

AVALIAÇÃO DA MEDIDA DE EVIDÊNCIA DE UMA MUDANÇA EM SÉRIES DE DADOS COM DISTRIBUIÇÃO POISSON

Flávio Bambirra Gonçalves, 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: 31270-901

E-mail: fbg@ufmg.br e fcruz@est.ufmg.br

Resumo: Neste trabalho avalia-se a medida de evidência de uma mudança ter ocorrido em cada instante proposta por Loschi e Cruz (2004). A avaliação desta medida é feita considerando diversas seqüências de dados, geradas independentemente de distribuições Poisson. Consideram-se seqüências ou cenários com e sem pontos de mudanças. Em cenários sem pontos de mudanças, consideram-se taxas pequenas e grandes para avaliar a eficiência da medida proposta. Em cenários com pontos de mudanças, consideram-se tanto cenários com mudanças estruturais quanto com observações atípicas. O objetivo dessa análise é avaliar o comportamento e a eficiência dessa medida na detecção de pontos de mudança. Concluiu-se que a medida proposta teve bom desempenho geral para estimar a evidência de mudança. E, além disso, percebeu-se que na maioria dos instantes em que ocorreu mudança, a probabilidade de tal instante ser um ponto de mudança foi maior que 0,50.

Palavras-chave: Modelo Partição Produto, Probabilidade de mudança, Distribuição de Poisson

Abstract: In this paper the evidence measured of a change occurring in each instant proposed by Loschi and Cruz (2004) is evaluated. It is done by considering several data sequences, independently generated from Poisson distributions. Sequences or scenes with and without change points are considered. In no change scenes, small and high rates are considered to evaluate the efficiency of the proposed measure. In scenes with change points, scenes with structural changes and with atypical observations are considered. The aim of this analysis is to evaluate the behavior and the measure's efficiency in change points detection. It was noticed that the proposed measure had a great performance on estimating the change evidence. Besides that, in great part of the instants where a change occurred, the probability of such instant being a change point was greater than 0,50.

Keywords: Product Partition Model, Probability of change, Poisson Distribution.

1. Introdução

Ao estudar problemas de pontos de mudanças utilizando o Modelo Partição Produto (MPP), Loschi e Cruz (2004) alegam que a probabilidade *a posteriori* da partição aleatória que indica os instantes onde as mudanças ocorrem pode não fornecer uma idéia precisa, uma vez que várias partições podem ocorrer com mesma probabilidade *a posteriori*. Em situações onde isto ocorre, a tarefa de decidir sobre os instantes onde as mudanças ocorreram é árdua. Quintana e Iglesias (2003) propõem uma maneira de escolher a partição ótima usando técnicas de teoria de decisão. Loschi e Cruz (2004) propõem avaliar a evidência de uma mudança calculando a probabilidade *a posteriori* de cada instante ser um ponto de mudança.

Neste trabalho, será feito um estudo sobre a eficiência da medida proposta por Loschi e Cruz (2004). Serão consideradas seqüências de dados gerados independentemente de distribuições Poisson. Seqüências ou cenários contendo e não pontos de mudanças são considerados. Em cenários sem pontos de mudanças, consideram-se seqüências com diferentes variabilidades e em cenários com uma ou mais mudanças estudam-se tanto cenários com mudanças estruturais quanto cenários contendo apenas observações atípicas.

O MPP definido por Hartigan (1990) 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 relatório está assim organizado: Na Seção 2 apresenta-se o MPP introduzido por Barry e Hartigan (1993) assumindo as coesões de Yao (1984); na Seção 3 apresenta-se o estudo para avaliar o comportamento e a eficiência da medida proposta para detecção de pontos de mudança, na Seção 4 apresentam-se as principais conclusões.

2. Modelo Partição Produto

2.1. 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 $B = b$ subsequências 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 as 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 $b \in I$, tem-se 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 defini-se 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 $\theta_k, k=1, \dots, n$, é dada por:

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

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

Assumir-se 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:

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

A distribuição *a posteriori* para p é dada por:

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

[2.6]

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 $0 = i_0 < i_1 < \dots < i_b = n$.

A probabilidade *a posteriori* de cada instante i ser um ponto de mudança é:

$$P(A_i | X_1, \dots, X_n) = \sum_{C_i} \prod_{j=1}^n f(X_{i_{j-1}i_j}) \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + b - 1)\Gamma(n + \beta - b)}{\Gamma(n + \beta + \alpha - 1)} \quad [2.7]$$

onde C_i é o conjunto de todas as partições que incluem o i -ésimo instante como ponto de mudança.

2.2. Estimativas produto para a taxa θ

Assuma que, dada a taxa em cada instante k , $\theta_1, \dots, \theta_n$, as observações, 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:

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

Como consequência tem-se 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, $\theta_{[ij]} | X_{[ij]} \sim \text{Gama}(\alpha^*, \beta^*)$.

Consequentemente, as estimativas *a posteriori* ou *estimativas produto* para a taxa θ no instante k , θ_k , para todo $k=1, \dots, n$ são dadas por :

$$\hat{\theta}_k = E(\theta_k | X_1, \dots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^n \frac{\alpha^*}{\beta^*} r^*([ij]). \quad [2.8]$$

2.3. Métodos Computacionais

Considere a quantidade aleatória auxiliar U_l , de forma que $U_l = 1$, se $\theta_l = \theta_{l+1}$ e $U_l = 0$, se $\theta_l \neq \theta_{l+1}$ para $l = 1, \dots, n-1$. Note que a partição aleatória ρ é imediatamente identificada ao considerar 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:

$$R_r = \frac{f_{[xy]}(X_{[xy]}) \int_0^1 p^{b-2} (1-p)^{n-b+1} d\pi(p)}{f_{[xr]}(X_{[xr]}) f_{[ry]}(X_{[ry]}) \int_0^1 p^{b-1} (1-p)^{n-b} d\pi(p)}, \quad r = 1, \dots, n-1 \quad [2.9]$$

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:

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

As estimativas produto podem ser obtidas computando a relevância *a posteriori* do bloco $[i, j]$, $i, j = 1, \dots, n$, $i < j$, como segue:

$$r^*_{[ij]} = \frac{M}{T}, \quad [2.10]$$

E a estimativa da probabilidade *a posteriori* de cada instante i ser um ponto de mudança utilizando os métodos computacionais citados acima é:

$$P(A_i) = \frac{N}{T} \quad [2.11]$$

para $i = 1, \dots, n-1$, onde N é o número de vetores U para os quais é observado que $U_{i-1} = 0$, 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 $U_j = 0$ e T é o número total de vetores U gerados.

Ver detalhes em Loschi, Cruz, Iglesias e Arellano-Valle (2003) e Loschi e Cruz (2004).

3. Avaliação da medida de evidência de uma mudança

Serão simuladas séries de dados de uma distribuição de Poisson de tamanho 50 com nenhum, um, dois e quatro pontos de mudanças. Como distribuição *a priori* de p para os casos em que se tem nenhuma ou uma mudança será considerada uma distribuição Beta (2, 98), para os casos em que se tem duas mudanças considera-se distribuição Beta (2, 48), para os casos em que se tem quatro mudanças, distribuição Beta (2, 23) será considerada. Como consequência destas suposições, espera-se *a priori* que ocorra 1, 2 e 4 pontos de mudanças respectivamente. Assume-se distribuição *a priori* do parâmetro θ em todos os casos distribuição Gama(2; 1), que tem esperança e variância iguais a 2.

Serão apresentados três gráficos para a análise de cada cenário. No primeiro são plotados a taxa real (linha cheia), a taxa estimada (linha tracejada) e os dados (pontos). No segundo são

apresentadas as distribuições *a priori* e *a posteriori* de p , e no terceiro apresenta-se a partição mais provável (pontos) e a probabilidade de cada instante ser um ponto de mudança (linha cheia).

A figura 3.1 apresenta os resultados para cenários sem mudança quando $\theta = 0,5$, $\theta = 2$ e $\theta = 10$ respectivamente.

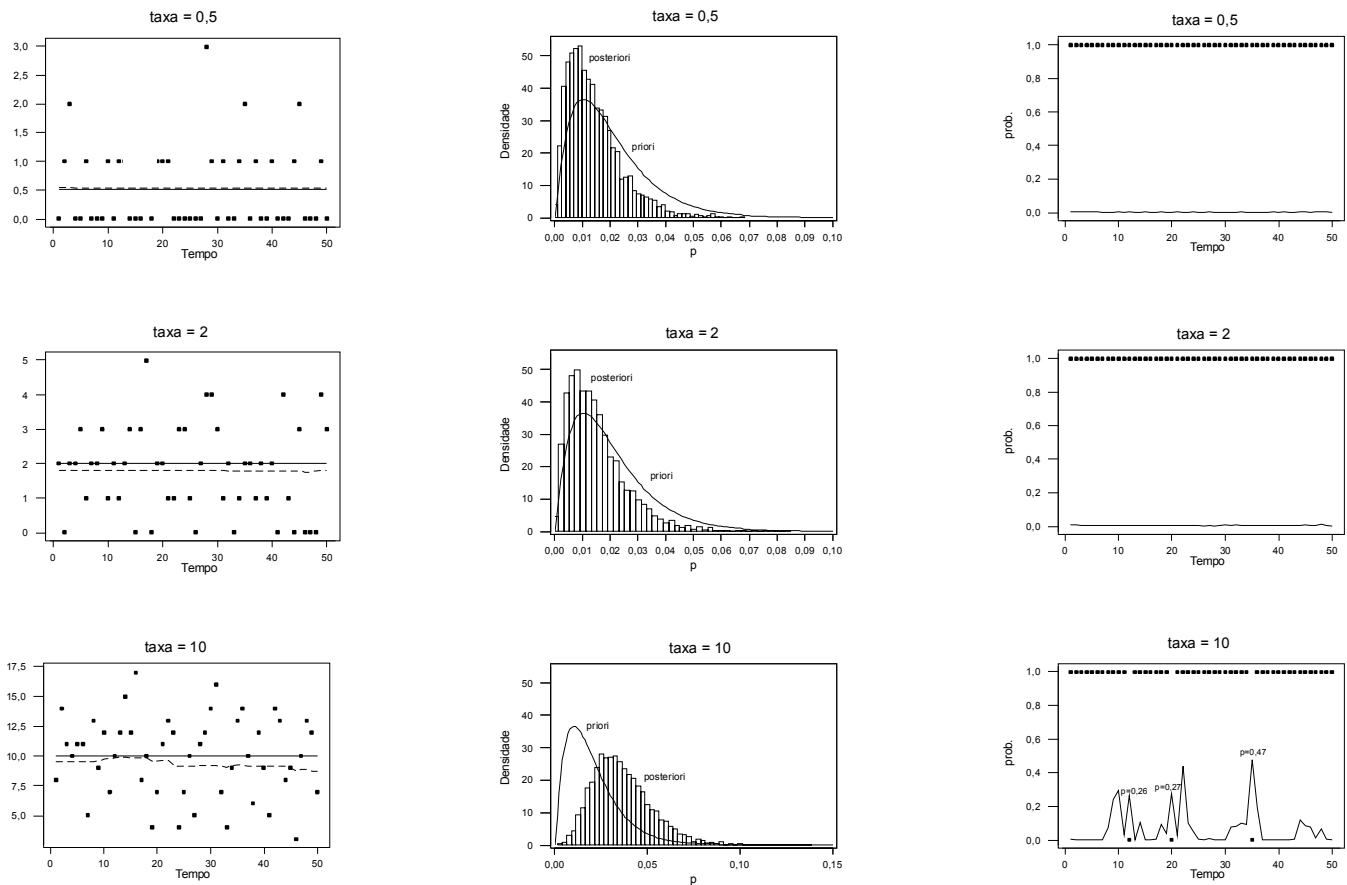


Figura 3.1. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário sem mudança.

Nos casos $\theta = 0,5$ e $\theta = 2$, nota-se que as probabilidades de cada instante ser um ponto de mudança são muito baixas em todos os instantes, evidenciando que θ realmente não sofre mudanças. Percebe-se também que as partições mais prováveis indicam que θ não sofre mudanças e tais partições têm alta probabilidade de ocorrência, 91% para $\theta = 0,5$ e 83% para $\theta = 2$.

No caso $\theta = 10$, a partição mais provável detectou três mudanças (instantes 12, 20 e 35), porém a probabilidade de ocorrência desta partição é de apenas 14,8%. Além disso, nos instantes em que a partição detectou mudanças, as probabilidades de tais instantes serem pontos de mudança não são altas (0,26 ; 0,28 ; 0,47). No instante 22, que não é um ponto de mudança, essa probabilidade é 45% e partição indica não ter havido mudança. Nos demais instantes, essa probabilidade é menor que 25%, indicando forte evidência que tais instantes não são pontos de mudança.

A figura 3.2 mostra o resultado obtido para três cenários contendo apenas uma mudança estrutural nos instantes 10, 25 e 40 respectivamente. Neste caso, as observações antes da mudança são geradas de uma distribuição Poisson com parâmetro $\theta = 1$ e as observações após a mudança são geradas de uma distribuição com taxa $\theta = 4$.

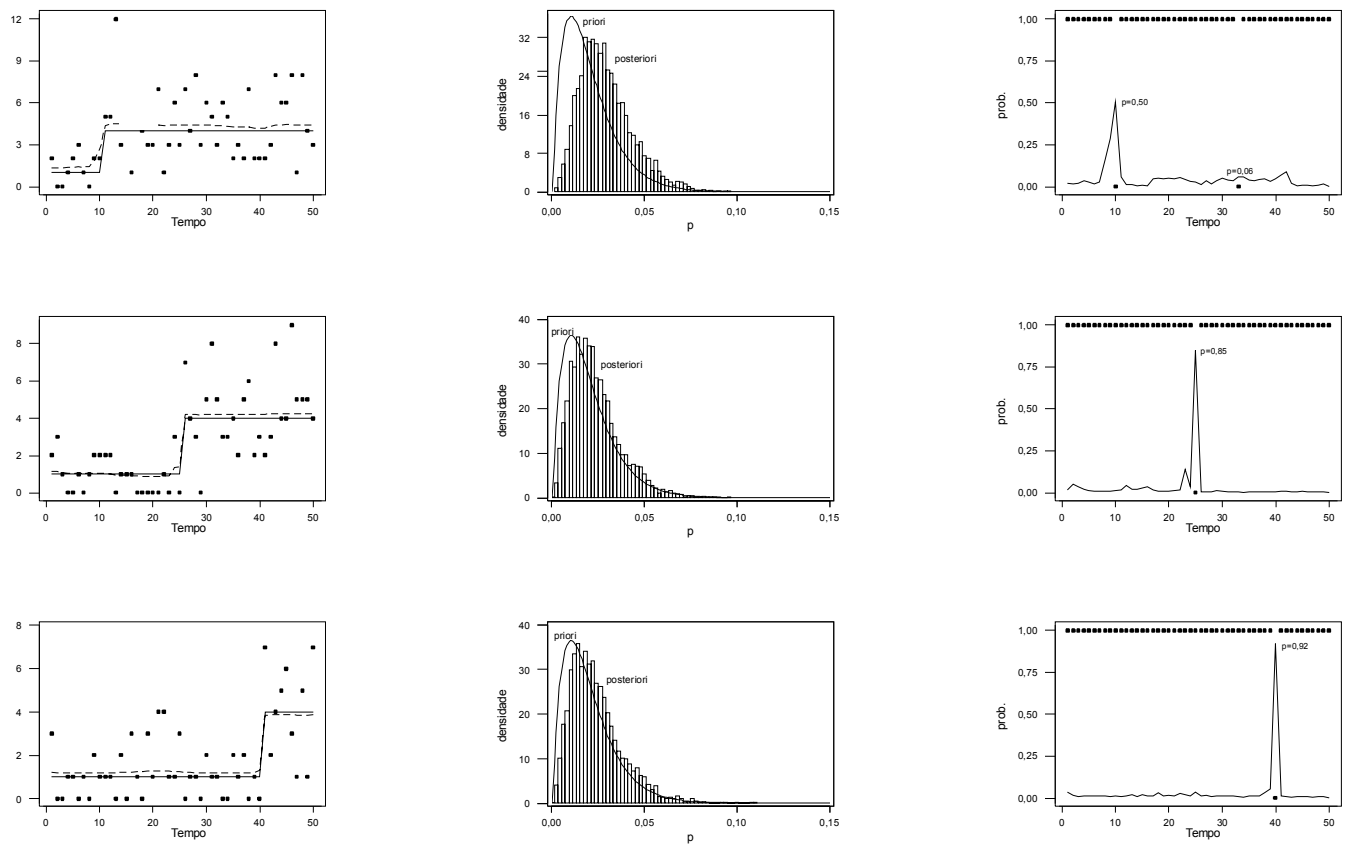


Figura 3.2. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com uma mudança.

No cenário em que θ sofre mudança no instante 10, a partição mais provável *a posteriori* indica erroneamente mudanças nos instantes 10 e 33. Porém tal partição tem probabilidade de ocorrência *a posteriori* muito baixa (0,02). No entanto, considerando a probabilidade *a posteriori* de cada instante ser ponto de mudança nota-se que o instante 10 é um ponto de mudança com probabilidade 50%, os demais instantes (inclusive o 33) tem probabilidade inferior 10% de ser ponto de mudança.

Para os cenários com mudança nos instantes 25 e 40 nota-se que a partição mais provável *a posteriori* indica corretamente os instantes de mudança e estas partições ocorrem com probabilidade 52% e 58% respectivamente. Nota-se também que para o cenário com mudança no instante 25, este instante tem probabilidade *a posteriori* de 85% de ser ponto de mudança e os demais instantes tem probabilidade *a posteriori* inferior a 15% de serem pontos de mudança. Similarmente, no cenário que apresenta mudança no instante 40, a probabilidade *a posteriori* de ele ser ponto de mudança é de 92% e os demais pontos apresentam probabilidade inferior a 5% de serem pontos de mudança. Ou seja, nestes dois últimos cenários a distribuição *a posteriori* de p e a probabilidade *a posteriori* de cada instante ser ponto de mudança conduzem à mesma decisão, porém no primeiro cenário, a probabilidade *a posteriori* de cada instante ser ponto de mudança fornece uma melhor informação.

Percebe-se que quando maior for a quantidade de observações antes das mudanças, mais eficientes na indicação de mudanças são os dois métodos considerados, a saber, a distribuição *a posteriori* da partição e a probabilidade *a posteriori* de cada instante ser ponto de mudança.

A figura 3.3 apresenta os resultados obtidos para cenários com apenas uma mudança estrutural nos instantes 10, 25 e 40, agora assumindo que após a mudança há mais variabilidade, ou seja, antes da mudança considera-se $\theta = 1$ e após a mudança considera-se $\theta = 10$.

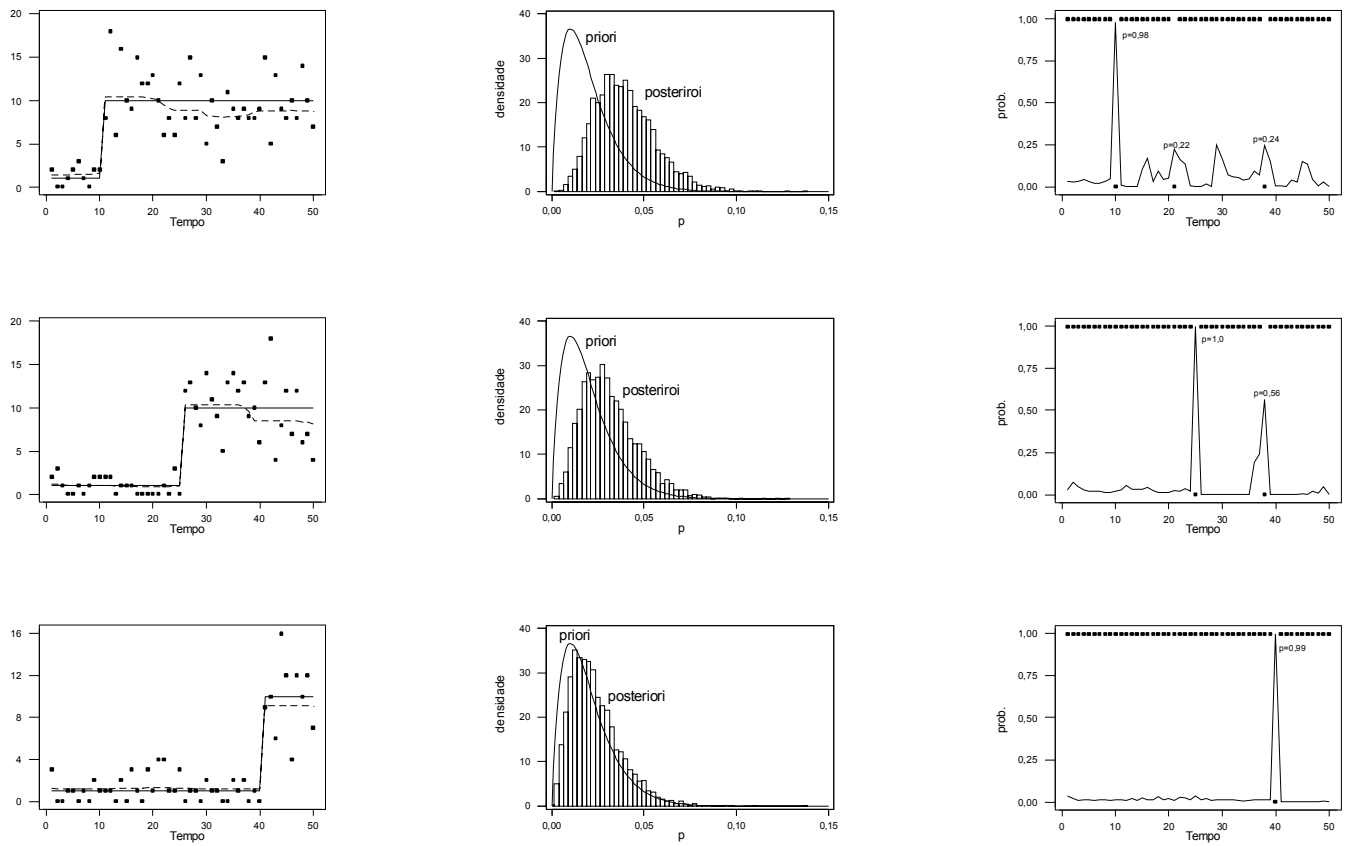


Figura 3.3. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com uma mudança.

No caso em que θ sofre mudança no instante 10, a partição detectou duas outras mudanças além do instante 10, porém, a probabilidade dessa partição é de apenas 9%. Já a probabilidade do instante 10 ser um ponto de mudança é bem alta (0,98) e dos outros dois instantes onde a partição detectou mudanças as probabilidades são bem mais baixas (0,22 e 0,24). Para os demais instantes tal probabilidade é menor 25%.

No caso em que θ sofre mudança no instante 25, a partição detectou uma outra mudança além do instante 25, porém, a probabilidade dessa partição não é alta (29%). A probabilidade do instante 25 ser um ponto de mudança é 100% e do outro instante é 56%. Nota-se que no instante em que a probabilidade é 0,56, a estimativa de θ diminui, isso pode ter ocorrido devido a maior variabilidade dos dados gerados com $\theta = 10$, que gerou dados menores que os anteriores.

No caso em θ sofre mudança no instante 40, a partição mais provável *a posteriori* (probabilidade 66%) detectou mudança apenas neste instante e a probabilidade deste instante ser ponto de mudança é 99%, nos demais instantes as probabilidades são menores que 5%.

Novamente, a partição e a probabilidade de mudança foram mais eficientes nos casos em que se têm mais observações antes da mudança.

A figura 3.4 mostra o resultado obtido para três cenários contendo duas mudanças estruturais nos instantes 10 e 20, 20 e 30, 30 e 40 respectivamente. Neste caso, as observações antes da primeira mudança são geradas de uma distribuição Poisson com parâmetro $\theta = 1$, as observações após a mudança são geradas de uma distribuição com taxa $\theta = 4$ e após a segunda mudança são geradas com $\theta = 1$.

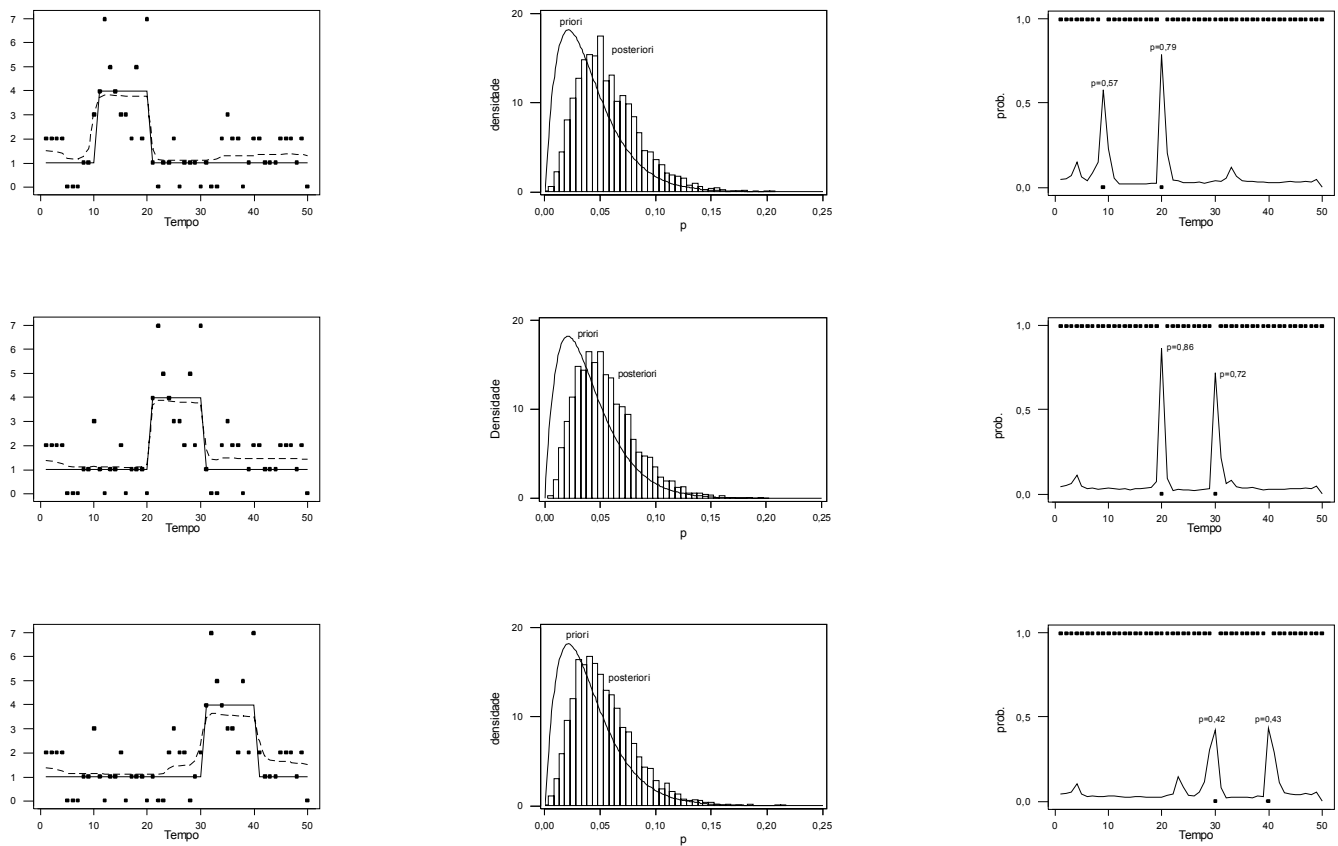


Figura 3.4. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com duas mudança.

No caso em que θ sofre mudança nos instantes 10 e 20, a partição mais provável (11%) detectou mudança nos instantes 9 e 20, onde as probabilidades de mudança são 57% e 79% respectivamente. No instante 10 essa probabilidade é 23% e nos demais instantes são menores que 15%.

Nos outros dois casos, a partição mais provável detectou mudanças apenas nos instantes em que essas realmente ocorrem com probabilidades *a posteriori* 19% e 6% respectivamente. No caso em que as mudanças ocorrem nos instantes 20 e 30, as probabilidades de mudança nesses instantes são 86% e 72% respectivamente, e menor que 0,15 nos demais instantes. No caso em que as mudanças ocorrem nos instantes 20 e 30, as probabilidades de mudança nesses instantes são 42% e 43% respectivamente, e menor que 15% nos demais instantes.

Nota-se que a partição e a probabilidade de mudança foram mais eficientes no caso em que se tem mudança nos instante 20 e 30, onde a quantidade de observações abaixo da primeira mudança e acima da segunda é a mesma.

A figura 3.5 mostra o resultado obtido para três cenários contendo um ponto atípico nos instantes 10, 25 e 40 respectivamente. Neste caso, o ponto atípico é gerado de uma distribuição Poisson com parâmetro $\theta = 10$ e as demais observações são geradas de uma distribuição com taxa $\theta = 1$.

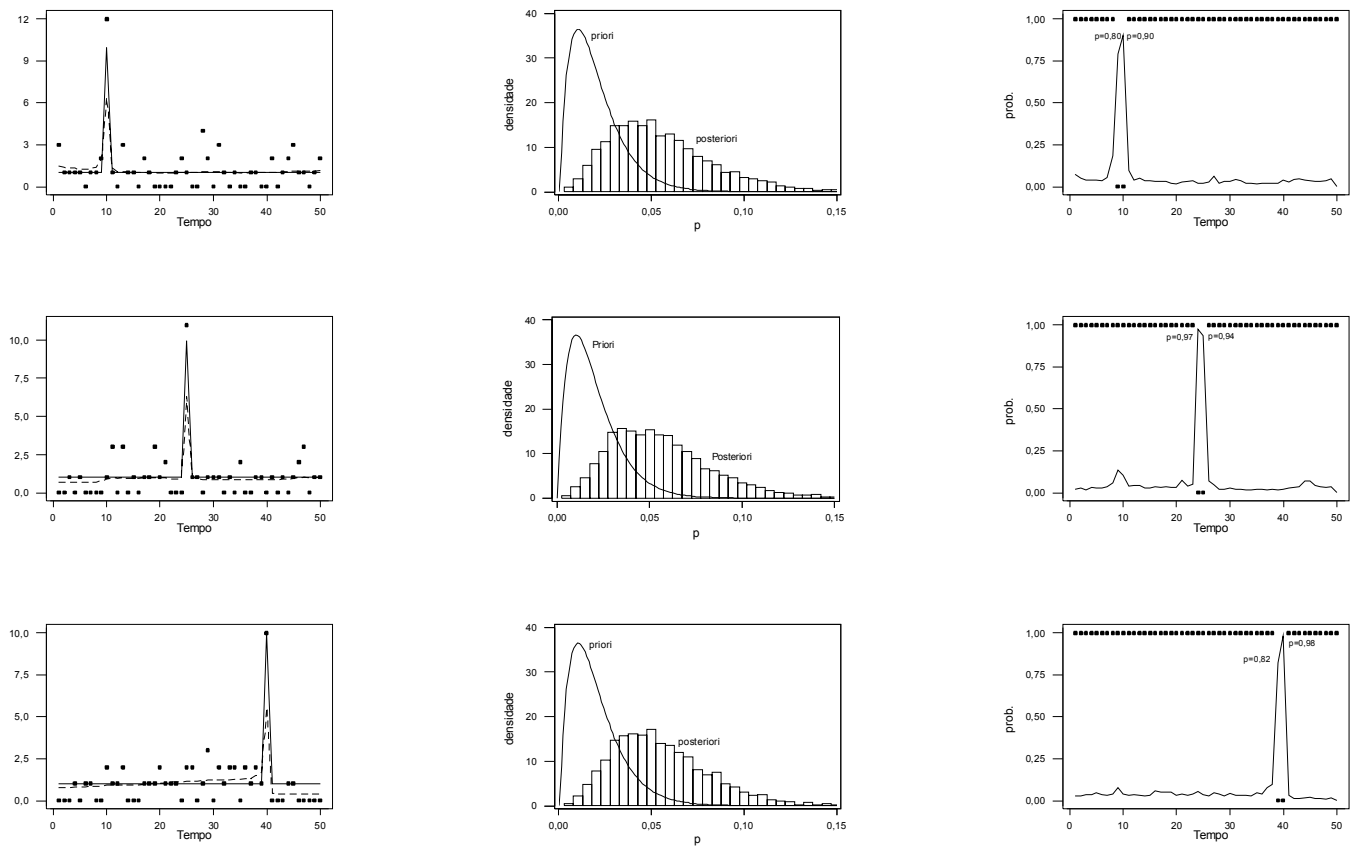
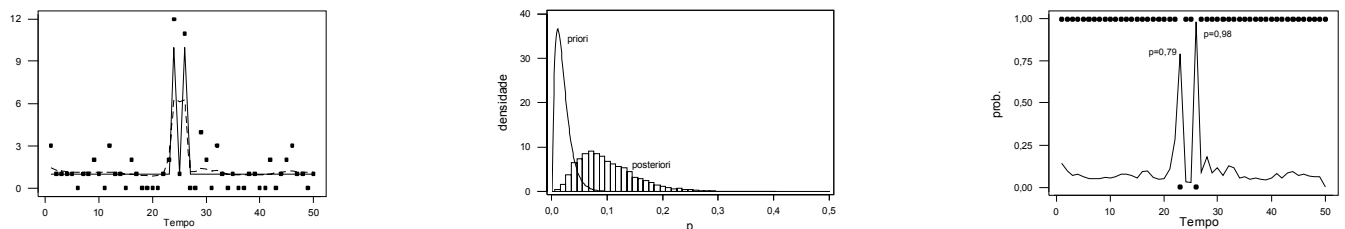


Figura 3.5. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com um ponto atípico.

Nos três casos em que se tem um ponto atípico com $\theta = 10$, a partição mais provável *a posteriori* detectou mudança nos exatos instantes em que essas ocorreram com probabilidades 25%, 27% e 18% e a probabilidade de mudança nesses instantes é sempre maior que 80%. Nos demais instantes essa probabilidade é sempre menor que 0,15.

Percebe-se que o modelo foi muito eficiente no caso em que se tem o ponto atípico.

A figura 3.6 mostra o resultado obtido para três cenários contendo dois pontos atípicos nos instantes 24 e 26, 23 e 27, 21 e 29 respectivamente. Os pontos atípicos são gerados de uma distribuição Poisson com parâmetro $\theta = 10$ e as demais observações são geradas de uma distribuição com taxa $\theta = 1$.



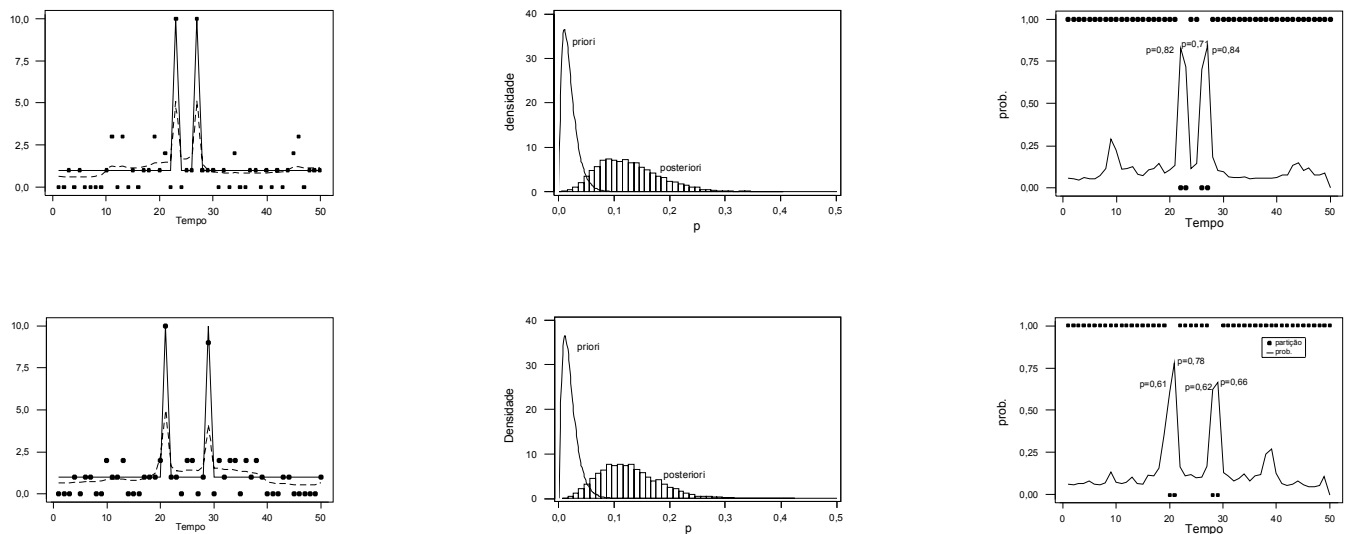


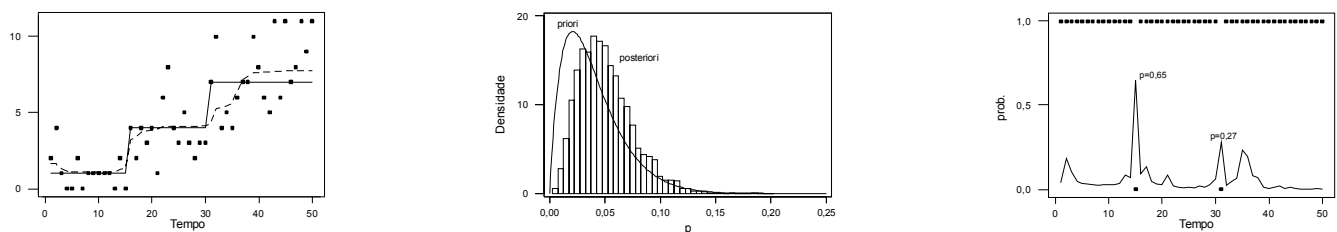
Figura 3.6. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com dois pontos atípicos.

No caso em que os pontos atípicos estão nos instantes 24 e 26, a partição mais provável detectou mudança apenas nos instantes 23 e 26 com probabilidade 9% e a probabilidade de mudança nesses instantes é de 79% e 98%. Nos instantes 24 e 25 onde também ocorrem mudanças, as probabilidades de mudança foram 3,1% e 2,7% respectivamente. Nos demais instantes a probabilidade de mudança é sempre menor que 20%. O modelo não conseguiu, então, detectar que θ é 1 no instante 25, pois essa é a única observação entre os dois pontos atípicos ($n = 24$ e 26).

Nos outros dois casos a partição mais provável detectou mudanças exatamente nos instantes em que elas ocorrem com probabilidade 1% e a probabilidade de mudança nesses instantes foi sempre maior que 71% no segundo caso e 61% no terceiro caso.

Observa-se então que são necessárias pelo menos três observações entre os pontos atípicos para se detectar todas as mudanças tanto com a partição como com a probabilidade de mudança em cada instante.

A figura 3.7 mostra o resultado obtido para três cenários contendo duas mudanças estruturais nos instantes 15 e 30. No primeiro cenário as observações antes da primeira mudança são geradas de uma distribuição Poisson com parâmetro $\theta = 1$, as observações após a mudança são geradas de uma distribuição com taxa $\theta = 4$ e após a segunda mudança são geradas com $\theta = 7$. No segundo cenário os valores de θ são 1, 4 e 10. No terceiro tais valores são 1, 6 e 12.



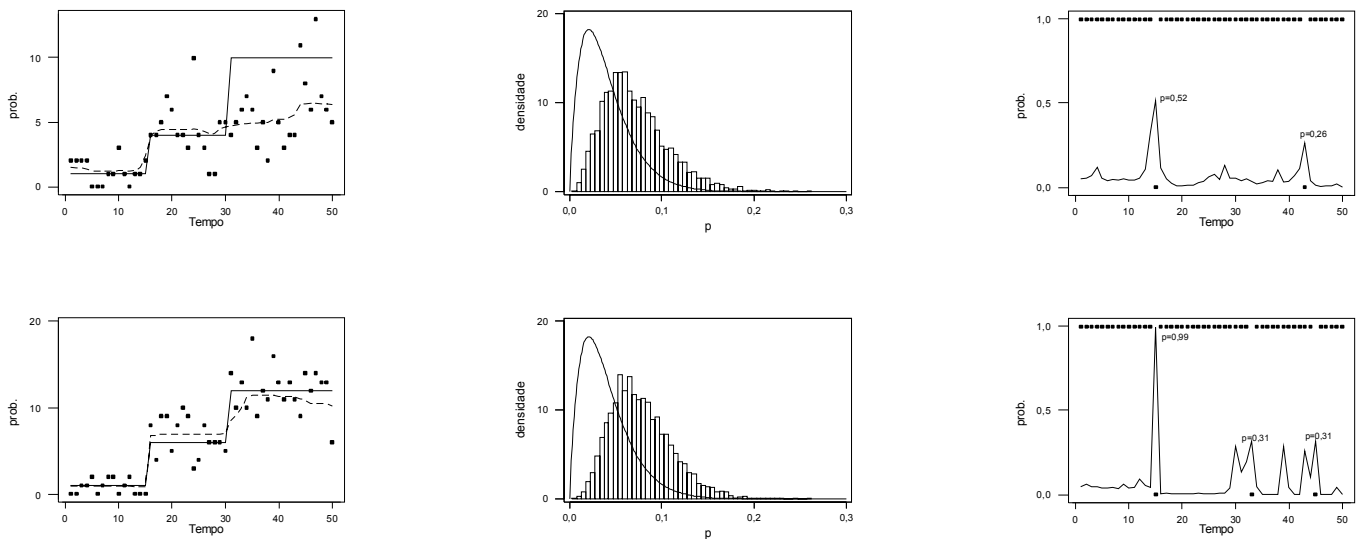


Figura 3.7. Estimativas produto de θ , distribuições *a posteriori* de p e ρ e probabilidade de mudança – cenário com duas mudanças.

No primeiro cenário a partição mais provável *a posteriori* detectou mudança nos instantes 15 e 31 com probabilidade 7%. A probabilidade de mudança nesses instantes é 65% e 27% respectivamente, no instante 30 é 6% e menor que 25% nos demais instantes. Nota-se que a partição detectou a segunda mudança com um pequeno atraso e a probabilidade de mudança neste instante não foi alto (0,27). Isso deve ter ocorrido pelo fato de as mudanças sofridas por θ terem sido pequenas (de 1 para 4 e de 4 para 7).

No segundo cenário as estimativas da partição (probabilidade 5%) e da probabilidade de mudança foram similares ao primeiro cenário. Porém o atraso na detecção da segunda mudança foi maior, já que se deu no instante 43. Já as probabilidades de mudança em cada instante foram bem parecida nos dois casos.

No terceiro cenário a partição detectou mudança nos instantes 15, 33 e 46 com probabilidade 13%. As probabilidade de mudança nesses instantes são 99% , 31% e 31% respectivamente e 29% no instante 30. Neste caso, a probabilidade da primeira mudança ocorre foi bem maior que nos outro dois casos, porém, além da segunda mudança ter sido detectada com atraso, foi detectada uma terceira mudança que não ocorreu na realidade.

Nota-se que para os casos em que θ sofre dois aumentos, a segunda mudança foi detectada com atraso e teve probabilidade de mudança relativamente baixa.

Observou-se através do estudo dos diferentes cenários simulados que a medida proposta (probabilidade de cada instante ser um ponto de mudança) teve um bom desempenho geral na detecção de pontos de mudança. Notou-se que na maioria dos instantes em que ocorreu mudança, a probabilidade de tal instante ser um ponto de mudança foi maior que 0,50.

4. Conclusão

Neste trabalho avaliou-se a medida de evidência de mudança proposta por Loschi e Cruz (2004) em uma série de dados com distribuição de Poisson. A avaliação dessa medida foi feita considerando seqüências de dados gerados independentemente.

Notou-se que a medida proposta teve bom desempenho geral para estimar a evidência de mudança. Percebeu-se que na maioria dos instantes em que ocorreu mudança, a probabilidade de tal instante ser um ponto de mudança foi maior que 0,50. Ou seja, pode-se dizer que 0,50 é um bom ponto de corte para decidir se um instante é ou não ponto de mudança ao avaliar-se a medida proposta.

Percebeu-se que, quanto maior for a quantidade de observações antes das mudanças, mais eficientes são a partição e a probabilidade de um instante ser ponto de mudança para detectá-las.

Observou-se que são necessárias pelo menos três observações entre dois pontos atípicos para se detectar todas as mudanças tanto com a partição como com a probabilidade de mudança em cada instante. Notou-se também que para os casos em que θ sofre dois aumentos, a segunda mudança foi detectada com atraso e teve probabilidade de mudança relativamente baixa.

Baseado nos resultados das simulações, recomenda-se o uso da medida proposta para detecção de pontos de mudança em séries de dados com distribuição de Poisson.

Agradecimentos

Flávio Bambirra Gonçalves é bolsista do programa Pibic-CNPq. A pesquisa de Rosângela Helena Loschi é parcialmente financiada pelo CNPq (300325/2003-7). Frederico R. B. da Cruz agradece ao CNPq (301809/96-8 e 201046/94-6) e FAPEMIG (CEX – 289/98 e CEX – 855/98).

Referências Bibliográficas

- [1] BARRY, D., AND HARTIGAN, J. A. A Bayesian analysis for change point problem. *Journal of the American Statistical Association* 88, 421 (1993), 309-319.
- [2] GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London, UK, 1997.
- [3] LOSCHI, R. H. *Imprevistos e suas Conseqüências (Unpredictabilities and Consequences)*. PhD thesis, Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil, 1998. (in Portuguese).
- [4] LOSCHI, R. H., AND CRUZ, F. R. B. An analysis of the influence of some prior specifications in the identification of the change points via product partition model. *Computational Statistics & Data Analysis* 39, 4 (2002), 477-501.
- [5] LOSCHI, R. H., AND CRUZ, F. R. B. Extension to the Product Partition Model: Computing the Probability of a Change. *Computational Statistics & Data Analysis*, (2004).
- [6] LOSCHI, R. H., IGLESIAS, P. L., AND ARELLANO-VALLE, R. B. Bayesian detection of the change points in the Chilean stock market. In *proceedings of the Section on Bayesian Statistical Science* (Baltimore, MD, USA, 1999), Annual Meeting of American Statistical Association, pp. 160-165.
- [7] GONÇALVES, F. B., LOSCHI, R. H., CRUZ, F. R. B. (2003) Análise Bayesiana da Taxa de criminalidade na região metropolitana de Belo Horizonte usando o modelo Partição Produto. XXXV Simpósio Brasileiro de Pesquisa Operacional, 715-725.