

# Introdução à Estatística Bayesiana

Carlos Daniel Paulino, Giovanni Silva e Paulo Soares  
Instituto Superior Técnico & CEAUL, ULisboa

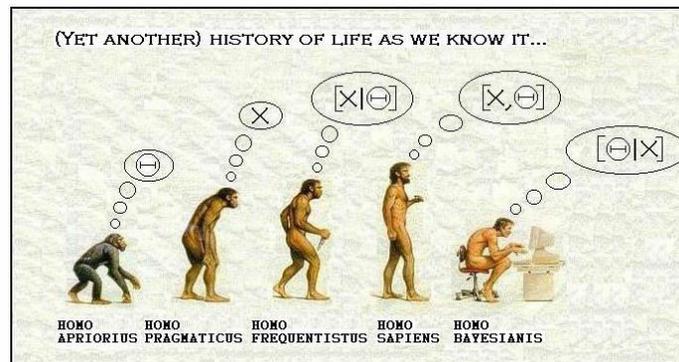


One Day Meeting on Statistics and Applied Probability  
Guimarães, 3 de outubro de 2019

## Conteúdo

- 1 Fundamentos e metodologia bayesiana
- 2 Representação da informação *a priori*
- 3 Aplicações básicas da metodologia bayesiana
- 4 Avaliação de modelos
- 5 Computação bayesiana
- 6 Aplicações

2



## Fundamentos e metodologia bayesiana Estatística Clássica versus Estatística Bayesiana

**Dados:** concretização de um v.a.  $X$ ,  $x \in \mathcal{X}$  /  
 $x = (x_1, \dots, x_n)$ ,  $\mathcal{X} \subseteq \mathbb{R}^n$

**Modelo estatístico:** Especificação com base em natureza do fenómeno, tratamento prévio de fenómenos análogos, evidência experimental, objetivos do estudo, requisitos de parcimónia e interpretabilidade.

$$\mathcal{F} = \{f(x|\theta), x \in \mathcal{X} : \theta \in \Theta\},$$

mas sem conhecimento do valor do índice  $\theta$  que produziu os dados;  
e.g.,  $\Theta \in \mathbb{R}^k$ ,  $f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$  – **modelo amostral**.

4

## Fundamentos e metodologia bayesiana

### Paradigma clássico

**Princípio da amostragem repetida:** avaliador dos procedimentos inferenciais através da análise do seu comportamento num número indefinido de hipotéticas repetições em condições essencialmente idênticas do esquema aleatório originador da amostra (pressuposto).

⇒ medição da incerteza baseada no **conceito frequentista** de probabilidade.

**Via inferencial:** Variáveis aleatórias observáveis (total ou parcialmente) e suas distribuições por amostragem associadas a  $\mathcal{F}$ , com base nas quais se avaliam as propriedades das inferências pré-experimentalmente.

- **Estimação** (estimadores pontuais, regiões de confiança).
- **Predição** (pontual e regional).
- **Testes de hipóteses** (comparação de modelos).

5

## Fundamentos e metodologia bayesiana

### Paradigma bayesiano

**Teorema de Bayes:** Instrumento inferencial fundamental

Ingredientes: **Dados** amostrais do **modelo**  $\{f(x|\theta) : \theta \in \Theta\}$  +

informação apriorística (anterior ou externa a tal amostra) sobre o que é desconhecido quantificada em **distribuição a priori** (inicial):  $h(\theta) : \theta \in \Theta$ ,

na base do argumento crucial:

Tudo o que é desconhecido é incerto e toda a incerteza é suscetível de ser quantificada probabilisticamente!

⇒ parâmetros dos modelos amostrais encarados como aleatórios, numa base tipicamente subjetiva:

6

## Fundamentos e metodologia bayesiana

### Paradigma bayesiano

**conceito subjetivista** de probabilidade – Grau de crença pessoal na ocorrência do evento/ veracidade da proposição, na base da evidência disponível.

⇒ **distribuição a posteriori** (final):

$$h(\theta|x) = \frac{h(\theta)f(x|\theta)}{p(x)} \equiv \frac{h(\theta)f(x|\theta)}{\int_{\Theta} f(x|\theta)h(\theta)d\theta} \quad \theta \in \Theta,$$

em que  $p(x), \forall x$  traduz a **distribuição marginal** dos dados observáveis  $X$ .

⇒ *Precisão das inferências:* pós-experimental (final)

7

## Fundamentos e metodologia bayesiana

### Paradigma bayesiano

**Exemplo I:** Ensaio clínico a uma amostra de  $n$  pacientes visando inferir a probabilidade  $\theta$  de controlo da respetiva doença por uma nova droga.

Dados  $(x_1, \dots, x_n) \leftarrow (X_1, \dots, X_n)$  tal que  $X_i, i = 1, \dots, n \stackrel{iid}{\sim} Ber(\theta)$

Distribuição *a priori*:  $\theta \sim Be(a, b)$ ,

$E(\theta) = 0,40; Var(\theta) = 0,10 \Rightarrow \theta \sim Be(9, 2; 13, 8)$ .

Nota:  $\theta \sim Be(a, b) \Leftrightarrow h(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} I_{(0,1)}(\theta)$

⇒  $E(\theta) = \frac{a}{a+b}; Var(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$ .

8

## Fundamentos e metodologia bayesiana

### Paradigma bayesiano

$$\Rightarrow h(\theta|x_1, \dots, x_n) \propto \theta^{a+\sum_i x_i-1} (1-\theta)^{b+n-\sum_i x_i-1}, \theta \in (0, 1)$$

$$\therefore \theta|x_1, \dots, x_n \sim Be(A, B), A = a + \sum_i x_i; B = b + n - \sum_i x_i$$

Supondo os dados  $x$  resumidos em  $n = 100$ ,  $\sum_i x_i = 30$ ,

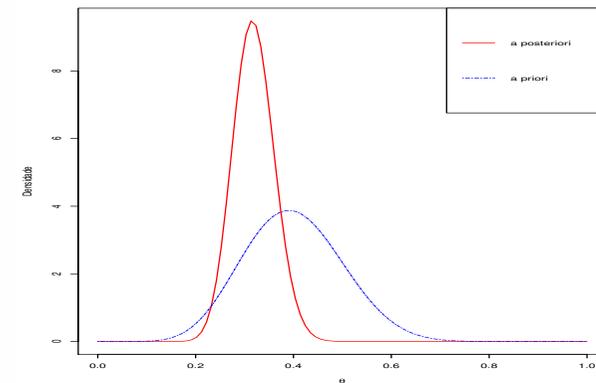
$$\theta|x_1, \dots, x_n \sim Be(A, B), A = 39, 2; B = 83, 8$$

$$\Rightarrow E(\theta|x) = 0, 319; \sigma(\theta|x) = 0, 04; P[\theta \in (0, 238; 0, 401)|x] = 0, 95.$$

9

## Fundamentos e metodologia bayesiana

### Paradigma bayesiano



Distribuições *a priori* Beta(9.2,13.8) e *a posteriori* Beta(39.2,83.8) ■

10

## Fundamentos e metodologia bayesiana

### Metodologia bayesiana

#### Caraterísticas da metodologia bayesiana:

- $h(\theta|x)$ ,  $\theta \in \Theta$  é a descrição completa do conhecimento corrente sobre  $\theta$ , obtido da quantificação da informação *a priori* (em  $h(\theta)$ ) e da informação amostral (em  $f(x|\theta)$ ).
- A parte relevante de  $f(x|\theta)$  para propósitos inferenciais é o fator que envolve  $\theta$ . Tomando-o como a função de verosimilhança  $L(\theta|x)$ , esta é encarada como o veículo de toda a informação amostral.
  - ⇒ Funções de verosimilhança proporcionais conduzem à mesma distribuição *a posteriori*, que é o fulcro de todas as inferências sobre o parâmetro.
  - ⇒ Satisfaz princípios (suficiência, condicionalidade e verosimilhança) que supostamente toda a Inferência Estatística deve respeitar.

11

## Fundamentos e metodologia bayesiana

### Metodologia bayesiana

- A operação bayesiana de atualização do conhecimento tem uma natureza sequencial:  $X = (X_1, X_2), X_1 \perp X_2 | \theta$

$$h(\theta|x) = \frac{h(\theta|x_1)f(x_2|\theta)}{\int_{\Theta} h(\theta|x_1)f(x_2|\theta)d\theta}, \quad p(x_1) > 0$$

∴  $h(\theta|x)$  é a atualização de  $h(\theta|x_1)$  como distribuição *a priori* pela verosimilhança  $f(x_2|\theta)$ .

⇒ Dada a concretização de  $n$   $X_i$  independentes,  $h(\theta|x_1, \dots, x_n)$  pode ser vista como resultante da atualização sucessiva, após observação de cada  $X_i$ , de  $h(\theta|x_1, \dots, x_{i-1})$ ,  $i = 1, \dots, n$  ( $h(\theta|x_0) \equiv h(\theta)$ ).

- Simplicidade conceptual e uniformidade na eliminação de parâmetros perturbadores,  $\theta = (\gamma, \phi) \in \Gamma \times \Phi$ ,  $\gamma$  parâmetro de interesse.

$$\Rightarrow \text{Cálculo de } h(\gamma|x) = \int_{\Phi} h(\gamma, \phi|x)d\phi$$

12

## Fundamentos e metodologia bayesiana

### Estimação pontual

Este problema consiste na determinação de um ponto típico da distribuição *a posteriori*. Escolhas possíveis:

- Moda *a posteriori*

$$\hat{\theta} : h(\hat{\theta}|x) = \max_{\theta \in \Theta} h(\theta|x) = \max_{\theta \in \Theta} [h(\theta)f(x|\theta)]$$

- Média *a posteriori*

$$\hat{\theta} = E[\theta|x] : E[\theta_i|x] = \int_{\Theta} \theta_i h(\theta|x) d\theta,$$

para toda a componente  $\theta_i$  de  $\theta$ .

- Vetor das medianas *a posteriori*

$$\hat{\theta} = (\hat{\theta}_i) : \begin{cases} P[\theta_i \geq \hat{\theta}_i, |x] \geq 1/2 \\ P[\theta_i \leq \hat{\theta}_i, |x] \geq 1/2, \forall i \end{cases}$$

13

## Fundamentos e metodologia bayesiana

### Estimação pontual

**Exemplo I** (revisitado-1): Modelo bayesiano Bernoulli  $\wedge$  Beta

Tomando  $A = a + \sum_i x_i = 39, 2$  e  $B = b + n - \sum_i x_i = 83, 8$ ,  
 $\theta|\{x_i\} \sim Be(A, B)$

Moda *a posteriori*:  $\theta_{mo} = \frac{A-1}{A+B-2} = 0, 316$

Média *a posteriori*:  $\theta_{ve} = \frac{A}{A+B} = 0, 319$

Mediana *a posteriori*:  $\theta_{me} = F_{Be(A,B)}^{-1}(1/2) = 0, 318$  ■

Como escolher entre estes e/ou  $\forall$  outro ponto típico de  $h(\theta|x)$ ?

Num quadro estritamente inferencial não há uma resposta inequívoca. A escolha poderá basear-se na relevância de cada quantidade para o problema em mão e/ou na facilidade do seu cálculo. Uma justificação cabal para uma dada opção exige a incorporação na análise de uma informação adicional sobre as consequências (custos) de cada alternativa  $\Rightarrow$  estimadores Bayes (**Teoria da Decisão Estatística**).

14

## Fundamentos e metodologia bayesiana

### Estimação por regiões

Um resumo de  $h(\theta|x)$  mais informativo do que qualquer estimativa pontual é obtido de uma região de  $\Theta$  que contenha uma parte substancial da massa probabilística *a posteriori* — o paralelo bayesiano da região de confiança:

**Definição:**  $R(x)$  é uma região de credibilidade  $\gamma$  para  $\theta$  se

$$P[\theta \in R(x)|x] \equiv \int_{R(x)} h(\theta|x) d\theta \geq \gamma.$$

Observações:

- Toda a região de credibilidade é definida numericamente (i.e., não é aleatória) e admite uma interpretação probabilística direta e inequívoca — contraste-se com a região de confiança clássica.
- Dada a infinidade de R.C. com a mesma credibilidade  $\gamma$ , interessa obviamente selecionar aquela que englobe todos os valores de  $\theta$  mais credíveis *a posteriori*, ou seja aquela que satisfaz a condição

$$h(\theta_1|x) \geq h(\theta_2|x), \quad \forall \theta_1 \in R(x), \theta_2 \notin R(x).$$

15

## Fundamentos e metodologia bayesiana

### Estimação por regiões

**Critério HPD** (ou de volume mínimo):  $R(x)$  é a região de credibilidade  $\gamma$  com densidade (probabilidade) *a posteriori* máxima se

$$R(x) = \{\theta : h(\theta|x) \geq c_\gamma\},$$

com  $c_\gamma > 0$  a maior constante tal que  $P[\theta \in R(x)|x] \geq \gamma$ .

- As RC HPD não são invariantes face a transformações paramétricas não lineares.
- A determinação das RC HPD na prática exige frequentemente o recurso a métodos numéricos, a não ser que para  $\theta \in \mathbb{R}$   $h(\theta|x)$  seja uma função simétrica. Para distribuições *a posteriori* contínuas em  $\mathbb{R}$ , o cálculo numérico da RC HPD  $R(x|c) = \{\theta : h(\theta|x) \geq c\}$  exige:
  - uma primeira sub-rotina que encontre as soluções das equações  $h(\theta|x) = c$  para  $c > 0$  variável, definidoras de  $R(x|c)$ ;
  - uma segunda sub-rotina que avalie as probabilidades  $P[\theta \in R(x|c)|x]$ .

Uma vez encontrado  $c$  tal que  $P[\theta \in R(x|c)|x] = \gamma$ , a região  $R(x|c)$  será HPD com credibilidade  $\gamma$ .

16

## Fundamentos e metodologia bayesiana

### Estimação por regiões

**Exemplo I** (revisitado-2): Modelo bayesiano Bernoulli  $\wedge$  Beta  
 $\Rightarrow \theta|x \sim Be(39, 2; 83, 8)$

IC HPD a 95%: (0,238; 0,401) - determinável pelo software FIRST BAYES (<http://tonyohagan.co.uk/1b/>)

IC central a 95%: (0,240; 0,403) - determinável pelo MS OFFICE EXCEL (função BETAINV) ■

**Exemplo II:** Modelo bayesiano Normal (d.p. conh.)  $\wedge$  "Uniforme"

$\{X_i\}$  i.i.d. de  $\{N(\mu, \sigma^2), \sigma^2 \text{ conh.}\}$  e  $h(\mu) = k \Rightarrow \mu|x \sim N(\bar{x}, \sigma^2/n)$ .

Est. pontual de  $\mu$ :  $\hat{\mu} = \bar{x}$ ; Reg. credib.  $\gamma$  HPD:

$$R(x) = \left\{ \bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \right\} \quad \blacksquare$$

Nota:  $P[\mu \in R(x)|x] = \gamma$  (medida de precisão final)

$P[\mu \in R(x)|\mu] = \gamma$  (medida de precisão inicial)

$$P[\mu \in R(x)|\mu] = I_{R(x)}(\mu)$$

17

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

O problema de testar  $H_0 : \theta \in \Theta_0$  contra  $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$  é também conceptualmente mais simples do que num contexto clássico. Atendendo à interpretação probabilística direta das hipóteses em confronto, não se tem mais do que calcular as respetivas probabilidades *a posteriori* e optar por uma delas em função de algum critério assente na sua grandeza relativa.

$\Rightarrow$  Cálculo das **chances a posteriori pró- $H_0$** :

$$O(H_0, H_1|x) = \frac{P[H_0|x]}{P[H_1|x]}$$

Com o objetivo de medir a influência dos dados  $x$  na alteração da credibilidade relativa de  $H_0$  e  $H_1$ , opta-se por contrapor as chances *a posteriori* a favor de  $H_0$  às respetivas chances *a priori*, através do

**Fator de Bayes pró- $H_0$** :

$$B(x) = \frac{P[H_0|x]/P[H_1|x]}{P[H_0]/P[H_1]} = \frac{\int_{\Theta_0} f(x|\theta)h_0(\theta)d\theta}{\int_{\Theta_1} f(x|\theta)h_1(\theta)d\theta}$$

onde  $h_i(\theta)$  é a f.(d.)p. *a priori* condicionada em  $H_i$ ,  $i = 0, 1$ .

18

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

Uma situação em que  $B(x) \gg 1$  ou  $B(x) \ll 1$  reflete uma tendência bastante forte nos dados a favor de uma hipótese contra a outra, entendida no sentido de que uma hipótese é muito mais ou muito menos provável *a posteriori* do que era *a priori*. A correspondente desigualdade entre as chances *a posteriori* e *a priori* conduz a que a hipótese favorecida tenha uma probabilidade *a posteriori* superior à da sua alternativa, a não ser que se verifique uma relação fortemente contrária entre as probabilidades *a priori*.

Na prática inferencial costuma-se usar regras orientadoras sobre a interpretação da evidência contida nos dados, como por exemplo estas

$B(x)$	$\ln B(x)$	Evidência
$< 1$	$< 0$	pró- $H_1$
1 – 3	0 – 2	fraca pró- $H_0$
3 – 20	2 – 6	pró- $H_0$
20 – 150	6 – 10	forte pró- $H_0$
$> 150$	$> 10$	muito forte pró- $H_0$

19

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

**Exemplo I** (revisitado-3): Modelo Bernoulli  $\wedge$  Beta

com  $\theta \sim Be(9, 2; 13, 8)$  e  $n = 100$ ,  $\sum_i x_i = 30 \Rightarrow \theta|x \sim Be(39, 2; 83, 8)$ .

Problema:  $\theta \geq 35\%$  versus  $\theta < 35\%$

$$O(H_0, H_1) = \frac{0,679}{0,321} = 2,115; \quad O(H_0, H_1|x) = \frac{0,225}{0,775} = 0,29$$

$$\Rightarrow B(x) = 0,137 \equiv 1/B(x) = 7,3$$

$\therefore$  A chance *a posteriori* pró- $H_1$  é mais de 7 vezes a respetiva chance *a priori*, implicando que  $H_1$  é ainda mais provável (mas apenas cerca de 3 vezes) que  $H_0$  como consequência de *a priori*  $H_1$  ser menos provável que  $H_0$ . ■

20

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

Observações (I):

- A forma dos testes bayesianos elimina a necessidade de uma distinção formal entre o que é a hipótese nula e o que é a hipótese alternativa e a natureza assimétrica do teste clássico.
- Um critério formal de rejeição de  $H_0$  só encontra plena justificação num quadro decisional ( $\Rightarrow$  testes Bayes).
- O valor-P de testes unilaterais pode ter justificação bayesiana como uma probabilidade *a posteriori* da hipótese nula, ainda que este resultado não possa ser generalizado — em outras situações, as duas quantidades podem ser semelhantes (mas não iguais) ou radicalmente diferentes.

21

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

Observações (II):

- Ainda que  $P[H_0|x]$  e o nível crítico coincidam, as conclusões do teste bayesiano e clássico podem ser contrárias.
- As probabilidades de erro de tipo I (cujo valor máximo é fixado) e de erro de tipo II não têm usualmente correspondência com as probabilidades *a posteriori* das hipóteses.
- Dada a forma essencial dos testes bayesianos, o problema de testar hipóteses múltiplas (duas ou mais) não acarreta dificuldades acrescidas relativamente ao problema usual de confronto de duas hipóteses.

22

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

*Explicitação dos testes bayesianos:*

O cálculo das chances *a posteriori* (ou do fator de Bayes quando  $h(\theta)$  é uma distribuição de probabilidades própria) não constitui problema quando as hipóteses em confronto são ou ambas simples ou ambas compostas.

Em problemas onde uma hipótese é simples, *e.g.*,  $H_0 : \theta = \theta_0$ , contra  $H_1 : \theta \neq \theta_0$ , e  $\theta$  tem uma gama não numerável de valores,  $P[H_0|x] = 0$  se for adotada uma distribuição contínua para  $\theta$ .

Vias propostas:

- Método da região de credibilidade
- Método de Jeffreys

23

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

#### MÉTODO DA REGIÃO DE CREDIBILIDADE

**A.** Pressuposto: A formulação de  $H_0$  não encerra a ideia de conhecimento *a priori* sobre  $\theta_0$  particularmente diferenciado daquele sobre os restantes valores de  $\theta$ .

Construção da menor RC HPD que contém  $\theta_0$ , definida por  $R_0(x) = \{\theta \in \Theta : h(\theta|x) \geq h(\theta_0|x)\}$ , e cálculo de  $P = P(\theta \notin R_0(x)|x)$ .

Valores grandes (pequenos) de  $P$  refletem evidência a favor de (contra)  $H_0 \Rightarrow P$  é um **nível de plausibilidade relativa a posteriori** de  $H_0$ .

**Exemplo I** (revisitado-4): Problema  $H_0 : \theta = 35\%$  vs  $H_0 : \theta \neq 35\%$

Menor IC HPD contendo  $\theta = 35\%$ :  $(0, 283; 0, 350) \Rightarrow P = 0, 425$

$\therefore$  Dada a relativa proximidade entre o valor hipotetizado e a moda *a posteriori* (31, 6%), não há evidência contra  $\theta = 35\%$ . ■

24

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

#### MÉTODO DE JEFFREYS

**B.** Pressuposto: A formulação de  $H_0$  tem implícito que  $\theta_0$  tem *a priori* uma ordem de importância diferente da que é atribuída aos restantes valores de  $\theta$ .

Adoção de uma distribuição *a priori* mista para  $\theta$ , que atribua uma probabilidade *a priori*  $h_0$  a  $H_0$  e distribua a probabilidade restante  $(1 - h_0)$  continuamente em  $\{\theta \neq \theta_0\}$  com uma densidade  $h_1(\theta)$ .

Desta forma,

$$h(\theta|x) = \begin{cases} h_0 f(x|\theta_0)/p(x) & , \theta = \theta_0 \\ (1 - h_0)h_1(\theta)f(x|\theta)/p(x) & , \theta \neq \theta_0 \end{cases}$$

com  $p(x) = h_0 f(x|\theta_0) + (1 - h_0)p^*(x)$ ,  $p^*(x) = \int_{\{\theta \neq \theta_0\}} h_1(\theta)f(x|\theta)d\theta$ .

25

## Fundamentos e metodologia bayesiana

### Testes de hipóteses

Como consequência,

$$O(H_0, H_1|x) = \frac{h(\theta_0|x)}{1 - h(\theta_0|x)} = \frac{h_0}{1 - h_0} \frac{f(x|\theta_0)}{p^*(x)} \equiv O(H_0, H_1)B(x),$$

aplicando-se então o procedimento usual de confronto das hipóteses em causa. Note-se ainda que, como  $p^*(x) \leq g(x) = \sup_{\theta \neq \theta_0} f(x|\theta)$ , então

$$B(x) \geq \frac{f(x|\theta_0)}{g(x)}.$$

**Observação adicional:** A aplicação do método de Jeffreys mostra que o conflito entre as respostas bayesiana e clássica ao problema de testar hipóteses bilaterais pode ser enorme.

26

## Fundamentos e metodologia bayesiana

### Predição

Frequentemente as inferências sobre os parâmetros do modelo postulado não são um fim em si, mas antes, um meio visando prever dados amostrais futuros.

Isto é, com base em observações  $x$  de um vetor aleatório  $X \sim f(x|\theta)$  (e eventualmente em todo o conhecimento acumulado sobre  $\theta$ ), pretende-se prever  $Y$  com distribuição amostral dependente de  $\theta$ .

Este problema, que não é resolvido de uma forma uniforme e incontroversa na abordagem clássica, tem também, na ótica bayesiana, uma solução conceptualmente (pelo menos) simples: cálculo da **distribuição preditiva a posteriori**

$$p(y|x) = \int_{\Theta} f(y|x, \theta)h(\theta|x)d\theta.$$

Uma vez obtida esta, podem determinar-se medidas que a sumariam, como predições pontuais (moda, predição média, etc.) e regionais (regiões de predição com a mais alta densidade preditiva) de  $Y$ .

27

## Fundamentos e metodologia bayesiana

### Predição

**Exemplo II** (revisitado-1): Modelo Normal (d.p. conh.)  $\wedge$  “Uniforme”

$$f(x|\mu) \propto \left(2\pi \frac{\sigma^2}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right\} \wedge h(\mu) = k$$

$$\Rightarrow h(\mu|x) = \text{f.d.p. de } N(\bar{x}, \sigma^2/n)$$

Predição da média  $\bar{Y}$  de  $m$  observações i.i.d. futuras da distribuição  $N(\mu, \sigma^2)$  e independentes de  $X$  dado  $\mu$ ,  $\bar{Y}|x, \mu \sim N(\mu, \sigma^2/m)$ :

$$\therefore \bar{Y}|x \sim N\left(\bar{x}, \sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)\right)$$

$\Rightarrow$  Predição pontual de  $\bar{Y}$ :  $\bar{x}$

$\Rightarrow$  Intervalo de predição HPD a 95% para  $\bar{Y}$ :  $\left(\bar{x} \pm 1.96\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$  ■

28

## Representação da informação *a priori*

Eliciação de uma distribuição que represente as crenças *a priori* de alguém: tarefa em geral particularmente difícil e rodeada de uma série de contingências.

Situações especiais:

- Estado de conhecimento *a priori* escasso (“vago”, “difuso”) ⇒ **Distribuições não informativas**
- Adoção de uma forma funcional adequada e especificação dos hiperparâmetros (através da sua relação com quantis e/ou momentos *a priori*) de acordo com as crenças apriorísticas eliciadas ⇒ **Distribuições conjugadas naturais**

29

## Representação da informação *a priori*

### Distribuições não informativas

Objetivos do uso de distribuições *a priori* não informativas:

- Descrição de situações onde o conhecimento *a priori* é pouco ou nada significativo relativamente à informação amostral;
- Desempenho de um papel de referência, ainda que se disponha de fortes crenças *a priori*, como forma de:
  - deduzir as crenças *a posteriori* para quem parte de um conhecimento escasso, *i.e.*, quando a amostra fornece o grosso da informação sobre o parâmetro;
  - permitir a comparação com os resultados da inferência clássica que “só” usa a informação amostral (no todo ou em parte);
  - averiguar a influência nas inferências da distribuição *a priori* que descreve a informação realmente existente, quando confrontada com as que resultam do uso da distribuição *a priori* de referência.

30

## Representação da informação *a priori*

### Distribuições *a priori* não informativas

Argumentos geradores de distribuições não informativas:

- **Princípio da razão insuficiente de Bayes-Laplace**

$\Theta$  finito  $\Rightarrow \theta \sim$  Uniforme discreta

$\therefore$  ideia considerada pacífica

$\Theta$  infinito não numerável  $\Rightarrow \theta \sim$  Uniforme contínua

$\therefore$  saída criticável (natureza imprópria se  $\Theta$  é não limitado; falta de invariância face a transformações não lineares)

31

## Representação da informação *a priori*

### Distribuições *a priori* não informativas

- **Invariância face a transformações apropriadas (Regra de Jeffreys)**

- $\theta$  é um parâmetro de localização ( $\Theta \in \mathbb{R}$ )

Invariância face a translações: para cada  $a$ , os intervalos  $(\theta_0, \theta_0 + a)$ ,  $\forall \theta_0 \in \mathbb{R}$  devem ter a mesma probabilidade.

$$\Rightarrow h(\theta) = c, \quad \theta \in \Theta \quad (\text{“uniforme contínua”})$$

- $\theta$  é parâmetro de escala ( $\Theta = \mathbb{R}^+$ )

Invariância face a transformações de escala: para cada  $b > 0$ , os intervalos  $(\theta_0, b\theta_0)$ ,  $\forall \theta_0 \in \mathbb{R}^+$  devem ter a mesma probabilidade  $\Rightarrow$  como  $\ln \theta$  é parâmetro de localização da transformação logarítmica dos dados

$$\begin{aligned} h^*(\ln \theta) = c &\Rightarrow h(\theta) \propto \theta^{-1}, \quad \theta > 0 \\ &\Rightarrow h_*(\theta^a) \propto (\theta^a)^{-1}, \quad \theta > 0, \quad \forall a \in \mathbb{Z} \end{aligned}$$

32

## Representação da informação *a priori* Distribuições *a priori* não informativas

- $\theta$  é um parâmetro escalar genérico  
Invariância que garanta a identidade das inferências resultantes do uso de qualquer transformação biunívoca – satisfeita com o uso da medida de informação de Fisher,  $I(\theta)$

$$\Rightarrow h(\theta) \propto [I(\theta)]^{1/2}$$

$$\Rightarrow h^*(\psi) = h(\theta) \left| \frac{d\theta}{d\psi} \right| \propto \left[ I(\theta) \left( \frac{d\theta}{d\psi} \right)^2 \right]^{1/2} \equiv [I(\psi)]^{1/2}$$

- $\theta$  é um parâmetro vetorial genérico  
Mesma justificação  $\Rightarrow h(\theta) \propto [I(\theta)]^{1/2}$

Nota: Sendo razoável admitir a independência *a priori*, (como pode acontecer em parâmetros de tipo distinto), a distribuição *a priori* deve verificar esta condição com as distribuições marginais definidas por aplicação da regra anterior.

33

## Representação da informação *a priori* Distribuições *a priori* não informativas

**Exemplo III:**  $\theta = (\mu, \sigma^2)$  do modelo  $\{N(\mu, \sigma^2)\}$

$$I(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$$

Regra de Jeffreys multiparamétrica

$$\Rightarrow h(\mu, \sigma^2) \propto 1/\sigma^3, \quad (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$$

Regra de Jeffreys uniparamétrica + independência *a priori*

$$\Rightarrow h(\mu, \sigma^2) \propto 1/\sigma^2, \quad (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ \quad \blacksquare$$

$\therefore$  Argumento sujeito a críticas (natureza frequentemente imprópria; dependência do modelo amostral) e contracríticas.

34

## Representação da informação *a priori* Distribuições conjugadas naturais

O sucesso da forma distribucional escolhida na quantificação das crenças *a priori* e no desencadeamento das inferências está naturalmente associado à:

- versatilidade da família;
- simplicidade da derivação analítica da distribuição *a posteriori*;
- facilidade de interpretação da operação bayesiana na conjugação das informações *a priori* e amostral.

**Exemplo I** (revisitado-5): Modelo bayesiano Bernoulli  $\wedge$  Beta

$$f(x|\theta) = \theta^{\sum_1^n x_i} (1-\theta)^{n-\sum_1^n x_i}, \quad 0 < \theta < 1$$

$$\equiv \text{núcleo de uma } Be(\sum x_i + 1, n - \sum x_i + 1)$$

35

## Representação da informação *a priori* Distribuições conjugadas naturais

Se  $h(\theta) = \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} I_{(0,1)}(\theta)$  (membro da família Beta)

$$\Rightarrow \theta|x \sim Be(a + \sum x_i, b + n - \sum x_i), \quad a, b > 0$$

Ora

- Família Beta bastante versátil.
- A atualização de  $h(\theta)$  faz-se dentro da família.
- Como  $Be(a, b)$  é uma distribuição *a posteriori* baseada numa amostra fictícia com  $a$  sucessos e  $b$  insucessos e na distribuição *a priori* imprópria “ $Be(0, 0)$ ”, a informação acumulada em  $h(\theta|x)$  é traduzível na soma dos sucessos e dos insucessos da amostra real com os da amostra fictícia.

36

## Representação da informação *a priori* Distribuições conjugadas naturais

Os três requisitos mencionados para a forma distribucional *a priori* são partilhados pela denominada família das distribuições conjugadas naturais a seguir definida:

**Definição:** A família  $\mathcal{H}$  diz-se **conjugada natural** do modelo  $\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}$  se  $h(\theta|x) \in \mathcal{H}$  sempre que a correspondente  $h(\theta) \in \mathcal{H}$ .

$\therefore$  A família conjugada natural de  $\mathcal{F}$  é a família fechada em relação à multiplicação para a qual existe um membro que é proporcional a  $L(\theta|x) \propto f(x|\theta), \forall x \in \mathcal{X}$ .

$\Rightarrow$  Pelo exposto no exemplo anterior, a família Beta é conjugada natural de uma amostragem aleatória do modelo  $\{Ber(\theta)\}$ .

**Nota:** A existência de  $\mathcal{H}$  é assegurada quando existem estatísticas suficientes de dimensionalidade fixa para  $\mathcal{F}$ .

37

## Aplicações básicas da metodologia bayesiana Comparação de duas populações Normais

### Modelo bayesiano

Dados:  $x_j = (x_{ji}, i = 1, \dots, n_j), j = 1, 2$  concretizações de duas amostras aleatórias independentes de populações  $N(\mu_j, \sigma_j^2)$ .

Distr. *a priori* de Jeffreys para os 4 parâmetros:  
 $h(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto (\sigma_1^2 \sigma_2^2)^{-1}$  no respetivo espaço paramétrico conjunto.

### Comparação de médias no caso heterocedástico

Distribuição *a posteriori*:  $(\mu_1, \sigma_1^2)$  e  $(\mu_2, \sigma_2^2)$  são *a posteriori* também independentes com as distribuições marginais univariadas

$$\mu_j | x_j \sim t_{(k_j)}(\bar{x}_j, s_j^2/n_j) \Leftrightarrow \nu_j = \frac{\mu_j - \bar{x}_j}{s_j/\sqrt{n_j}} | x_j \sim t_{(k_j)}$$

$$\sigma_j^2 | x_j \sim GaI\left(\frac{k_j}{2}, \frac{k_j s_j^2}{2}\right)$$

em que  $k_j = n_j - 1$  e  $k_j s_j^2 = \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2$ .

38

## Aplicações básicas da metodologia bayesiana Comparação de duas Normais

Uma alternativa ao uso da aproximação de Patil consiste na geração de uma amostra da distribuição *a posteriori* de  $\tau$  através de simulação a partir das distribuições *a posteriori* **t-Student** de  $\nu_1$  e  $\nu_2$ , com base na qual se podem calcular empiricamente estimativas pontuais e intervalares e testar hipóteses pontuais sobre a diferença de médias.

### Comparação de variâncias

Tomando como parâmetro de interesse  $\psi = \frac{\sigma_1^2}{\sigma_2^2}$ , conclui-se das distribuições *a posteriori* Gama independentes de  $\{1/\sigma_j^2\}$  que

$$\psi | x_1, x_2 \stackrel{d}{=} \frac{s_1^2}{s_2^2} F_{(k_2, k_1)},$$

o que permite realizar facilmente inferências básicas sobre  $\psi$ .

39

## Aplicações básicas da metodologia bayesiana Comparação de duas Normais

### Comparação de médias no caso homocedástico

Contexto:  $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$  e  $h(\mu_1, \mu_2, \sigma^2) \propto \sigma^{-2}, \mu_1, \mu_2 \in \mathbb{R}, \sigma^2 > 0$ .

Tendo em conta

$$\lambda = \mu_1 - \mu_2 | \sigma^2, x_1, x_2 \sim N(\bar{x}_1 - \bar{x}_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}));$$

$$\sigma^2 | x_1, x_2 \sim GaI\left(\frac{k}{2}, \frac{k s^2}{2}\right);$$

onde  $k = n_1 + n_2 - 2$  e  $s^2 = k^{-1} \sum_j (n_j - 1) s_j^2$  é a variância empírica combinada, obtém-se

$$\lambda = \mu_1 - \mu_2 | x_1, x_2 \sim t_{(k)}\left(\bar{x}_1 - \bar{x}_2, s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \Leftrightarrow \frac{\lambda - (\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} | x_1, x_2 \sim t_{(k)},$$

que é o resultado básico para o traçado das inferências de interesse sobre a comparação das duas populações Normais.

40

## Aplicações básicas da metodologia bayesiana

### Comparação de duas populações Binomiais

#### Modelo bayesiano

Dados:  $t_j, j = 1, 2$  contagens observadas de  $T_j, j = 1, 2 | \theta_j \underset{ind}{\sim} Bi(m_j, \theta_j)$ , com  $\{m_j\}$  conhecidos

Distr. *a priori* :  $\theta_j, j = 1, 2 \underset{ind}{\sim} Be(a_j, b_j)$ .

⇒ Distr. *a posteriori* resultantes:

$$\begin{aligned} \theta_j | t_j, j = 1, 2 &\underset{ind}{\sim} Be(A_j, B_j), \quad A_j = a_j + t_j, \quad B_j = b_j + m_j - t_j \\ &\Leftrightarrow (B_j/A_j) \frac{\theta_j}{1-\theta_j} | t_j, j = 1, 2 \underset{ind}{\sim} F_{(2A_j, 2B_j)} \\ &\Leftrightarrow \left[ (1/2) \ln(B_j/A_j) + (1/2) \ln \frac{\theta_j}{1-\theta_j} \right] | t_j, j = 1, 2 \underset{ind}{\sim} Z_{(2A_j, 2B_j)}. \end{aligned}$$

41

## Aplicações básicas da metodologia bayesiana

### Comparação de duas Binomiais

**Testes de homogeneidade das Binomiais**  $H_0 : \theta_1 = \theta_2$  vs.  $H_1 : \theta_1 \neq \theta_2$

Como  $H_0 : \pi = 0 \Leftrightarrow \ln \Delta = 0$  usando  $\pi = \theta_1 - \theta_2$  e  $\Delta = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}$ , o recurso a simulação a partir das distribuições *a posteriori*  $Be(A_j, B_j)$  de amostras de  $\pi$  ou de  $\Delta$  (ou  $\ln \Delta$ ) permite obter boas aproximações empíricas do nível de plausibilidade relativa *a posteriori* de  $H_0$  ou de intervalos HPD.

Esta via é também aplicável a hipóteses unilaterais por meio do cálculo de apropriadas proporções com base nas simulações.

No caso de grandes valores observados de sucessos e insucessos, a utilização de aproximações assintóticas à distribuição  $Z$  de Fisher possibilita recorrer à distribuição *a posteriori* aproximada

$$\ln \Delta | t_1, t_2 \overset{approx}{\sim} N \left[ \ln \frac{(A_1 - 1/2)/(B_1 - 1/2)}{(A_2 - 1/2)/(B_2 - 1/2)}, \sum_{j=1,2} (A_j^{-1} + B_j^{-1}) \right]$$

para a construção de testes bayesianos unilaterais ou bilaterais das hipóteses em confronto.

42

## Avaliação de modelos

### Crítica e adequação

#### MEIOS DE DIAGNÓSTICO DO MODELO

Medidas de discrepância entre o observado ( $x$ ) e observável de acordo com o modelo ( $y$ ) assentes em

► Variáveis  $V(x, \theta)$ , como e.g.  $\ln f(x|\theta)$  ou  $\sum_i [x_i - E(X_i|\theta)]^2 / \text{Var}(X_i|\theta)$  e;

► Dados simulados  $\{(y^{(j)}, \theta^{(j)}), j = 1, \dots, m\}$  da distribuição conjunta de  $(y, \theta)$  condicional em  $x$ .

- *Diagrama de dispersão* de  $\{(V(y^{(j)}, \theta^{(j)}), V(x, \theta^{(j)})), k = 1, \dots, m\}$  ou *histograma* dos valores  $\{V(y^{(j)}, \theta^{(j)}) - V(x, \theta^{(j)}), j = 1, \dots, m\}$ ;
- *Valores-P bayesianos* (preditivos *a posteriori*)

$$P_B = P[V(Y, \theta) \geq V(x, \theta) | x] \simeq \frac{\#\{(y^{(j)}, \theta^{(j)}) : V(y^{(j)}, \theta^{(j)}) \geq V(x, \theta^{(j)})\}}{m}$$

43

## Avaliação de modelos

### Crítica e adequação

- Resíduos bayesianos em validação cruzada

Dados decompostos em amostra de treino  $x$  e amostra de validação  $y = \{y_j\}$

*Resíduos preditivos padronizados*:  $d_j = \frac{y_j - E(Y_j|x)}{\sqrt{\text{var}(Y_j|x)}}$ ,  $j = 1, \dots, l$

Critério: Quanto menor  $\sum_{j=1}^l |d_j|$ , tanto mais adequado o modelo.

Validação cruzada com um de fora: Para cada  $i = 1, \dots, n$  amostra de treino  $x_{(-i)} = (x_j, j \neq i)$ ; amostra de validação  $x_i$

*Resíduos de eliminação padronizados*:

$$d'_i = \frac{x_i - E(Y_i|x_{(-i)})}{\sqrt{\text{var}(Y_i|x_{(-i)})}}, \quad i = 1, \dots, n$$

calculados das distribuições preditivas condicionais  $p(y_i|x_{(-i)})$ ,

44

## Avaliação de modelos

### Crítica e adequação

$$\begin{aligned} p(y_i|x_{(-i)}) &= \int f(y_i|\theta, x_{(-i)}) h(\theta|x_{(-i)}) d\theta \\ &\simeq \frac{1}{m} \sum_{j=1}^m f(y_i|\theta_{(-i)}^{(j)}, x_{(-i)}), \quad \{\theta_{(-i)}^{(j)}\} \leftarrow h(\theta|x_{(-i)}) \\ &\simeq \frac{1}{\frac{1}{m} \sum_{j=1}^m \frac{1}{f(y_i|x_{(-i)}, \theta^{(j)})}}, \quad \{\theta^{(j)}\} \leftarrow h(\theta|x). \end{aligned}$$

- **Ordenadas preditivas condicionais (CPO):**  $\forall i, p(y_i|x_{(-i)})$  com  $y_i = x_i$

Critério: Quanto maior  $\sum_{i=1}^n \ln CPO_i = \ln \prod_{i=1}^n p(x_i|x_{(-i)})$  tanto mais adequado o modelo.

- **Fator pseudo-Bayes:**  $PBF(M_1/M_2) = \prod_{i=1}^n \frac{p(x_i|x_{(-i)}; M_1)}{p(x_i|x_{(-i)}; M_2)}$  para comparação dos modelos  $M_1$  e  $M_2$ .

45

## Avaliação de modelos

### Seleção e comparação

#### Medidas de desempenho preditivo

Ideia: Refletir a acurácia preditiva extra-amostra com correção do duplo uso da amostra de modo a que quanto menor for o seu valor tanto melhor será o desempenho do modelo.

- Critério de informação de Akaike (AIC)  
 $AIC = -2 \ln f(x|\hat{\theta}) + 2p$   
onde  $\hat{\theta}$  - EMV;  $p = \dim \Theta$ .
- Critério de informação de Schwarz/Bayes (SIC/BIC)  
 $BIC = -2 \ln f(x|\hat{\theta}) + p \ln n$

Variantes:

► Carlin-Louis (2000):  $BIC_{CL} = -2E[\ln f(x|\theta)|x] + p \ln n$

► Raftery *et al.* (2007):  $BIC_R = -2(\bar{l} + s_l^2) + \tilde{p} \ln n$   
onde  $\bar{l}$  and  $s_l^2$  são a média e variância empíricas dos valores simulados de  $l = \ln f(x|\theta)$ ,  $\{l^{(j)} = \ln f(x|\theta^{(j)})\}$ ;  $\tilde{p}$  é igual a  $p$  se este for conhecido ou estimado por  $2s_l^2$ , de outro modo.

46

## Avaliação de modelos

### Seleção e comparação

- Critério de informação pela desviância (DIC)

$$DIC = \overline{D(\theta)} + p_D \equiv E_{\theta|x} [D(\theta)] + \left[ \overline{D(\theta)} - D(\bar{\theta}) \right],$$

com  $D(\theta) = -2 \ln [f(x|\theta)/g(x)]$  e  $\bar{\theta} = E(\theta|x)$ .

Na generalidade dos casos, os valores esperados em  $p_D$  (número efetivo de parâmetros) são calculados por Monte Carlo a partir de uma amostra simulada de  $h(\theta|x)$ .

Uma proposta alternativa para o termo de complexidade do modelo garantindo a sua positividade é  $p_D^* = 2Var[\ln f(x|\theta)|x]$ .

- Critério de informação amplamente aplicável [de Watanabe] (WAIC)

$$WAIC = -2 \sum_{i=1}^n \ln E_{\theta|x} [f(x_i|\theta)] + 2p_W,$$

com duas propostas para a “dimensão efetiva do modelo”,  $p_W$ :

- análoga de algum modo a  $p_D$  usado no DIC

$$p_{W_1} = -2 \sum_{i=1}^n \{E_{\theta|x} [\ln f(x_i|\theta)] - \ln E_{\theta|x} [f(x_i|\theta)]\}$$

47

## Avaliação de modelos

### Seleção e comparação

$$p_{W_1} = -2 \sum_{i=1}^n \{E_{\theta|x} [\ln f(x_i|\theta)] - \ln E_{\theta|x} [f(x_i|\theta)]\} \quad (1)$$

$$\simeq -2 \sum_{i=1}^n \left\{ \frac{1}{m} \sum_{j=1}^m \ln f(x_i|\theta^{(j)}) - \ln \left[ \frac{1}{m} \sum_{j=1}^m f(x_i|\theta^{(j)}) \right] \right\}. \quad (2)$$

- parecida de algum modo com  $p_D^*$  usado alternativamente no DIC

$$p_{W_2} = \sum_{i=1}^n Var_{\theta|x} [\ln f(x_i|\theta)] \quad (3)$$

$$\simeq \sum_{i=1}^n \left\{ \frac{1}{m-1} \sum_{j=1}^m [l^{(j)}(x_i) - \bar{l}(x_i)]^2 \right\}, \quad (4)$$

em que  $l^{(j)}(x_i) = \ln f(x_i|\theta^{(j)})$  e  $\bar{l}(x_i) = \frac{1}{m} \sum_{j=1}^m l^{(j)}(x_i)$ .

48

## Avaliação de modelos

### Seleção e comparação

#### Análise por pares via fator Bayes

$$B_{kl}(x) = \frac{p(x|M_k)}{p(x|M_l)} \equiv \frac{p(M_k|x)/p(M_l|x)}{p(M_k)/p(M_l)}$$

em que para cada modelo

$$p(x|M_r) = \int f_r(x|\theta_r)h_r(\theta_r)d\theta_r$$

$$p(M_r|x) = \frac{p(M_r)p(x|M_r)}{p(x)}, \forall r$$

Opções de cálculo sem nenhuma impropriedade distribucional:

- Método Monte Carlo simples (geralmente ineficiente)

Simulação de  $h_r(\theta_r) \rightarrow (\theta_r^{(j)}, j = 1, \dots, m)$

$$p(x|M_r) \simeq \frac{1}{m} \sum_{j=1}^m f_r(x|\theta_r^{(j)})$$

49

## Avaliação de modelos

### Seleção e comparação

- Método de Newton-Raftery (geralmente instável)

Simulação de  $h_r(\theta_r|x) \rightarrow (\theta_r^{(j)}, j = 1, \dots, m)$

$$p(x|M_r) = \left[ \int \frac{h_r(\theta_r|x)}{f_r(x|\theta_r)} d\theta_r \right]^{-1} \simeq \left( \frac{1}{m} \sum_{j=1}^m [f_r(x|\theta_r^{(j)})]^{-1} \right)^{-1}$$

- Método de Gelfand-Dey

Seja  $g_r(\theta_r)$  uma boa aproximação da  $h_r(\theta_r|x)$ . Com os valores simulados da verdadeira distribuição *a posteriori*,

$$p(x|M_r) = \left[ \int \frac{g_r(\theta_r)}{f_r(x|\theta_r)h(\theta_r)} h_r(\theta_r|x) d\theta_r \right]^{-1} \\ \simeq \left[ \frac{1}{m} \sum_{j=1}^m \frac{g_r(\theta_r^{(j)})}{f_r(x|\theta_r^{(j)})h(\theta_r^{(j)})} \right]^{-1}$$

50

## Computação bayesiana

### Introdução

Se se tiver dados  $x = (x_1, \dots, x_n)$  obtido a partir de um modelo indexado pelo parâmetro  $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$ , pode-se obter a distribuição *a posteriori* conjunta, dada por

$$h(\theta|x) = \frac{f(x|\theta)h(\theta)}{\int f(x|\theta)h(\theta)d\theta}, \quad (5)$$

tendo-se usualmente dificuldades em

- Obter explicitamente as suas distribuições marginais e condicionais da distribuição *a posteriori*.
- Encontrar as quantidades sumárias sobre  $\theta$ , tais como momentos, quantis e graus de credibilidade de regiões HPD.

51

## Computação bayesiana

Inferências sumárias sobre  $\theta$  podem usualmente ser expressas em termos do valor esperado *a posteriori* de apropriadas funções de  $\theta$ , i.e.,

$$E(g(\theta)|x) = \int g(\theta) h(\theta|x) d\theta. \quad (6)$$

- Momentos *a posteriori*: Uma estimativa pontual para  $\theta_i$ , e.g., é obtida fazendo  $g(\theta) = \theta_i$ .
- Probabilidades *a posteriori*:

$$P(\theta \in B|x) = \int I_B(\theta) h(\theta|x) d\theta,$$

52

## Computação bayesiana

- Se  $y$  for uma observação futura de um modelo indexado pelo  $\theta$ , inferências preditivas sobre  $y$  são baseadas em

$$p(y|x) = \int f(y|x, \theta)h(\theta|x)d\theta, \quad (7)$$

onde  $f(y|x, \theta)$  é a distribuição de  $y$  sob o modelo paramétrico considerado, dependente ou não estocasticamente de  $x$ . Note-se que a distribuição preditiva *a priori* é  $p(x) = \int f(x|\theta)h(\theta)d\theta$ .

- Novamente, previsões sumárias podem ser obtidas na forma de esperanças preditivas,

$$E[g(y)|x] = \int g(y) p(y|x) dy. \quad (8)$$

53

## Computação bayesiana

Contudo, raramente é possível obter expressões explícitas para os integrais envolvidos.

Várias estratégias foram sugeridas para ultrapassar os problemas técnicos de cálculo dos integrais necessários. Nomeadamente,

- Aproximação à distribuição Normal multivariada,
- Método clássico de Laplace,
- Métodos de quadratura numérica,
- Métodos de Monte Carlo clássicos,
- Métodos de Monte Carlo em Cadeias de Markov (MCMC).

54

## Computação bayesiana

### Métodos de Monte Carlo

#### I. Método de Monte Carlo simples

Considere-se então o problema de aproximar um integral da forma

$$\int g(\theta)h(\theta|x)d\theta = E[g(\theta)|x], \quad (9)$$

onde  $\theta$  e  $x$  podem ser vetores, cuja existência se admite. Muitas quantidades *a posteriori* de interesse são expressáveis por (9) para algum tipo de função  $g(\theta)$  integrável. Por exemplo, covariâncias *a posteriori* de componentes de  $\theta$ , em que  $g(\theta)$  é representada por  $[\theta_i - E(\theta_i | x)][\theta_j - E(\theta_j | x)]$  para cada  $i$  e  $j$ .

55

## Computação bayesiana

Se se puder simular uma amostra aleatória  $(\theta_{(1)}, \dots, \theta_{(m)})$  da densidade *a posteriori*  $h(\theta | x)$ , o método de Monte Carlo simples aproxima o integral (9) pela média empírica

$$\hat{E}[g(\theta) | x] = \frac{1}{m} \sum_{i=1}^m g(\theta_{(i)}) \quad (10)$$

a qual, pela Lei Forte dos Grandes Números, converge quase certamente para  $E[g(\theta) | x]$ . A precisão deste estimador pode ser medida pelo erro padrão (estimado) de Monte Carlo dado por

$$\frac{1}{\sqrt{n(n-1)}} \left\{ \sum_{i=1}^m \left[ g(\theta_{(i)}) - \frac{1}{m} \sum_{i=1}^m g(\theta_{(i)}) \right]^2 \right\}^{1/2}, \quad (11)$$

quando a quantidade  $E\{[g(\theta)]^2|x\}$  é finita.

56

## Computação bayesiana

Os estimadores Monte Carlo associados com as diversas representações apresentam precisões variáveis, com implicações no esforço computacional requerido para obtenção de estimativas fiáveis.

Isto sugere a opção por vias de maior eficiência de modo a obter estimadores altamente precisos com um número relativamente baixo de valores simulados.

Para conhecimento de técnicas de redução de variância na estimação por Monte Carlo veja-se, *e.g.*, Rubinstein (1981) e Robert e Casella (2004).

Para ultrapassar dificuldades em gerar uma amostra diretamente a partir de uma distribuição (5), usam-se geralmente os métodos de Monte Carlo em cadeias de Markov **MCMC**.

57

## Computação bayesiana

### Métodos de Monte Carlo via cadeias de Markov

#### II. Métodos de Monte Carlo via cadeias de Markov

Os métodos de Monte Carlo em cadeias de Markov **MCMC** são usados de acordo com o seguinte procedimento:

- 1 Construção de uma cadeia de Markov com espaço de estados  $\Theta$ , que seja simples de simular e cuja distribuição estacionária seja a distribuição *a posteriori*  $h(\theta|x)$ ;
- 2 Simulação desta cadeia por um longo período, usando os valores simulados da cadeia para traçar inferências sobre as quantidades *a posteriori* (6) através do método de integração Monte Carlo,

$$E(g(\theta)|x) \approx \frac{1}{m} \sum_{j=1}^m g(\theta^j),$$

onde  $\theta^j$  é o  $j$ -ésimo valor para  $\theta$  numa cadeia com  $m$  iterações.

58

## Computação bayesiana

#### Noções e resultados básicos sobre cadeias de Markov:

Um **processo estocástico** é qualquer coleção de variáveis aleatórias definidas sobre o mesmo espaço de probabilidade,  $\{U(t), t \in T\}$ , onde  $T$  é um subconjunto de  $\mathbb{R}$  que, por comodidade, é entendido como uma classe de instantes de tempo.

Quando esta classe é o conjunto discreto de inteiros positivos  $T = \{0, 1, 2, \dots\}$ , o processo estocástico dito a tempo discreto é usualmente denotado por  $\{U_n, n \geq 0\}$ , sendo esta a situação típica no contexto de um esquema de simulação estocástica. O conjunto  $\mathcal{U}$  de valores das variáveis é denominado **espaço de estados**.

59

## Computação bayesiana

O processo  $\{U_n, n \geq 0\}$  satisfazendo a propriedade de Markov de independência condicional é denominado **cadeia de Markov**, podendo ser definido através de

$$P(U_{n+1} \in A | U_0 = u_0, \dots, U_n = u) = P(U_{n+1} \in A | U_n = u) \equiv P_n(u, A),$$

para todo o acontecimento  $A$  e  $n \geq 0$ , onde o símbolo  $P_n(u, A)$  denota a chamada **função de transição** (em um passo) quando parte do instante  $n$ .

Quando a função de transição é invariável com  $n$ , sendo então denotada por  $P(u, A)$ , a cadeia de Markov diz-se **homogénea**.

60

## Computação bayesiana

O estudo do comportamento assintótico ( $n \rightarrow \infty$ ) das cadeias é fundamental para os métodos MCMC e nele desempenha um papel crucial o seguinte conceito. Diz-se que uma distribuição de probabilidade  $\pi(u)$ ,  $u \in \mathcal{U}$  é **estacionária** se

$$\pi(v) = \sum_u \pi(u) p(u, v).$$

Em particular, a distribuição inicial  $P(U_0 = u) = \pi(u)$  é estacionária se a distribuição de  $U_n$  é invariante com  $n$ , *i.e.*

$$P(U_n = u) = \pi(u), \forall n \geq 0.$$

A convergência para a distribuição estacionária  $\pi$  depende de a cadeia possuir algumas propriedades de estabilidade conhecidas como **irredutibilidade** e **recorrência** e mais amplamente ser **ergódica**, envolvendo a lei forte dos grandes números (vide Paulino *et al.*, 2018).

61

## Computação bayesiana

### Amostrador de Gibbs

Entre várias formas de construção das cadeias de Markov nos métodos MCMC, destaca-se o método de amostragem Gibbs (*Gibbs sampling*) introduzido por Geman and Geman (1984) para simular distribuições multivariadas em modelos de **processamento de imagens**.

O algoritmo de Gibbs baseia-se no facto de que uma distribuição *a posteriori* conjunta  $h(\theta|x)$  com  $\theta = (\theta_1, \dots, \theta_p)^T$  pode ser em condições gerais determinada unicamente pelas  $p$  distribuições *a posteriori* condicionais completas

$$h(\theta_q | \theta_{-q}, x), \quad q = 1, \dots, p, \quad (12)$$

onde  $\theta_{-q}$  é o vetor  $\theta$  sem a componente  $\theta_q$ .

62

## Computação bayesiana

O algoritmo de **Gibbs** apresenta o seguinte esquema markoviano dinâmico para a amostragem das distribuições condicionais (12):

- 1 Escolhe-se um conjunto de valores iniciais para  $\theta$ ,

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})^T;$$

- 2 Geram-se as componentes de  $\theta$  a partir de (12), considerando o procedimento iterativo abaixo para a  $l$ -ésima iteração

$$\begin{array}{ll} \theta_1^{(l)} & \text{de } h(\theta_1 | \theta_2^{(l-1)}, \dots, \theta_p^{(l-1)}, x), \\ \theta_2^{(l)} & \text{de } h(\theta_2 | \theta_1^{(l)}, \theta_3^{(l-1)}, \dots, \theta_p^{(l-1)}, x), \\ \theta_3^{(l)} & \text{de } h(\theta_3 | \theta_1^{(l)}, \theta_2^{(l)}, \theta_4^{(l-1)}, \dots, \theta_p^{(l-1)}, x), \\ \vdots & \vdots \\ \theta_{p-1}^{(l)} & \text{de } h(\theta_{p-1} | \theta_1^{(l)}, \dots, \theta_{p-2}^{(l)}, \theta_p^{(l-1)}, x), \\ \theta_p^{(l)} & \text{de } h(\theta_p | \theta_1^{(l)}, \dots, \theta_{p-1}^{(l)}, x); \end{array}$$

63

## Computação bayesiana

- 3 A etapa anterior é repetida  $s$  vezes até à geração de  $m$  **amostras** independentes de  $\theta$ . Note-se que cada elemento da sucessão  $\theta^{(1)}, \dots, \theta^{(s)}, \dots$  é uma realização de uma cadeia de Markov com espaço de estados  $\Theta$  e probabilidades de transição dadas por

$$p(\theta^{(s)}, \theta^{(s+1)}) = \prod_{q=1}^p h(\theta_q^{(s+1)} | \theta_{l>q}^{(s)}, \theta_{l<q}^{(s+1)}, x).$$

Quando  $s \rightarrow \infty$  no procedimento acima,  $\theta^{(s)} = (\theta_1^{(s)}, \dots, \theta_p^{(s)})^T$  tende em distribuição para um vetor aleatório com f.d.p.  $h(\theta|x)$  (Tanner, 1996). Veja-se também Casella e George (1992) e Gelfand e Smith (1990).

64

## Computação bayesiana

Em particular, a  $q$ -ésima distribuição *a posteriori* marginal pode ser obtida usando a sua distribuição empírica com as  $m$  amostras *i.e.*

$$h(\theta_q|x) \approx \frac{1}{m} \sum_{l=1}^m h(\theta_q | \theta_{-q}^{(l)}, x), \quad (13)$$

onde  $h(\theta_q | \theta_{-q}^{(l)}, x)$  é a distribuição (12) com os  $\theta_{q'}$ ,  $q' \neq q = 1, \dots, p$ , substituídos pelos seus respectivos valores na iteração  $l$ ,  $l = 1, \dots, m$ .

Observe-se que as  $s - m$  iterações,  $m < s$ , do procedimento em causa são ignoradas na estimação das quantidades de interesse, visto que elas fazem parte do período de aquecimento (**burn-in**) da cadeia, onde se acredita haver uma maior correlação entre os vetores  $\theta^{(s)}$ ,  $s = 1, 2, \dots$

65

## Computação bayesiana

### Distribuição preditiva

Os métodos MCMC são também usados na predição de uma observação futura  $y$  de um modelo indexado pelo parâmetro  $\theta$  em causa. Por exemplo, a distribuição preditiva

$$p(y|x) = \int f(y|\theta, x)h(\theta|x)d\theta, \quad (14)$$

onde  $f(y|\theta, x)$  é a distribuição de  $y$  sob esse modelo paramétrico, pode ser estimada por

$$\tilde{p}(y|x) = \frac{1}{m} \sum_{l=1}^m f(y|\theta^{(l)}, x),$$

onde  $\theta^{(l)}$ ,  $l = 1, \dots, m$ , são os valores obtidos para  $\theta$  nas  $m$  amostras referidas acima.

66

## Computação bayesiana

Gelfand (1996) sugere, para estimar  $p(x_i|x_{(-i)})$ , a utilização da média harmónica de  $\{f(x_i|x_{(-i)}, \theta_{(j)}), j = 1, \dots, m\}$ . Considerando que

$$p(x)h(\theta|x) = h(\theta)f(x|\theta) = h(\theta)f(x_i|x_{(-i)}, \theta)f(x_{(-i)}|\theta),$$

tem-se

$$\begin{aligned} p(x_i|x_{(-i)}) &= \frac{p(x)}{p(x_{(-i)})} = \frac{1}{\int \frac{f(x_{(-i)}|\theta) h(\theta) h(\theta|x)}{p(x) h(\theta|x)} d\theta} \\ &= \frac{1}{\int \frac{1}{f(x_i|x_{(-i)}, \theta)} h(\theta|x) d\theta}, \end{aligned}$$

e portanto, se  $\{\theta_{(j)}; j = 1, \dots, m\}$  é uma amostra de  $h(\theta|x)$  tem-se

$$\hat{p}(x_i|x_{(-i)}) = \frac{1}{\frac{1}{m} \sum_{j=1}^m \frac{1}{f(x_i|x_{(-i)}, \theta_{(j)})}}. \quad (15)$$

67

## Computação bayesiana

**Exemplo IV:** Seja  $(Y, X)$  um par aleatório em que  $Y$  condicional a  $X = x$  segue uma distribuição de Poisson com valor médio  $\lambda(x) = \delta^x$  e  $X$  apresenta uma distribuição Normal com valor médio  $\mu$  e precisão  $\tau = 1/\sigma^2$ . A verosimilhança relativa a dados constituídos por  $n$  observações i.i.d. deste par aleatório,  $\mathcal{D} = \{(y_1, x_1), \dots, (y_n, x_n)\}$ , é

$$f(x, y|\theta) = \prod_{i=1}^n \frac{[\delta^{x_i}]^{y_i}}{y_i!} e^{-\delta^{x_i}} \left[ \frac{\tau}{2\pi} \right]^{1/2} \exp \left\{ -\frac{\tau}{2} (x_i - \mu)^2 \right\}$$

onde  $\theta = (\delta, \mu, \tau)$  para  $\delta, \tau > 0$  e  $-\infty < \mu < +\infty$ . Se se considerar uma distribuição *a priori* não informativa  $h(\delta, \mu, \tau) \propto (\delta\tau)^{-1}$ , a função densidade de probabilidade *a posteriori* para  $\theta$  é:

$$\begin{aligned} h(\theta|\mathcal{D}) &\propto \tau^{n/2-1} \delta^{\sum_i x_i y_i - 1} \exp \left\{ -\sum_i \delta^{x_i} \right\} \times \\ &\times \exp \left\{ -\frac{\tau}{2} \left[ \sum_i (x_i - \bar{x})^2 + n(\mu - \bar{x})^2 \right] \right\}, \end{aligned}$$

68

## Computação bayesiana

sendo as correspondentes distribuições condicionais completas identificadas facilmente por

$$\begin{aligned} h(\delta|\mathcal{D}, \mu, \tau) &\propto \delta^{\sum_i x_i y_i - 1} \exp\{-\sum_i \delta^{x_i}\}, \quad \delta > 0; \\ h(\mu|\mathcal{D}, \delta, \tau) &= N(\bar{x}, (\tau n)^{-1}), \quad -\infty < \mu < +\infty; \\ h(\tau|\mathcal{D}, \delta, \mu) &= Ga\left(\frac{n}{2}, \frac{A}{2}\right), \quad \tau > 0, \end{aligned} \quad (16)$$

onde  $A = \sum_i (x_i - \bar{x})^2 + n(\mu - \bar{x})^2$ .

As distribuições respeitantes a  $\mu$  e  $\tau$  são familiares, processando-se a amostragem nos correspondentes passos Gibbs através de conhecidos e eficientes algoritmos de simulação. A situação relativa a  $\delta$  já não é trivial requerendo o uso de outros métodos como o de rejeição (e.g. método de rejeição adaptativa proposto por Wild & Gilks, 1993). ■

69

## Computação bayesiana

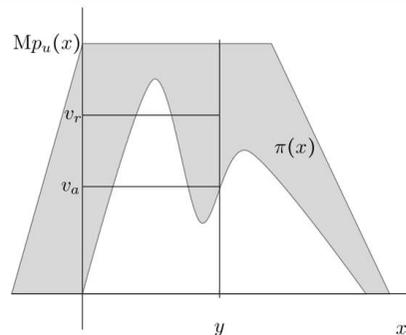
### Métodos de rejeição

Seja  $\pi(x) = c\pi^*(x)$  uma função densidade de probabilidade (f.d.p.), onde  $c$  é a constante de normalização. Suponha-se que é difícil amostrar diretamente de  $\pi$ , mas que no entanto há um modo de simular de uma f.d.p.  $p_u(x)$  com base na qual se cria uma função que delimita superiormente  $\pi$  (conhecida como *envelope*), ou seja, tal que para qualquer  $x$  no suporte de  $\pi$  se tem  $\pi(x) \leq Mp_u(x)$ , onde  $M > 1$  é uma constante especificada.

O **método básico de rejeição** que devolve um valor  $x$  de uma distribuição  $X \sim \pi(x)$  explicita-se no seguinte algoritmo que se ilustra esquematicamente na Figura 1.

70

## Computação bayesiana



**Figura:** Representação esquemática do algoritmo básico de rejeição em que  $v = uMp_u(y)$ ,  $r \equiv$  rejeitado e  $a \equiv$  aceite.

71

## Computação bayesiana

### Algoritmo de rejeição básico

- 1 Gera-se  $y$  da f.d.p.  $p_u$ .
- 2 Gera-se  $u$  de uma distribuição Uniforme em  $(0, 1)$ .
- 3 Se  $u \leq \frac{\pi(y)}{Mp_u(y)}$  tome-se  $x = y$ ; se não, volta-se a 1.

Com efeito, tomando  $X$  como a variável  $Y$  quando esta é aceite, tem-se  $\forall v$

$$P\left[Y \leq v, U \leq \frac{\pi(Y)}{Mp_u(Y)}\right] = \int_{-\infty}^v p_u(y) \int_0^{\frac{\pi(y)}{Mp_u(y)}} du dy = M^{-1} \int_{-\infty}^v \pi(y) dy,$$

donde

$$P(X \leq v) = P(Y \leq v | U \leq \frac{\pi(Y)}{Mp_u(Y)}) = \int_{-\infty}^v \pi(y) dy.$$

72

## Computação bayesiana

**Exemplo V:** Um método simples de rejeição para amostrar da densidade *a posteriori*  $h_x(\theta)$  (basta conhecer o seu núcleo  $h_x^*(\theta)$  pelo resultado acima) quando a densidade *a priori* é própria é tomar esta como função envelope e  $M$  igual à verossimilhança máxima, já que

$$h_x^*(\theta) = h(\theta)f(x|\theta) \leq Mh(\theta).$$

Assim, gerando  $\theta_0 \sim h(\cdot)$  e  $u_0 \sim Unif(0, 1)$ , tem-se  $\theta_0 \sim h_x^*(\cdot)$  se  $u_0 < \frac{f(x|\theta_0)}{M}$ . ■

A aplicação deste método exige que se ache uma densidade instrumental  $p_u$  que se adapte bem a  $\pi$ , devendo ter caudas mais pesadas do que esta, e um gerador simples para ela. Além disso, a constante  $M$  deve ser escolhida de modo a ser o mais pequena possível para que o algoritmo seja eficiente.

73

## Computação bayesiana

### Algoritmo de rejeição adaptativo

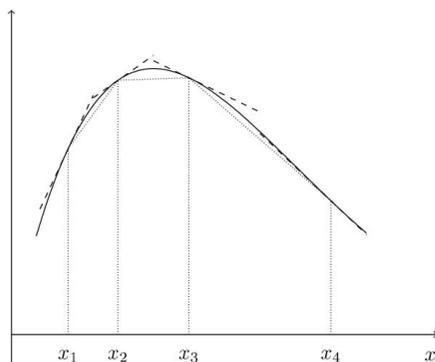
Gilks e Wild (1992) sugerem um método automático de geração de funções delimitantes para amostragem de densidades-alvo  $\pi(x)$  (ou do seu fator relevante  $\pi^*(x)$ ) logaritmicamente côncavas, isto é, cujo logaritmo é uma função côncava.

Sabe-se que qualquer função côncava pode ser limitada superiormente e inferiormente por invólucros formados por troços lineares.

Para os construir consideram-se pontos sobre o gráfico da função e fazem-se passar por e entre esses pontos, respetivamente, tangentes e cordas ao gráfico - veja-se a Figura 2.

74

## Computação bayesiana



**Figura:** Invólucros lineares, superior e inferior, por troços para delimitar a função  $L(x) = \ln \pi(x)$  no método de rejeição adaptativa.

75

## Computação bayesiana

Seja então  $\pi(x) \propto \exp(L(x))$  uma função densidade de probabilidade univariada log-côncava com suporte  $D \subset \mathbb{R}$  e  $T_k = \{x_i, i = 1, \dots, k\}$  um conjunto de  $k$  pontos ordenados,  $x_1 \leq x_2 \leq \dots \leq x_k$ , para os quais se calcula  $L(x)$  e  $L'(x) = dL(x)/dx$ , se  $\pi$  for contínua e diferenciável em  $D$ .

Defina-se a função envelope em  $T_k$  para  $\pi(x)$  como  $\exp[u_k(x)]$  onde  $u_k(x)$  é o **invólucro superior**, linear por troços, de  $L(x)$

$$u_k(x) = L(x_j) + (x - x_j)L'(x_j),$$

para  $x \in [z_{j-1}, z_j]$  e  $j = 1, \dots, k - 1$  com

$$z_j = \frac{L(x_{j+1}) - L(x_j) - x_{j+1}L'(x_{j+1}) + x_jL'(x_j)}{L'(x_j) - L'(x_{j+1})}$$

76

## Computação bayesiana

o ponto de interseção das tangentes à curva  $l(x)$  em  $x_j$  e  $x_{j+1}$ . Os pontos  $z_0$  e  $z_k$  são tomados, respetivamente, como o limite inferior de  $D$  (ou  $-\infty$  se  $D$  não for limitado inferiormente) e superior de  $D$  (ou  $+\infty$  se  $D$  não for limitado superiormente).

Defina-se ainda a **função de enquadramento inferior** em  $T_k$  de  $\pi(x)$  como  $\exp[l_k(x)]$  onde  $l_k(x)$  é o invólucro inferior, linear por troços, de  $L(x)$

$$l_k(x) = \frac{(x_{j+1} - x)L(x_j) + (x - x_j)L(x_{j+1})}{x_{j+1} - x_j}$$

para  $x \in [x_j, x_{j+1}]$  e  $j = 1, \dots, k - 1$ . Para  $x < x_1$  ou  $x > x_k$ ,  $l_k(x) = -\infty$ .

Como se admite que  $L(x)$  é côncava tem-se que  $l_k(x) \leq L(x) \leq u_k(x)$  para todo o  $x$  em  $D$ .

77

## Computação bayesiana

**Algoritmo de rejeição adaptativa** (uma amostra de  $n$  pontos de  $\pi(x)$ )

- 1 Obtém-se  $x$  do envelope normalizado  $S_k(x) = \exp[u_k(x)] / \int_D \exp[u_k(y)] dy$ .
- 2 Obtém-se  $u$  de uma distribuição Uniforme em  $(0, 1)$  e
  - se  $u \leq \exp\{l_k(x) - u_k(x)\}$ , aceita-se  $x$  sem fazer qualquer cálculo da função  $L(x)$  nesse ponto; caso contrário, calcula-se  $L(x)$  e faz-se o teste de rejeição seguinte;
  - se  $u \leq \exp\{L(x) - u_k(x)\}$ , aceita-se  $x$ ; caso contrário rejeita-se  $x$ ;
  - retome-se os passos anteriores até se aceitar o candidato gerado.
- 3 Uma vez findo o ciclo anterior com aceitação do valor candidato, atualizam-se os invólucros superiores e inferiores juntando  $x$  a  $T_k$ , e aumentando  $k$  de uma unidade.
- 4 Volta-se a 1.
- 5 Termina-se quando se atingir o número de pontos que se decidiu inicialmente amostrar.

78

## Computação bayesiana

### Métodos de diagnóstico da convergência

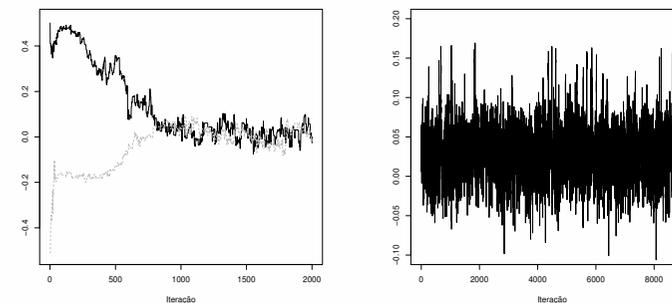
São vários os instrumentos e métodos disponíveis para monitorização e diagnóstico dos dois tipos de convergência, alguns dos quais estão automaticamente incluídos em software mais específico ou mais geral de análise bayesiana.

O instrumento mais conhecido para monitorização da convergência para a distribuição estacionária é a representação gráfica para cada quantidade escalar dos valores simulados da cadeia ao longo das sucessivas iterações, ligados por uma linha contínua.

A Figura 3 retrata aspetos típicos do gráfico de traços em zonas iniciais (esquerda) e finais (direita) da sequência de valores simulados.

79

## Computação bayesiana



**Figura:** Gráficos dos traços relativos a um mesmo parâmetro de 2 cadeias ao longo das primeiras 1000 iterações (esquerda) e 1 cadeia nas últimas 9000 iterações (direita).

80

## Computação bayesiana

### Alguns métodos de diagnóstico da convergência

- Método de Gelman e Rubin.
- Método de Geweke.
- Método de Raftery e Lewis.
- Método de Heidelberg e Welch

81

## Computação bayesiana

### Software

O software **R** possui uma variedade de pacotes que podem ser utilizados para fazer inferência bayesiana. Aconselha-se a consulta da página <http://cran.r-project.org/web/views/Bayesian.html>, onde se pode encontrar, e.g., o pacote **DPpackage** que contém funções para fazer inferência bayesiana não paramétrica, o pacote **bayesSurv** específico para fazer inferência bayesiana em modelos de sobrevivência, etc.

Do software que implementa métodos baseados em simulação estocástica, podem ser utilizados através de ligação ao **R** nomeadamente o **OpenBUGS** (Thomas et al., 2006), **JAGS** (Plummer, 2003), **INLA** (Rue et al., 2009), **BayesX** (Belitz et al., 2013) e **Stan** (Carpenter et al., 2017).

A monitorização da convergência das cadeias pode ser feita com recurso ao software **CODA** e **BOA**, ambos pacotes do **R**.

82

## Aplicações

### Modelo hierárquico Poisson-Gama

Seja  $Y_i$  o total observável da ocorrência de um evento de interesse (sucesso) em  $n_i$  indivíduos (independentes) em risco da população na região  $i$ ,  $i=1, \dots, r$ . Se  $\theta_i$  é probabilidade de sucesso para qualquer indivíduo da região  $i$ , pode-se adotar o modelo probabilístico Binomial, *i.e.*,

$$Y_i \sim \text{Binomial}(n_i, \theta_i).$$

Usando o método de estimação da máxima verossimilhança (MV), a chance de sucesso (*odds*), definida por  $\frac{\theta}{1-\theta}$ , tem estimador de MV dado por  $\frac{Y}{n-Y}$ .

Satisfeitas algumas condições, pode-se aproximar a distribuição Binomial pela distribuição Poisson (e.g. doenças raras). Portanto,

$$Y_i \sim \text{Poisson}(E_i \eta_i),$$

onde  $E_i$  é o total esperado da população em risco e  $\eta_i$  é conhecido como o risco relativo da região  $i$ ,  $i=1, \dots, r$  (reparametrização:  $n_i, \theta_i = E_i \eta_i$ ).

83

## Aplicações

### Modelo hierárquico Poisson-Gama

A razão entre os valores observáveis e os valores esperados da doença (ou morte) na região  $i$  é conhecida por taxa de mortalidade padronizada (*Standardized Mortality Ratio/SMR*), *i.e.*,

$$SMR_i = \frac{Y_i}{E_i}.$$

Um modelo hierárquico simples para estes tipos de dados (regiões independentes) com distribuição Poisson é obtido com o uso de uma distribuição *a priori* Gama para os riscos relativos  $\eta_i$ , *i.e.*,

$$Y_i | \eta_i \sim \text{Poisson}(E_i \eta_i) \quad \text{e} \quad \eta_i \sim \text{Gama}(c, d),$$

onde  $c$  (parâmetro de forma) e  $d$  (parâmetro de escala) são quantidades conhecidas.

84

## Aplicações

### Modelo hierárquico Poisson-Gama

Assim, a distribuição *a posteriori* do modelo Poisson/Gama é proporcional a

$$L(\boldsymbol{\eta}|\mathbf{y})p(\boldsymbol{\eta}) \propto \prod_{i=1}^r e^{-E_i \eta_i} (E_i \eta_i)^{y_i} \eta_i^{c-1} e^{-d \eta_i},$$

onde  $L(\cdot|\mathbf{y})$  e  $p(\cdot)$  denotam a função de verossimilhança e a distribuição *a priori*, respectivamente. Ou seja,  $\eta_i|\mathbf{y}$  tem distribuição Gama ( $y_i + c, E_i + d$ ) e portanto a média *a posteriori* de  $\eta_i$  é

$$\nu_i \equiv E(\eta_i|\mathbf{y}) = \frac{y_i + c}{E_i + d} = w_i SMR_i + (1 - w_i)\mu,$$

onde  $SMR_i$  é a taxa padronizada,  $w_i = \frac{E_i}{E_i + d}$  tal que  $0 \leq w_i \leq 1$  e  $\mu = E(\eta_i) = \frac{c}{d}$ ,  $i=1, \dots, r$ . Esta estimativa é uma média ponderada da proporção  $SMR_i$  da região  $i$  e a média *a priori*  $\mu$ . Quando  $w_i \rightarrow 1$ ,  $\nu_i \rightarrow SMR_i$ , revelando que os dados são altamente informativos quando  $E_i$  é grande, enquanto  $w_i \rightarrow 0$ ,  $\nu_i \rightarrow \mu$ , mostrando como a distribuição *a priori* altamente informativa no caso de  $E_i$  pequeno.

85

## Aplicações

### Regressão Binomial

Um dos objectivos de muitos estudos estatísticos é avaliar a relação entre as variáveis observadas num conjunto de dados, *i.e.*, estudar a influência que uma ou mais variáveis explicativas (**covariáveis**) têm sobre uma variável de interesse (**resposta**).

O **modelo de regressão** relaciona a variável resposta com as covariáveis.

O modelo de regressão **linear normal**, introduzido por Legendre e Gauss em 1805-1809, dominou a modelação estatística até meados do século XX, mesmo havendo modelos mais adequados para certas situações: **log-log complementar** para ensaios de diluição (Fisher, 1922), **probit** (Bliss, 1935) para proporções, **log-linear** para dados de contagens (Birch, 1963), etc.

86

## Aplicações

### Modelos Lineares Generalizados

Os Modelos Lineares Generalizados (**MLG**), introduzidos por Nelder and Wedderburn (1972), apresentam uma estrutura de regressão **linear** e têm em comum o facto da variável resposta pertencer à família de distribuições **exponencial**. Casos particulares de MLG:

- Modelo de regressão linear normal;
- Modelo de análise de variância;
- Modelo de regressão logística;
- Modelos log-lineares para tabelas de contingência.

**Notação:** Os dados  $\mathcal{D} = \{(y_i, \mathbf{z}_i), i=1, \dots, n\}$  são realizações da variável resposta (v.a.)  $Y$  e um vetor de covariáveis  $\mathbf{z} = (z_1, \dots, z_p)'$  em  $n$  indivíduos, sendo as componentes  $Y_i$  do vetor aleatório  $\mathbf{Y} = (Y_1, \dots, Y_n)'$  independentes (Amaral-Turkman e Silva, 2000).

87

## Aplicações

### Modelos Lineares Generalizados

Os **MLG** são caracterizados por:

- 1 **Componente aleatória:** Dadas as covariáveis  $\mathbf{z}_i$ , as variáveis  $Y_i$  são (condicionalmente) independentes com distribuição pertencente à família exponencial, com  $E(Y_i|\mathbf{z}_i) = \mu_i$  e, possivelmente, um parâmetro de dispersão  $\phi$  não dependente de  $i$ ,  $i=1, \dots, n$ .
- 2 **Componente estrutural ou sistemática:** O valor esperado  $\mu_i$  está relacionado com o preditor linear  $\eta_i = \mathbf{z}_i' \boldsymbol{\beta}$  através da relação

$$\mu_i = h(\eta_i) = h(\mathbf{z}_i' \boldsymbol{\beta}), \quad \eta_i = g(\mu_i),$$

onde  $h$  é uma função monótona e diferenciável,  $g = h^{-1}$  é a **função de ligação**,  $\boldsymbol{\beta}$  é um vetor de parâmetros e  $\mathbf{z}_i = (1, z_{i1}, \dots, z_{i,p-1})'$ , por vezes.

88

## Aplicações

### Modelos Logístico, Probit and Log-log complementar

Consideram-se  $n$  v.a. independentes  $Y_i \sim Bin(1, \pi_i)$ ,  $i=1, \dots, n$ , com respectivo vetor de covariáveis  $\mathbf{z}_i$ . Note-se que  $E(Y_i) = \pi_i$  e  $\theta_i = \ln(\frac{\pi_i}{1-\pi_i})$  (função de ligação canónica *logit*), pelo que o MLG resultante é o modelo de regressão **logística**

$$\pi_i = \exp(\mathbf{z}'_i \boldsymbol{\beta}) / (1 + \exp(\mathbf{z}'_i \boldsymbol{\beta})). \quad (17)$$

Analogamente, se a relação entre  $\pi_i$  e  $\mathbf{z}_i$  é da forma

$$\pi_i = \Phi(\eta_i) = \Phi(\mathbf{z}'_i \boldsymbol{\beta}), \quad (18)$$

onde  $\Phi(\cdot)$  é a função de distribuição de uma v.a.  $N(0, 1)$ , obtém-se o modelo **probit**. Se a função inversa da função de ligação é a função de distribuição de Gumbel, tem-se o modelo **log-log complementar**,

$$\pi_i = 1 - \exp(-\exp(\mathbf{z}'_i \boldsymbol{\beta})). \quad (19)$$

89

## Aplicações

### Dados 1: Mortalidade de besouros

Bliss (1935) estudou o comportamento de besouros adultos à exposição ao gás dissulfureto de carbono ( $CS_2$ ) durante cinco horas observando 481 besouros divididos em 8 grupos.

Tabela 4: Mortalidade de besouros (Bliss, 1935).

$i$	$x_i$	$n_i$	$y_i$	$i$	$x_i$	$n_i$	$y_i$
1	1.6907	59	6	5	1.8113	63	52
2	1.7242	60	13	6	1.8369	59	52
3	1.7552	62	18	7	1.8610	62	61
4	1.7842	56	28	8	1.8839	60	60

**Variáveis:**  $n$  (número de besouros expostos),  $y$  (número de besouros mortos) e  $x$  (dosagem de  $\log_{10} CS_2$  (mg/litro)).

**Objectivo:** Estimar a curva de dose-resposta quanto à mortalidade de besouros a partir de diferentes dosagens.

90

## Aplicações

### Código do JAGS

#### Modelo:

```
model {
  for(i in 1 : N) {
    y[i] ~ dbin(p[i], n[i])
    logit(p[i]) <- alpha.star + beta*(x[i]-mean(x[]))
    # alternative link functions:
    # p[i] <- phi(alpha.star + beta*(x[i]-mean(x[])))
    # p[i] <- 1 - exp(-exp(alpha.star + beta*(x[i]-mean(x[]))))
    y.hat[i] <- n[i] * p[i]
  }
  alpha <- alpha.star - beta * mean(x[])
  beta ~ dnorm(0,0.001)
  alpha.star ~ dnorm(0,0.001)
}
```

#### Dados:

```
list(n = c(59, 60, 62, 56, 63, 59, 62, 60), y = c(6, 13, 18, 28, 52, 53, 61, 60),
     x = c(1.6907, 1.7242, 1.7552, 1.7842, 1.8113, 1.8369, 1.8610, 1.8839), N = 8)
```

#### Valores iniciais:

```
list(alpha.star=0, beta=0)
```

91

## Aplicações

### Resultados – Besouros

Equações de regressão estimadas (*DIC*) para os modelos:

- Logístico:  $\ln[\hat{\pi}(x)/(1-\hat{\pi}(x))] = -60.87 + 34.36x$ .
- Probit:  $\Phi^{-1}(\hat{\pi}(x)) = -35.04 + 19.79x$ .
- Log-log complementar:  $\ln(-\ln(1-\hat{\pi}(x))) = -39.73 + 22.13x$ .

Tabela 5: Proporções estimadas de besouros mortos (Bliss, 1935).

dosagem $x$	observada	logístico	probit	clog-log
		DIC=41.39	DIC=40.31	DIC=33.60
1.6907	0.1017	0.0605	0.0585	0.0956
1.7242	0.2167	0.1658	0.1798	0.1884
1.7552	0.2903	0.3629	0.3794	0.3376
1.7842	0.5000	0.6054	0.6045	0.5416
1.8113	0.8254	0.7945	0.7878	0.7578
1.8369	0.8983	0.9021	0.9032	0.9165
1.8610	0.9839	0.9540	0.9614	0.9842
1.8839	1.0000	0.9781	0.9863	0.9986

92

## Aplicações

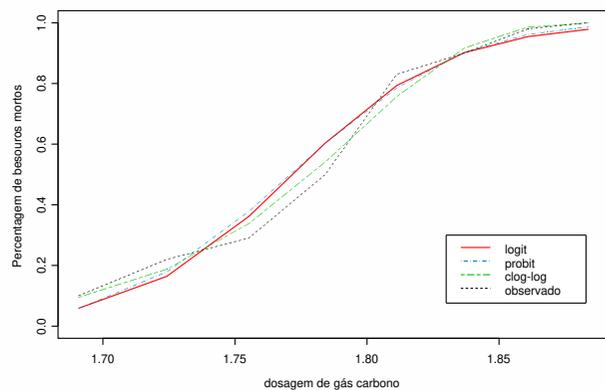


Figura: Gráfico das proporções ajustados com base nos 3 modelos

93

## Aplicações

### Dose Letal

Os modelos de dose-resposta visam não só a predição da probabilidade de sucesso para uma dosagem específica

$$\pi(x) = g^{-1}(\eta) = F(\eta), \quad \eta = \beta_0 + \beta_1 x,$$

mas também a determinação da dosagem necessária para se atingir uma probabilidade de sucesso  $P$ . Essa dosagem é chamada de **dose letal**. Seja  $DL_{100P}$  a notação para uma dose letal de  $100P\%$  de sucesso, tal que  $P = F(\beta_0 + \beta_1 DL_{100P})$ ,  $0 < P < 1$ . Note-se que a dosagem de gás carbono que mata  $100P\%$  de besouros no modelo logístico é estimada por:  $\widehat{DL}_{100P} = (34.36)^{-1}[\ln(P/(1-P)) + 60.87]$ .

Tabela 6: Estimativas de três doses letais (Bliss, 1935).

dose letal	logístico	probit	clog-log
$DL_{50}$	1.771	1.771	1.779
$DL_{99}$	1.907	1.889	1.865
$DL_1$	1.637	1.652	1.587

94

## Aplicações

### Análise de Sobrevivência

- O estudo da análise de sobrevivência centra-se num conjunto de unidades que são observadas até à ocorrência de algum evento de interesse, *e.g.*, a falha das unidades (morte). Frequentemente esse evento não chega a ocorrer para algumas dessas unidades durante o período de observação (censura).
- Os dados de sobrevivência podem ser **univariados**, se o evento de interesse ocorre quando muito uma vez para cada indivíduo (unidade), ou **multivariados** quando o indivíduo pode experimentá-lo mais do que uma vez (eventos múltiplos) ou os indivíduos estão agrupados em blocos (famílias).
- Para maiores detalhes, veja-se. *e.g.*, Rocha e Silva (2008).

95

## Aplicações

### Análise de Sobrevivência

- Se  $T$  representa o tempo de sobrevivência de um indivíduo, a **função de sobrevivência** no instante  $t$  é dada por

$$S(t) \equiv P(T \geq t).$$

- A **função risco** é a taxa de ocorrência do evento no instante  $t$ ,

$$\lambda(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt | T \geq t)}{dt}.$$

- Sem perda de generalidade, pode-se mostrar que a função de verosimilhança de  $\theta$  dado o conjunto  $\mathcal{D}$  é expressa por

$$L(\theta | \mathcal{D}) = \prod_{i=1}^n f(t_i | \theta)^{\gamma_i} S(t_i | \theta)^{1-\gamma_i}. \quad (20)$$

onde  $t_i$  é o valor observado do tempo de sobrevivência  $T_i$  com f.d.p.  $f(\cdot)$  e  $\gamma_i$  é a função indicadora de não censura da unidade  $i$ .

96

## Aplicações

### Regressão Weibull

Os modelos de sobrevivência **paramétricos** são construídos a partir de famílias de distribuições de probabilidade específicas. Por exemplo, a distribuição de Weibull com f.d.p, função de sobrevivência e função risco dadas, respectivamente, por

$$f(t) = \delta \nu t^{\nu-1} \exp[-\delta t^\nu], \quad S(t) = \exp[-\delta t^\nu], \quad \lambda(t) = \delta \nu t^{\nu-1}. \quad (21)$$

Perante populações heterogêneas, o **modelo Weibull** (21) com covariáveis  $\mathbf{z}$  é definido pela seguinte função risco condicional

$$\lambda(t|\mathbf{z}) = \nu t^{\nu-1} \psi(\mathbf{z}), \quad t \geq 0, \quad (22)$$

onde  $\psi(\mathbf{z}) = \exp[\mathbf{z}'\boldsymbol{\beta}]$  e  $\boldsymbol{\beta}$  é o vetor de coeficientes desconhecidos associado ao vetor de covariáveis  $\mathbf{z}$ . Note-se que  $\lambda(t|\mathbf{z}_1)/\lambda(t|\mathbf{z}_2)$  não depende de  $t$  (**funções risco proporcionais**).

97

## Aplicações

### Dados 2 – Cancro da laringe

- Num hospital holandês 90 pacientes do sexo masculino com cancro de laringe foram diagnosticados e tratados durante o período de 1970 a 1978 (Kardaun, 1983).
- Os tempos de sobrevivência observados neste estudo foram os tempos decorridos entre o primeiro tratamento de cada paciente e a sua morte ou o fim do estudo (1/3/81).
- Para cada paciente, observaram-se também a idade no momento do diagnóstico, o ano do diagnóstico e o estágio da doença.
- Esses estádios estão ordenados do menos grave (estádio 1) ao mais grave (estádio 4).

98

## Aplicações

**Tabela 7:** Dados de sobrevivência de pacientes com cancro de laringe (Kardaun, 1983).

Estádio	Tempos de vida em anos (* censurado), Idade, Ano do diagnóstico					
1	0.6,77,76	1.3,53,71	2.4,45,71	2.5*,57,78	3.2,58,74	3.2*,51,77
	3.3,76,74	3.3*,63,77	3.5,43,71	3.5,60,73	4.0,52,71	4.0,63,76
	4.3,86,74	4.5*,48,76	4.5*,68,76	5.3,81,72	5.5*,70,75	5.9*,58,75
	5.9*,47,75	6.0,75,73	6.1*,77,75	6.2*,64,75	6.4,77,72	6.5,67,70
	6.5*,79,74	6.7*,61,74	7.0*,66,74	7.4,68,71	7.4*,73,73	8.1*,56,73
	8.1*,73,73	9.6*,58,71	10.7*,68,70			
2	0.2,86,74	1.8,64,77	2.0,63,75	2.2*,71,78	2.6*,67,78	3.3*,51,77
	3.6,70,77	3.6*,72,77	4.0,81,71	4.3*,47,76	4.3*,64,76	5.0*,66,76
	6.2,74,72	7.0,62,73	7.5*,50,73	7.6*,53,73	9.3*,61,71	
3	0.3,49,72	0.3,71,76	0.5,57,74	0.7,79,77	0.8,82,74	1.0,49,76
	1.3,60,76	1.6,64,72	1.8,74,71	1.9,72,74	1.9,53,74	3.2,54,75
	3.5,81,74	3.7*,52,77	4.5*,66,76	4.8*,54,76	4.8*,63,76	5.0,59,73
	5.0,49,76	5.1*,69,76	6.3,70,72	6.4,65,72	6.5*,65,74	7.8,68,72
	8.0*,78,73	9.3*,69,71	10.1*,51,71			
4	0.1,65,72	0.3,71,76	0.4,76,77	0.8,65,76	0.8,78,77	1.0,41,77
	1.5,68,73	2.0,69,76	2.3,62,71	2.9*,74,78	3.6,71,75	3.8,84,74
	4.3*,48,76					

99

## Aplicações

### Código JAGS

```

model {
  for(i in 1:N) {
    t[i] ~ dweib(r,mu[i]) I(t.cen[i],)      # se o paciente i morre
    log(mu[i]) <- beta[estadio[i]]        # modelo de risco relativo
  }
  for(j in 1:M) {
    beta[j] ~ dnorm(0.0, 0.0001)          # distribuição a priori
    mediana[j] <- pow(log(2) *exp(-beta[j]),1/r) # tempo mediano
  }
  r ~ dgamma(1.0,0.0001)                  # decrescimento lento
  alpha ~ dnorm(0.0, 0.0001)              # distribuição a priori
  contra[1] <- beta[2]-beta[1]             # reparameterização
  contra[2] <- beta[3]-beta[1]
  contra[3] <- beta[4]-beta[1]
}

```

100

## Aplicações

### Comentários

- O parâmetro de escala da distribuição dos tempos de vida (Weibull) é  $\mu_i = \exp(-\beta_{j_i})$ ,  $i=1, \dots, 90$ ,  $j=1, \dots, 4$ ;
- Os tempos de morte dos pacientes censurados são substituídos por NA, enquanto os tempos de vida censurados para os pacientes mortos são iguais a 0.
- Como a distribuição dos tempos de vida censurados é a distribuição Weibull truncada nesses tempos, deve-se introduzir uma variável indicadora,  $I(a, b)$ , para garantir que  $t_i \in (a, b)$ ;
- Os parâmetros de forma da Weibull ( $r$ ) e de regressão ( $\beta_i$ ) têm valores iniciais iguais a 1 e zero, respectivamente;
- A distribuição a priori não informativa para  $r$  pode ser a distribuição gama com parâmetro de forma 1 e de escala 0.0001, visto que esta decresce lentamente em  $\mathbb{R}^+$ .

A análise destes dados via OpenBugs encontra-se tutorialmente em

[https://www.math.tecnico.ulisboa.pt/~gsilva/EBed2/2-Software\\_Ex5.html](https://www.math.tecnico.ulisboa.pt/~gsilva/EBed2/2-Software_Ex5.html)

101

## Aplicações

### Resultados

```

...
model is syntactically correct
...
Node statistics
node      mean      sd      MC error  2.5% median 97.5% start sample
contra[1] -0.109 0.4988 0.005026 -1.148 -0.088 0.823 5001 10000
contra[2]  0.689 0.3500 0.003546  0.0025 0.6891 1.377 5001 10000
contra[3]  1.798 0.4183 0.004474  0.9595 1.8030 2.611 5001 10000
mediana[1] 8.137 1.9890 0.020630  5.2000 7.8200 12.95 5001 10000
mediana[2] 9.425 4.1660 0.043550  4.5550 8.4440 20.01 5001 10000
mediana[3] 4.447 0.9656 0.008804  2.9090 4.3240 6.716 5001 10000
mediana[4] 1.740 0.5052 0.004979  0.9867 1.6640 2.933 5001 10000
r          1.165 0.1397 0.001811  0.9029 1.1590 1.454 5001 10000
dic.stats()
DIC
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
      Dbar      Dhat      pD      DIC
t      288.318    283.414    4.903    293.221
total 288.318    283.414    4.903    293.221
clear(mediana)
coda(*,laringe-boa)
save(laringe-log.txt)

```

102

## Aplicações

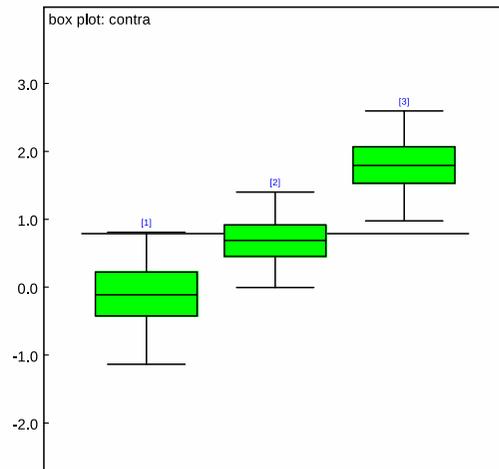


Figura: Box plot dos contrastes  $\beta_2 - \beta_1$ ,  $\beta_3 - \beta_1$  e  $\beta_4 - \beta_1$ .

103

## Aplicações

### Resultados: CODA output

```

HIGHEST PROBABILITY DENSITY INTERVALS:
=====
Alpha level = 0.05
Chain: laringe-boa
      Lower Bound Upper Bound
contra[1] -1.10000    0.8529
contra[2]  0.01479    1.3840
contra[3]  1.01300    2.6500
r          0.89500    1.4430

```

```

RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:
=====
Quantile = 0.025
Accuracy = +/- 0.005
Probability = 0.95
Chain: laringe-boa
      Thin Burn-in Total Lower Bound Dependence Factor
contra[1]  1      2 3874    3746    1.0341698
contra[2]  1      2 3650    3746    0.9743727
contra[3]  1      2 3710    3746    0.9903897
r          2      4 7448    3746    1.9882541

```

104

**Bibliografia principal:**

Paulino, C.D., Amaral Turkman, A., Murteira, B., Silva, G.L. (2018). *Estatística Bayesiana*, 2ª edição. Fundação Calouste Gulbenkian, Lisboa.

**Bibliografia complementar:**

Amaral Turkman, A., Paulino, C.D., Müller, P. (2019). *Computational Bayesian Statistics: An Introduction*. Cambridge University Press, Cambridge.

Banerjee, S., Carlin, B.P., Gelfand, A.E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd Edition. Chapman and Hall/CRC Press, London.

Carlin, B.P., Louis, T.A. (2008). *Bayesian Methods for Data Analysis*, 3rd Edition. Chapman and Hall/CRC, London.

Gamerman, D., Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd Edition. Taylor and Francis, London.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. (2014). *Bayesian Data Analysis*, 3rd Edition. Chapman and Hall/CRC, London.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1995). *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC Press, London.

Ibrahim, J.G., Chen, M.H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.

Pole, A., West, M., Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Chapman and Hall/CRC Press, London.

Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd Edition. Springer, New York.

■ Carlos Daniel Paulino  
<https://www.math.tecnico.ulisboa.pt/~dpaulino/>

■ Giovani Loiola da Silva  
<https://www.math.tecnico.ulisboa.pt/~gsilva/>

■ Paulo Soares  
<https://www.math.tecnico.ulisboa.pt/~psoares/>

■ Webpage do livro **Estatística Bayesiana**  
<https://www.math.tecnico.ulisboa.pt/~gsilva/EBed2/>

■ Exemplo tutorial usando **OpenBugs** in R  
[https://www.math.tecnico.ulisboa.pt/~gsilva/EBed2/2-Software\\_Ex1.html](https://www.math.tecnico.ulisboa.pt/~gsilva/EBed2/2-Software_Ex1.html)