

ANÁLISE BAYESIANA DA TAXA DE CRIMINALIDADE NA REGIÃO METROPOLITANA DE BELO HORIZONTE USANDO O MODELO PARTIÇÃO PRODUTO

Flávio B. Gonçalves, Rosângela H. Loschi e Frederico R. B. da Cruz

Departamento de Estatística – ICEX - Universidade Federal de Minas Gerais
Av. Presidente Antônio Carlos, 6627, Campus Pampulha, Belo Horizonte, Minas Gerais, CEP:
31270-010

Email: *bambirra@mail.com*, { *loschi, fcruz* }@*est.ufmg.br*

Resumo: Neste trabalho aplicaremos o Modelo Partição Produto (MPP) para analisar dados de crimes violentos na cidade de Belo Horizonte no período compreendido entre janeiro de 1998 e setembro de 2001. Nosso objetivo é estudar o comportamento, ao longo do tempo, das taxas de criminalidade em três regiões da cidade - a região central, uma região comercial próxima do centro (Santo Agostinho, Barro Preto) e uma região afastada do centro de classe média-alta (Mangabeiras, Cruzeiro, etc.)- e compará-las. Além da taxa de criminalidade em cada mês, também exibiremos as distribuições *a posteriori* do número de mudanças nesta taxa, dos instantes onde as mudanças ocorreram e da probabilidade de ocorrer uma mudança em um instante qualquer.

Palavras Chaves: amostrador de Gibbs, distribuição de Poisson, distribuição Gama.

Abstract: The aim of this paper is to apply the product partition model to obtain the product estimates of the rate of violent crimes in three neighbourhoods of Belo Horizonte city - the downtown, a commercial area near downtown (Santo Agostinho, Barro Preto) and a rich area far from the central neighbourhood (Mangabeiras, Cruzeiro, etc.). Besides the criminality rate in each month, we will provide the posterior distribution of the number of changes in this rate, the posterior distribution of the instants in which the changes occurred and the posterior distribution of the probability that a change occurs at any instant in the sequence. A comparative analysis of the three areas will be presented.

Keywords: Gamma distribution, Gibbs sampling, Poisson distribution.

1. Introdução

Neste trabalho aplicaremos o Modelo Partição Produto (MPP) para analisar algumas séries formadas pelo número de crimes violentos registrados mês a mês em alguns bairros de Belo Horizonte-MG. Nosso objetivo principal é avaliar se a taxa de criminalidade nestas regiões sofreram mudanças ao longo do tempo. Além da estimativa para a taxa de criminalidade em cada mês, também exibiremos as distribuições *a posteriori* do número de mudanças nesta taxa, dos instantes onde as mudanças ocorreram e da probabilidade de ocorrer uma mudança em um instante qualquer.

O MPP introduz mais flexibilidade na análise do problema de pontos de mudança por considerar como aleatório o número de pontos de mudança na série. Apesar disto as distribuições *a posteriori* são fortemente influenciadas pelas distribuições *a priori* (Loschi e Cruz, 2002). Outras referências sobre análise de pontos de mudança usando o MPP são Barry e Hartigan (1993), Crowley (1997), Quintana e Iglesias (2003), Loschi, Cruz, Iglesias e Arellano Valle (2003) entre outras.

Este artigo está assim organizado: Na Seção 2 apresentamos o MPP assumindo as coesões de Yao (1984); na Seção 3 apresentamos o MPP para dados que dadas as taxas, são distribuídos segundo uma distribuição de Poisson; na Seção 4 descrevemos alguns métodos computacionais propostos na literatura para estimar as distribuições *a posteriori* envolvidas no MPP. Finalmente, na Seção 5 apresentamos uma análise comparativa das séries de crimes violentos para a região central, uma região comercial próxima do centro (Santo Agostinho, Barro Preto) e uma região afastada do centro de classe média-alta (Mangabeiras, Cruzeiro, etc.) da cidade de Belo Horizonte–MG no período de janeiro de 1998 a setembro de 2001.

2. Modelo Partição Produto para Coesões de Yao

Seja X_1, \dots, X_n uma seqüência de dados seqüencialmente observados e considere o conjunto de índices $I = \{1, \dots, n\}$. Considere uma partição aleatória $\rho = \{i_0, \dots, i_b\}$ do conjunto de índice I tal que, $0 = i_0 < i_1 < \dots < i_b = n$, e uma variável aleatória B que denota o número de blocos em ρ . Considere que cada partição divide a seqüência X_1, \dots, X_n em $B = b$ subseqüências contíguas, as quais serão denotadas aqui por $X_{[ij]} = (X_{i+1}, \dots, X_j)$. Seja C_{ij} a coesão *a priori* associada com o bloco $[ij] = \{i+1, \dots, j\}$, para $i, j \in I \cup \{0\}$, e $j > i$, que representa o grau de similaridade entre as observações em $X_{[ij]}$ e pode ser interpretada aqui como a probabilidades de transição na cadeia de Markov definida por pontos de mudança ((Barry e Hartigan, 1992)).

Considere p , para $0 \leq p \leq 1$, a probabilidade de que uma mudança ocorra em um instante qualquer. A coesão *a priori* para o bloco $[ij]$ proposta por Yao (1984) e dada por:

$$C_{ij} = \begin{cases} p (1 - p)^{j-i-1}; & j < n \\ (1 - p)^{j-i-1}; & j = n \end{cases} \quad [2.1]$$

para todo $i, j \in I$, $i < j$. Essas coesões *a priori* implicam que a seqüência de pontos de mudança estabelece um processo de renovação, com tempos de ocorrência idêntica e geometricamente distribuídos.

Seja $\theta_1, \dots, \theta_n$ uma seqüência de parâmetros desconhecidos, de forma que condicionalmente em $\theta_1, \dots, \theta_n$ a seqüência de variáveis aleatórias X_1, \dots, X_n têm densidades condicionais marginais $f_1(X_1 | \theta_1), \dots, f_n(X_n | \theta_n)$ respectivamente. A distribuição *a priori* de $\theta_1, \dots, \theta_n$ é construída como segue. Dado uma partição $\rho = \{i_0, \dots, i_b\}$, para $b \in I$, temos que $\theta_i = \theta_{[i_{r-1}, i_r]}$, para todo $i_{r-1} < i \leq i_r$, $r = 1, \dots, b$, e que $\theta_{[i_0, i_1]}, \dots, \theta_{[i_{b-1}, i_b]}$ são independentes entre si e de p , com $\theta_{[ij]}$ tendo distribuição *a priori* $\pi_{[ij]}(\theta)$, $\theta \in \Theta_{[ij]}$, onde $\Theta_{[ij]}$ é o espaço paramétrico correspondente ao parâmetro comum, ou seja, $\theta_{[ij]} = \theta_{i+1} = \dots = \theta_j$, que indexa a densidade condicional de $X_{[ij]}$. Então, seguindo Barry e Hartigan (1992) nós definimos o modelo partição produto para coesões *a priori* de Yao como segue:

i) dado p , a distribuição a priori de ρ é a seguinte distribuição produto:

$$P(\rho = \{i_0, i_1, \dots, i_b\} | p) = p^{b-1} (1-p)^{n-b} \quad [2.2]$$

para toda partição $\{i_0, \dots, i_b\}$, satisfazendo $0 = i_0 < i_1 < \dots < i_b = n$.

ii) condicionalmente em $\rho = \{i_0, \dots, i_b\}$ e p , a seqüência X_1, \dots, X_n são independentes de p e tem a densidade conjunta dada por:

$$f(X_1, X_2, \dots, X_n | \rho = \{i_0, \dots, i_b\}) = \prod_{j=1}^b f_{i_{(j-1)}, i_j}(X_{i_{(j-1)}, i_j}) \quad [2.3]$$

onde $f_{[ij]}(X_{[ij]}) = \int_{\Theta_{[ij]}} f_{[ij]}(X_{[ij]} | \theta) \pi_{[ij]}(\theta) d\theta$ é chamada de fator dado.

Barry e Hartigan (1992) também mostram que a esperança a posteriori (ou estimativa produto) para θ_k , $k=1, \dots, n$, é dada por:

$$E(\theta_k | X_1, \dots, X_n) = \sum_{i=0}^{s-1} \sum_{j=0}^n r_{[ij]}^* E(\theta^s | X_{[ij]}), \quad s=1, \dots, n, \quad [2.4]$$

onde $r_{[ij]}^* = P([ij] \in \rho | X_0, \dots, X_n)$ denota a relevância a posteriori para o bloco $[ij]$.

Vamos assumir que p tem distribuição a posteriori Beta com parâmetros α e β . Como consequência dessa consideração, segue que as distribuições a posteriori de ρ e B , são respectivamente dadas por:

$$P(\rho = \{i_0, \dots, i_b\} | X_1, \dots, X_n) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + b - 1)\Gamma(n + \beta - b)}{\Gamma(n + \beta + \alpha - 1)} \prod_{j=1}^b f_{i_{(j-1)}, i_j}(X_{i_{(j-1)}, i_j}) \quad [2.5]$$

$$P(B = b | X_1, \dots, X_n) = \binom{n-1}{b-1} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + b - 1)\Gamma(n + \beta - b)}{\Gamma(n + \beta + \alpha - 1)} \prod_{j=1}^b f_{i_{(j-1)}, i_j}(X_{i_{(j-1)}, i_j}) \quad [2.6]$$

A distribuição a posteriori para p é dada por:

$$\pi(p | X_1, \dots, X_n) \propto \sum \prod_{j=1}^b f(X_{i_{j-1}}, i_j) p^{b+\alpha-2} (1-p)^{n-b+\beta-1} \quad [2.7]$$

onde a soma é sobre todas as partições de $\{1, \dots, n\}$ em b blocos com pontos finais em i_0, i_1, \dots, i_b , satisfazendo a condição $0 = i_0 < i_1 < \dots < i_b = n$.

3. Estimativas produto para a taxa de criminalidade θ

Assuma que, dados a taxa de criminalidade em cada mês, $\theta_1, \dots, \theta_n$, o número de crimes em cada mês, X_1, \dots, X_n , são independentes e tais que $X_k | \theta_k \sim \text{Poisson}(\theta_k)$, para $k=1, \dots, n$. Assuma também que, *a priori*, o parâmetro comum $\theta_{[ij]}$ que indexa a distribuição das observações do bloco $[ij]$, tem distribuição gama com parâmetros $\alpha > 0$ e $\beta > 0$, denotada por $\theta_{[ij]} \sim \text{Gama}(\alpha, \beta)$, com função densidade dada por:

$$f(\theta_{[ij]} | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_{[ij]})^{\alpha-1} \exp(-\beta \theta_{[ij]}).$$

Como conseqüência temos que a distribuição preditiva *a priori* de $X_{[ij]}$ é a seguinte:

$$f(X_{[ij]}) = \prod_{k=i+1}^j \frac{1}{X_k!} \frac{\Gamma(\alpha^*)}{\Gamma(\alpha)} \left(\frac{\beta}{\beta^*} \right)^\alpha \left(\frac{1}{\beta^*} \right)^{\sum_{k=i+1}^j X_k}$$

onde

$$\begin{cases} \beta^* = \beta + j - 1, \\ \alpha^* = \alpha + \sum_{k=i+1}^j X_k \end{cases}$$

para todo $i=0, \dots, n-1$ e $j=i+1, \dots, n$, e que, dado $X_{[ij]}$, a distribuição condicional de $\theta_{[ij]}$ é a distribuição gama com parâmetro β^* e α^* , isto é, $\theta_{[ij]} | X_{[ij]} \sim \text{Gama}(\alpha^*, \beta^*)$.

Conseqüentemente, as estimativas *a posteriori* ou *estimativas produto* para a taxa de criminalidade no mês k , θ_k , para todo $k=1, \dots, n$ são dadas por :

$$\hat{\theta}_k = E(\theta_k | X_1, \dots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^n \frac{\alpha^*}{\beta^*} r^*([ij]).$$

[3.1]

4. Métodos Computacionais

Como descrito em Loschi (1998) e Loschi e Cruz (2003) a distribuição a posteriori de ρ e, conseqüentemente, a distribuição a posteriori de p , B e θ_l são facilmente obtidas levando em conta a seguinte transformação sugerida por Barry e Hartigan (1993).

Considere a quantidade aleatória auxiliar U_l , de forma que $U_l = 1$, se $\theta_l = \theta_{l+1}$ e $U_l = 1$, se $\theta_l \neq \theta_{l+1}$ para $l = 1, \dots, n-1$. Note que a partição aleatória ρ é imediatamente identificada ao considerarmos vetores $U = (U_1, \dots, U_{n-1})$ dessas quantidades aleatórias.

Cada partição $(U_1^s, \dots, U_{n-1}^s)$, $s \geq 1$, é gerada usando o amostrador de Gibbs como segue. Iniciando do valor inicial $(U_1^0, \dots, U_{n-1}^0)$, o r -ésimo elemento no passo s , U_r^s , é gerado considerando a seguinte razão:

$$R_r = \frac{f_{[xy]}(X_{[xy]}) \int_0^1 p^{b-2} (1-p)^{n-b+1} d\pi(p)}{f_{[xr]}(X_{[xr]}) f_{[ry]}(X_{[ry]}) \int_0^1 p^{b-1} (1-p)^{n-b} d\pi(p)},$$

[4.1]

para $r = 1, \dots, n-1$, e x e y como definido em Loschi, Cruz, Iglesias e Arellano Valle (2003). Então, cada amostra da distribuição a posteriori de p é gerada da distribuição seguinte:

$$p^s | X_1, \dots, X_n \sim p^{b^s-1} (1-p)^{n-b^s} \pi(p),$$

para $s \geq 1$, onde b^s é o número de blocos no s -ésimo vetor U e é obtido notando-se que o número de blocos em ρ é dado por:

$$B = 1 + \sum_{l=1}^{n-1} (1 - U_l)$$

Consequentemente, a distribuição a posteriori de B (ou da número de pontos de mudança $B-1$) é estimada por:

$$P(B = b | X_1, \dots, X_n) = \frac{\sum_{s=1}^T 1_{\{B^s = b\}}}{T},$$

onde T é número total de vetores U considerados para a análise, $1_{\{B\}}$ denota a função indicadora de B . O somatório presente no numerador representa o total de vetores U que estão indicando a formação de b blocos.

Loschi, Cruz, Iglesias e Arellano-Valle (2003) obtém as estimativas produto computando a relevância a posteriori do bloco $[i, j]$, $i, j = 1, \dots, n$, $i < j$, como segue:

$$r^*_{[ij]} = \frac{M}{T},$$

[4.2]

onde M é o número de vetores U para os quais é observado que $U_i = 0, U_{i+1} = \dots = U_{j-1} = 1$, e $U_j = 0$.

5. Aplicação

Aplicaremos o MPP aos dados referentes ao número de crimes violentos notificados mês a mês no período de janeiro de 1998 a setembro de 2001 nas 5ª, 6ª e 127ª Cia's da Polícia Militar da cidade de Belo Horizonte. A 5ª Cia é responsável pelo registro dos crimes cometidos nos Bairros Santo Agostinho e Barro Preto – região comercial próxima ao centro da cidade, o centro da cidade tem os crimes registrados pela 6ª Cia, a 127ª Cia é responsável pelo registro dos crimes cometidos em alguns dos bairros nobres da cidade (Mangabeiras, Cruzeiro, etc.).

Para estimar as relevâncias *a posteriori* $r^*_{[ij]}$, as distribuições *a posteriori* de ρ , B e p , foram geradas 4600 amostras de valores 0-1 com a dimensão da série (45), começando de uma seqüência de zeros. As primeiras 100 iterações foram descartadas para *burn-in* e um lag de um foi selecionado.

Assumimos que a distribuição *a priori* para p (probabilidade de ocorrer uma mudança em um instante qualquer) é uma distribuição Beta com parâmetros $\alpha=1,5$ e $\beta=28,5$. Essa distribuição reflete que, *a priori*, achamos que a probabilidade de ocorrer uma mudança em um instante qualquer é pequena (média = 0,05 e moda = 0,0178) e temos uma grande certeza sobre isso

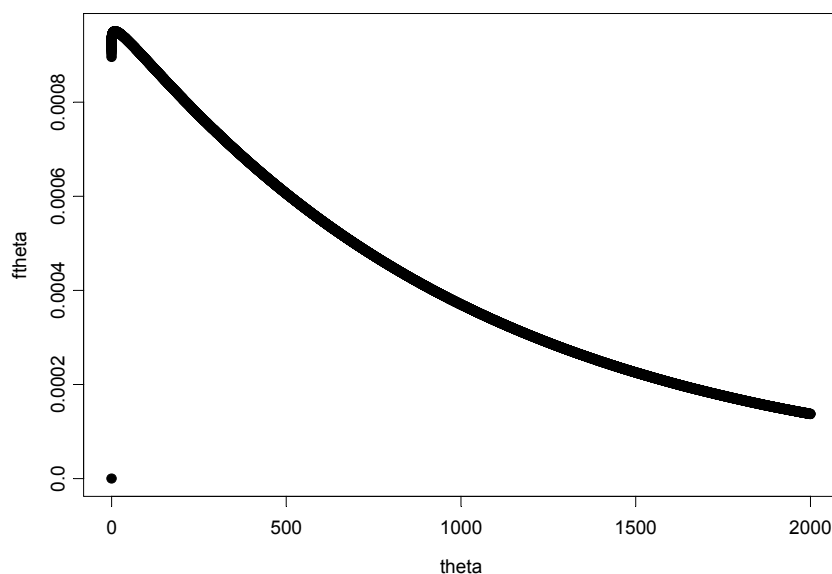
(variância = 0,0015). Conseqüentemente, estamos assumindo que, *a priori*, o número esperado de blocos é 3,2 e a variância é 4,989. Isso significa dizer que, *a priori*, se considerarmos a função de perda quadrática, estimamos que o número de blocos é pequeno (em torno de 3) e conseqüentemente o número de pontos de mudança esperado na série é dois. Também estamos assumindo que é mais provável que nenhuma mudança seja observada (moda de B é 1).

Assumiremos que, dado a taxa de criminalidade θ , o número de crimes cometidos nestas regiões possuem uma distribuição de Poisson e que θ possui uma distribuição Gama com parâmetros $\alpha = 1,01$ e $\beta = 0,001$. Pode-se notar da Tabela 5.1 e da Figura 5.1 que esta distribuição é pouco informativa a respeito da taxa de criminalidade em cada região, uma vez que a variância da distribuição é muito grande. Considerando a função de perda quadrática (Bernardo e Smith, 1994), a estimativa *a priori* para a taxa de criminalidade em cada região é 1010 crimes por mês.

Tabela 5.1 – Distribuição *a priori* da taxa de criminalidade e suas respectivas estatísticas descritivas.

Distribuição de θ	$E(\theta)$	$Var(\theta)$	Moda(θ)
Gama(1,01 ; 0,001)	1010	1010000	10

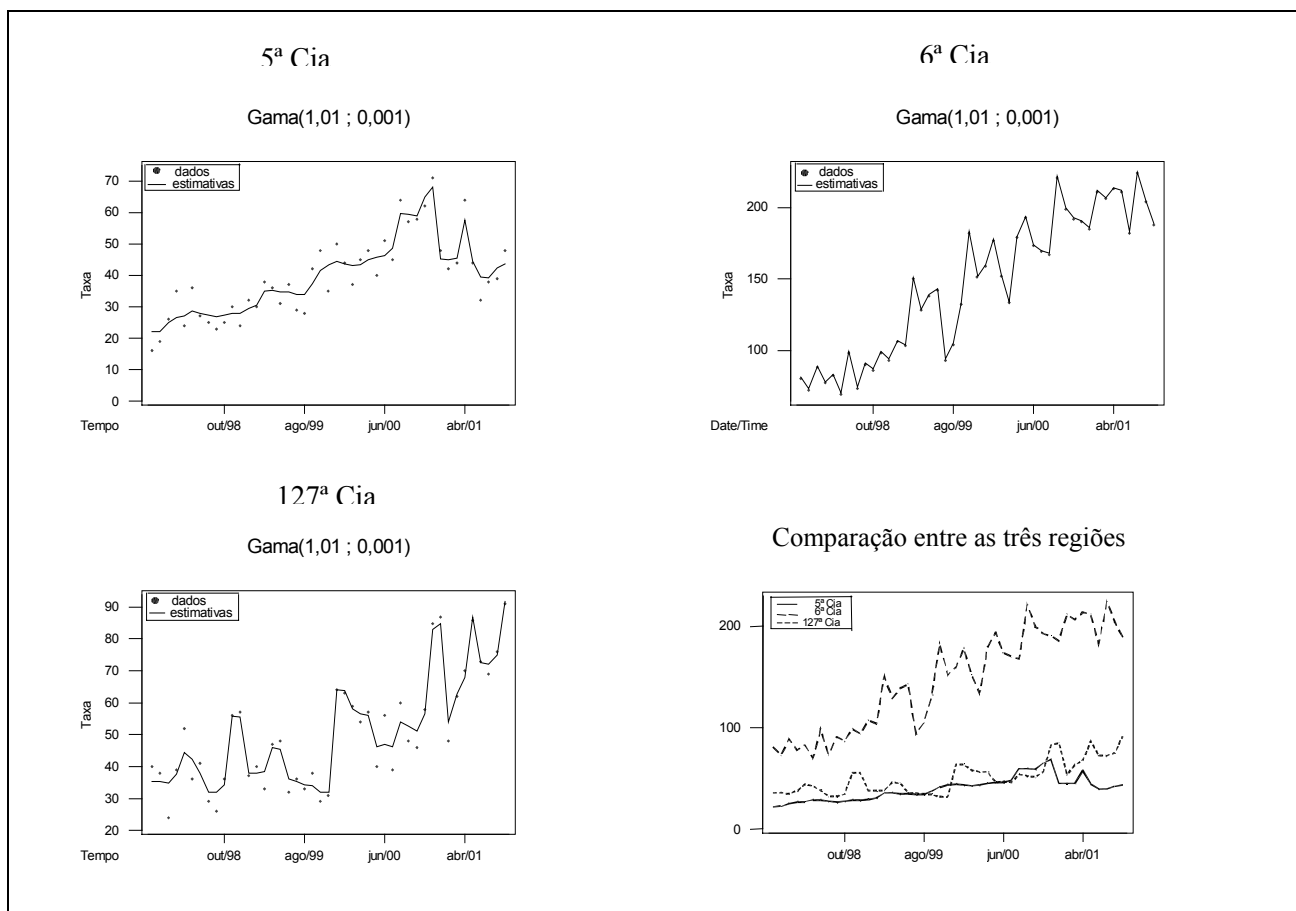
Figura 5.1 Função de densidade da distribuição *a priori* da taxa de criminalidade



A Figura 5.2 apresenta um gráfico conjunto das estimativas produto para a taxa de criminalidade em cada região (linha cheia) e os dados (pontos) bem como uma comparação entre as estimativas. Nota-se que as estimativas produto para θ na 5ª Cia apresentam comportamento crescente no período de janeiro de 1998 a dezembro de 2000, atingindo seu máximo neste último mês. Nos períodos de junho de 98 a fevereiro de 99, março a agosto de 99 e de novembro de 99 a julho de 2000, a taxa de criminalidade apresenta crescimento menos acentuado, sendo que, no período de junho de 98 a fevereiro de 99 a taxa é de aproximadamente

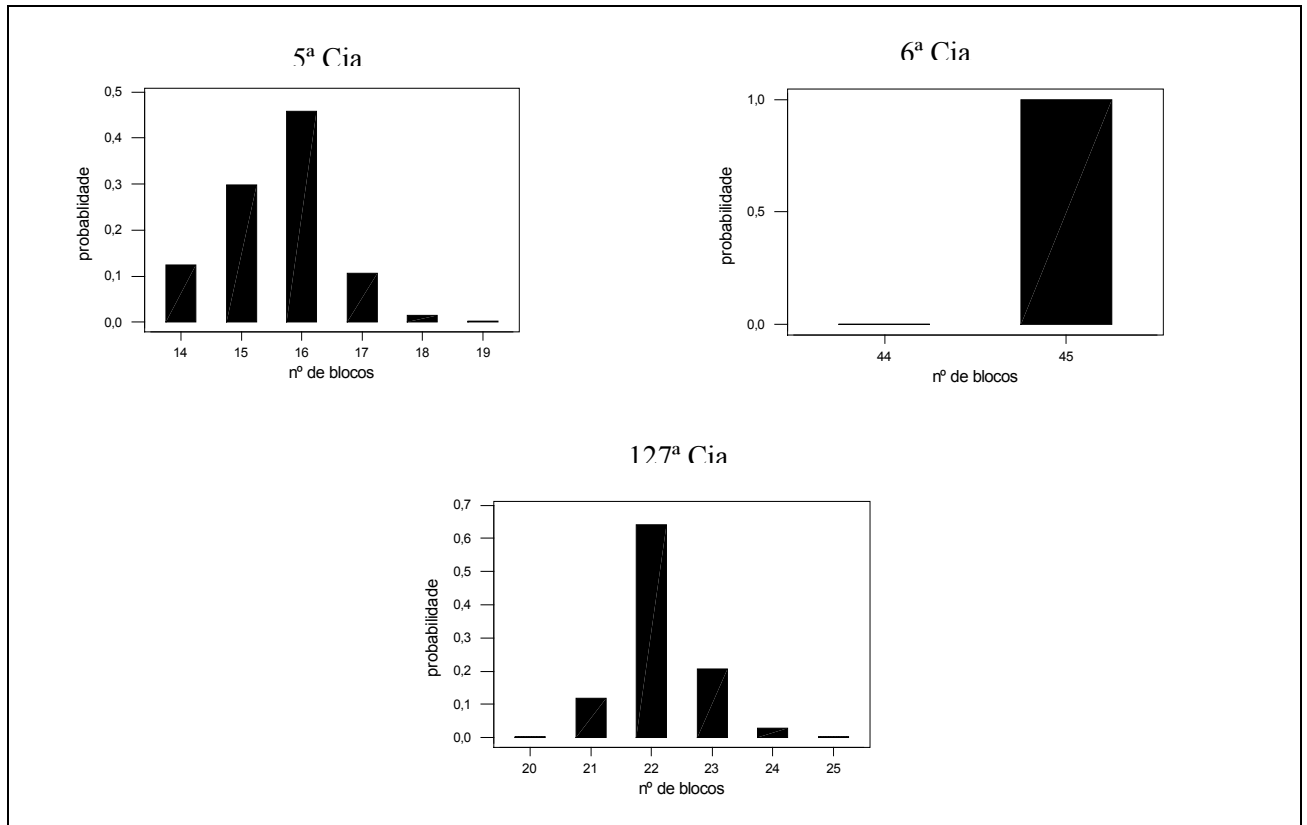
28 crimes por mês, no período de março a agosto de 99 esta taxa é de aproximadamente 35 crimes por mês e no período de novembro de 99 a julho de 2000 a taxa é de aproximadamente 45 crimes por mês. Mudanças importantes na taxa podem ser observadas em fevereiro e agosto de 99, julho, outubro e dezembro de 2000 e março e abril de 2001. A partir de dezembro de 2000 a taxa de criminalidade começa a diminuir. Após dezembro de 2000 a maior taxa é observada em abril de 2001. Para os dados da 6ª Cia, as estimativas produto para a taxa de criminalidade apresentam comportamento crescente no período de janeiro de 1998 a setembro de 2000, atingindo seu máximo em julho de 2001. Pode-se perceber que a taxa de criminalidade apresenta muitas mudanças em seu comportamento ao longo do tempo e não apresenta nenhum período com taxa constante. Mudanças importantes na taxa podem ser observadas em fevereiro, junho, agosto e setembro de 99, janeiro, março, agosto e setembro de 2000 e janeiro, maio, junho e julho de 2001. As estimativas produto para θ na 127ª Cia apresentam comportamento crescente no período de novembro de 1999 a setembro de 2001, atingindo seu máximo neste último mês. Nos períodos de janeiro a abril de 98 e junho a novembro de 99, a taxa de criminalidade apresenta crescimento menos acentuado, sendo que, no período de janeiro a abril de 98 a taxa é de aproximadamente 35 crimes por mês, no período de junho a novembro de 99 esta taxa é de aproximadamente 33 crimes por mês. Mudanças importantes na taxa podem ser observadas em outubro e dezembro de 98, novembro de 99, novembro de 2000 e abril e agosto de 2001. De janeiro de 98 a novembro de 99 a taxa de criminalidade tem um pico em novembro e dezembro de 98 com valores em torno de 55 crimes por mês.

Figura 5.2 Estimativas produto para a taxa de criminalidade em cada região



Note também da Figura 5.2 que, ao compararmos o comportamento das estimativas da taxa de criminalidade nas três regiões, percebemos que a região com maior taxa de criminalidade foi o centro (6ª Cia), seguido pela região nobre (127ª Cia) e finalmente a região comercial (5ª Cia).

Figura 5.3 Distribuições *a posteriori* para o número de Blocos



Perceba da Figura 5.3 que as distribuições *a posteriori* para o número de blocos (ou número de pontos de mudança $B-1$) nas três regiões são unimodais e tendem a serem assimétricas. Para a 5ª Cia observa-se que esta distribuição concentra maior massa entre 14 e 17 blocos, sendo que a moda é 16 blocos e ocorre com probabilidade 45,80%. Para a 6ª Cia percebe-se que a distribuição de B concentra praticamente toda sua massa 45 blocos, o que ocorre com probabilidade 99,98%. A distribuição *a posteriori* de B para dados da 127ª Cia concentra maior massa entre 21 e 23 blocos, sendo que o valor modal é 22 blocos e ocorre com probabilidade 64,20%. Comparando as distribuições *a posteriori* de B nas três regiões, confirmamos o que foi dito na análise das estimativas de θ , ou seja, o centro da cidade é a região que apresenta maior mudanças na taxa, seguido pela região nobre e depois da região comercial. Repare que o centro da cidade apresenta 45 blocos com probabilidade praticamente 1, ou seja a taxa sofreu mudanças em todos os meses da série. Isto também pode ser observado da Tabela 5.2 a seguir. Note que considerando a função de perda quadrática, a priori, esperávamos que existissem 2,2 pontos de mudança. *A posteriori* as estimativas para o número de mudanças foram bem maiores nas três regiões - 14,59, 43,99 e 21,15 para as 5ª, 6ª e 127ª Cias, respectivamente.

Tabela 5.2 – Estatísticas descritivas das distribuições *a priori* e *a posteriori* de B.

Região	Distribuição <i>a priori</i> de B			Distribuição <i>a posteriori</i> de B					
	Média	Moda	Var	Média	Moda	Var	Med	Q1	Q3
5ª Cia	3,2	1	4,989	15,59	16	0,79	16	15	16
6ª Cia	3,2	1	4,989	44,99	45	4.10^{-5}	45	45	45
127ª Cia	3,2	1	4,989	22,15	22	0,44	22	22	22

A Tabela 5.3 apresenta a partição mais provável *a posteriori* com a sua probabilidade de ocorrência *a priori* e *a posteriori* para cada uma das regiões.

 Tabela 5.3 – Probabilidades *a priori* e *a posteriori* de p .

Região	Partição	prob. <i>priori</i>	prob. <i>posteriori</i>
5ª Cia	{0,4,9,14,18,22,25,28,31,33,35,36,39,42,45}	1.10^{-14}	2.10^{-2}
6ª Cia	{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45}	3.10^{-20}	0,999778
127ª Cia	{0,4,7,10,12,15,17,21,23,25,26,28,31,33,35,36,37,38,40,41,43,44,45}	2.10^{-18}	0,011333

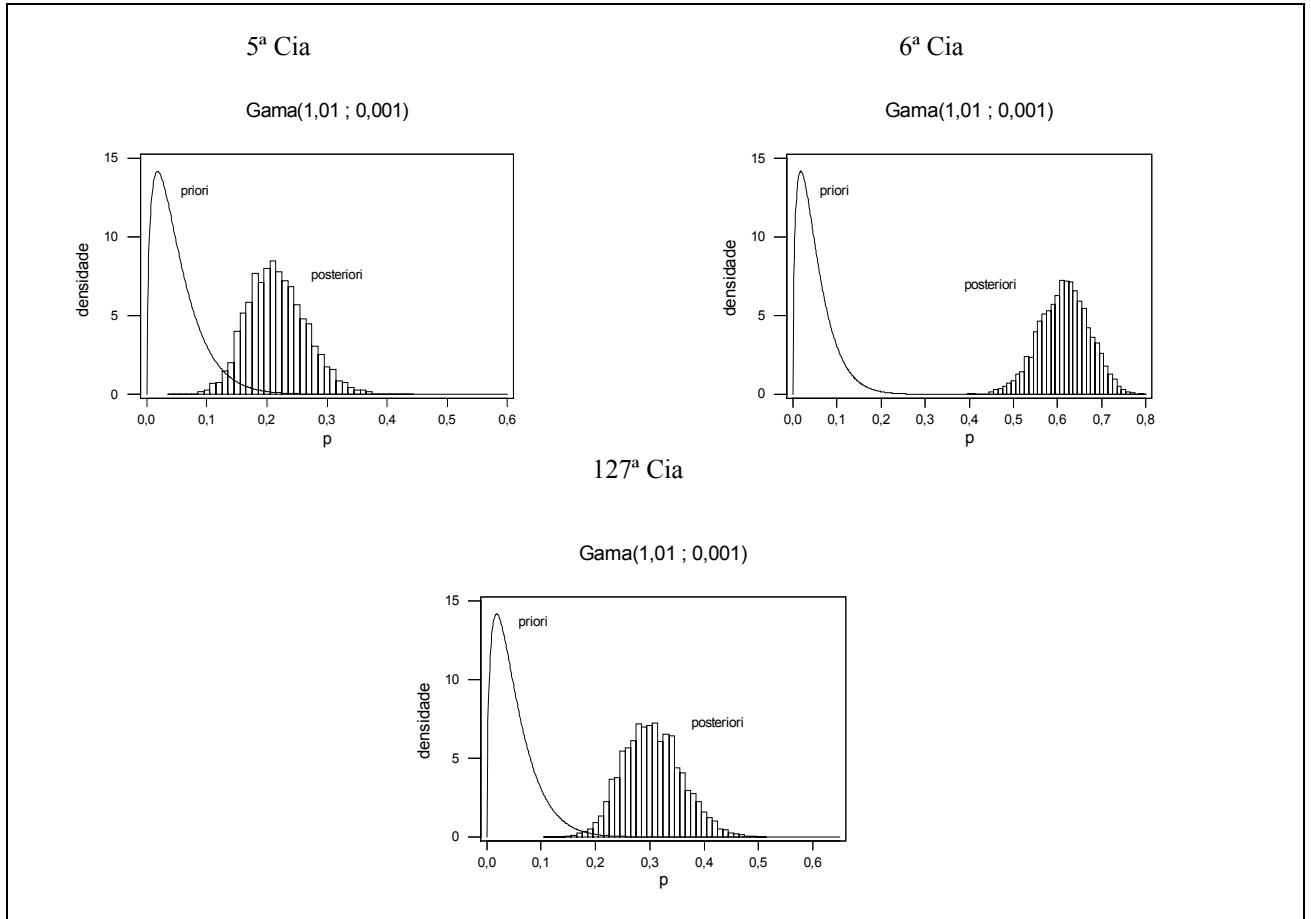
Percebe-se que as probabilidades *a priori* são bem pequenas, pois a distribuição *a priori* de p depende apenas do tamanho da série e do número de blocos na partição e a distribuição *a priori* de B põe maior massa em valores baixos de b , indicando que partições com muitos blocos são pouco prováveis. A probabilidade *a posteriori* da partição mais provável para a 5ª Cia é também pequena, mas não tanto quanto a probabilidade *a priori*. Já as probabilidades *a posteriori* das outras duas partições são bem maiores. A partição mais provável *a posteriori* revela que a taxa de criminalidade na região comercial (5ª Cia) sofreu mudanças em abril e setembro de 98, fevereiro, junho e outubro de 99, janeiro, abril, julho, setembro, novembro e dezembro de 2000, março e junho de 2001 com probabilidade 2%; que a taxa de criminalidade no centro da cidade (6ª Cia) sofreu mudanças em todos os meses da série com probabilidade 99%; e que a taxa de criminalidade na região nobre (127ª Cia) sofre mudanças em abril, julho, outubro e dezembro de 98, março, maio, setembro e novembro de 99, janeiro, fevereiro, abril, julho, setembro, novembro e dezembro de 2000, janeiro, fevereiro, abril, maio, julho e agosto de 2001 com probabilidade 1%.

Perceba da Figura 5.4 que tanto a distribuição *a priori* como as distribuição *a posteriori* de p das três regiões são unimodais e assimétricas. No entanto, a distribuição *a priori* concentra maior massa em valores menores que as distribuições *a posteriori*. Da Tabela 1.3, nota-se que, considerando a função de perda quadrática, a estimativa *a priori* para p é 0,05. *A posteriori* este valor é 0,21 para a região comercial (5ª Cia); 0,61 para o centro (6ª Cia); e 0,30 para a região nobre (127ª Cia). Ou seja, probabilidade de ocorrer uma mudança em um instante qualquer é maior do que esperávamos nas três regiões e, além disso, essa probabilidade é maior no centro, seguido da região nobre e depois pela região comercial. Além disso, observamos que há maior incerteza sobre p *a posteriori* nas três regiões pois as variâncias das distribuições *a posteriori* são 0,002 0,003 e 0,003 para a região comercial, o centro e a região nobre, respectivamente.

 Tabela 5.4 Estatísticas descritivas das distribuições *a priori* e *a posteriori* de p

Região	Distribuição <i>a priori</i> de p			Distribuição <i>a posteriori</i> de p					
	Média	Moda	Var	Média	Moda	Var	Med	Q1	Q3
5ª Cia	0,05	0,017	0,001	0,21	0,24	0,002	0,21	0,18	0,24
6ª Cia	0,05	0,017	0,001	0,61	0,61	0,003	0,61	0,57	0,65
127ª Cia	0,05	0,017	0,001	0,30	0,28	0,003	0,30	0,26	0,34

Figura 5.4 Distribuições *a priori* e *a posteriori* para p



Conclusões

O modelo partição produto (Barry e Hartigan, 1992) para coesões de Yao foi definido e aplicado para a identificação de múltiplos pontos de mudança na taxa de criminalidade de algumas regiões de Belo Horizonte-MG. Extensões do modelo partição produto (Loschi e Cruz, 2003) e alguns métodos computacionais (Loschi *et al.*, 1999, Loschi, Cruz, Iglesias and Arellano-Valle, 2003) propostos na literatura foram considerados. Vimos que a taxa de criminalidade nas regiões analisadas apresenta, em geral, tendência ao crescimento. Como era previsto, a região central apresenta maior taxa de criminalidade. Também observamos que, para todas as regiões, a taxa de criminalidade apresenta mudanças ao longo do tempo e vimos que esta taxa apresenta menos mudanças na região comercial próxima ao centro.

Concluimos que o modelo partição produto apresentou bom desempenho para a análise de pontos de mudança em dados de criminalidade.

Agradecimentos: R.H.Loschi recebe apoio financeiro da PRPq-UFMG (fundos 3893-UFMG/RTR/PRPq/FUNDO/98 e 4801-UFMG/ RTR/ FUNDO/ PRPq/ RECEM-DOUTORES/00).F.R.B. da Cruz, agradece à FAPEMIG, CNPq e PRPq-UFMG pelo apoio financeiro à sua pesquisa. Os dados foram cedidos pela Polícia Militar do Estado de Minas Gerais através do CRISP –UFMG.

Referências Bibliográficas

- BARRY, D., AND HARTIGAN, J. A. Product partition models for change point problems. *The Annals of Statistics* 20, 1(1992), 260-279.
- BARRY, D., AND HARTIGAN, J. A. A Bayesian analysis for change point problem. *Journal of the American Statistical Association* 88 , 421 (1993), 309 – 319.
- BERNARDO, J.M. AND SMITH, A. F. M. *Bayesian Theory*. John Wiley e Sons, New York, U.S.A. 1994.
- CROWLEY, E. M. Product partition models for normal means. *Journal of the American Statistical Association* 92. 437 (1997). 192 – 198.
- GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London, U.K., 1997.
- LOSCHI, R. H. *Imprevistos e suas Conseqüências* . Tese de Doutorado, Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil , 1998.
- LOSCHI, R. H. , AND CRUZ, F. R. B. An analysis of the influence of some prior specifications in the identification of change points via product partition model. *Computational Statistics & Data Analysis* 39, 4 (2002) , 477 – 501.
- LOSCHI, R. H. , CRUZ, F.R.B., IGLESIAS, P. L., AND ARELLANO – VALLE, R. B. A Gibbs sampling scheme to the product partition model: An application to change point problems. *Computers and Operations Research* 30, 3 (2003),463-482.
- LOSCHI, R. H. , AND CRUZ, F. R. B. Extension to the product partition model: Computing the probability of a change.(2003) (submetido para publicação).
- O'HAGAN, A . Bayesian Inference. *Kendall's Advanced Theory of Statistics 2A*. John Wiley eSons, New York, USA, 1994.
- QUINTANA, F. A. AND IGLESIAS, P.L. Nonpaametric Bayesian clustering and product partition model. *Journal of the Royal Statistical Society B*. 2003. (a ser publicado).
- YAO, Y. estimation of a noisy discrete-time step function:Bayes and empirical Bayes approaches. *The Annals of Statistics* 12, 4(1984), 1434-1447.