado pode ser considerado uma amostra D.C.C.P

- **se sim:** considera-se  $\theta_1^{(k)}$  como sendo o valor gerado para  $\theta_1$  na iteração k
- **se não:** considera-se  $\theta_1^{(k)} = \theta_1^{(k-1)}$  (ou seja, o valor gerado na iteração k é o mesmo valor da iteração anterior)

Faz-se o mesmo para cada parâmetro!!!

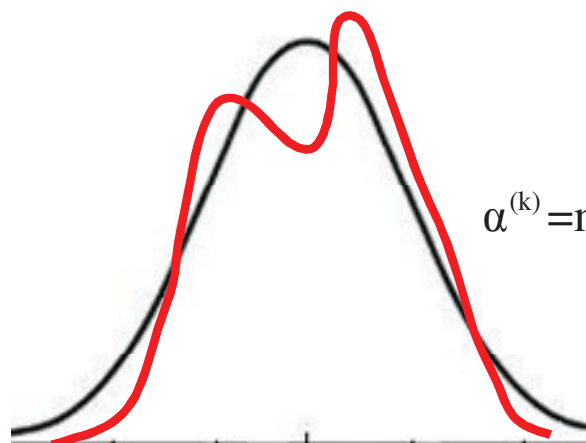
**OBS.:**

Valores gerados indiretamente por meio de dist. Candidatas (simulação Monte Carlo)



## ➤ Algoritmo Metropolis-Hastings:

- Alvo (D.C.C.P): não se caracteriza como uma dist de prob conhecida
- Candidata



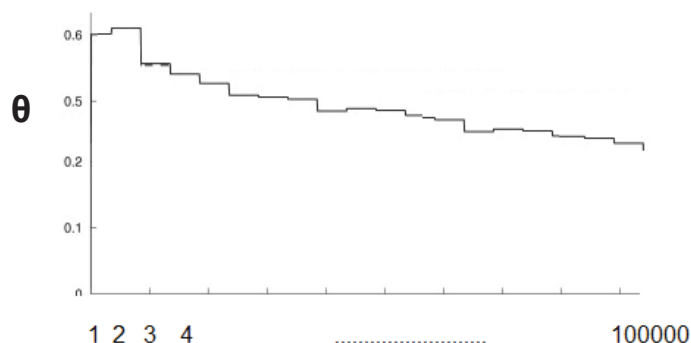
$$\alpha^{(k)} = \min \left( 1, \frac{P(\theta^{(k)} | \mathbf{y}, \text{demais parâmetros})}{P(\theta^{(k-1)} | \mathbf{y}, \text{demais parâmetros})} \right)$$

Densidade da D.C.C.P de  $\theta$  ao substituir o valor de  $\theta$  pelo valor gerado na iter  $k$  (atual)

Densidade da D.C.C.P de  $\theta$  ao substituir o valor de  $\theta$  pelo valor gerado na iter  $k-1$  (anterior)

- Se  $\alpha^{(k)} \geq u^{(k)}$  aceita-se  $\theta^{(k)}$  (pertence a dist alvo)
- Se  $\alpha^{(k)} < u^{(k)}$  não aceita-se  $\theta^{(k)}$  (gera-se um novo valor)

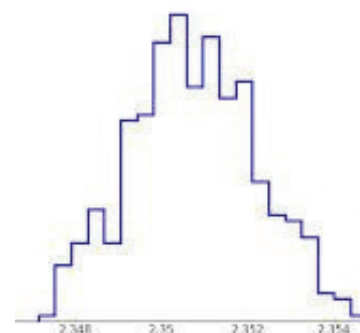
$$u^{(k)} \sim U[0,1]$$



Iterações

Se  $N \longrightarrow \infty$

Dist. marginal a posteriori de  $\theta$



### ➤ 5.3 Verificação da convergência: algoritmos MCMC

são algoritmos iterativos, sendo necessário avaliar a convergência

- Número de iterações
- “Bur-in” (eliminação das primeiras iterações)
- “Thin” (intervalo de amostragem)

**Crítérios:** Geweke (1992), Gelman e Rubin (1992), Raftery-Lewis (1992b) e Heidelberg e Welch (1983)

**Protocolo de avaliação:** combinação de vários métodos (Nogueira, 2004)

- 1- aplicar Raftery e Lewis em uma amostra piloto e determinar o tamanho ideal
- 2 - monitorar a convergência pelos critérios de Raftery e Lewis, Gelman e Rubin (cadeias múltiplas!!!), Geweke e Heidelberger e Welch.

**Software R**  $\left\{ \begin{array}{l} \textbf{Coda}(\text{Convergence Diagnosis and Output Analysis}) \\ \textbf{BOA}(\text{Bayesian Output Analysis}) \end{array} \right.$

## ➤ Principais referências sobre algoritmos MCMC

Casella, G.; George, E. (1992) Explaining the Gibbs sampler.

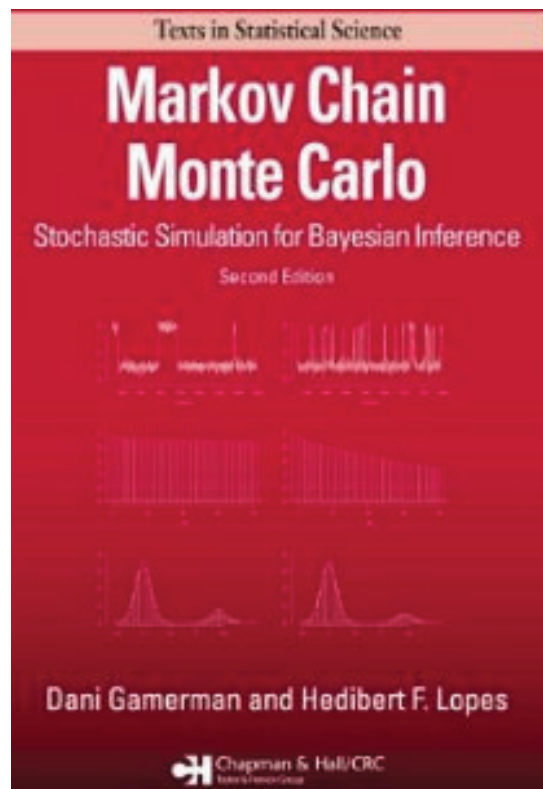
*The American Statistician* 46 (3): 167–174. [doi:10.2307/2685208](https://doi.org/10.2307/2685208)

Chib, S.; Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm.

*The American Statistician* 49 (4): 327– 335. [doi: 10.2307/2684568](https://doi.org/10.2307/2684568)

Smith, B.J. (2007) boa: An R Package for MCMC Output Convergence

Assessment and Posterior Inference. *Journal of Statistical Software*, v.21, Issue 11



Gamerman, D.; Lopes, H.F. 2006 **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.** Second Edition. Chapman & Hall/CRC.

## 6. Inferência Bayesiana: modelo de regressão linear simples

$$y_i = a + bx_i + e_i, \quad e_i \stackrel{iid}{\sim} N(0, \sigma^2) \rightarrow y_i \stackrel{iid}{\sim} N(a + bx_i, \sigma^2)$$

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} [y_i - (a + bx_i)]^2\right\}$$

$$p(\mathbf{y} | a, b, \sigma^2) = \prod_{i=1}^n p(y_i) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (a + bx_i)]^2\right\}$$

Função de verosimilhança

$$a \sim U[a_1, a_2], \quad b \sim U[b_1, b_2], \quad \sigma^2 \sim U[\sigma_1^2, \sigma_2^2]$$

Dist. a priori não informativas  
(Dist uniforme: constantes)

$$p(a, b, \sigma^2 | \mathbf{y}) \propto p(\mathbf{y} | a, b, \sigma^2) p(a) p(b) p(\sigma^2) \propto p(\mathbf{y} | a, b, \sigma^2)$$

Teorema de Bayes

➤ Dist. conjunta a posteriori

$$p(a, b, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (a + bx_i)]^2\right\}$$

$$p(a, b, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i^2 - 2y_i(a + bx_i) + (a + bx_i)^2]\right\}$$

$$p(a, b, \sigma^2 | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i^2 - 2y_i a - 2by_i x_i + a^2 + 2abx_i + b^2 x_i^2] \right\}$$

$$p(a, b, \sigma^2 | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n y_i x_i + na^2 + 2ab \sum_{i=1}^n x_i + b^2 \sum_{i=1}^n x_i^2 \right] \right\}$$

➤ **D.C.C.P. para a**

$$p(a, \sigma^2, b | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n y_i x_i + na^2 + 2ab \sum_{i=1}^n x_i + b^2 \sum_{i=1}^n x_i^2 \right] \right\}$$

$$p(a | b, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ -2a \sum_{i=1}^n y_i + na^2 + 2ab \sum_{i=1}^n x_i \right] \right\} \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ na^2 - 2a \left( \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i \right) \frac{n}{n} \right] \right\}$$

$$p(a | b, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2} n \left[ a^2 - 2a \frac{\left( \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i \right)}{n} \right] \right\}$$

**OBS.**

$$a^2 - 2ak + k^2 - k^2 = (a - k)^2 - k^2$$

$$p(a | b, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2 \frac{1}{n}} \left[ a - \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \right]^2 \right\} \propto \exp \left\{ -\frac{1}{2\sigma_a^{2*}} [a - \mu_a^*]^2 \right\}$$

Gibbs ou MH?

$$a | b, \sigma^2, y \sim N(\mu_a^*, \sigma_{a36}^{2*})$$

➤ D.C.C.P. para b

$$p(a, \sigma^2, b | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n y_i x_i + na^2 + 2ab \sum_{i=1}^n x_i + b^2 \sum_{i=1}^n x_i^2 \right] \right\}$$

$$p(b | a, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ -2b \sum_{i=1}^n y_i x_i + 2ab \sum_{i=1}^n x_i + b^2 \sum_{i=1}^n x_i^2 \right] \right\}$$

$$p(b | a, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ b^2 \sum_{i=1}^n x_i^2 - 2b \left( \sum_{i=1}^n y_i x_i - a \sum_{i=1}^n x_i \right) \right] \right\}$$

↘ K
↘ w

**OBS.**

$$k \left( b - \frac{w}{k} \right)^2 - \frac{w^2}{k} = \left( b^2 k - \frac{2bkw}{k} \right) + \frac{w^2}{k^2} k - \frac{w^2}{k}$$

$$p(b | a, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma^2} k \left[ b - \frac{w}{k} \right]^2 \right\} \propto \exp \left\{ -\frac{1}{\frac{2\sigma^2}{k}} k \left[ b - \frac{w}{k} \right]^2 \right\}$$

↘  $\sigma_b^{2*}$ 
↘  $\mu_b^*$

**OBS.**

$$\frac{k}{2\sigma^2} = \frac{1}{2\sigma^2} \cdot \frac{1}{\frac{1}{k}} = \frac{1}{\frac{2\sigma^2}{k}}$$

$$p(b | a, \sigma^2, y) \propto \exp \left\{ -\frac{1}{2\sigma_b^{2*}} [a - \mu_b^*]^2 \right\}$$

Gibbs ou MH?

$$b | a, \sigma^2, y \sim N(\mu_b^*, \sigma_b^{2*})$$

➤ **D.C.C.P. para  $\sigma^2$**

$$p(a, \sigma^2, b | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (a + bx_i)]^2 \right\}$$

$$p(\sigma^2 | a, b, y) \propto (\sigma^2)^{\frac{-n}{2} - 1 + 1} \exp \left\{ -\frac{1}{\sigma^2} \frac{\sum_{i=1}^n [y_i - (a + bx_i)]^2}{2} \right\}$$

$$p(\sigma^2 | a, b, y) \propto (\sigma^2)^{\alpha^* + 1} \exp \left\{ -\frac{1}{\sigma^2} \beta^* \right\}$$

Gibbs ou MH?

$$\sigma^2 | a, b, y \sim GI(\alpha^*, \beta^*)$$

OBS.: para todos os parâmetros tem-se D.C.C.P. conhecidas, portanto, deve-se utilizar o algoritmo **Gibbs Sampler**

## ➤ Algoritmo Gibbs Sampler

$$a | b, \sigma^2, \mathbf{y} \sim N(\mu_a^*, \sigma_a^{2*}) \sim N\left(\frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}, \frac{\sigma^2}{n}\right)$$

$$b | a, \sigma^2, \mathbf{y} \sim N(\mu_b^*, \sigma_b^{2*}) \sim N\left(\frac{\left(\sum_{i=1}^n y_i x_i - a \sum_{i=1}^n x_i\right)}{\sum_{i=1}^n x_i^2}, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$$

$$\sigma^2 | a, b, \mathbf{y} \sim Gl(\alpha^*, \beta^*) \sim Gl\left(\frac{n}{2} + 1, \frac{\sum_{i=1}^n [y_i - (a + bx_i)]^2}{2}\right)$$

OBS.:

Notar dependência entre os parâmetros!

Ao gerar valores aleatórios destas distribuições tem-se um processo iterativo de atualização



## ➤ Algoritmo Gibbs Sampler

1º Iteração

$$a^{(1)} | b^{(0)}, \sigma^{2(0)}, y \sim N \left( \frac{\sum_{i=1}^n y_i - b^{(0)} \sum_{i=1}^n x_i}{n}, \frac{\sigma^{2(0)}}{n} \right)$$

$$b^{(1)} | a^{(1)}, \sigma^{2(0)}, y \sim N \left( \frac{\left( \sum_{i=1}^n y_i x_i - a^{(1)} \sum_{i=1}^n x_i \right)}{\sum_{i=1}^n x_i^2}, \frac{\sigma^{2(0)}}{\sum_{i=1}^n x_i^2} \right)$$

$$\sigma^{2(1)} | a^{(1)}, b^{(1)}, y \sim GI \left( \frac{n}{2} + 1, \frac{\sum_{i=1}^n [y_i - (a^{(1)} + b^{(1)} x_i)]^2}{2} \right)$$

2º Iteração

$$a^{(2)} | b^{(1)}, \sigma^{2(1)}, y \sim N \left( \frac{\sum_{i=1}^n y_i - b^{(1)} \sum_{i=1}^n x_i}{n}, \frac{\sigma^{2(1)}}{n} \right)$$

$$b^{(2)} | a^{(2)}, \sigma^{2(1)}, y \sim N \left( \frac{\left( \sum_{i=1}^n y_i x_i - a^{(2)} \sum_{i=1}^n x_i \right)}{\sum_{i=1}^n x_i^2}, \frac{\sigma^{2(1)}}{\sum_{i=1}^n x_i^2} \right)$$

$$\sigma^{2(2)} | a^{(2)}, b^{(2)}, y \sim GI \left( \frac{n}{2} + 1, \frac{\sum_{i=1}^n [y_i - (a^{(2)} + b^{(2)} x_i)]^2}{2} \right)$$

kº Iteração

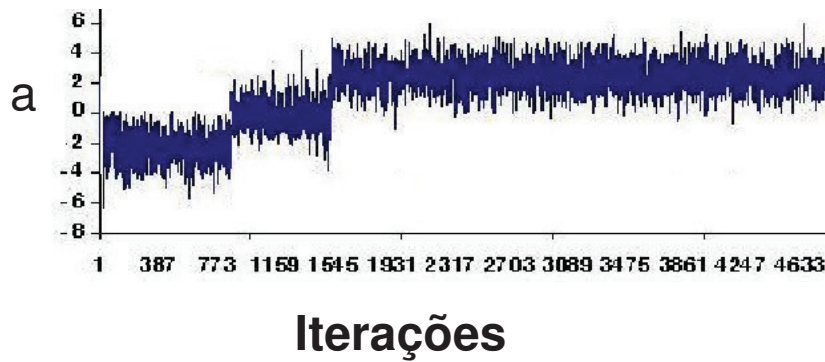
$$a^{(k)} | b^{(k-1)}, \sigma^{2(k-1)}, y \sim N \left( \frac{\sum_{i=1}^n y_i - b^{(k-1)} \sum_{i=1}^n x_i}{n}, \frac{\sigma^{2(k-1)}}{n} \right)$$

■ ■ ■

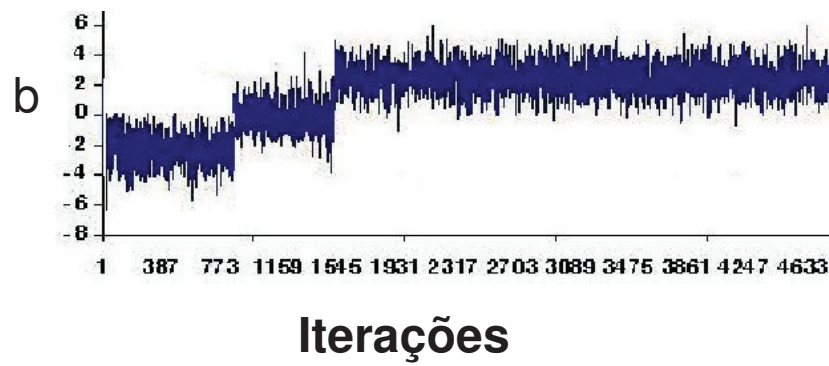
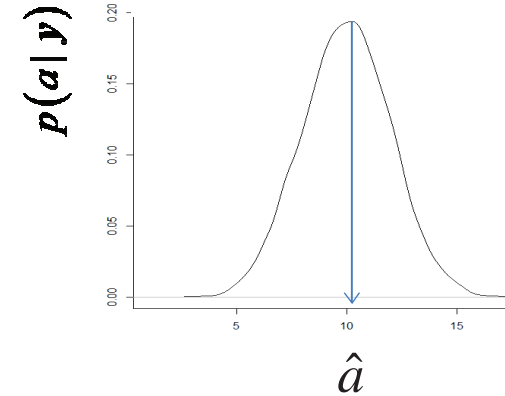
$$b^{(k)} | a^{(k)}, \sigma^{2(k-1)}, y \sim N \left( \frac{\left( \sum_{i=1}^n y_i x_i - a^{(k)} \sum_{i=1}^n x_i \right)}{\sum_{i=1}^n x_i^2}, \frac{\sigma^{2(k-1)}}{\sum_{i=1}^n x_i^2} \right)$$

$$\sigma^{2(k)} | a^{(k)}, b^{(k)}, y \sim GI \left( \frac{n}{2} + 1, \frac{\sum_{i=1}^n [y_i - (a^{(k)} + b^{(k)} x_i)]^2}{2} \right)$$

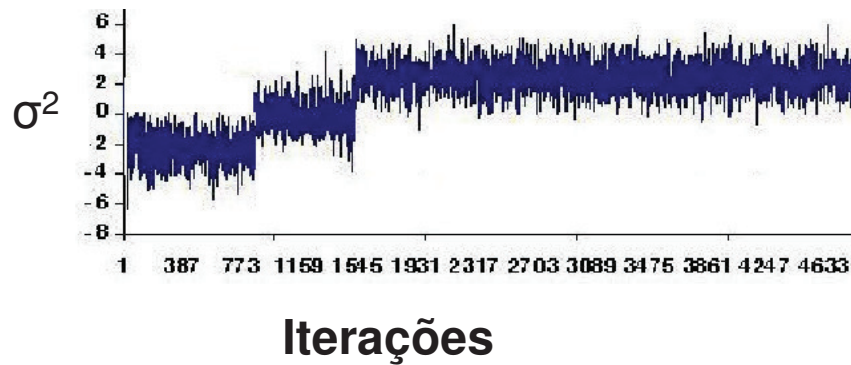
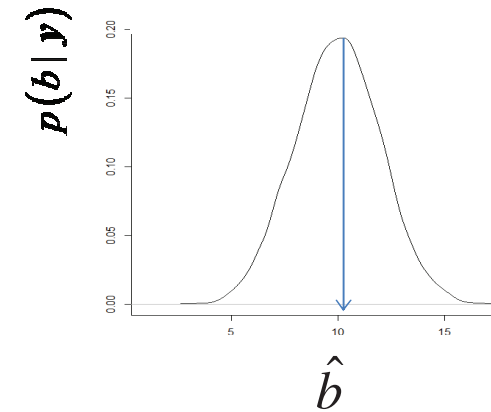
## ➤ Algoritmo Gibbs Sampler



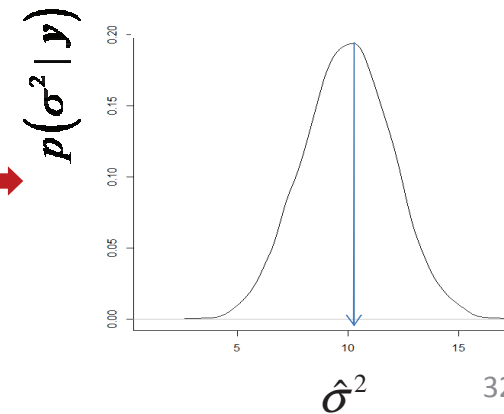
Dist. marginal a posteriori de a



Dist. marginal a posteriori de b



Dist. marginal a posteriori de  $\sigma^2$



```
x=seq(10,100,5)
n=length(x) #número de observações
set.seed(12345) #fixando semente aleatória
y=15 + 2.5*x + rnorm(n,0,sqrt(10))

#componentes das DCCP

soma_y=sum(y)
soma_x=sum(x)
soma_x2=sum(x^2)
soma_xy=sum(x*y)

#Gibbs sampler

Niter=50000 #número de iterações

a=matrix(0,Niter,1) #vetor que vai armazenar valores de a
b=matrix(0,Niter,1) #vetor que vai armazenar valores de b
sig2=matrix(0,Niter,1) #vetor que vai armazenar valores de sig2
mu_star_a=matrix(0,Niter,1) #media da DCCP normal para a
var_star_a=matrix(0,Niter,1) #var da DCCP normal para a
mu_star_b=matrix(0,Niter,1) #media da DCCP normal para b
var_star_b=matrix(0,Niter,1) #var da DCCP normal para b
alfa_star=matrix(0,Niter,1) #alfa da DCCP gama inv para sig2
beta_star=matrix(0,Niter,1) #beta da DCCP gama inv para sig2
```

```

for(i in 1:1) #valores iniciais para os parâmetros
{
a[i]=10
b[i]=5
sig2[i]=20
}
for(i in 2:Niter)
{
#cond completa posteriori de a (algoritmo GS)
mu_star_a[i]=(soma_y-b[i-1]*soma_x)/n
var_star_a[i]=sig2[i-1]/n
a[i]=rnorm(1,mu_star_a[i],sqrt(var_star_a[i]))

#cond completa posteriori de b (algoritmo GS)
mu_star_b[i]=(soma_xy-a[i]*soma_x)/soma_x2
var_star_b[i]=sig2[i-1]/soma_x2
b[i]=rnorm(1,mu_star_b[i],sqrt(var_star_b[i]))

#cond completa posteriori de sig2 (algoritmo GS)
alfa_star[i]= (n/2) +1
beta_star[i]= sum( (y-(a[i]+b[i]*x))^2 )/2
library(MCMCpack) # função de geração de valores aleatórios da gama inversa
sig2[i]=rinvgamma(1,alfa_star[i],beta_star[i])
}
result=cbind(a,b,sig2)
colnames(result)=c("a","b","sig2")      #usar pacote boa para inferência

```