MAT02034 - Métodos bayesianos para análise de dados

Modelos bayesianos hierárquicos

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

- Em muitos problemas estatísticos, estamos interessados em aprender sobre **muitos parâmetros** que estão conectados de alguma forma.
  - Taxas (λ_i) de ocorrência de eventos em diferentes unidades (hospitais, municípios, usinas, etc.).
  - Proporções (p_i) ou probabilidades de "sucesso" em diferentes unidades (fábricas, escolas, etc.).
  - Médias (μ_i ou "estados latentes" em diferentes unidades (modelos dinâmicos no tempo).
- Em situações de muitos parâmetros é comum construir uma distribuição *a priori* de forma **hierárquica**.

Introdução

- Suponha um conjunto de observações $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
- Suponha que a distribuição conjunta de $\mathbf{x}$ é dada por

$$p(\mathbf{x}|\boldsymbol{\theta}) = p(x_1, x_2, \ldots, x_n|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

em que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ é um vetor de parâmetros ($\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ também podem ser vetores).

Introdução

- Agora suponha que a distribuição *a priori* (**conjunta**) é especificada da seguinte forma:

$$g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)g_2(\boldsymbol{\theta}_2).$$

- Uma forma alternativa de descrição deste modelo é dada por

$$\begin{aligned} (\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\sim p(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\ (\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) &\sim g_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \\ \boldsymbol{\theta}_2 &\sim g_2(\boldsymbol{\theta}_2) \end{aligned}$$

Introdução

- Note que, em geral, $g_2(\boldsymbol{\theta}_2)$ depende de parâmetros a serem especificados pelo pesquisador (especialista) da área.
  - Em muitas situações não há interesse direto em $\boldsymbol{\theta}_2$.
- Ainda, de acordo com o problema estatístico, $\boldsymbol{\theta}$ pode ser fatorado em mais de duas partes, formulando mais estágios na hierarquia.

Introdução

- A distribuição *a posteriori* de um **modelo bayesiano hierárquico** pode ser expressa como

$$g(\boldsymbol{\theta}|\mathbf{x}) = g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \times g_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \times g_2(\boldsymbol{\theta}_2).$$

- Note que, de acordo com as suposições do modelo, a distribuição $p(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ pode não depender de $\boldsymbol{\theta}_2$, o que simplifica o modelo:

$$g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta}_1) \times g_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \times g_2(\boldsymbol{\theta}_2).$$

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

- 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 $i = 1, \ldots, 66$ e ano codificado $j = 1, \ldots, 10$.

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

	1984		1985		1986		1987		1988		1989		1990		1991		1992		1993	
Plant	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.
- Considerando apenas os dados do ano de 1984, condicional à verdadeira taxa de scram desconhecida λ_i, os autores assumem que x_i segue uma distribuição de Poisson com parâmetro $\lambda_i T_i$. Além disso, assumimos que todos os valores x_i são condicionalmente independentemente, dadas as taxas de scram verdadeiras (mas desconhecidas).

Exemplo: Taxa de Scram em Usinas Nucleares

- Ou seja,

$$x_i|\lambda_i \overset{ind.}{\sim} Poi(\lambda_i T_i),\ \lambda_i > 0,\ i = 1, \ldots, n.$$

- Suponha que assumimos uma distribuição Gama *a priori*

$$g_1(\lambda_i) \overset{iid}{\sim} Gama(\alpha, \beta),\ \alpha, \beta > 0,\ i = 1, \ldots, n.$$

- Ainda, suponha que

$$g_2(\beta) \sim Gama(\gamma, \delta),\ \alpha = 1.4, \gamma = 0.01, \delta = 1.$$

Exemplo: Taxa de Scram em Usinas Nucleares

- A distribuição *a posteriori* deste modelo, $g(\boldsymbol{\lambda}, \beta|\mathbf{x})$, é proporcional à

$$
\begin{aligned}
& p(\mathbf{x}|\boldsymbol{\lambda}, \beta) \times g_1(\boldsymbol{\lambda}|\beta) \times g_2(\beta) \\
= & \left\{\prod_{i=1}^{n} p(x_i|\lambda_i)\right\} \times \left\{\prod_{i=1}^{n} g_1(\lambda_i|\beta)\right\} \times g_2(\beta) \\
= & \left\{\prod_{i=1}^{n} p(x_i|\lambda_i) \times g_1(\lambda_i|\beta)\right\} \times g_2(\beta) \\
\propto & \left\{\prod_{i=1}^{n} (\lambda_i T_i)^{x_i} e^{-\lambda_i T_i} \beta^{\alpha} \lambda_i^{\alpha-1} e^{-\lambda_i \beta}\right\} \times \beta^{\gamma-1} e^{-\delta\beta} \\
\propto & \left\{\prod_{i=1}^{n} \lambda_i^{\alpha+x_i-1} e^{-\lambda_i(\beta+T_i)}\right\} \times \beta^{n\alpha+\gamma-1} e^{-\delta\beta}.
\end{aligned}
$$

Exercício

- A partir da distribuição *a posteriori*, encontre as distribuições **condicionais completas** dos parâmetros do modelo do exemplo apresentado em aula.
- A partir das condicionais completas, escreva o pseudo-código do amostrador de Gibbs para obter uma amostra da distribuição *a posteriori* deste modelo.

Exemplo: Taxa de Scram em Usinas Nucleares

Da distribuição *a posteriori* do modelo

$$g(\boldsymbol{\lambda}, \beta|\mathbf{x}) \propto \left\{ \prod_{i=1}^{n} \lambda_i^{\alpha+x_i-1} e^{-\lambda_i(\beta+T_i)} \right\} \times \beta^{n\alpha+\gamma-1} e^{-\delta\beta}.$$

temos que as distribuições condicionais completas são dadas por

$$\begin{aligned} p(\lambda_i|\boldsymbol{\lambda}_{-i}, \beta, \mathbf{x}) &\propto \lambda_i^{\alpha+x_i-1} e^{-\lambda_i(\beta+T_i)}, \ \ i = 1, \ldots, n, \\ p(\beta|\boldsymbol{\lambda}, \mathbf{x}) &\propto \beta^{n\alpha+\gamma-1} e^{-\beta(\delta+\sum_{i=1}^{n}\lambda_i)}. \end{aligned}$$

Exemplo: Taxa de Scram em Usinas Nucleares

Exemplo: Taxa de Scram em Usinas Nucleares

- Assim, notamos que as condicionais completas possuem uma forma familiar,

$$
\begin{aligned}
(\lambda_i|\beta, x_i, T_i) &\sim \mathit{Gama}(\alpha + x_i, \beta + T_i),\ i = 1, \ldots, n, \\
(\beta|\boldsymbol{\lambda}, \mathbf{x}) &\sim \mathit{Gama}(n\alpha + \gamma, \delta + \sum_{i=1}^{n} \lambda_i),
\end{aligned}
$$

e em vez de amostrar diretamente o vetor $\theta = (\lambda_1, \ldots, \lambda_n, \beta)$ de uma só vez, pode-se sugerir uma amostragem progressiva e iterativa, começando, por exemplo, com os λ_i para um dado valor inicial de β, seguido por uma atualização de β dadas as novas amostras $\lambda_1, \ldots, \lambda_n$.

Exemplo: Taxa de Scram em Usinas Nucleares

Mais precisamente, dada uma amostra, na iteração t, $\theta^t = (\lambda_1^t, \ldots, \lambda_n^t, \beta^t)$ pode-se proceder da seguinte forma na iteração $t+1$,

1. $\lambda_i^{t+1} | (\beta^t, x_i, T_i) \sim Gama(\alpha + x_i, \beta^t + T_i), \; i = 1, \ldots, n,$
2. $\beta^{t+1} | (\lambda_1^{t+1}, \ldots, \lambda_n^{t+1}, \mathbf{x}) \sim Gama(n\alpha + \gamma, \delta + \sum_{i=1}^n \lambda_i^{t+1})$.

Exemplo: Taxa de Scram em Usinas Nucleares

```r
library(readr)
library(dplyr)

scram <- 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

```r
# Configuração

alpha <- 1.4
gamma <- 0.01
delta <- 1

n <- length(scram$x)
M <- 10000

theta <- matrix(0, nrow = M, ncol = n + 1)
colnames(theta) <- c("beta", paste0("lambda_", 1:n))
```

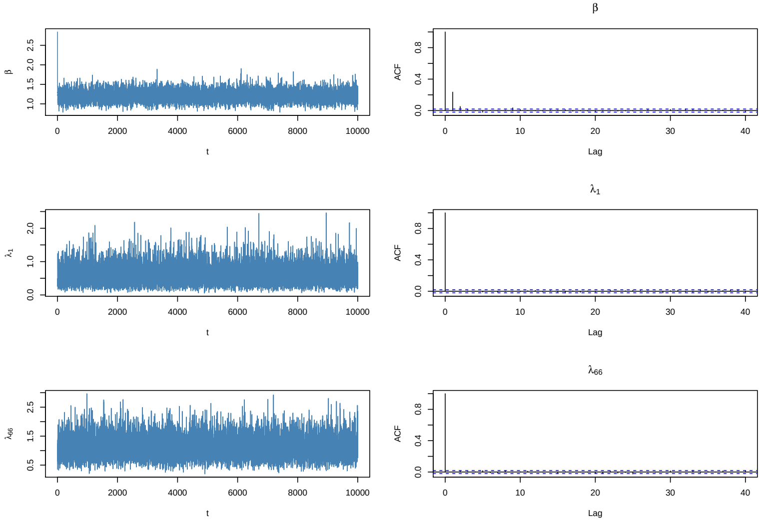
Exemplo: Taxa de Scram em Usinas Nucleares

```r
# Amostrador de Gibbs (valor inicial e passo 1)

beta <- 10
lambda <- rgamma(n = n,
                 shape = alpha + scram$x,
                 rate = beta + scram$T)
beta <- rgamma(n = 1,
               shape = n*alpha + gamma,
               rate = delta + sum(lambda))
theta[1,] <- c(beta, lambda)
```

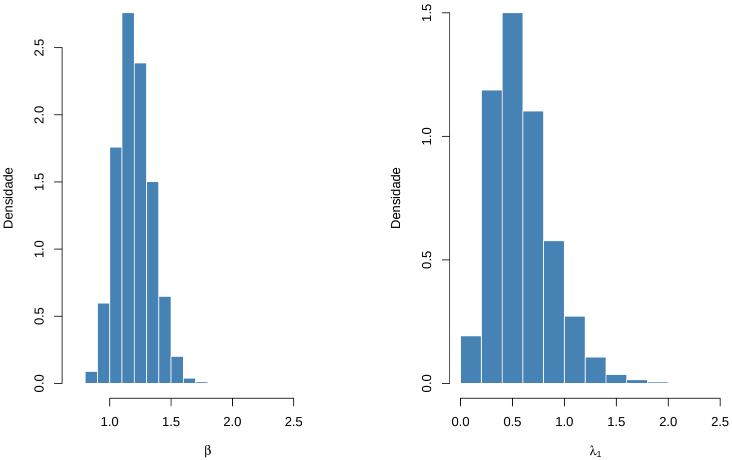
Exemplo: Taxa de Scram em Usinas Nucleares

```r
# Amostrador de Gibbs (passo t)

for (t in 2:M){

  theta[t,2:(n+1)] <- rgamma(n = n,
                              shape = alpha + scram$x,
                              rate = theta[(t-1),1] + scram$T)

  theta[t,1] <- rgamma(n = 1,
                        shape = n*alpha + gamma,
                        rate = delta + sum(theta[t,2:(n+1)]))

}
```

Exemplo: Taxa de Scram em Usinas Nucleares



Exemplo: Taxa de Scram em Usinas Nucleares



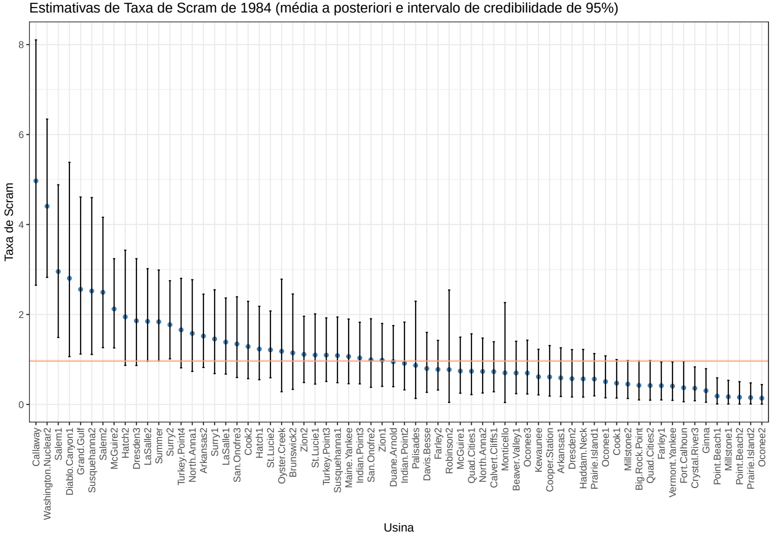
Exemplo: Taxa de Scram em Usinas Nucleares

```
# Resumo da distribuição
# a posteriori das taxas de scram

lambda_post <- data.frame(
  usina = scram$Plant,
  x = scram$x,
  T = scram$T,
  EMV = scram$x/scram$T,
  media = colMeans(theta[,2:(n+1)]))

lambda_post$li95 <- apply(X = theta[,2:(n+1)],
                          MARGIN = 2,
                          FUN = quantile, probs = 0.025)

lambda_post$ls95 <- apply(X = theta[,2:(n+1)],
                          MARGIN = 2,
                          FUN = quantile, probs = 0.975)

head(lambda_post)
```

Exemplo: Taxa de Scram em Usinas Nucleares

```
##                     usina  x      T       EMV     media      li95      ls95
## lambda_1       Arkansas1  3 6.2500 0.4800000 0.5909799 0.1774041 1.2590230
## lambda_2       Arkansas2 12 7.6433 1.5700025 1.5196746 0.8223605 2.4511172
## lambda_3 Beaver.Valley1  4 6.4516 0.6200012 0.6994506 0.2395248 1.4044072
## lambda_4 Big.Rock.Point  2 6.8966 0.2899980 0.4237414 0.1010126 0.9683057
## lambda_5      Brunswick2  3 2.6549 1.1299861 1.1453272 0.3351259 2.4553522
## lambda_6        Callaway 12 1.5038 7.9797845 4.9689558 2.6508726 8.1038016
```

Exemplo: Taxa de Scram em Usinas Nucleares



Exercícios

1. Mude o valor inicial da cadeia, gere uma nova trajetória e aplique os métodos de avaliação de convergência.
2. Rode uma cadeia de trajetória com 100.000 passos; avalie o traço da cadeia (para cada parâmetro); pode-se concluir que a cadeia convergiu?
3. Com base nos exercícios anteriores, adapte o código apresentado em aula considerando o período de burn-in e thinning.

Para casa

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

Próxima aula

- Modelos bayesianos hierárquicos (continuação).

Por hoje é só!

Bons estudos!