

# MAT02034 - Métodos bayesianos para análise de dados

Modelos de parâmetro único e modelos de múltiplos parâmetros

Rodrigo Citton P. dos Reis  
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2022

# Introdução

# Introdução

- ▶ Nesta aula, descreveremos o uso do R para **resumir** distribuições *a posteriori* para modelos de parâmetro único e com múltiplos parâmetros.
- ▶ Em ambos os casos, a sumarização da distribuição *a posteriori* é facilitada pelo uso de funções R para calcular e simular distribuições da família exponencial.
- ▶ Através de um exemplo, vamos explorar uma abordagem conjugada e um método de aproximação da distribuição *a posteriori*.

## Exemplo: Taxa de Scram em Usinas Nucleares

# Estimation of Trends in the Scram Rate at Nuclear Power Plants

p. 352

**Harry F. MARTZ**

Los Alamos National Laboratory  
Los Alamos, NM 87545  
(hfm@lanl.gov)

**Robert L. PARKER**

Intel Corporation  
Chandler, AZ 85226  
(robert.L.parker@intel.com)

**Dale M. RASMUSON**

Office of Nuclear Regulatory Research  
U.S. Nuclear Regulatory Commission  
Washington, DC 20555  
(dmr@nrc.gov)

An important task of the U.S. Nuclear Regulatory Commission is to examine annual operating data from the nation's population of nuclear power plants for trends over time. We are interested here in trends in the scram rate at 66 commercial nuclear power plants based on annual observed scram data from 1984–1993. For an assumed Poisson distribution on the number of unplanned scrams, a gamma prior, and an appropriate hyperprior, a parametric empirical Bayes (PEB) approximation to a full hierarchical Bayes formulation is used to estimate the scram rate for each plant for each year. The PEB-estimated prior and posterior distributions are then subsequently smoothed over time using an exponentially weighted moving average. The results indicate that such bidirectional shrinkage is quite useful for identifying reliability trends over time.

**KEY WORDS:** Data smoothing; Gamma prior distribution; Hierarchical Bayes; Parametric empirical Bayes; Poisson sampling; Reliability; Scram rate estimation; Shrinkage.

Source: *Technometrics*, Nov., 1999, Vol. 41, No. 4 (Nov., 1999), pp. 352–364

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality

Stable URL: <https://www.jstor.org/stable/1271351>

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ Uma tarefa importante na regulação de usinas nucleares é examinar dados operacionais anuais de usinas para tendências de confiabilidade e segurança ao longo do tempo.
- ▶ Ao nível da planta (usina), uma tendência decrescente na **taxa anual de falhas** de importantes sistemas de segurança é uma indicação de melhorias de confiabilidade sustentadas ao longo do tempo, mas uma tendência crescente geralmente indica a necessidade de ações corretivas.

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ Quando o processo de fissão ou reatividade de um reator de energia nuclear está acima de um certo nível limite, o reator é dito estar em estado crítico ou, simplesmente, crítico.
- ▶ Um importante sistema de segurança em uma usina nuclear é o sistema de proteção do reator. Este sistema é projetado para mudar rapidamente o reator de um estado crítico para um não crítico após algum evento transitório, como perda de energia externa.
- ▶ Em resposta a tal evento, hastes de controle, que são feitas de um material apropriado para absorção de nêutrons, são inseridas no núcleo do reator, desligando assim a reatividade.

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ Isso é conhecido como trip ou **scram** do reator e pode ser realizado de forma automática ou manual. Os scrams do reator podem resultar de eventos iniciais que variam de incidentes relativamente menores a eventos que são precursores de acidentes.
- ▶ A **taxa na qual ocorrem scrams não planejados** é, portanto, um indicador importante do desempenho e da confiabilidade geral da planta.



## Exemplo: Taxa de Scram em Usinas Nucleares

- A Tabela 1 contém dados de scrams de 1984-1993 para 66 usinas nucleares comerciais dos EUA, cada uma com horas críticas diferentes de zero para cada ano desse período. Os dados foram obtidos de uma série de relatórios anuais e consistem no número anual de scrams não planejados  $x_{ij}$  no total de horas críticas  $T_{ij}$  para a planta  $i = 1, \dots, 66$  e ano codificado  $j = 1, \dots, 10$ .

Table 1. U.S. Commercial Nuclear Power Plant Scram Data from 1984-1993

Plant	1984		1985		1986		1987		1988		1989		1990		1991		1992		1993	
	x	T	x	T	x	T	x	T	x	T	x	T	x	T	x	T	x	T	x	T
Arkansas 1	3	6,250.0	8	7,017.5	2	5,536.7	2	7,855.7	1	6,156.6	5	5,999.1	0	6,500.2	2	8,149.8	0	7,137.8	1	7,599.4
Arkansas 2	12	7,643.3	9	6,383.0	5	6,370.0	2	7,715.4	2	6,032.0	2	6,610.1	4	8,246.6	1	7,341.1	0	6,454.2	0	8,390.4
Beaver Valley 1	4	6,451.6	8	8,247.4	3	6,243.8	4	7,339.4	3	7,066.7	4	5,887.6	1	8,155.9	2	5,029.2	1	8,226.7	1	5,980.6
Big Rock Point	2	6,896.6	3	6,521.7	1	8,387.3	1	6,215.5	3	6,394.2	1	6,920.8	0	6,759.0	0	7,460.5	3	4,790.5	0	6,958.8
Brunswick 2	3	2,654.9	3	7,142.9	2	4,232.4	2	8,328.5	2	5,645.8	1	5,779.9	6	5,926.6	2	5,236.2	1	2,378.3	0	5,915.3
Callaway	12	1,503.8	19	8,154.5	7	7,307.6	1	6,227.7	6	8,202.1	2	7,481.6	4	7,365.0	1	8,734.1	3	7,289.2	0	7,569.0
Calvert Cliffs 1	5	7,575.8	6	5,357.1	4	6,906.2	6	6,615.6	3	6,398.5	0	1,806.6	0	1,924.5	1	6,687.0	1	5,050.2	2	8,619.0
Cook 1	3	8,108.1	1	2,564.1	5	7,536.4	2	6,012.2	3	8,433.8	2	6,169.8	0	6,944.8	1	7,754.3	1	5,752.1	0	8,760.0
Cook 2	7	5,303.0	4	5,970.1	2	5,560.5	5	6,290.3	0	2,715.5	1	6,580.9	3	4,958.9	3	8,053.2	1	3,169.4	2	8,491.5

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ **Martz, Parker e Rasmuson (1999)**, consideraram a distribuição de Poisson um modelo apropriado para descrever a variabilidade de amostragem condicional nos dados individuais específicos da planta na Tabela 1.
- ▶ Ou seja, condicional à verdadeira taxa de scram desconhecida  $\lambda_{ij}$ , os autores assumem que  $x_{ij}$  segue uma distribuição de Poisson com parâmetro  $\lambda_{ij} T_{ij}$ . Além disso, assumimos que todos os valores  $x_{ij}$  são distribuídos condicionalmente independentemente, dadas as taxas de scram verdadeiras (mas desconhecidas), ou seja,

$$x_{ij} | \lambda_{ij} \stackrel{\text{ind.}}{\sim} \text{Poi}(\lambda_{ij} T_{ij}), \lambda_{ij} > 0.$$

## Exemplo: Taxa de Scram em Usinas Nucleares

- Note que o **estimador de máxima verossimilhança (EMV)** de  $\lambda_{ij}$  é simplesmente  $x_{ij} / T_{ij}$ . Isso pode ser visto, escrevendo a (contribuição individual para a) **função de verossimilhança** para  $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \dots, \lambda_{21}, \dots, \lambda_{66,10})$

$$L(\lambda_{ij}) = \frac{1}{x_{ij}!} (\lambda_{ij} T_{ij})^{x_{ij}} \exp\{-\lambda_{ij} T_{ij}\}.$$

- Desta forma, a **função de log-verossimilhança** é obtida

$$\ell(\lambda_{ij}) \propto x_{ij} \log(\lambda_{ij} T_{ij}) - \lambda_{ij} T_{ij}.$$

e a derivada parcial de  $\ell(\boldsymbol{\lambda})$  é dada por

$$\frac{\partial}{\partial \lambda_{ij}} \ell(\boldsymbol{\lambda}) = \frac{x_{ij}}{\lambda_{ij}} - T_{ij}.$$

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ Ao igualar esta última expressão a zero, obtemos o EMV de  $\lambda_{ij}$

$$\hat{\lambda}_{ij} = \frac{x_{ij}}{T_{ij}}.$$

- ▶ Usando os dados da Tabela 1, cada um deles é plotado na Figura 1. Os EMVs correspondentes da taxa média de scram para as 66 plantas também são apresentadas na Figura 1 por ano. Observe a tendência decrescente na taxa média de scram ao longo do tempo, particularmente entre 1984 e 1988.

## Exemplo: Taxa de Scram em Usinas Nucleares

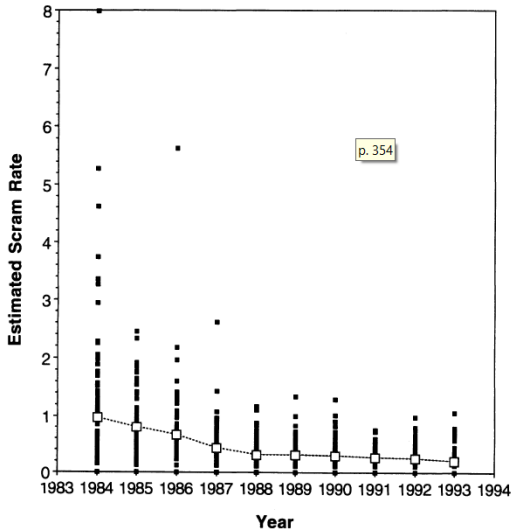


Figure 1. Maximum Likelihood Estimates of the Scram Rate Per 1,000 Critical Operating Hours.

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ **Martz, Parker e Rasmuson (1999)** utilizam um **modelo hierárquico** para levar em consideração aspectos **individuais** e **comuns** das usinas nucleares para estimar as **taxas de scrams** de cada usina em cada ano de observação.
- ▶ Voltaremos a esta abordagem quando formos estudar os **modelos bayesianos hierárquicos**.

## Exemplo: Taxa de Scram em Usinas Nucleares

- ▶ Nesta aula, vamos considerar apenas os dados do ano de 1984 (podemos descartar o uso do índice *j*)

```
library(dplyr)

scram <- readr::read_table2(
  file = here::here("dados", "table76.txt"))

scram <- scram %>%
  filter(Year == 1) %>%
  select(-X5, -Year) %>%
  rename(x = y) %>%
  mutate(T = T/1000)

scram
```

## Exemplo: Taxa de Scram em Usinas Nucleares

```
## # A tibble: 66 x 3
##   Plant          x      T
##   <chr>      <dbl> <dbl>
## 1 Arkansas1      3  6.25
## 2 Arkansas2     12  7.64
## 3 Beaver.Valley1  4  6.45
## 4 Big.Rock.Point  2  6.90
## 5 Brunswick2     3  2.65
## 6 Callaway      12  1.50
## 7 Calvert.Cliffs1  5  7.58
## 8 Cook1         3  8.11
## 9 Cook2         7  5.30
## 10 Cooper.Station  3  6
## # ... with 56 more rows
```



## Exemplo: Taxa de Scram em Usinas Nucleares

- Vamos supor que os dados de ocorrências de scrams são **independentes**, distribuídos conforme uma distribuição de **Poisson com parâmetro comum**  $\lambda$ . Desta forma, a função de verossimilhança é dada por

$$L(\lambda) = \prod_{i=1}^{66} \frac{1}{x_i!} (\lambda T_i)^{x_i} \exp\{-\lambda T_i\} \propto \lambda^{\sum_{i=1}^{66} x_i} \exp\left\{-\lambda \sum_{i=1}^{66} T_i\right\}.$$

- Note que o EMV de  $\lambda$  é dado por  $\hat{\lambda} = \sum_{i=1}^{66} x_i / \sum_{i=1}^{66} T_i$ .

```
soma.x <- sum(scram$x)
soma.T <- sum(scram$T)
lambda.chap <- soma.x/soma.T
soma.x
```

```
## [1] 361
```

## Exemplo: Taxa de Scram em Usinas Nucleares

```
soma.T
```

```
## [1] 374.2288
```

```
lambda.chap
```

```
## [1] 0.9646505
```

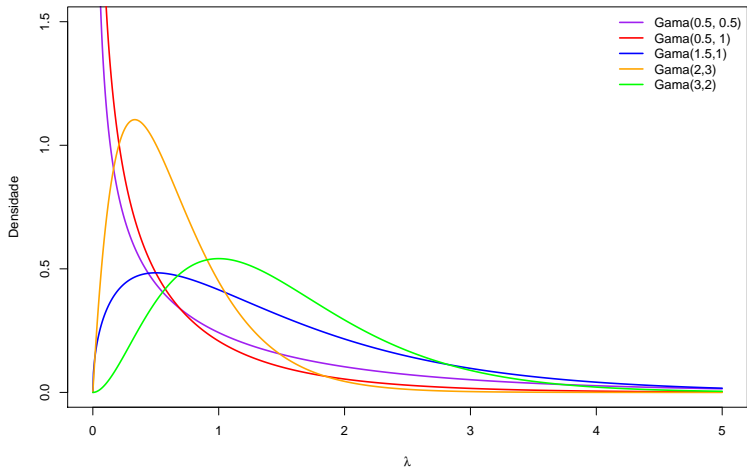
## Modelos de parâmetro único

# Modelos de parâmetro único

- ▶ Uma abordagem bayesiana para o problema de estimar o parâmetro  $\lambda$  consiste em especificar uma distribuição *a priori* para  $\lambda$ .
- ▶ O uso de uma **análise conjugada** proporciona facilidade algébrica para obtenção da distribuição *a posteriori* de  $\lambda$ .
- ▶ Suponha que assumimos uma distribuição Gama *a priori*

$$g(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \alpha, \beta > 0.$$

# Modelos de parâmetro único



## Modelos de parâmetro único

- Combinando esta distribuição com a verossimilhança, temos a distribuição *a posteriori*

$$g(\lambda|\mathbf{x}) \propto \lambda^{\alpha + \sum_{i=1}^{66} x_i - 1} \exp \left\{ -\lambda \left( \beta + \sum_{i=1}^{66} T_i \right) \right\},$$

que também é Gama com parâmetros  $\alpha + \sum_{i=1}^{66} x_i$  e  $\beta + \sum_{i=1}^{66} T_i$ .

- Dado os hiperparâmetros  $\alpha$  e  $\beta$ , podemos realizar inferências para  $\lambda$  a partir das propriedades da distribuição Gama.

## Modelos de parâmetro único

- Note, por exemplo, que a média *a posteriori* é dada por

$$\frac{\alpha + \sum_{i=1}^{66} x_i}{\beta + \sum_{i=1}^{66} T_i} = \frac{\alpha}{\beta} \times \left( \frac{1}{1 + \sum_{i=1}^{66} T_i / \beta} \right) + \hat{\lambda} \times \left( \frac{\sum_{i=1}^{66} T_i / \beta}{1 + \sum_{i=1}^{66} T_i / \beta} \right)$$

é uma **média ponderada** entre a **média *a priori*** e o **EMV** de  $\lambda$ .

## Modelos de parâmetro único

- ▶ Assumindo uma **informação *a priori* “vaga”**, e especificando os hiperparâmetros  $\alpha = \beta = 0.01$  (a média e a variância *a priori* são 1 e 100, respectivamente), temos que a distribuição *a posteriori* de  $\lambda$  é

$$\lambda | \mathbf{x} \sim \text{Gama}(0.01 + 361, 0.01 + 374.2288).$$

- ▶ Um estimador bayesiano com base na média *a posteriori* nos fornece a seguinte estimativa:  $\tilde{\lambda} = 0.9646514$ .
  - ▶  $\tilde{\lambda} \approx \hat{\lambda}$ , como era de se esperar, devido a informação *a priori* vaga.
  - ▶ A partir da distribuição *a posteriori* podemos construir intervalos de credibilidade e avaliar probabilidades com respeito a  $\lambda$ .



## Modelos de múltiplos parâmetros

## Modelos de múltiplos parâmetros

- Um modelo alternativo à distribuição de Poisson para os dados é dada pela seguinte função de probabilidade<sup>1</sup>

$$p(x_i|\alpha, \beta) = \frac{\Gamma(\alpha + x_i)}{\Gamma(\alpha)\Gamma(x_i + 1)} \left( \frac{\beta}{\beta + T_i} \right)^\alpha \left( \frac{T_i}{\beta + T_i} \right)^{x_i}, \quad \alpha, \beta > 0.$$

---

<sup>1</sup>Esta distribuição pode ser obtida da mistura da distribuição Poisson por Gama e pode ser vista como a generalização da distribuição Binomial Negativa.

# Modelos de múltiplos parâmetros

- ▶ Supondo que
  - ▶  $x_i \stackrel{\text{ind.}}{\sim} p(x_i|\alpha, \beta)$ ,  $i = 1, \dots, n$ ;
  - ▶ e que a distribuição *a priori* é especificada como  $g(\alpha, \beta) \propto \alpha^{-1}(\beta + 1)^{-2}$ ,  $\alpha, \beta > 0$  (também pode ser vista como uma informação vaga para os parâmetros  $\alpha$  e  $\beta$ ).
- ▶ Então a distribuição *a posteriori* de  $\alpha$  e  $\beta$  é obtida por

$$g(\alpha, \beta|\mathbf{x}) \propto g(\alpha, \beta) \prod_{i=1}^n p(x_i|\alpha, \beta).$$

- ▶ Embora a distribuição *a priori*,  $g(\alpha, \beta)$  seja **imprópria**<sup>2</sup>, a distribuição *a posteriori* é própria.
- ▶ Note que a distribuição *a posteriori* não pode ser obtida analiticamente.

---

<sup>2</sup>Quando a integral da densidade é diferente de 1.

# Aproximação por grade

- Imagine que há uma imagem que você não pode ver em sua totalidade; você apenas observa trechos ao longo de uma grade que varre da esquerda para a direita na imagem.



- Quanto mais fina a grade, mais nítida é a imagem.

# Aproximação por grade

- ▶ Esta é a ideia geral por trás da aproximação bayesiana por grade, caso em que a “imagem” alvo é distribuição *a posteriori*.
  - ▶ Assumindo  $\theta$  como um parâmetro (ou vetor de parâmetros) genérico, não precisamos observar  $g(\theta|\mathbf{x})$  em todos os possíveis  $\theta$  para ter uma noção de sua estrutura.
  - ▶ Podemos avaliar  $g(\theta|\mathbf{x})$  em uma grade finita e discreta de possíveis valores  $\theta$ .
  - ▶ Posteriormente, podemos obter **amostras aleatórias** desta distribuição discretizada para **aproximar** a distribuição *a posteriori*.

# Aproximação por grade

A aproximação de grade produz uma amostra de  $N$  valores independentes de  $\theta$ ,  $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}\}$ , a partir de uma aproximação discretizada da distribuição *a posteriori*  $g(\theta|\mathbf{x})$ . Este algoritmo evolui em quatro etapas:

1. Defina uma grade discreta de possíveis valores  $\theta$ .
2. Avalie a distribuição *a priori*  $g(\theta)$  e a função de verossimilhança  $L(\theta)$  em cada valor de  $\theta$  da grade.
3. Obtenha uma aproximação discreta da distribuição *a posteriori*  $g(\theta|\mathbf{x})$  por: (a) calcular o produto  $g(\theta)L(\theta)$  em cada valor de  $\theta$  da grade; e então (b) **normalizar** os produtos para que somem 1 em todos os  $\theta$ .
4. Obter uma amostra aleatória de  $N$  valores da grade de  $\theta$  em relação às suas correspondentes probabilidades *a posteriori* normalizadas.

# Aproximação por grade

```
# Passo 1: criar a grade
```

```
alpha_grade <- seq(from = 0.01, to = 4, length = 200)
```

```
beta_grade <- seq(from = 0.01, to = 4, length = 200)
```

```
grade_df <- expand.grid(alpha_grade, beta_grade)
```

```
names(grade_df) <- c("alpha_g", "beta_g")
```

## Aproximação por grade

*# Passos 2: avaliar a priori e vero. em cada ponto da grade*

```
priori_ab <- function(a, b){  
  a(-1)*(b + 1)(-2)  
}
```

```
veros_ab <- function(a, b, dados){  
  x <- dados$x  
  t <- dados$T  
  
  prod( (gamma(a + x) / (gamma(x + 1) * gamma(a)))  
        * ((b/(b + t))a * ((t/(b + t))x) )  
}
```



## Aproximação por grade

```
grade_df$priori <- 0
grade_df$veros <- 0

for(i in 1:length(grade_df$priori)){

  grade_df$priori[i] <- priori_ab(a = grade_df$alpha_g[i],
                                  b = grade_df$beta_g[i])
  grade_df$veros[i] = veros_ab(a = grade_df$alpha_g[i],
                               b = grade_df$beta_g[i],
                               dados = scram)

}
```

## Aproximação por grade

```
# Passo 3: aproximar a posteriori
```

```
library(dplyr)
```

```
grade_df <- grade_df %>%  
  mutate(post_naonorm = priori * veros,  
         posteriori = post_naonorm/sum(post_naonorm))
```

```
# Confermando que a posteriori aproximada soma 1
```

```
grade_df %>%  
  summarize(sum(post_naonorm), sum(posteriori))
```

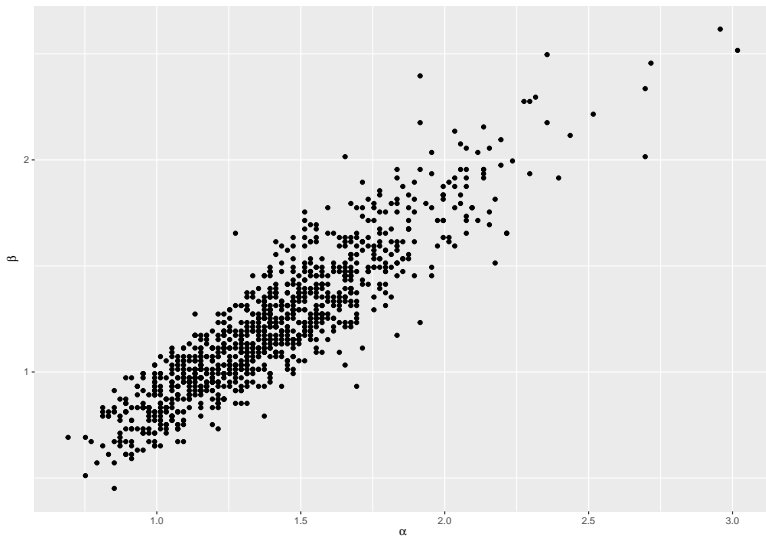
```
##    sum(post_naonorm) sum(posteriori)  
## 1      4.850068e-81              1
```

## Aproximação por grade

*# Passo 4: amostrar da posteriori discretizada*

```
post_amostra <- sample_n(grade_df, size = 1000,  
                          weight = posteriori,  
                          replace = TRUE)
```

# Aproximação por grade



## Aproximação por grade

- Podemos obter estimativas de  $\alpha$  e  $\beta$  a partir da amostra da distribuição *a posteriori*.

```
mean(post_amostra$alpha_g)
```

```
## [1] 1.399803
```

```
sd(post_amostra$alpha_g)
```

```
## [1] 0.3317416
```

```
mean(post_amostra$beta_g)
```

```
## [1] 1.217827
```

```
sd(post_amostra$beta_g)
```

```
## [1] 0.336224
```

# Considerações finais

## Considerações finais

- ▶ Limitações no método de aproximação por grade se apresentam rapidamente à medida que nossos modelos se tornam mais complicados (que apresentam muitos parâmetros).
- ▶ Em tais configurações, a aproximação por grade sofre com a **“maldição da dimensionalidade”**.
- ▶ Analogamente, ao usar a aproximação de grade para simular posteriores multivariadas, precisamos dividir o espaço amostral multidimensional em uma grade muito, muito fina para evitar grandes lacunas em nossa aproximação.
- ▶ Na prática, isso pode não ser viável.
- ▶ Quando avaliado em grades cada vez mais finas, o método de aproximação de grade torna-se computacionalmente caro.
- ▶ Os métodos **MCMC** fornecem uma alternativa mais flexível.

## Para casa

- ▶ Rodar os códigos dos exemplos de aula.
  - ▶ Trazer as dúvidas para o Fórum Geral do Moodle e para a próxima aula.
- ▶ Exercício (Moodle).



# Próxima aula

- ▶ Introdução a computação bayesiana.

# Por hoje é só!

Bons estudos!

