

Max Sousa de Lima

*Métodos Adaptativos para Detecção de Clusters no
Espaço-tempo*

Belo Horizonte - MG

Agosto / 2011

Max Sousa de Lima

*Métodos Adaptativos para detecção de Clusters no
Espaço-tempo*

Tese apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Estatística.

Orientador: Prof.Dr. Luiz Henrique Duczmal

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Belo Horizonte – MG

Agosto / 2011

*Às mulheres de minha vida,
A minha mãe pelo exemplo de vida que me representa.
A minha eterna namorada e esposa Milena pelos anos de amor e dedicação.
A minha filha Gabriela, razão da minha alegria e viver.*

Agradecimentos

- À Deus pela força, saúde e família que me concedeu.
- Aos meus pais pela educação que me deram e que mesmo na dificuldade sempre souberam cuidar de mim com muito amor e carinho.
- Aos meus irmãos que mesmo distante sempre me apoiaram nos momentos em que precisei.
- Ao meu orientador Luiz Duczmal, pelo incentivo, pela paciência, mas acima de tudo pela oportunidade de desenvolver com ele este trabalho.
- Aos meus fiéis escudeiros no doutorado, Rodrigo e Markus.
- Aos meus amigos Fabio, Carlito, Luciano e Ronaldo ceará, pela convivência e também por terem me suportado nos momentos mais difíceis.
- A todos os meus professores e colegas de doutorado.
- Ao Departamento de Estatística da UFMG pela oportunidade.
- A minha professora, orientador e amiga de graduação Maria Ivanilde por acreditar no meu potencial e ser a maior incentivadora da minha carreira acadêmica.
- A Fundação de Amparo à Pesquisa do Estado do Amazonas-FAPEAM, pelo apoio financeiro.

Resumo

Nesta tese de doutorado, nós desenvolvemos novos métodos baseados em uma abordagem adaptativa para detecção e monitoramento estatístico de mudanças no padrão espaço-temporal de um processo estocástico. Essencialmente desenvolvemos: Razão de Verossimilhança Adaptativa, Fator de Bayes Adaptativo e processo a *posteriori* adaptativo. O Foco das aplicações é a detecção de *cluster* emergentes no espaço-tempo, onde a classe dos possíveis candidatos a *cluster* é excessivamente grande o que pode dificultar a performance de qualquer método desenvolvido com este propósito. Nossos resultados mostraram que a abordagem adaptativa melhora os métodos em dois aspectos. Primeiro aumenta a velocidade na detecção e segundo diminui o espaço dos candidatos a *cluster* tornando o algoritmo mais eficiente. A vantagem da abordagem adaptativa no espaço dos cluster é que a cada tempo precisamos avaliar um número pequeno de candidatos e ainda assim controlamos a taxa de falsos alarmes. Os métodos propostos são aplicados a dados reais e vários estudos com dados simulados também são conduzidos. Os resultados destes estudos sugerem que a abordagem adaptativa é bastante eficiente para solucionar o problema de detecção de *cluster* emergentes.

Abstract

New adaptive based methods to the detection and statistical monitoring of changes in the spatial-temporal pattern of a stochastic process are developed in this thesis. Namely, this study focuses on the following methodologies: Adaptive Likelihood Ratio, Adaptive Bayes Factor, and Adaptive Posterior Process. The applications aim to detect emerging space-time clusters, where the collection of possible cluster candidates is excessively large, which could result in a very inefficient method. Results are presented, showing that the adaptive approach improves the performance in two aspects: first, decreasing the computation to detect emerging clusters at each time, and second, reducing the size of the candidate clusters' configuration space. Using the adaptive approach, the evaluation of only a relatively small number of candidates is necessary. Additionally, the false alarm rate can be controlled. Real data and simulated data are used to demonstrate the usefulness and the practicality of the methods. Those results confirm the theoretical advantages of the proposed methodologies to detect emerging clusters.

Sumário

Lista de Figuras		p. x
Lista de Tabelas		p. xii
1 Introdução		p. 1
1.1 Aspectos Gerais		p. 1
1.2 Justificativa e Importância do Trabalho		p. 4
1.3 Objetivos		p. 4
1.4 Estrutura do Trabalho		p. 5
2 Conceitos e métodos para a detecção de mudanças ou cluster no espaço, tempo e espaço-tempo		p. 6
2.1 Detecção de clusters no espaço		p. 6
2.1.1 Método GAM-Geographical Analysis Machine		p. 7
2.1.2 Método de Besag e Newel		p. 8
2.1.3 Método de Cuzick e Edwards		p. 9
2.1.4 Método de Varredura Espacial-Scan Espacial		p. 9
2.1.5 Método Bayesiano		p. 11
2.2 Detecção de mudanças no tempo		p. 12
2.2.1 Método de Somas Cumulativas (CUSUM)		p. 14
2.2.2 Métodos de Shiryaev(S) e Shiryaev-Roberts (SR)		p. 14
2.2.3 Mistura da Razão de Verossimilhanças (MRV)		p. 15
2.2.4 Métodos Adaptativos		p. 16
2.2.5 Método Bayesiano		p. 17
2.3 Detecção de mudanças no espaço-tempo		p. 18

2.3.1	Formulação do Problema de detecção de clusters no Espaço-Tempo	p. 18
2.3.2	Estatística Scan espaço-tempo	p. 19
2.3.3	Estatística Scan espaço-tempo baseada no valor esperado	p. 21
2.3.4	Scan espaço-tempo baseada no teste score	p. 21
3	Algumas Abordagens para Detecção de Cluster no Espaço-Tempo Usando	
	A Razão de Verossimilhança Adaptativa	p. 24
3.1	Introdução	p. 24
3.2	Formulação, Notação, LR e ALR para detecção de clusters no espaço-tempo	p. 24
3.2.1	Formulação do problema de detecção de cluster no espaço-tempo	p. 24
3.2.2	Notação	p. 25
3.2.3	Detecção de clusters no espaço-tempo usando LR	p. 26
3.2.4	LR Adaptativa para detecção de clusters no espaço-tempo	p. 30
3.2.5	Representação Espacial dos Clusters	p. 31
3.2.6	Algoritmo para detecção de clusters emergentes no espaço tempo	p. 33
3.3	Aplicação	p. 33
3.4	Extensões para abordagem Adaptativa no espaço dos clusters	p. 37
3.5	Estudo com dados simulados	p. 40
3.5.1	Performance dos métodos	p. 40
3.5.2	Resultados obtidos na simulação	p. 41
3.6	Discussão	p. 43
4	Fator de Bayes Adaptativo para detecção de cluster no espaço-tempo	p. 45
4.1	Introdução	p. 45
4.2	Análise Sequencial de Modelos	p. 45
4.2.1	Distribuições preditivas e Fator de Bayes	p. 45
4.2.2	Fator de Bayes Local	p. 47
4.3	Detecção de cluster no espaço-tempo via Fator de Bayes Adaptativo	p. 48
4.3.1	Tempo de alarme para detecção de cluster no espaço-tempo	p. 48

4.3.2	Probabilidade de um falso alarme para o MAX-FB	p. 49
4.3.3	Fator de Bayes Adaptativo	p. 50
4.4	Estrutura Geral dos Modelos	p. 52
4.4.1	Modelo Padrão	p. 52
4.4.2	Modelo Alternativo	p. 54
4.5	Representação Espacial do cluster	p. 56
4.6	Exemplo	p. 60
4.7	Uma aplicação a dados reais: Detecção de cluster de Hanseníase no Amazonas- Brasil	p. 64
4.7.1	Análise exploratória dos dados	p. 64
4.7.2	Aplicação do Fator de Bayes Adaptativo para os dados de Hanseníase	p. 65
4.8	Resultados da Aplicação	p. 67
4.9	Estudo com dados simulados	p. 68
4.9.1	Dados Simulados	p. 68
4.9.2	Análise dos resultados obtidos com dados simulados	p. 71
4.10	Discussão	p. 73
5	Detecção de clusters emergentes via análise a posteriori adaptativa	p. 75
5.1	Introdução	p. 75
5.2	Construção hierárquica do modelo	p. 75
5.3	Distribuições a posteriori	p. 77
5.4	Tempo de alarme baseado no processo a posteriori	p. 80
5.5	Detecção de cluster emergente usando processo a posteriori adaptativo	p. 84
5.6	Escolha da distribuição a priori para região e ponto de mudança	p. 85
5.7	Aplicação	p. 87
5.7.1	Análise dos dados de câncer da tireóide	p. 88
5.7.2	Obtenção do processo a posteriori adaptativo	p. 91
5.7.3	Análise dos resultados da aplicação	p. 92

5.8	Estudo com dados simulados	p. 95
5.8.1	Dados simulados	p. 95
5.8.2	Análise dos resultados obtidos com dados simulados	p. 96
5.9	Discussão	p. 97
6	Conclusão	p. 99
	Referências Bibliográficas	p. 101

Lista de Figuras

2.1	Exemplo visual do uso do método GAM mostrando clusters de áreas por emaranhados de círculos	p. 8
2.2	Exemplo visual do uso do método de Varredura Espacial - Scan circular	p. 10
2.3	Exemplo visual do uso do método de Varredura espacial-temporal - Scan espaço-tempo	p. 21
3.1	Mapa do Novo México. Fonte: US Census.	p. 34
3.2	Monitoramento on-line usando MAX, MIX and WEIGHT-ALR com B(threshold)=20 p. 37	
3.3	Comparação do monitoramento On-line usando o MAX-ALR com GMAX-ALR em (a) e MIX-ALR com GMIX-ALR em (b) com B(threshold)=20.	p. 39
3.4	Amostra de uma simulação do processo para GMIX and GWEIGHT-ALR na escala logartímica com $\log(B)=(\text{threshold})=\log(20)$	p. 43
3.5	Histograma Bivariado para o GMIX-ALR: Sensitividade (Sensitivity) versus atraso (Delay) em (A) e Valor Predito Positivo (PPV-Positive Predicted Value) versus Delay(atraso) em (B).	p. 43
3.6	Histograma Bivariado para o WEIGHT-ALR: Sensitividade (Sensitivity) versus atraso (Delay) em (A) e Valor Predito Positivo (PPV-Positive Predicted Value) versus Delay(atraso) em (B).	p. 43
4.1	<i>Uma possível representação para as densidades preditivas em relação aos modelos padrão e alternativo.</i>	p. 57
4.2	Exemplo de uma Elipse	p. 58
4.3	Formato de Clusters que podem ser detectados	p. 58
4.4	<i>Coefficiente de Detecção de Hanseníase (por 10000 ml hab) no Estado do Amazonas(em azul), segundo o mês, nas cidades: Manaus(em verde), Humaitá (em vermelho) e Itacoatiara (em preto)</i>	p. 65
4.5	Mapa do Estado do Amazonas com os municípios espacialmente representados pelas coordenadas de seus centróides.	p. 66

4.6	Monitoramento on-line usando Fator de Bayes Sequencial Adaptativo com janela $w = 3, \alpha = 0,05$	p. 68
4.7	<i>cluster detectado usando o Fator de Bayes Sequencial Elíptico</i> $t \in [15, 17]$	p. 68
4.8	<i>cluster detectado usando o Fator de Bayes Sequencial Elíptico</i> $t \in [16, 18]$	p. 69
4.9	Mapa do Novo México. Fonte: US Census.	p. 70
4.10	Máximo do Fator de Bayes Adaptativo em (A), Cardinalidade J^n de Ξ^n em (B) para 1000 simulações do processo sobre o modelo padrão com $\rho = 0.2$; Máximo do Fator de Bayes Adaptativo em (C), Cardinalidade J^n de Ξ^n em (D) para 1000 simulações do processo sobre o modelo padrão com $\rho = 0.4$	p. 71
4.11	Histograma Bivariado para o Fator de Bayes Adaptativo com $\rho = 0.2$: Sensitividade versus atraso e Valor Predito Positivo versus atraso	p. 73
4.12	Histograma Bivariado para o Fator de Bayes Adaptativo com $\rho = 0.4$: Sensitividade versus atraso e Valor Predito Positivo versus atraso	p. 73
4.13	Distribuições Preditivas e Padrão após o ponto de mudança nas áreas que compõe o <i>cluster</i> : Chaves em (A), Eddy em (B), Lea em (C) e Otero em (D).	p. 74
5.1	Uma Classe de clusters com formato circular e tamanho variável	p. 86
5.2	Mapa do Novo México. Fonte: US Census.	p. 88
5.3	função de densidade weibull (f.d), função de densidade do máximo de m-variáveis weibull (f.d. max), Função de distribuição weibull (F.d), Função de distribuição do máximo de m-variáveis weibull (F.d. max), 1-Função de distribuição weibull (1-F.d), 1-Função de distribuição do máximo de m-variáveis weibull (1-F.d. max)	p. 90
5.4	A esquerda: Máxima probabilidade a posteriori adaptativa para $m=2$, ponto $k=15$ e região da mudança; A direita: Máxima probabilidade a posteriori adaptativa para $m=8$, ponto $k=19$ e região da mudança.	p. 94
5.5	Em (A): Máxima probabilidade a posteriori adaptativa para $m=2, 4$ e 8 ; Em (B) Razão valor observado por valor esperado $x_t(s_l)/e_t(s_l), t \geq k = 1, 2, \dots, n$ para as regiões do cluster emergente detectado e considerado significativo	p. 94
5.6	Amostra de uma simulação do processo a posteriori adaptativo para distribuições de referência Weibull e Uniforme com: $\delta = 2, B=0.5, m=2, 4, 8$	p. 96

Lista de Tabelas

- 3.1 *Detecção de clusters emergentes de casos de câncer da tireóide em homens que ocorreram no Novo México, usando o MIX-ALR e WEIGHT-ALR nos últimos cinco anos.* p. 37
- 3.2 Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) com diferentes valores para $\theta_{t,j}$, diferentes taxas aceitáveis de alarmes falsos (α) usando o GMIX e GWEIGHT-ALR. p. 42
- 4.1 *Interpretação do Fator de Bayes (Kass e Raftery, 1995)* p. 52
- 4.2 *Percentual de área de uma elipse que é também parte de qualquer outra elipse com o mesmo centro, forma e tamanho, porém com quantidades (#) diferentes de ângulos* p. 59
- 4.3 Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) com diferentes valores para δ , diferentes probabilidades de alarmes falsos (α) usando o Fator de Bayes Adaptativo Eliptico com janela $w=3$. p. 72
- 5.1 *Valores para p de acordo com o valor de w e $p_0 = 0.5$.* p. 91
- 5.2 Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) usando o processo a posteriori adaptativo com janela $w=6$, diferentes valores para δ e duas distribuições para G_0 : Weibull(2,1.414) e Uni(0,2) p. 97

1 *Introdução*

1.1 Aspectos Gerais

Métodos estatísticos aplicados á análise de dados, obtidos periodicamente pelos sistemas de vigilância (em saúde pública, crime ou ambiental) são importantes para detectar *cluster* de eventos, os quais podem indicar uma rápida mudança no padrão dos dados observados. Geralmente, um *cluster* é uma inesperada aglomeração de eventos no espaço, tempo ou no espaço-tempo. Quando o período de vigilância é pré-especificado e os eventos são avaliados de acordo com sua distribuição espacial, um *cluster* (*cluster* espacial) é representado por um sub-conjunto de localizações espaciais da região em estudo onde a taxa de ocorrência de tais eventos é elevada de forma significativa. Se o espaço é ignorado, representamos o *cluster* pelo grupo (ou sequência) de observações que causam uma mudança no padrão temporal dos dados. No entanto, se o espaço e o tempo são monitorados simultaneamente, um *cluster* espaço-temporal é identificado por um par (localizações espaciais, intervalo de tempo) para o qual ocorreu a mudança. Detecção de *cluster* tem sido o foco de uma variedade de sistemas de vigilância: Sistema de vigilância em crimes (Gorr and Harries, 2003) para detectar cluster emergentes de eventos relacionados a algum tipo de crime; Sistema de vigilância em doenças para monitorar dados de saúde pública como por exemplo visitas em hospitais (Sabhnani et al., 2005) ou *cluster* de casos de doenças (Kulldorff, 2001); Sistemas de monitoramento ambiental para detectar altos níveis de poluição no ar e água (Ailamaki et al., 2003) ou detectar *cluster* emergentes de queimadas em florestas (Tuia et al., 2008).

Na escala espacial, o problema de detecção de *clusters* é abordado através de testes de hipóteses. Nestes testes, primeiramente é assumido que a região em estudo é dividido em l áreas contíguas. A hipótese nula é de que a taxa de ocorrência dos eventos é constante sobre toda a região. A hipótese alternativa é de que existe uma elevada taxa em alguma sub-região conexa de tamanho menor ou igual a l . A localização e o tamanho do *cluster* são assumidos desconhecidos. Logo, existem $2^l - 1$ subconjuntos de áreas candidatos a *cluster* dentre os quais teremos que verificar quais são *clusters* conexos. Muitos testes baseados em algoritmos computacionalmente intensivos, (Turnbull et al., 1990; Cuzick and Edwards, 1990; Besag and Newell, 1991; Walter, 1994; Tango, 1995; Kulldorff, 1997; Held and Raber, 2000; Patil and Tallie 2003; Duczmal and

Assunção, 2004; Tango, 2005; Kulldorff et al., 2006; Gangon and Klayton, 2007; Duczmal et al., 2007), têm sido desenvolvidos para detectar e estimar a localização do *cluster*. Estes métodos não levam em consideração a incerteza sobre o tamanho e localização do *cluster*.

Na escala temporal, a detecção de uma mudança na taxa de ocorrência dos eventos em um *cluster* (grupo) de observações pode ser realizada usando métodos desenvolvidos para monitoramento de processos industriais como por exemplo o método CUSUM (Page, 1954), Shirayayev-Roberts (Roberts, 1966) ou mistura de razões de verossimilhanças (Pollak, 1987). Uma boa revisão sobre estes métodos e suas variações pode ser vista em Lai (1995). Usualmente esses métodos assumem que no estado de controle do processo, existe uma taxa padrão conhecida θ_0 de eventos, e sob a suposição de um ponto de mudança no processo, a taxa muda para θ_1 também assumida conhecida. Na prática essas suposições nunca são satisfeitas e uma alternativa é usar métodos adaptativos (Lorden e Pollak, 2005; Tartakovsky, 2006) ou dinâmicos (West, 1986).

Na escala espaço-temporal, a detecção de *cluster* pode ser entendido como um problema de detecção de região e ponto de mudança no espaço-tempo. Estatisticamente este problema é descrito da seguinte forma: Denote por \mathcal{S} uma região de estudo particionado em L -áreas A_l contíguas identificadas por um ponto s_l . Onde, por exemplo, \mathcal{S} é um mapa e s_l representa o centróide do polígono que delinea a área A_l . Consider sobre \mathcal{S} o monitoramento (ou a vigilância) estatístico(a) de um processo estocástico $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots, L\}$. A cada tempo discreto $t \geq 1$ observamos um vetor L -variado $\mathbf{X}_t = (X_t(s_1), X_t(s_2), \dots, X_t(s_L))'$, em que $X_t(s_l)$ representa alguma valor observado na l -ésima área. Sejam $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$ um conjunto conexo de localizações espaciais em \mathcal{S} , $\mathcal{H}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ o conjunto de dados acumulados até o instante n . O objetivo da monitoramento é verificar se existe algum *cluster* emergente em \mathcal{S} . O interesse é detectar um *cluster* que começou em um tempo desconhecido $k \leq n$ e que está presente até o estágio atual n . O *cluster* no espaço-tempo é expresso em função de uma mudança no padrão da distribuição do sub-processo $\{X_t(s_l) \in \mathbf{X} : s_l \in \xi, t \geq k\}$ tal que $\xi^{n,k} = \{s_l \in \mathcal{S} : s_l \in \xi\} \times [k, n]$ representa o cluster no espaço-tempo também chamado de *cluster* emergente.

Para solucionar o problema de detecção de *cluster* emergente no espaço-tempo alguns métodos têm sido propostos. Kulldorff(2001) usa a estatística scan espaço-tempo para monitoramento de casos de doenças. No entanto, essa estatística não controla adequadamente o erro tipo I, para todas as análises sequenciais realizadas. Rogerson(2001), propôs o uso de uma estatística de Knox local (Knox, 1963) combinada com métodos de somas cumulativas (CUSUM) usando aproximações pela distribuição normal para monitoramento de cluster em processos pontuais. Marshall(2007) discute vários problemas ao se utilizar estas aproximações e sugere que este método não seja usado. Variantes da Estatística Scan espaço-temporal são apresentadas em Neill (2006). Sonesson (2007) usa a estatística CUSUM para detecção de cluster no espaço-tempo em dados de área. Porter (2007) utiliza versões adaptativas dos métodos CUSUM e Shirayayev-Roberts

para detecção de mudanças no espaço-tempo em processos de Poisson Homogêneo. Assunção e Corrêa (2009) monitoram a função de intensidade de um processo de Poisson heterogêneo no espaço-tempo através da estatística de Shiriyayev-Roberts. Lima e Duczmal(2009,2011) propõem o Fator de Bayes Cumulativo para monitoramento e detecção de *cluster* em dados de área com taxa padrão conhecida. Tango (2010) Usa uma estatística scan combinada com um teste score para detecção de *cluster* no espaço-tempo em uma janela temporal variável.

No geral, poucos métodos para detecção e monitoramento de *clusters* têm sido propostos com o intuito de desenvolver um sistema de vigilância que soe um alarme assim que um cluster no espaço-tempo seja detectado. Uma revisão dos métodos sugeridos na literatura para detecção sequencial de mudanças no monitoramento de dados de saúde pública na escala temporal é apresentada por Sonesson e Bock (2003). Frisen (2003) discute as propriedades ótimas de um bom sistema de vigilância e reforça a necessidade de métodos estatísticos e sistemas apropriados para a detecção prospectiva de cluster no espaço-tempo de forma rápida e eficiente.

A escassez de métodos para monitoramento e detecção de *cluster* no espaço tempo, pode ser justificada pelo fato que para detectar qualquer mudança inesperada na taxa de certos tipos de eventos, faz-se necessária uma análise repetida de dados acumulados ao longo do tempo e, qualquer método desenvolvido tem que ser ajustado para o problema de $t \times (2^l - 1)$ comparações no espaço-tempo, para cada tempo t . Para isso, neste trabalho nós desenvolvemos novos métodos para detecção de cluster e que podem ser usados como ferramenta auxiliar em sistemas de vigilância em que o foco principal é a detecção de cluster emergentes no espaço-tempo. Os métodos desenvolvidos nesta trabalho são baseados em uma abordagem adaptativa para os parâmetros. Em essência isso significa que os parâmetros desconhecidos são estimados a cada tempo t com os dados acumulados até o instante anterior $t - 1$. É óbvio que a medida que aumentamos a classe de candidatos a cluster e o tempo de vigilância, nós aumentamos o número de teste múltiplos no espaço-tempo. No entanto, ao aplicamos uma estimação adaptativa também no espaço dos clusters, melhoramos os métodos de detecção em pelo menos dois aspectos: 1) aumento na velocidade de detecção; 2) diminui o espaço dos cluster tornando o algoritmo mais rápido e eficiente. A vantagem da abordagem adaptativa no espaço dos cluster é que a cada tempo precisamos avaliar poucos candidatos e ainda assim controlamos a taxa de falsos alarmes. Ao combinarmos técnicas de vigilância na escala temporal com estimação adaptativa e algoritmos para detecção de *cluster* espacial, nós propomos uma vigilância no espaço-tempo usando: Razão de verossimilhança adaptativa, Fator de Bayes Adaptativo, distribuição a posteriori adaptativa. Os métodos desenvolvidos neste trabalho não requerem o uso de simulações de Monte Carlo para validar a significância estatística do procedimento. Fato esse que é comum aos procedimentos existentes na literatura e pode tornar a análise muito demorada dependendo do tamanho da partição da região em estudo.

1.2 Justificativa e Importância do Trabalho

Devido ao avanço computacional, têm surgido um grande interesse na construção de sistemas de informações governamentais que funcionem como uma importante ferramenta de auxílio no controle da administração pública. Um exemplo é o e-siga (Sistema de Informações Governamental do Amazonas). O e-siga é um sistema *on-line* operacional que permite, entre outros recursos, o acompanhamento dos problemas epidemiológicos e monitoramento em tempo real dos fatores adversos a saúde do amazonense. É nesse sentido, e tendo em vista o crescente interesse pelo estudo de agregados (*cluster*) espaço-temporais de doenças, que nós propomos novas metodologias para detecção e vigilância de *clusters* em mapas de doenças, criminalidade, etc.... Uma lacuna na literatura, que justifica apresentação deste trabalho, é a ausência de uma sistematização crítica que amplie o horizonte de aplicações dos métodos de detecção de *clusters* para além da investigação etiológica de doenças raras, ou endêmicas, e oriente o investigador na escolha da decisão mais adequado aos seus objetivos.

1.3 Objetivos

Esta tese de doutorado teve como objetivo geral o estudo dos métodos estatísticos para detecção de *clusters* no espaço-tempo, assim como o desenvolvimento de novos métodos. Como objetivos específicos tivemos:

1. Desenvolvimento de um novo método estatístico baseado em estatística bayesiana ou clássica para detecção e vigilância de *cluster* no espaço-tempo em dados de área e processos pontuais.
2. Utilizar o método desenvolvido como parte de um sistema de vigilância em alguns tipos de eventos como, por exemplo, os casos de Hanseníase no estado do Amazonas.

Com o desenvolvimento destas novas metodologias para detecção de *cluster* no espaço-tempo, esperamos dar as seguintes contribuições:

- Estabelecer novas metodologias técnico-científicas para vigilância no espaço-tempo de eventos.
- Desenvolvimento de métodos estatísticos que ajudem e identifiquem rapidamente as áreas prioritárias onde ocorram *clusters* de eventos para que ações preventivas sejam tomadas o mais rápido possível.

1.4 Estrutura do Trabalho

No Capítulo 2 apresentamos uma breve revisão sobre o problema de detecção de *cluster* no espaço, tempo e no espaço-tempo. Alguns métodos "clássicos" na literatura são descritos. Uma definição formal e a questão de como avaliar a performance de um método para a detecção de *cluster* no espaço-tempo também é discutida. No Capítulo 3, Ao combinarmos algoritmos usados para detecção de *clusters* espaciais com técnicas de estimação adaptativa, nós generalizamos para detecção de *clusters* no espaço-tempo algumas funções de monitoramento baseadas na razão de verossilhança e que antes eram usadas para monitoramento somente na escala temporal. No Capítulo 4, nós extendemos para detecção de *cluster* no espaço-tempo com janela temporal variável o Fator de Bayes Sequencial. Neste caso nós impomos também uma estimação adaptativa no espaço dos *cluster* e realizamos uma varredura no espaço-tempo através de cilindros com base elíptica. No Capítulo 5, incorporamos a incerteza sobre a localização espacial e sobre o tempo de mudança através de uma distribuição a *priori* para o *cluster*, e realizamos uma análise a *posteriori* no espaço-tempo para detecção de *cluster* emergentes. Finalmente estabelecemos nossas conclusões no Capítulo 6.

2 *Conceitos e métodos para a detecção de mudanças ou cluster no espaço, tempo e espaço-tempo*

2.1 Detecção de clusters no espaço

A análise de *clusters* espaciais tem recebido uma atenção considerável em diversas áreas do conhecimento. O objetivo básico em um problema de detecção de *cluster* é determinar automaticamente regiões do espaço onde tenha ocorrido uma mudança não esperada no padrão espacial do processo estocástico observado. Essas mudanças no padrão espacial podem corresponder a uma variedade de fenômenos, dependendo do campo de aplicação como, por exemplo, detecção de *cluster* de doenças, de estrelas ou galáxias, de focos de queimadas em florestas, de incidências de algum tipo de crime, de alguma espécie de animais.

Os estudos de *clusters* são baseados em localizações de eventos (processo pontual) ou agregado de eventos (dados de área). Em ambos os casos a heterogeneidade da população estudada é assumida conhecida e em certas situações algumas medidas de covariáveis também podem ser incorporadas. Um processo pontual pode ser transformado em dados de área. Por isso, nosso foco são processos espaciais modelados como processos medidos em áreas (ou dados de área). Neste caso, supomos que existe um processo estocástico $\mathbf{X}(s) = \{X(s_l) : l = 1, 2, \dots, L\}$. Onde $X(s_l)$ é a realização do processo na área A_l completamente identificada por um ponto $s_l \in \mathcal{S} = \{s_1, \dots, s_L\}$. Tipicamente, \mathcal{S} é um mapa particionado em L áreas e s_l é (não necessariamente) o centro do polígono que delinea a área A_l .

Os métodos usuais utilizados para a detecção de *clusters* espaciais são baseados em testes de hipóteses, utilizam algoritmos computacionalmente intensivos que procuram identificar um conjunto conexo de localizações (ou área geográficas) com uma ocorrência de eventos acima (ou abaixo) do esperado sem pré-especificar quais e quantas áreas são. A hipótese nula do teste representa a completa aleatoriedade espacial dos eventos implicando que a taxa de ocorrência unitária de eventos (por pessoa, animal, árvore) é a mesma em todas as áreas de modo que o número esperado de eventos em um local é proporcional ao número de pessoas (ou animais,

árvores) habitando neste local. Geralmente, sob a hipótese nula

$$H_0 : X(s_l) \sim \text{Poisson}(\lambda N(s_l)).$$

Onde λ é uma taxa de ocorrência global dos eventos e $N(s_l)$ representa o número total da população em risco habitando em A_l .

Muitos métodos para a detecção de *clusters* espaciais têm sido desenvolvidos com diferentes propósitos. Uns avaliam a existência de um *cluster* global, sem especificar sua localização. Outros determinam a localização e avaliam a significância estatística do *cluster*. Estes métodos usam técnicas computacionalmente intensivas como permutação aleatória, Monte Carlo, etc. Os mais usuais, entre estes métodos, assumem que temos a disposição um mapa \mathcal{S} de áreas, cada uma com população de risco e um certo número de eventos observados. Eles utilizam janelas espaciais que superpõe conjuntos conexos, com algum formato geométrico (círculo, quadrado, elipse), sobre as áreas de estudo e contam o número de eventos que ocorrem nas regiões cujos os centróides caem dentro de cada janela espacial. Um problema comum na detecção de *cluster* é o ajuste para múltiplas comparações. Este problema acontece quando listados todas os conjuntos conexos que em princípio podem representar um *cluster* (ou candidato a *cluster*) e para cada um deles é testado se sua taxa de ocorrência de eventos difere de forma significativa daquela associada com o restante do mapa. Aqueles conjuntos em que o teste fosse significativo seriam considerados um agregado de taxa acima (ou) abaixo do esperado. Este procedimento é inadequado para esta situação pelo seguinte fato: suponha que vários testes são realizados com nível de significância global α . Então, vários teste serão significativos mesmo que a hipótese nula seja verdadeira em todos eles. Isto ocorre porque achamos que o valor α usado individualmente continua o mesmo. Na verdade, para o conjunto de todos os testes, este nível de significância será muito maior que o valor nominal para os testes individuais. Esta situação só é válida, quando consideramos todos os testes simultaneamente. O resultado da prática desse fato é que muitos falsos *cluster* seriam detectados por este método. Vejamos, agora, alguns métodos para detecção de *cluster*.

2.1.1 Método GAM-Geographical Analysis Machine

Para a detecção de *cluster* espacial, Openshaw et al. (1988) propuseram um método computacionalmente intensivo com grande apelo visual chamado Geographical Analysis Machine, abreviado por **GAM**. Neste método, assumamos que $X(s_l)$ representa o número de eventos que ocorrem na área A_l com o valor esperado, sob H_0 dado por $\lambda N(s_l)$. Associe os valores observados de cada área aos seus centróides (centro do polígono que delinea a área A_l) denotados por s_l . O procedimento **GAM** usa o seguinte algoritmo:

- Selecione um raio r (por exemplo, $r=1,2$ ou 4 km).
- Em cada centróide s_l fixe um círculo $C_{l,r}$ de raio r .

- Calcule

$$X_{l_r} = \sum_{l=1}^L X(s_l) I_{\{s_l \in C_{l_r}\}} \quad \text{e} \quad N_{l_r} = \sum_{l=1}^L N(s_l) I_{\{s_l \in C_{l_r}\}}$$

o número de eventos e o número total da população em risco habitando no círculo C_{l_r} , onde $I_{\{B\}}$ é a função indicadora do evento B .

- Calcule o valor-p, p_{l_r} , do teste associado a $X(s_l)$, sob a hipótese nula, $H_0 : X(s_l) \sim \text{Poisson}(\lambda N(s_l))$.
- Desenhe o círculo C_{l_r} se $p_{l_r} \leq 0.002$
- Repita o procedimento acima aumentando (ou escolhendo) outro raio para o círculo.

O resultado final é a identificação de *clusters* de áreas por emaranhados de círculos sobrepostos como mostrado na figura (2.1).

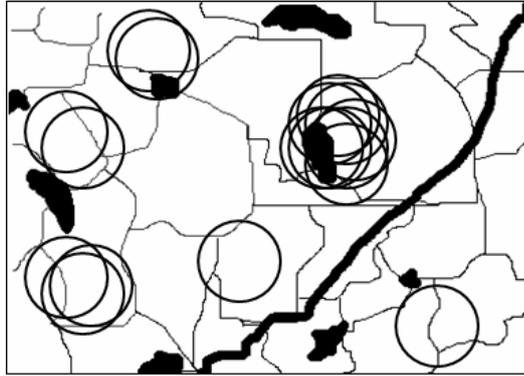


Figura 2.1: Exemplo visual do uso do método GAM mostrando clusters de áreas por emaranhados de círculos

Embora cada círculo seja julgado individualmente significativo ao nível 0.002, um nível de significância para todos os círculos simultaneamente não é conhecido. O motivo é o uso de testes simultâneos não independentes que, neste caso, são devidos a comparação de um imenso número de círculos sobrepostos.

2.1.2 Método de Besag e Newel

No método **GAM** de Openshaw fixa-se o raio r do círculo e calcula-se o valor-p sobre o número de eventos que ocorrem no círculo C_{l_r} . Em termos gerais, no método Besag e Newel (1991) fixa-se o número y de eventos que devem ser buscados e calcula-se o raio necessário para englobá-los. No círculo resultante, calcula-se o p-valor e procedendo como o **GAM**, desenha círculos significativos (valor-p ≤ 0.002). Repete o procedimento variando o valor de y .

Para computar o valor-p, seja $X = \sum_{l=1}^L X(s_l)$ e $N = \sum_{l=1}^L N(s_l)$. Centrado em s_l , assumamos que a área A_l possui pelo menos um caso. Seja \tilde{L} a variável aleatória que conta o número de outras

áreas (ou centróides) necessárias para acumular os y primeiros casos mais próximos de s_l . Seja \tilde{l} o valor observado de \tilde{L} e $N_{\tilde{l}}$ o total da população nessas \tilde{l} áreas. Sob a hipótese nula (a mesma do método **GAM**), $X_{\tilde{l}}$ o número de eventos nessas \tilde{l} áreas segue uma distribuição de Poisson com valor esperado dado por $N_{\tilde{l}}X/N$. Agora, notando que $P(\tilde{L} \leq \tilde{l}) = 1 - P(\tilde{L} > \tilde{l} + 1)$ representa 1 menos probabilidade de as \tilde{l} primeiras áreas possuírem menos que y eventos. Então o valor-p para um círculo C_{l_y} centrados em s_l contendo y eventos é dado por,

$$p_{l_y} = 1 - \sum_{j=1}^{y-1} P(X_{\tilde{l}} = j) = 1 - \sum_{j=1}^{y-1} \frac{(N_{\tilde{l}}X/N)^j}{j!} e^{-N_{\tilde{l}}X/N}.$$

2.1.3 Método de Cuzick e Edwards

Cuzick e Edwards(1990) fizeram uma proposta que representa uma pequena variação em relação aos métodos de Besag e Newel (1991). Como em Besag e Newel (1991), inicia-se fixando o número de eventos y . A seguir, em torno do centróide de cada área A_l que possui pelo menos um evento, traça-se um círculo que vai se expandindo por aumentarmos o seu raio até que o círculo contenha uma população para qual espera-se observar y eventos. Depois, verifica-se quantos eventos X_l foram de fato observados e calcula-se a estatística

$$U_y = \sum_{l=1}^L (X_l - y) I_{\{x(s_l) > 0\}}.$$

Cuzick e Edwards (1990) derivaram as fórmulas dos momentos desta estatística sob a hipótese nula e mostraram que ela possui uma distribuição assintoticamente normal possibilitando assim calcular o valor-p para o teste.

2.1.4 Método de Varredura Espacial-Scan Espacial

Kulldorff e Nagarwalla (1995) e Kulldorff(1997) apresentam um método que generaliza os anteriores e permite resolver o problema de teste múltiplos. Denote por Ξ uma classe de subconjuntos conexos candidatos a formarem *cluster*. Neste texto, definimos $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$. Suponha por exemplo que ξ é um círculo de raio r arbitrário centrado em cada um dos L centróides s_l . Em tese, existe um número infinito de tais círculos mas, na prática, estes círculos podem ser restritos apenas aqueles centrados em s_i e com raios iguais as distâncias (distância Euclidiana) entre s_i e o demais centróides. O motivo é que círculos com raios ligeiramente diferentes vão conter, em geral as mesmas áreas e portanto definem o mesmo possível conjunto de localizações espaciais. A cardinalidade de Ξ pode ser reduzida se colocarmos um limite no raio de modo que nenhum ξ contenha mais do que uma certa porcentagem de centróides, tal como 20% de todos os centróides ou de toda a população em risco.

O método de varredura espacial de Kulldorff, também conhecido como Scan, é baseado na razão de verossilhança generalizada. O parâmetro neste caso é (ξ, p_0, p_1) onde ξ denota um conjunto de localizações espaciais, parametrizado pelo seu raio e coordenadas do centro, p_0 é a probabilidade de ocorrer um evento em ξ e p_1 a probabilidade de ocorrer um evento no conjunto complementar de ξ em \mathcal{S} denotado por $\bar{\xi} = \{s_l \in \mathcal{S} : s_l \notin \xi\}$. As hipóteses testadas são;

$$H_0 : p_0 = p_1, \quad \text{para todo } s_i \in \mathcal{S}$$

e

$$H_1 : p_0 > p_1, \quad \text{para todo } s_i \in \xi$$

Observe que sob H_0 a probabilidade de ocorrência de um evento é a mesma em qualquer área.

Sejam, respectivamente, $N(\xi)$ o número de indivíduos e $X(\xi)$ o número de eventos em ξ , e ainda, N e X o número total de indivíduos e eventos na região. A função de verossilhança para o parâmetro (ξ, p_0, p_1) é dada por,

$$\mathcal{L}(\xi, p_0, p_1) = p_0^{x(\xi)} (1 - p_0)^{N(\xi) - x(\xi)} p_1^{x - x(\xi)} (1 - p_1)^{N - N(\xi) - x + x(\xi)}.$$

A razão de verossilhança generalizada para um fixo ξ é

$$RV(\xi) = \frac{\sup_{p_0 > p_1} \mathcal{L}(\xi, p_0, p_1)}{\sup_{p_0 = p_1} \mathcal{L}(\xi, p_0, p_1)}, \quad p_0, p_1 \in (0, 1).$$

Então a estatística Scan é definida como $T = \max_{\xi \in \Xi} RV(\xi)$, e o *cluster* espacial estimado é $\hat{\xi} = \arg[\max_{\xi \in \Xi} RV(\xi)]$. A distribuição de T depende da distribuição da população e é difícil de ser obtida analiticamente. Recetemente, algumas aproximações têm sido propostas em alguns casos particulares. Assim, a solução proposta é utilizar métodos de simulação de Monte Carlo para obter a distribuição empírica de T condicionada no número total de eventos observados X . Desta forma o valor-p empírico é obtido após um grande número de simulações do processo (por exemplo, 1000 vezes) sob a hipótese nula. Existem várias extensões para este método como por exemplo: Scan Eliptico (Kulldorf et al, 2006), Scan irregular (tango e Takahashi, 2005). Uma representação da varredura espacial realizada pela estatística Scan é mostrada na figura (2.2)

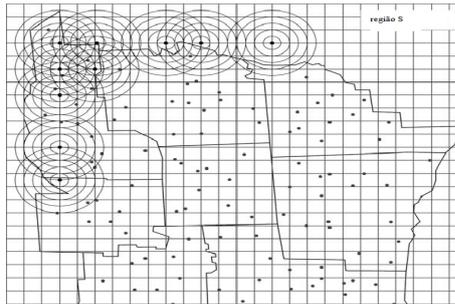


Figura 2.2: Exemplo visual do uso do método de Varredura Espacial - Scan circular

2.1.5 Método Bayesiano

Os métodos descritos anteriormente são usados para detectar um único *cluster*, e não levam em consideração a incerteza sobre a estrutura (forma e tamanho) do *cluster*. Com intuito de solucionar este problema do ponto de Bayesiano, Gangnon e Clayton (2000) utilizam uma modelagem bayesiana para o problema de detecção de *cluster*. Basicamente, os autores assumem que $X(s_l)|\lambda(s_l) \sim \text{Poisson}(\lambda(s_l)N(s_l))$, em que a incerteza sobre $\lambda(s_l)$ e o *cluster* é descrita através de uma distribuição de probabilidade. Suponha que o interesse é detectar *cluster* através de mudanças na distribuição do vetor $(\lambda(s_1), \dots, \lambda(s_L))$. Então uma abordagem bayesiana com distribuições *a priori* hierárquicas é construída. Primeiramente, o espaço é dividido em $c + 1$ grupos com vários componentes ou áreas. denote um desses grupos por grupo padrão ξ_0 (no sentido de que estes componentes não formam um *cluster*) e os demais b grupos de *clusters*. Identifique um modelo com b *clusters* por um vetor $\xi = (\xi_1, \dots, \xi_L)$ onde $\xi_l = 0$ se $s_l \in \xi_0$ e $\xi_l = j$ se s_l pertence ao *cluster* $j = 1, 2, \dots, c$. Esta modelagem é importante pois permite uma única representação para qualquer ξ . Agora, dado um ξ assuma que cada s_l pertencente ao *cluster* j possui taxa de eventos λ_j e seja $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_c)$.

Assuma uma distribuição *a priori* permutável para $(\lambda_1, \dots, \lambda_c)$. Mais especificamente, assuma que $\lambda_0|\xi \sim \text{gama}(a_0, b_0)$ e $\lambda_j|\xi \sim \text{gama}(a, b)$ onde $\text{gama}(a, b)$ representa a distribuição gama com média a/b e variância a/b^2 . Dado $\mathbf{X}(s) = (X(s_1), X(s_2), \dots, X(s_L))$ e ξ , $\lambda_0, \lambda_1, \dots, \lambda_c$ são independentes com distribuições

$$\lambda_0|\mathbf{X}(s), \xi \sim \text{gama}(a_0 + X_0, b_0 + N_0)$$

e

$$\lambda_j|\mathbf{X}(s), \xi \sim \text{gama}(a + X_j, b + N_j), j = 1, 2, \dots, c.$$

onde, $X_j = \sum_{l=1}^L X(s_l)I_{\{\xi_l=j\}}$ é o número total de eventos no *cluster* j , $N_j = \sum_{l=1}^L N(s_l)I_{\{\xi_l=j\}}$ é o total de indivíduos no *cluster* j , a verossimilhança marginal de $\mathbf{X}(s)|\xi$ é dada por,

$$\mathbf{p}(\mathbf{X}(s)|\xi) = \int_{\Lambda} f(\mathbf{X}(s)|\Lambda)\pi(\Lambda|\xi)d\Lambda$$

onde f é função de densidade de $\mathbf{X}(s)|\Lambda$ e π é a distribuição condicional de $\Lambda|\xi$. A distribuição *a priori* para ξ é da forma

$$\pi(\xi) \propto \exp\left(-\sum_{j=1}^c S_j\right),$$

onde S_j é escore para o cluster j dependendo somente das características(forma, tamanho) do *cluster*. Desta forma a distribuição *a posteriori* de ξ é obtida por,

$$\pi(\xi|\mathbf{X}(s)) \propto \mathbf{p}(\mathbf{X}(s)|\xi)\pi(\xi).$$

um algoritmo de varredura no espaço dos candidatos a *cluster* é construído e selecionamos o

modelo de *cluster* com máxima probabilidade a *posteriori*.

2.2 Detecção de mudanças no tempo

Detecção rápida de uma mudança na distribuição de uma sequência $(\mathbf{X}_t)_{t \geq 1}$ de variáveis (ou vetores) aleatórios é um problema muito importante em muitas áreas, como controle de qualidade e vigilância epidemiológica. Começando com o trabalho pioneiro de Shewhart (1931), uma vasta literatura foi produzida com este objetivo e vários métodos competitivos foram desenvolvidos. Além do método de Shewhart, o CUSUM (Page, 1954) e Shiriyayev-Roberts (Roberts, 1966) são os métodos mais conhecidos para detecção de uma mudança no tempo. Esses métodos têm sido extensivamente discutidos, e suas performances comparadas de acordo com algum critério de otimalidade (Lai, 1995; Srivastava e Wu, 1993).

Para utilizar esses métodos, realizamos um monitoramento estatístico de um processo estocástico $\{X_t : t = 1, 2, \dots\}$ com o objetivo de detectar uma mudança importante no processo, em um tempo desconhecido k , tão rápida e precisa quanto possível. Especificamente, seja x_1, x_2, \dots uma sequência de observações do processo em monitoramento com densidade (ou função de probabilidade) $f(x_t, x_{t-1}, \dots, x_1 | \theta)$. A cada tempo t deve-se discriminar entre dois estados do processo: fora-de-controle ou sob-controle. A detecção de uma mudança no tempo é fornecida através da avaliação de todos os dados passados. O momento de mudança é o primeiro ponto no tempo para o qual existem evidências suficientes de que uma mudança importante tenha ocorrido.

Matematicamente, o monitoramento é feito usando uma função de alarme $\Psi_t = \Psi_t(x_1, \dots, x_t)$ e um limite de controle B , tais que o tempo de alarme, τ , de uma possível mudança é dada por,

$$\tau = \min\{t : \Psi_t > B\}.$$

Se Ψ_t excede B , então há uma indicação que o processo está no estado fora-de-controle e um alarme é soado. Se o processo está sob-controle, todos os alarmes são falsos. O valor de B é geralmente especificado de forma a minimizar a ocorrência de alarmes falsos e Ψ_t pode ser, por exemplo, o $\max_{1 \leq k \leq t} \log \Lambda_k^t$ onde Λ_k^t é a razão de verossimilhança para as observações entre os tempos k e t , ou simplesmente o Fator de Bayes.

Em situações de monitoramento, realizamos análises repetidas sobre o estado do processo. Desde que o processo pode eventualmente mudar durante o monitoramento, o nível de significância e poder do teste não são medidas apropriadas para a avaliação. Algumas medidas especiais tem sido sugeridas por vários autores (Sonesson, 2007). Dentre elas temos:

1. ARL^0 (*average in-control run length*): É a principal medida utilizada para descrever a performance do método quando o processo está sob controle. Este é definido como o número médio de observações até que uma mudança no processo seja detectada, sob a

hipótese de que não houve nenhuma mudança. Quando o processo está sob controle, todos os alarmes são falsos. A distribuição do alarme falso é geralmente resumida pelo ARL^0 , denificado por

$$ARL^0 = \mathbb{E}_0[\tau].$$

Em que \mathbb{E}_0 é o valor esperado sobre a hipótese de que a mudança não ocorreu. A taxa de falsos alarmes (FAR-*False alarm rate*) é geralmente expressa por $FAR = 1/ARL^0 = 1/\mathbb{E}_0[\tau]$. Quando o valor do ARL^0 é alto, teremos poucos alarmes falsos, mas pequenas mudanças reais podem não ser muito bem detectadas. Da mesma forma, para valores pequenos de ARL^0 , as mudanças reais geralmente serão detectadas, mas alarmes falsos também serão mais frequentes.

2. *Delay* (atraso): expressa a velocidade com que um método detecta uma mudança no processo. O atraso pode ser expresso de duas formas: CED (*conditional expected delay*) - o tempo médio de atraso até que o alarme soe dado que realmente ocorreu uma mudança no processo, é uma medida equivalente ao poder de detecção. Esta medida caracteriza o comportamento de um processo fora de controle. É uma medida pontual, depende do ponto em que ocorreu a mudança, é definido por

$$CED(k) = \mathbb{E}_k[\tau - \nu | \tau \geq \nu = k],$$

em que \mathbb{E}_k é o valor esperado sobre a hipótese de que ocorreu uma mudança no ponto $\nu = k$. Se ν possui uma distribuição de probabilidade, então podemos considerar o atraso esperado (*ED-expected delay*) que é dado por,

$$ED_\nu = \sum_{k=1}^{\infty} \pi(\nu = k) P(\tau \geq k) CED(k).$$

Existe uma extensa literatura estatística sobre métodos de monitoramento no tempo com importantes aplicações, onde o interesse é a detecção rápida e, a prevenção de vários eventos adversos ao interesse do processo em estudo. Nas seções seguintes, nós descreveremos alguns deles. Uma lista de referências sobre suas aplicações em variados problemas envolvendo vigilância epidemiológica podem ser encontrados em Sonesson and Book, 2003. Chamamos a atenção para nossa notação sobre o tempo de alarme τ para os métodos descritos a seguir. Por exemplo, denotamos por τ_{CUSUM} o tempo de alarme associado ao procedimento conhecido na literatura como somas cumulativas ou método CUSUM. A mesma notação é feita para o tempo de alarme dos outros procedimentos apresentados.

Considere nesta seção que $\{X_t\}_{t \geq 1}$ é uma sequência de variáveis aleatórias tal que para $t < \nu$, $X_1, X_2, \dots, X_{\nu-1}$ são i.i.d com densidade completamente especificada f_0 , e depois de ν , isto é $t \geq \nu$, $X_\nu, X_{\nu+1}, \dots$ são i.i.d com densidade, também completamente especificada, f_1 . ν é chamado ponto de mudança e f_1 é a densidade da sequência pós-mudança. Seja n o tempo

corrente e se não houve mudança na sequência dizemos que $\nu > n$ e denotamos pelo evento $\{\nu = \infty\}$.

2.2.1 Método de Somas Cumulativas (CUSUM)

O método CUSUM para monitoramento do processo é definido por (Page, 1954),

$$\tau_{\text{CUSUM}} = \inf \left\{ n \geq 1 : \text{Max}_{1 \leq k \leq n} \sum_{t=k}^n \log(f_1(X_t)/f_0(X_t)) > B_{\text{CUSUM}} \right\} \quad (2.1)$$

onde B_{CUSUM} é selecionado de modo que $ARL^0(\tau_{\text{CUSUM}}) = B_{\text{CUSUM}}$. Se não há mudanças definimos $\inf\{\emptyset\} = \infty$. Quando f é parcialmente especificada, pertencente a uma família paramétrica, Lorden (1971) propõe a generalização do CUSUM,

$$\tau_{\text{GCUSUM}} = \inf \left\{ n \geq 1 : \text{Max Sup}_{1 \leq k \leq n, \theta \in \Theta} \sum_{t=k}^n \log(f_\theta(X_t)/f_{\theta_0}(X_t)) > B_{\text{GCUSUM}} \right\}. \quad (2.2)$$

Neste caso, apenas propriedades assintóticas podem ser derivadas, tal como, $ARL^0(\tau_{\text{GCUSUM}}) \geq B_{\text{CUSUM}}$.

2.2.2 Métodos de Shiryaev(S) e Shiryaev-Roberts (SR)

Shiryaev (1963, 1978), formula o problema de detecção sequencial de um ponto de mudança ν em uma estrutura Bayesiana por descrever sua incerteza sobre ν através de uma distribuição a *priori* geométrica $P(\nu = k) = p(1-p)^{k-1}$, $k = 1, 2, \dots$, em que p é a probabilidade de que ocorra uma mudança no tempo $\nu = k$. Assumindo uma perda com valor igual a c para um atraso na detecção da mudança e uma perda igual a 1 para um falso alarme, Shiryaev mostra que a regra de Bayes, no sentido que minimiza o risco de Bayes para o problema (que é o valor esperado da função perda), soa um alarme de fora-de-controle, quando a probabilidade a *posteriori* de que tenha ocorrido uma mudança, excede algum valor fixado $0 < B_s < 1$. O tempo de alarme para este procedimento é dado por,

$$\tau_s = \inf \{ n \geq 1 : P(\nu \leq n | X_1, \dots, X_n) > B_s \} \quad (2.3)$$

Pode ser mostrado (Pollak, 1985) que a probabilidade a *posteriori* de que tenha ocorrido uma mudança até o tempo corrente n , é dada por

$$P(\nu \leq n | X_1, \dots, X_n) = \frac{S_{p,n}}{S_{p,n} + p^{-1}},$$

em que,

$$S_{p,n} = \sum_{k=1}^n \prod_{t=k}^n \frac{f_1(X_t)}{(1-p)f_0(X_t)}.$$

de modo que o procedimento pode ser redefinido, da seguinte forma

$$\tilde{\tau}_s = \inf \left\{ n \geq 1 : S_{p,n} > \tilde{B}_s \right\} \quad (2.4)$$

Seja,

$$SR_n = \lim_{p \rightarrow 0} S_{p,n} = \sum_{k=1}^n \prod_{t=k}^n \frac{f_1(X_t)}{f_0(X_t)},$$

Roberts(1966) modifica a regra (2.4) para,

$$\tau_{SR} = \inf \left\{ n \geq 1 : SR_n > B_{SR} \right\} \quad (2.5)$$

Este procedimento é chamado de Shiryaev-Roberts (SR). Observa-se nesse procedimento que $p \rightarrow 0$ implica que não há mudanças no processo e isto causa uma contradição. No entanto, Pollak (1985) mostra que este procedimento atinge, assintoticamente, o risco de Bayes quando $p \rightarrow 0$. Borovkov (1998) mostra que este resultado pode ser obtido para uma classe de distribuições a priori, na qual a geométrica com parâmetro p pequeno, mas $p > 0$, está inserida.

2.2.3 Mistura da Razão de Verossimilhanças (MRV)

Em vez de maximizar o logaritmo da razão de verossimilhança no procedimento CUSUM, Pollak and Siegmund (1975) integram a razão de verossimilhanças com respeito a alguma distribuição de probabilidade $H(\theta)$. Especificamente, eles propõem o procedimento chamado de "mistura de razão de verossimilhanças", dado por

$$\tau_{MRV} = \inf \left\{ n \geq 1 : \text{Max}_{1 \leq k \leq n} \left\{ \int_{\Theta} \left(\prod_{t=k}^n \frac{f_{\theta}(X_t)}{f_{\theta_0}(X_t)} \right) dH(\theta) \right\} > e^{B_{CUSUM}} \right\} \quad (2.6)$$

Estes autores mostram que $ARL^0(\tau_{MLR}) \geq e^{B_{CUSUM}}$.

Pollak (1987) assume no procedimento Shiryaev-Roberts que $f_1 = f_{\theta}$ com θ desconhecido e estende a regra de mistura de razão de verossimilhanças para o procedimento de Shiryaev-Roberts definindo um novo procedimeto,

$$\tau_{MRVSR} = \inf \left\{ n \geq 1 : \sum_{k=1}^n \left\{ \int_{\Theta} \left(\prod_{t=k}^n \frac{f_{\theta}(X_t)}{f_{\theta_0}(X_t)} \right) dH(\theta) \right\} > B_{SR} \right\} \quad (2.7)$$

Uma nota importante sobre as expressões (2.6) e (2.7) é que se substituirmos a notação $f_{\theta}(X_i)$ por $f(X_i|\theta)$ e $dH(\theta)$ por $h(\theta)d\theta$. Então, para um k fixo temos que

$$\left\{ \int_{\Theta} \left(\prod_{t=k}^n \frac{f(X_i|\theta)}{f(X_t|\theta_0)} \right) h(\theta)d\theta \right\} \quad (2.8)$$

representa do ponto de vista bayesiano o Fator de Bayes para compararmos os modelos

$$M_0 : f(X|\theta_0)$$

$$M_1 : \{f(X|\theta), h(\theta)\}.$$

2.2.4 Métodos Adaptativos

É bem conhecido na literatura,(veja por exemplo, Lai(1998)) que quando a sequência X_1, X_2, \dots é formada por variáveis i.i.d., sob a hipótese de que não houve mudança a razão de verossimilhanças dada por,

$$\prod_{t=k}^n \frac{f_{\theta_1}(X_t)}{f_{\theta_0}(X_t)} \quad (2.9)$$

é um martingal com valor esperado igual a 1 para todo $n \geq k$. Esta propriedade é preservada pela mistura de razão de verossimilhanças

$$\left\{ \int_{\Theta} \left(\prod_{t=k}^n \frac{f_{\theta}(X_t)}{f_{\theta_0}(X_t)} \right) dH(\theta) \right\}$$

Isto permite mostrar que $ARL^0(\tau_{MLRSR}) \geq B_{SR}$. No entanto, Lorden e Pollak (2005) argumentam que é difícil uma escolha para θ no caso multidimensional, então eles sugerem uma abordagem alternativa definindo,

$$\Lambda_{n,k} = \prod_{t=k}^n \frac{f_{\hat{\theta}_{t-1}}(X_t)}{f_{\theta_0}(X_t)}$$

em que $\hat{\theta}_{t-1}$ é um estimador para θ baseado nas observações X_1, \dots, X_{t-1} . Desde que $\hat{\theta}_{t-1}$ é um valor conhecido no tempo t , então são válidas as propriedades de (2.9). O método é adaptativo, no sentido de que para cada tempo t , θ é estimado com os dados obtidos até o tempo $t-1$. Desta forma, são definidos dois novos procedimentos de vigilância. O CUSUM-Adaptativo (AdC) dado por,

$$\tau_{AdC} = \inf \left\{ n \geq 1 : \text{Max}_{1 \leq k \leq n} \sum_{t=k}^n \log(f_{\hat{\theta}_{t-1}}(X_t)/f_{\theta_0}(X_t)) > B_{CUSUM} \right\} \quad (2.10)$$

e o Shiryaev-Robert Adaptativo (ASR),

$$\tau_{SRSR} = \inf \left\{ n \geq 1 : \sum_{k=1}^n \left(\prod_{i=k}^n \frac{f_{\hat{\theta}_{i-1}}(X_i)}{f_{\theta_0}(X_i)} \right) > B_{SR} \right\}. \quad (2.11)$$

Este último também é conhecido como Shiryaev-Robert-Robbins-Siegmund (SRSR), devido ao fato que Robbins e Siegmund (1970) foram os primeiros a tratar esse problema em testes de hipóteses.

2.2.5 Método Bayesiano

Nesta seção, nós apresentamos um método bayesiano para a detecção de mudanças em um grupo de observações usando um Fator de Bayes sequencial (ou razão de verossimilhanças preditivas). Este método, proposto por West(1986), realiza o monitoramento de uma série de dados através da comparação entre as predições para o modelo padrão contra as predições para um modelo alternativo sendo estes modelos atualizados no tempo. A idéia consiste em encarar o problema de detecção de mudança no processo como um problema de comparação entre dois modelos. Considere o processo $\{X_t : t = 1, 2, \dots\}$ em monitoramento com um modelo estatístico padrão M_0 no sentido de que não há mudanças. Este modelo é composto por:

$$M_0 = \{f(X_t|\theta_t), g_0(\theta_t|\mathcal{H}_{t-1}), p_0(X_t|\mathcal{H}_{t-1})\}$$

onde, $f(X_t|\theta_t)$ é a densidade condicional para a distribuição de $(X_t|\theta_t)$; $g_0(\theta_t|\mathcal{H}_{t-1})$ representa a distribuição a *priori* para θ_t dado o histórico da série \mathcal{H}_{t-1} , com $\mathcal{H}_t = \{\mathcal{H}_{t-1}, x_t\}$; $p_0(X_t|\mathcal{H}_{t-1})$ denota a densidade para a distribuição preditiva de $X_t|\mathcal{H}_{t-1}$ e,

$$p_0(X_t|\mathcal{H}_{t-1}) = \int f(X_t|\theta_t)g_0(\theta_t|\mathcal{H}_{t-1})d\theta_t.$$

Após observar $\mathbf{X}_t = \mathbf{x}_t$, a quantidade $p_0(x_t|\mathcal{H}_{t-1})$ é o valor da verossimilhança do modelo padrão e representa sua habilidade preditiva. Suponha que no tempo t , M_0 é comparado com um modelo alternativo

$$M_1 = \{f(X_t|\theta_t), g_1(\theta_t|\mathcal{H}_{t-1}), p_1(\mathbf{X}_t|\mathcal{H}_{t-1})\}.$$

Para verificar o ajuste preditivo dos modelos M_0 e M_1 , usamos a razão de verossimilhanças preditivas ou Fator de Bayes (FB)

$$FB_t = \frac{p_1(X_t|\mathcal{H}_{t-1})}{p_0(X_t|\mathcal{H}_{t-1})}. \quad (2.12)$$

Um valor $FB_t > 1$ implica em uma melhor performance preditiva de M_1 indicando uma possível mudança no processo. Chamamos a atenção para o fato que em West(1986) o Fator de Bayes é definido como $FB_t = p_0(X_t|\mathcal{H}_{t-1})/p_1(X_t|\mathcal{H}_{t-1})$. No entanto, para mantermos uma certa coerência nas definições usadas neste texto, nós realizamos esta inversão na notação de West(1986). Para detecção de mudanças no grupo formado pelas últimas k observações, West(1986) usa o Fator de Bayes sequencial,

$$FB_{n,k} = FB_n FB_{n-1} \dots FB_{n-k+1} = \prod_{t=n-k+1}^n FB_t, \quad k = 1, 2, \dots, n.$$

e o tempo de alarme pode ser definido como,

$$\tau_{FB} = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} FB_{n,k} \geq B_{FB} \right\}.$$

Maiores detalhes sobre propriedades, escolha das distribuições *a priori* para θ_t e uma variedade de aplicações pode ser vista em West(1986) e West e Harrison(1986).

2.3 Detecção de mudanças no espaço-tempo

Existem situações para as quais há uma necessidade de se realizar por um longo período de tempo uma vigilância espacial (por exemplo, os sistemas de vigilância discutimos na seção 1.1). Neste caso, podemos aplicar conjuntamente as técnicas descritas nas seções 2.1 e 2.2. Muitos métodos desenvolvidos para este propósito são de natureza retrospectiva, isto é, avaliam evidências em favor de *cluster* em um conjunto de dados com um número fixo de eventos passados. No entanto, com o surgimento de novos eventos, a maioria dos sistemas de vigilância são constantemente atualizados e a ocorrência de um aumento repentino na ocorrência de certos tipos de eventos em uma determinada região geográfica pode indicar, por exemplo o início de uma epidemia ou surto se os eventos estão relacionados à ocorrência de algum tipo de doença. Neste caso, o ideal seria usar um método de monitoramento capaz de detectar este provável *cluster* emergente (também chamado de *cluster* no espaço-tempo) o mais rápido possível, independentemente de sua localização e tamanho. Métodos para este tipo de problema são de natureza prospectiva, realizam análises repetidas sobre o conjunto de dados e detectam *cluster* em tempo real ou on-line. Ao mesmo tempo estes métodos devem ajustar para problemas de testes múltiplos realizados sobre o espaço e o tempo. A extensão de métodos de natureza puramente espacial para detectar *clusters* emergentes, pode dificultar o problema de ajuste para teste múltiplos. Isso porque ao usarmos a análise repetida de dados acumulados ao longo do tempo, qualquer método desenvolvido tem que ser ajustado para o problema de $t \times |\Xi|$ comparações no espaço-tempo. Onde $|\Xi|$ é a cardinalidade do espaço dos candidatos a *clusters* no espaço. Para contornar este fato, nós formulamos o problema de detecção de *cluster* emergente como um problema de detecção de ponto de mudança no espaço-tempo e, então, combinamos os métodos de natureza puramente temporal com os algoritmos para detecção de *clusters* espaciais formando assim uma classe de métodos para detecção de *cluster* emergentes.

2.3.1 Formulação do Problema de detecção de clusters no Espaço-Tempo

Considere que a área de estudo é representada por um mapa \mathcal{S} particionado em L -áreas A_l contíguas identificadas por um ponto s_l . Vamos assumir a partir de agora que s_l é o centróide do polígono que delinea a área A_l . Considere sobre \mathcal{S} o monitoramento (ou a vigilância) estatístico(a) de um processo estocástico $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots L\}$. A cada tempo discreto $t \geq 1$ observamos um vetor L -variado $\mathbf{X}_t = (X_t(s_1), X_t(s_2), \dots, X_t(s_L))'$, em que $X_t(s_l)$ representa alguma valor observado na l -ésima área. Sejam $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$ um conjunto conexo de localizações espaciais em \mathcal{S} , $\mathcal{H}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ o conjunto de dados acumulados

até o instante n . O objetivo do monitoramento é verificar se existe algum *cluster* emergente em \mathcal{S} . O interesse é detectar um *cluster* que começou em um tempo desconhecido $k \leq n$ e que está presente até o estágio atual n . O *cluster* no espaço-tempo é expresso em função de uma mudança no padrão da distribuição do sub-processo $\{X_t(s_l) \in \mathbf{X} : s_l \in \xi, t \geq k\}$ tal que $\xi^{n,k} = \{s_l \in \mathcal{S} : s_l \in \xi\} \times [k, n]$ representa o *cluster* no espaço-tempo também chamado de *cluster* emergente.

Definição 1 *Seja τ um tempo de alarme e ξ um conjunto desconhecido de áreas do mapa. Então, um procedimento para detecção e localização do *cluster* no espaço-tempo é representado por um par (τ, ξ) onde $\{\tau = \tilde{\tau} < \infty, \xi = \tilde{\xi}\}$ representa o evento que uma possível mudança ocorreu no tempo $\tilde{\tau}$ e localizada em $\tilde{\xi}$.*

Pela definição 1, um algoritmo para detecção e localização de um *cluster* no espaço-tempo tem que computar um par $\{\tau, \xi\}$ baseado em \mathcal{H}_n , e três erros podem ocorrer:

1. **Atraso:** Se o *cluster* é detectado e localizado corretamente ($\tilde{\xi} = \xi$) depois de ν tal que o tempo de alarme para $\tilde{\xi}$, $\tilde{\tau} \geq \nu$ é verdadeiro. Então o atraso na detecção localizada é $[\tilde{\tau} - \nu]$.
2. **Falso alarme:** Se o tempo da mudança detectado é $\tau < \nu$. Então teremos uma decisão final incorreta com $\tilde{\xi} \neq \xi$.
3. **Falsa localização:** O *cluster* é detectado e localizado depois de ν ($\tau \geq \nu$ é verdadeiro) porém $\tilde{\xi} = \xi^*$ mas $\xi \neq \xi^*$.

Dado estes erros, é muito difícil (ou até impossível) determinar um método que possa minimizar uma perda baseada nestes três tipos de erros. Na próxima seção, apresentamos alguns métodos propostos na literatura que usam como medida de evidência em favor do *cluster* emergente o valor-p do teste que não é ajustado para o problema de testes múltiplos. Estes testes representam extensões e variações da estatística Scan descrita na seção 2.1.4.

2.3.2 Estatística Scan espaço-tempo

Para detecção de *cluster* emergente, Kulldorff (2001) usa uma versão estendida da estatística Scan espacial descrita na seção 2.1.4. Para construir a estatística de Kulldorff(2001) considere que o interesse é a detecção de *cluster* em intervalo temporal $I_k = [k, n], k = 1, 2, \dots, n$, onde n é chamado de tempo corrente ou atual. Assuma que $X_t(s_l) \sim Poisson(\theta_t(s_l) \times \mu_t(s_l))$, Onde $\theta_t(s_l)$ denota uma medida desconhecida de risco relativo na l -ésima área no tempo t e $\mu_t(s_l)$ representa o número de eventos esperados condicional ao número total de eventos observados em

\mathcal{S} . Este valor esperado condicional é geralmente expresso por $\theta_t(s_l) = N_t(s_l)\lambda$, onde λ é taxa global de eventos e $N_t(s_l)$ representa o total de indivíduos em risco no tempo t na l -ésima. Está implícito neste modelo que o processo é mutuamente independente no espaço-tempo. Ou seja eventos que ocorrem no tempo t são considerados independentes de qualquer outro evento que tenha ocorrido no tempo $t - 1$, mesmo que tenham ocorrido na mesma área. Seja ξ um conjunto de localizações espaciais, como definido na seção 2.3.1. Para a estatística scan espaço-tempo é assumido que,

$$\theta_t(s_l) = \begin{cases} \theta(W) = \theta_0 & \text{se } (s_l, t) \in W = \xi \times I_k, \\ \theta(\bar{W}) = \theta_1 & \text{caso contrário.} \end{cases}$$

onde \bar{W} denota todo o domínio de estudo no espaço-tempo exceto W que representa uma janela espaço-temporal. As hipóteses consideradas são que para todo $W = \xi \times I_k$,

$$H_0 : \theta_0 = \theta_1 \quad \text{e} \quad H_1 : \theta_0 > \theta_1.$$

Para estas hipóteses, a estatística scan espaço-tempo T baseada na razão de verossimilhança generalizada sobre as hipóteses nula e alternativa e sobre todo W , é dada por (Kulldorff, 2001):

$$\begin{aligned} T &= \sup_{W \in \mathcal{W}} \left(\frac{X(W)}{\mu(W)} \right)^{X(W)} \left(\frac{X - X(W)}{X - \mu(W)} \right)^{X - X(W)} \\ &\times I \left\{ \left(\frac{X(W)}{\mu(W)} \right) > \left(\frac{X - X(W)}{X - \mu(W)} \right) \right\}, \end{aligned} \quad (2.13)$$

onde $I\{\}$ é a função indicadora, X é o número total de eventos observados no espaço tempo e $X(W)$ e $\mu(w)$ denotam, respectivamente, o número de eventos e o valor esperado para W dados por:

$$X(W) = \sum_{(s_l, t) \in W} X_t(s_l) \quad \text{e} \quad \mu(W) = \sum_{(s_l, t) \in W} \mu_t(s_l).$$

O conjunto $\hat{W} = \hat{\xi} \times I_{\hat{k}}$ para o qual a razão de verossimilhança é maximizada é identificado como o *cluster* emergente mais provável. Como no caso de detecção de *clusters* espaciais, a significância estatística do *cluster* detectado é feita através da simulação de Monte Carlo da seguinte forma: A cada tempo de análise t , um conjunto (por exemplo, 1000 réplicas) de dados aleatórios são gerados sob a hipótese H_0 com os eventos simulados distribuídos de forma independente no espaço e no tempo. Então o valor-p empírico (Dwass, 1957) para o teste é calculado e no momento atual, um *cluster* emergente observado é estatisticamente significativo ao nível α se seu valor-p é menor que α . Desta forma o ajuste é feito somente para os cilindros $W = \xi \times I_k$ atuais. No entanto se o monitoramento for continuado por um longo período de tempo, o teste não controla corretamente o valor de α para todas as análises anteriores. Com isso o método scan espaço-tempo para detecção de *clusters* emergentes pode apresentar alguns cilindros significativos mesmo que a hipótese nula seja verdadeira. Para entendermos uma varredura espaço-temporal

através de cilindros, considere a figura 2.3. Imagine uma pilha de mapas onde cada camada da pilha representa um tempo diferente. A base da varredura do cilindro representa espaço geográfico formado pelo conjunto de localizações espaciais em ξ e a altura do cilindro representa o tempo.

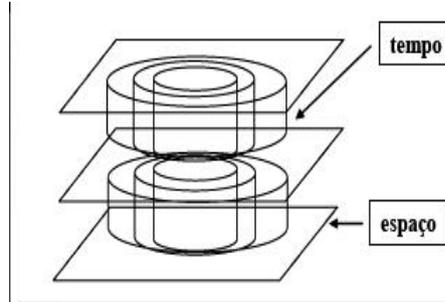


Figura 2.3: Exemplo visual do uso do método de Varredura espacial-temporal - Scan espaço-tempo

2.3.3 Estatística Scan espaço-tempo baseada no valor esperado

A estatística scan espço-tempo de kulldorff (2001) usa um valor esperado $\mu_t(s_l)$ condicional que é estimado com os dados do tempo atual. Neill et al.(2005) observou empiricamente que este fato pode diminuir o poder do teste para detectar *cluster* grandes (com muitas localizações espaciais) e também esta estimativa não é robusta na presença de tendência temporal ou sazonalidade. Para corrigir este problema, Neill (2009) propôs o uma estatística scan espaço-tempo baseada em uma janela temporal limitada $I_w = [n - w + 1, n]$ onde w é o máximo valor pré especificado para a altura do *cluster*. Assim, o scan espaço-tempo baseado no valor esperado é usado em dois passo: 1) O valor esperado $\mu_t(s_l)$ é estimado para cada localização espacial s_l com os dados acumulados até $n - w$. 2) utiliza a estatística scan espaço-tempo sobre o intervalo $I_w = [n - w + 1, n]$ para detectar mudanças no valor esperado (daí o nome scan espaço-tempo baseado no valor esperado) em um conjunto ξ . A vantagem deste procedimento é que o valor $\mu_t(s_l)$ pode ser ajustado para possíveis tendências ou sazonalidades observadas na série.

2.3.4 Scan espaço-tempo baseada no teste escore

Considere, novamente, que o interesse é a detecção de *clusters* emergentes em uma janela temporal limitada $I_w = [n - w + 1, n]$. Para corrigir um possível problema de sobredispersão temporal, Tango et al.(2010) derivam uma estatística scan espaço-tempo baseada em um teste escore usando a distribuição Binomial Negativa (BN). As hipóteses consideradas são as seguintes:

$$H_0 : X_t(s_l) \sim BN(\mu_t(s_l), \phi_t(s_l)) \quad \text{e} \quad H_1 : X_t(s_l) \sim BN(\theta_t(s_l)\mu_t(s_l), \phi_t(s_l))$$

onde,

$$P_0(X_t(s_l)) = P(X_t(s_l) = x_t(s_l)) = \frac{\Gamma(x_t(s_l) + \phi_t(s_l))}{\Gamma(\phi_t(s_l))x_t(s_l)!} \left(\frac{\phi_t(s_l)}{\mu_t(s_l) + \phi_t(s_l)} \right)^{\phi_t(s_l)} \left(\frac{\mu_t(s_l)}{\mu_t(s_l) + \phi_t(s_l)} \right)^{x_t(s_l)} \quad (2.14)$$

e por analogia,

$$P_1(X_t(s_l)) = \frac{\Gamma(x_t(s_l) + \phi_t(s_l))}{\Gamma(\phi_t(s_l))x_t(s_l)!} \left(\frac{\phi_t(s_l)}{\theta_t(s_l)\mu_t(s_l) + \phi_t(s_l)} \right)^{\phi_t(s_l)} \left(\frac{\theta_t(s_l)\mu_t(s_l)}{\theta_t(s_l)\mu_t(s_l) + \phi_t(s_l)} \right)^{x_t(s_l)} \quad (2.15)$$

Propriedades para $X_t(s_l)$ sob H_0 são,

$$\mathbb{E}_0(X_t(s_l)) = \mu_t(s_l), \quad \mathbb{V}_0(X_t(s_l)) = \mu_t(s_l) + \mu_t(s_l)^2/\phi_t(s_l) = \mu_t(s_l)\kappa_t(s_l)$$

onde a sobredispersão temporal é $\kappa_t(s_l) = 1 + \mu_t(s_l)/\phi_t(s_l)$. O par $(\mu_t(s_l), \phi_t(s_l))$ é estimado com os dados acumulados até o tempo $n - w$ e assumido conhecido em $I_w = [n - w + 1, n]$, o parâmetro $\theta_t(s_l)$. Para detectar e localizar *clusters* emergentes, é assumido que

$$\theta_t(s_l) = \begin{cases} h(u + \beta_W(t - n + w)) & \text{se } (s_l, t) \in W = \xi \times I_w \\ 1 & \text{caso contrário} \end{cases}$$

onde $h(u)$ denota o risco relativo no tempo $t = n - w$ é um risco inicial. Depois do tempo $n - w$ temos,

$$\left[\frac{\partial \theta_t(s_l)}{\partial t} \right]_{t=n-w} = \beta_W h'(u).$$

com $h(\cdot)$ denotando qualquer função monotona crescente e satisfazendo, $h(u) = 1$ e o par $(h'(\cdot), h''(\cdot))$ considerado finito. Deste modo o teste de hipótese fica a reduzido a testar sobre todos os possíveis conjuntos $W = \xi \times I_w$ as hipóteses:

$$H_0 : \beta_W = 0, \quad H_1 : \beta_W > 0.$$

A verossimilhança para β_W expressa em função de $\theta_t(s_l)$ é dada por

$$\mathcal{L}(\beta_W) = \prod_{t \in I_w} \prod_{s_l \in \xi} P_1(X_t(s_l)) \times \prod_{(s_l, t) \notin W} P_0(X_t(s_l)).$$

Com $P_0(X_t(s_l))$ e $P_1(X_t(s_l))$ dados, respectivamente, em 2.14 e 2.15. A estatística scan baseada no teste escore para a hipótese nula $H_0 : \beta_W = 0$ é obtido da seguinte forma: 1) Encontra-se a função escore e a informação de Fisher observada no ponto $\beta_W = 0$, ou seja

$$U(0) = \left[\frac{\partial \log \mathcal{L}(\beta_W)}{\partial \beta_W} \right]_{\beta_W=0}, \quad J(0) = -\mathbb{E} \left[\frac{\partial^2 \log \mathcal{L}(\beta_W)}{\partial \beta_W^2} \right]_{\beta_W=0}$$

2) Determine $S(W) = U(0)/\sqrt{J(0)}$. Então a estatística de teste é,

$$S = \sup_{W \in \mathcal{W}} S(W)$$

Pode-se mostrar (Tango, 2010) que

$$S(W) = \frac{\sum_{t \in I_w} \sum_{s_l \in \xi} (X_t(s_l) - \mu_t(s_l))(t - n + w)/\kappa_t(s_l)}{\sqrt{\sum_{t \in I_w} \sum_{s_l \in \xi} \mu_t(s_l)(t - n + w)^2/\kappa_t(s_l)}}$$

Para calcular o valor-p do teste use a aproximação (Tango, 2010) $S \approx N(0, 1)$.

3 *Algumas Abordagens para Detecção de Cluster no Espaço-Tempo Usando A Razão de Verossimilhança Adaptativa*

3.1 Introdução

Uma forma possível de monitoramento e detecção de *cluster* no espaço-tempo, é definir uma ampla classe de possíveis candidatos a *cluster* espacial e aplicar os métodos de natureza puramente temporal em cada um desses possíveis candidatos. No entanto, este procedimento pode levar ao problema de múltiplos testes e qualquer método desenvolvido deve ser ajustado para as múltiplas comparações no espaço-tempo. Com este propósito, nós desenvolvemos e apresentamos neste capítulo alguns novos métodos baseados nas razões de verossimilhanças (LR) adaptativas para a hipótese nula de ausência de *cluster* contra a hipótese alternativa que reflete a presença de um *cluster* emergente na região. Ao usarmos o fato de que a razão de verossimilhança adaptativa (ALR) preserva a estrutura martingal da razão de verossimilhança regular, nós encontramos uma cota superior para a taxa de falsos alarmes dos métodos propostos que depende somente da quantidade de candidatos avaliados. Isto implica que estes métodos não requerem procedimentos computacionais intensivos para determinação do ponto de corte que controla a taxa de alarmes falsos e por isso podem ser aplicado para detecção de *clusters* emergentes em qualquer mapa. Os métodos desenvolvidos podem ser entendido como uma extensão para o espaço-tempo dos métodos descritos na seção 4.4.2.

3.2 **Formulação, Notação, LR e ALR para detecção de clusters no espaço-tempo**

3.2.1 **Formulação do problema de detecção de cluster no espaço-tempo**

Considere que a área de estudo é representada por um mapa \mathcal{S} particionado em L -áreas A_l contíguas identificadas por um ponto s_l . Assuma que s_l é o centróide do polígono que delinea a área A_l . Seja $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots, L\}$ o processo sob vigilância. A cada tempo discreto $t \geq 1$ observamos um vetor L -variado $\mathbf{X}_t = (X_t(s_1), X_t(s_2), \dots, X_t(s_L))'$, em que $X_t(s_l)$

representa algum valor observado na l -ésima área. Sejam $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$ um conjunto conexo de localizações espaciais em \mathcal{S} , $\mathcal{H}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ o conjunto de dados acumulados até o instante n e $f(\mathbf{X}_t | \boldsymbol{\theta}^t) \triangleq f_{\boldsymbol{\theta}^t}(\mathbf{X}_t)$ uma função de densidade conhecida para \mathbf{X}_t , dado um valor selecionado para o parâmetro $\boldsymbol{\theta}^t$, para cada t , tal que $\boldsymbol{\theta}^t$ pertence a algum conjunto aberto Θ do espaço finito dimensional. O objetivo da vigilância é verificar se existe um *cluster* emergente no mapa. O interesse aqui, é detectar um *cluster* que começou em um tempo desconhecido $\nu \leq n$ e que está presente até o estágio atual n . O *cluster* no espaço-tempo é expresso em função de uma mudança no padrão espaço-temporal de eventos em ξ que emerge depois de um tempo desconhecido ν .

A detecção da mudança é obtida através da avaliação dos dados em \mathcal{H}_n sobre todos os candidatos a *cluster* ξ . O instante da mudança, é o primeiro momento no tempo para o qual existem evidências suficientes de que uma mudança no padrão de casos da doença tenha ocorrido. Matematicamente, o monitoramento é feito usando uma função de alarme $\Psi_n = \Psi(\mathcal{H}_n)$ e um limite de controle B , tais que o tempo de alarme, τ , de uma possível mudança é dada por,

$$\tau(B) = \min\{n : \Psi_n \geq B\}.$$

Se Ψ_n excede B , então há uma indicação de uma possível mudança no processo e um alarme é soado. Se não há mudanças no processo, todos os alarmes são falsos. O valor de B é especificado de forma a minimizar a ocorrência de alarmes falsos.

Em geral, ξ representa alguma forma geométrica regular como círculo (Kulldorff, 1997), elipse (kulldorff et al, 2006) ou irregular (Duczmal and Assunção, 2004) de modo que a cardinalidade de Ξ depende da forma e tamanho de ξ . Essa cardinalidade, pode ser excessivamente grande, e uma boa função para monitoramento de *clusters* emergentes deve ser ajustada para o problema de comparação múltiplas, ter um pequeno atraso na detecção e uma boa capacidade para localizar corretamente todo ou pelo menos uma grande parte do *cluster* emergente.

3.2.2 Notação

Para uma notação conveniente neste capítulo, denote o evento $\{\nu > n\} = \{\nu = \infty\}$, $\Xi = \{\xi_1, \xi_2, \dots, \xi_J\}$ onde cada ξ_j ($j = 1, 2, \dots, J$), definido em na seção 3.2.1, representa um conjunto de localizações espaciais em \mathcal{S} com os ξ'_j s todos distintos. Suponha que $\{\mathbf{X}_t\}_{t=1}^{\infty}$ é uma sequência de vetores independentes, com componentes mutuamente independentes no sentido de que $X_t(s_l)$ é independente de $X_t(s'_l)$ quando $l \neq l'$ e, ainda, $X_t(s_l)$ é independente de $X_{t-1}(s_l)$. Estas suposições são sempre satisfeitas quando \mathbf{X}_t é um processo de Poisson em dados de áreas e também podem ser relaxadas para outros modelos probabilísticos. Introduza as hipóteses:

$$H_0 : \nu = \infty \quad \text{e} \quad H_{k,j} : \nu = k, \xi = \xi_j$$

em que a hipótese H_0 é tal que o padrão normal de eventos corresponde a $\boldsymbol{\theta}^t = \{\theta_{0,l}^t : s_l \in \mathcal{S}\}$, e sob $H_{k,j}$ a distribuição dos eventos depois do tempo de mudança ν em ξ_j corresponde a $\boldsymbol{\theta}^t = \{\theta_{1,l}^t : s_l \in \xi\} \cup \{\theta_{0,l}^t : s_l \notin \xi\}$, com $\theta_{0,l}^t$ e $\theta_{1,l}^t$ fixados no tempo t . Isto é, são conhecidos no tempo t .

3.2.3 Detecção de clusters no espaço-tempo usando LR

Sejam \mathbb{F}_0 e $\mathbb{F}_{k,j}$ distribuições de probabilidades que correspondem, respectivamente, as hipóteses H_0 e $H_{k,j}$ tal que sob \mathbb{F}_0 a densidade de \mathbf{X}_t é dada por

$$f_0(\mathbf{X}_t) = \prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)), \quad \text{para todo } t \geq 1.$$

e sob $\mathbb{F}_{k,j}$, para qualquer $1 \leq k < \infty$, a densidade de \mathbf{X}_t é

$$f_{k,j}(\mathbf{X}_t) = \begin{cases} \prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)) & \text{se } t < k \\ \prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{X}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)) & \text{se } t \geq k \end{cases}$$

Ou seja, se a mudança ocorre no conjunto de localizações ξ_j no tempo $\nu = k$, o padrão de casos muda de $\theta_{0,l}^t$ para $\theta_{1,l}^t$ para todo $s_l \in \xi_j$. Como os vetores são mutuamente independentes e para qualquer ξ_j

$$\prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)) = \prod_{s_l \in \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)),$$

temos que a razão de verossimilhanças (LR) para os dados \mathcal{H}_n entre as hipóteses $H_{k,j}$ e H_0 é dada por

$$LR_{n,k}^j = \frac{\mathbf{P}(\mathcal{H}_n | H_{k,j})}{\mathbf{P}(\mathcal{H}_n | H_0)} = \frac{\prod_{t=k}^n \prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{X}_t(s_l))}{\prod_{t=k}^n \prod_{s_l \in \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l))}, \quad k \leq n. \quad (3.1)$$

Uma propriedade importante de $LR_{n,k}^j$, é que para quaisquer ξ_j e $(\theta_{0,l}^t, \theta_{1,l}^t)$ fixos, sob \mathbb{F}_0 a sequência $\{LR_{n,k}^j, n \geq k\}$ é um processo martingale com valor esperado igual a 1. Para mostrar essa propriedade, observe primeiramente que $LR_{n,k}^j$ pode ser escrita como

$$LR_{n,k}^j = \frac{\prod_{t=k}^n \prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{X}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l))}{\prod_{t=k}^n \prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l))} = \frac{\prod_{t=k}^n f_{k,j}(\mathbf{X}_t)}{\prod_{t=k}^n f_0(\mathbf{X}_t)}$$

sejam \mathbb{E}_0 o valor esperado sobre \mathbb{F}_0 e

$$\Lambda_t^j = \frac{\prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{X}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l))}{\prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l))}.$$

Note que para todo t ,

$$\mathbb{E}_0[\Lambda_t^j] = \sum_{\mathbf{x}_t} \left(\frac{\prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{x}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{x}_t(s_l))}{\prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{x}_t(s_l))} \times f_0(\mathbf{x}_t) \right). \quad (3.2)$$

onde a soma em (3.2) é realizada sobre todos os possíveis valores para o vetor \mathbf{X}_t . Mas,

$$f_0(\mathbf{X}_t) = \prod_{s_l \in \mathcal{S}} f_{\theta_{0,l}^t}(\mathbf{X}_t(s_l)).$$

de modo que,

$$\mathbb{E}_0[\Lambda_t^j] = \sum_{\mathbf{x}_t} \prod_{s_l \in \xi_j} f_{\theta_{1,l}^t}(\mathbf{x}_t(s_l)) \times \prod_{s_l \notin \xi_j} f_{\theta_{0,l}^t}(\mathbf{x}_t(s_l)) = \sum_{\mathbf{x}_t} f_{k,j}(\mathbf{x}_t) = 1, \quad t \geq k. \quad (3.3)$$

e pela mutua independência dos vetores \mathbf{X}_t^l s, segue que

$$\mathbb{E}_0[LR_{n,k}^j] = \mathbb{E}_0 \left[\prod_{t=k}^n \Lambda_t^j \right] = \prod_{t=k}^n \mathbb{E}_0[\Lambda_t^j] = 1.$$

Agora, temos que

$$\begin{aligned} \mathbb{E}_0[LR_{n+1,k}^j | \mathbf{X}_n, \dots, \mathbf{X}_k] &= \mathbb{E}_0[LR_{n,k}^j \times \Lambda_{n+1}^j | \mathbf{X}_n, \dots, \mathbf{X}_k] \\ &= LR_{n,k}^j \mathbb{E}_0[\Lambda_{n+1}^j | \mathbf{X}_n, \dots, \mathbf{X}_k] \\ &= LR_{n,k}^j \mathbb{E}_0[\Lambda_{n+1}^j] \quad (\text{pela independência}) \\ &= LR_{n,k}^j, \quad (\text{por (3.3)}). \end{aligned} \quad (3.4)$$

Portanto, por (3.3) e (3.4), segue que $\{LR_{n,k}^j\}_{n \geq k}$ é um processo martingal com média 1. Esta propriedade de (3.1) será importante na elaboração do procedimento desenvolvido para detecção de *clusters* emergentes.

Para construirmos o algoritmo para detecção e localização do *cluster*, primeiramente definimos,

$$LR_n^j = \max_{1 \leq k \leq n} LR_{n,k}^j \quad (3.5)$$

de modo que LR_n^j representa o máximo para a LR sobre todos os possíveis pontos no tempo para o qual tenha ocorrido uma mudança em ξ_j . Seja $k_j = \arg[LR_n^j]$, então os possíveis candidatos a *clusters* emergentes são $\xi_1^{n,k_1}, \dots, \xi_J^{n,k_J}$, e como estratégia para detecção considere os seguintes procedimentos:

- Procedimento 1: MAX-LR.

$$\tau_1(B) = \min \left\{ n \geq 1 : J^{-1} \max_{\{\xi_1^{n,k_1}, \dots, \xi_J^{n,k_J}\}} [LR_n^j] \geq B \right\}. \quad (3.6)$$

Ou seja, o tempo de alarme, é o primeiro momento no tempo tal que a estatística

$$J^{-1} \max_{\{\xi_1^{n,k_1}, \dots, \xi_J^{n,k_J}\}} [LR_n^j] = J^{-1} \max_{\xi_j \in \Xi} \max_{1 \leq k \leq n} LR_{n,k}^j$$

excede o valor B que controla a taxa de alarmes falsos (**FAR**). É bom ressaltar que $\tau_1(B)$ é uma natural escolha para detecção de *clusters* emergentes (Sonesson, 2007) pois diferentes regiões poderão ter diferentes pontos de mudança. Note que o limite de controle B é assumido igual para todos os candidatos a *cluster* emergente. O fato é que o valor de B depende da taxa de alarmes falsos e um valor B comum para todos os candidatos garante que esta taxa seja igual para todos. Se existe um n tal que o evento $\left\{ J^{-1} \max_{\{\xi_1^{n,k_1}, \dots, \xi_J^{n,k_J}\}} [LR_n^j] \geq B \right\}$ ocorre, estime o *cluster* emergente por

$$\xi^{n,\hat{k}} = \arg \left(J^{-1} \max_{\{\xi_1^{n,k_1}, \dots, \xi_J^{n,k_J}\}} [LR_n^j] \right).$$

Caso contrário, defina $\tau_1(B) = \infty$ para $\min\{\emptyset\}$. A escolha $\xi^{n,\hat{k}}$ é baseada no fato de que, se o *cluster* começou no tempo \hat{k} e terminou em $n' < n$ ele deveria ter sido detectado pelo método em n' . Desde de que $\tau_1(B)$ é baseado no valor máximo de LR_n^j , ele é denominado MAX-LR.

- Procedimento 2: MIX-LR.

$$\tau_2(B) = \min \left\{ n \geq 1 : J^{-1} \sum_{j=1}^J LR_n^j \geq B \right\}. \quad (3.7)$$

Diferentemente de $\tau_1(B)$, o procedimento $\tau_2(B)$ não requer que todos os candidatos tenham o mesmo limite de controle. Mais precisamente, $\tau_2(B)$ é baseado em uma mistura de LR_n^j com respeito à distribuição de probabilidade $\pi_j = P(\xi = \xi_j) = J^{-1}$ e por isso é denominado MIX-LR. Note ainda que

$$\begin{aligned} \sum_{j=1}^J LR_n^j &= \max_{1 \leq k \leq n} LR_{n,k}^1 + \max_{1 \leq k \leq n} LR_{n,k}^2 + \dots + \max_{1 \leq k \leq n} LR_{n,k}^J. \\ &= LR_{n,k_1}^1 + LR_{n,k_2}^2 + \dots + LR_{n,k_J}^J. \end{aligned} \quad (3.8)$$

onde $k_j = \arg[LR_n^j]$.

Então, grandes valores de $LR_n^j = \max_{1 \leq k \leq n} LR_{n,k}^j$ possuem maior contribuição para que soe o alarme de modo que se existe um n tal que o evento $\left\{ J^{-1} \sum_{j=1}^J LR_n^j \geq B \right\}$ ocorre, estime o *cluster* por

$$\xi^{n,k_j} = \operatorname{argmax} \left(J^{-1} \sum_{j=1}^J LR_n^j \right).$$

Caso contrário, defina $\tau_2(B) = \infty$ para $\min\{\emptyset\}$.

- Procedimento 3: WEIGHT-LR

$$\tau_3(B) = \min \left\{ n \geq 1 : \sum_{j=1}^J w_j LR_n^j \geq B \right\}. \quad (3.9)$$

Caso contrário, defina $\tau_3(B) = \infty$ para $\min\{\emptyset\}$. Onde w_j são valores conhecidos, representam pesos de evidências em favor do candidato a *cluster* ξ^{n,k_j} e $\sum_{j=1}^J w_j = 1$. Neste caso podemos usar a idéia de atualização de informação. Diferentes pesos podem ser atribuídos baseado na informação a prior que dispomos sobre todos os possíveis candidatos. Os pesos podem ser atualizados quando uma nova informação é obtida. Observe que $\tau_2(B)$ é um caso de particular de $\tau_3(B)$ com $w_j = J^{-1}$. Neste caso, o *cluster* emergente é estimado por

$$\xi^{n,k_j} = \arg \left(J^{-1} \sum_{j=1}^J w_j LR_n^j \right).$$

Observe que este *cluster* estimado pode ser diferente do obtido com o tempo de alarme τ_2 .

Taxa de Falso Alarme para MAX, MIX and WEIGHT-RL

Os procedimentos para detecção e monitoramento de *cluster* no espaço-tempo baseada em (3.6), (3.7) e (3.9) representam um tempo aleatório de alarme que dependem de \mathcal{H}_n . Um bom procedimento deve ter uma pequena taxa de alarmes falsos (**FAR**) e um pequeno atraso na detecção do *cluster*. Seja $\mathbb{E}_0[\tau(B)]$ o valor esperado para $\tau(B)$ sobre \mathbb{F}_0 , temos que a **FAR** $[\tau(B)]$ é expressa por (Tartakovsky, 2006),

$$\mathbf{FAR}[\tau(B)] = 1/\mathbb{E}_0[\tau(B)]. \quad (3.10)$$

A quantidade **FAR** $[\tau(B)]$ é interpretada como a frequência média de alarmes falsos.

Theorem 2 *No tempo atual n , seja J fixo e conhecido a cardinalidade de Ξ . Então*

$$\mathbf{FAR}[\tau_i(B)] \leq 1/B, \text{ com } i = 1, 2, 3. \quad (3.11)$$

Para mostrar (3.11), defina

$$\Psi_n^j = \sum_{k=1}^n LR_{n,k}^j, \quad j = 1, 2, \dots, J$$

Note que Ψ_n^j é a estatística de Shiriyayev-Roberts definida na seção 2.2.2 para cada ξ_j . Como sob \mathbb{F}_0 a sequência $\{LR_{n,k}^j, n \geq k\}$ é um martingal com valor esperado igual a 1, temos que $\{(\Psi_n^j - n)\}_{n \geq 1}$ é um martingal com média zero. Isto implica que sob \mathbb{F}_0 a sequência $\{J^{-1} \sum_{j=1}^J \Psi_n^j - n\}_{n \geq 1}$ também é um martingal com média zero. Agora, considere dois outros procedimentos

para detecção de *clusters* emergentes dados por,

$$\tau_4(B) = \min \left\{ n \geq 1 : J^{-1} \sum_{j=1}^J \Psi_n^j \geq B \right\}. \quad (3.12)$$

e,

$$\tau_5(B) = \min \left\{ n \geq 1 : \sum_{j=1}^J w_j \Psi_n^j \geq B \right\}. \quad (3.13)$$

Então,

$$0 = \mathbb{E}_0 \left(\sum_{j=1}^J w_j \Psi_{\tau_5(B)}^j - \tau_5(B) \right) \geq B - \mathbb{E}_0[\tau_5(B)]$$

portanto $\mathbb{E}_0[\tau_5(B)] \geq B$ e por analogia $\mathbb{E}_0[\tau_4(B)] \geq B$. No entanto, $\tau_1(B) \geq \tau_4(B)$ e $\tau_3(B) \geq \tau_5(B)$. Logo

$$\mathbb{E}_0[\tau_1(B)] \geq \mathbb{E}_0[\tau_4(B)] \geq B \quad \text{e} \quad \mathbb{E}_0[\tau_3(B)] \geq \mathbb{E}_0[\tau_5(B)] \geq B$$

e por (3.10) obtemos

$$\mathbf{FAR}[\tau_i(B)] \leq 1/B, \text{ com } i = 1, 2, 3.$$

Para mostrar que este resultado é válido também para $\tau_2(B)$, basta observar que $\tau_2(B)$ é um caso particular de $\tau_3(B)$. O importante desse resultado é apresentado no lema a seguir:

Lemma 3 *Seja $\Delta = \{\tau : \mathbf{FAR}[\tau] \leq \alpha\}$, para algum $0 < \alpha < 1$. Se $B = 1/\alpha$. Então $\tau_i(B) \in \Delta$, $i = 1, 2, 3$.*

onde $0 < \alpha < 1$ é a taxa aceitável de alarmes falsos para o procedimento baseado em $\tau_i(B)$. Por exemplo, se o interesse é detectar *clusters* emergentes usando o procedimento $\tau_i(B)$ com uma taxa de alarmes falsos inferior a 0.05, faça $B = 20$.

3.2.4 LR Adaptativa para detecção de clusters no espaço-tempo

Embora na seção 3.2.3 nós assumimos que o par $(\theta_{0,l}^t, \theta_{1,l}^t)$ é conhecido, isso pode não ocorrer na prática, pois uma especificação a priori para $(\theta_{0,l}^t, \theta_{1,l}^t)$ sobre todos os possíveis ξ'_j s é praticamente impossível. Para contornar este problema, usamos uma abordagem adaptativa (Lorden and Pollak, 2005). Seja $(\hat{\theta}_{0,l}^{[k,t-1]}, \hat{\theta}_{1,l}^{[k,t-1]})$ uma estimativa razoável para $(\theta_{0,l}^t, \theta_{1,l}^t)$ baseado em $\{\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_{t-1} = \mathbf{x}_{t-1}\}$ e defina a razão de verossimilhança adaptativa (ALR) por

$$ALR_{n,k}^j = \frac{\prod_{t=k}^n \prod_{s_l \in \xi_j} f_{\hat{\theta}_{1,l}^{[k,t-1]}}(X_t(s_l))}{\prod_{t=k}^n \prod_{s_l \in \xi_j} f_{\hat{\theta}_{0,l}^{[k,t-1]}}(X_t(s_l))}. \quad (3.14)$$

Por requerer que $(\hat{\theta}_{0,l}^{[k,n]}, \hat{\theta}_{1,l}^{[k,n]})$ não dependa de \mathbf{X}_n temos que (Lorden and Pollak, 2003) sob \mathbb{F}_0 , a sequência

$$\{ALR_{n,k}^j, n \geq k\} \quad (3.15)$$

preserva a estrutura martingal de (3.1), de forma que o limite superior para a taxa de alarmes falsos dado em (3.11) pode ser utilizado pelos procedimentos para detecção de *clusters* emergentes representados por $\tilde{\tau}_i(B)$ com $i = 1, 2, 3$. Onde o procedimento $\tilde{\tau}_i(B)$ é obtido através de $\tau_i(B)$ pela simples substituição de $LR_{n,k}^j$ por $ALR_{n,k}^j$. O fato importante desta abordagem é que para cada valor fixado do parâmetro no tempo t , assumimos que o par $(f_{\hat{\theta}_{0,l}^{[k,t-1]}}, f_{\hat{\theta}_{1,l}^{[k,t-1]}})$ é uma boa aproximação para $(f_{\theta_{0,l}^t}, f_{\theta_{1,l}^t})$ de modo que $LR_{n,k}^j$ é aproximada por $ALR_{n,k}^j$.

Desde de que $\tilde{\tau}_i(B)$ é função da razão de verossimilhanças adaptativas (ALR), os métodos $\tilde{\tau}_1(B)$, $\tilde{\tau}_2(B)$ e $\tilde{\tau}_3(B)$ são denominados, respectivamente, MAX-ALR, MIX-ALR e WEIGHT-ALR. Esses métodos são adaptativos no sentido de que para cada tempo t , os parâmetros $(\theta_{0,l}^t, \theta_{1,l}^t)$ são estimados com os dados obtidos até o tempo $t - 1$. Essa abordagem adaptativa não é apenas um artifício teórico, pois adicionando a n -ésima observação do vetor \mathbf{X}_n na estimação de $\theta_{1,l}^n$, podemos ultrapassar o limite superior dado em (3.11), necessitando de um maior nível de corte, o que evitaria qualquer ganho em tempo de detecção devido à estimativa um pouco melhor de $(\theta_{0,l}^t, \theta_{1,l}^t)$. Embora em $\tau_3(B)$ os valores para w_j sejam assumidos conhecidos, na prática eles são desconhecidos. Porém, defina $w_j^{[0]} = w_j^{[1]} = J^{-1}$ e, para $n \geq 2$, considere uma estimativa adaptativa para w_j dada por

$$w_j^{[n-1]} = (1 - \epsilon)w_j^{[n-2]} + \epsilon \frac{ALR_{n-1}^j}{\sum_{j=1}^J ALR_{n-1}^j}$$

onde $0 < \epsilon \leq 1$ é um fator adaptativo usado para aumentar a sensibilidade do método quando pequenas mudanças surgem em um passado bem recente. Desde de que w_j é conhecido no tempo n , o procedimento

$$\tilde{\tau}_3(B) = \min \left\{ n \geq 1 : \sum_{j=1}^J w_j ALR_n^j \geq B \right\}. \quad (3.16)$$

também pertence a classe Δ com w_j substituído por $w_j^{[n-1]}$. Note que sob esta formulação, os pesos w_j são atualizados sequencialmente a medida que dispomos de uma nova informação.

3.2.5 Representação Espacial dos Clusters

Em uma estrutura geral, considere que $S = \{s_1, s_2, \dots, s_L\}$ é um mapa particionado (como na figura 3.1) em L áreas conexas em que $s_l = (s_{1,l}, s_{2,l})$ é um vetor de coordenadas geográficas do centróide da l -ésima área. Para construirmos o espaço dos *clusters* Ξ , primeiramente considere uma matriz de adjacências,

$$I_{l,j} = \begin{cases} l & \text{se } j = 1 \\ c & \text{se } s_c \text{ é o } j\text{-ésimo centróide mais próximo de } s_l \end{cases}$$

Suponha que a proximidade é medida pela distância Euclidiana entre os centróides s_l e s_c . Cada linha desta matriz representa um vetor de adjacências (l_1, l_2, \dots, l_L) em que, $l_1 = j$ e nas outras celas desta matriz, os elementos são alocados de acordo com a proximidade espacial em relação ao centróide s_l . Por exemplo, suponha que $L = 5$. Se a linha 1 é o vetor $(1, 5, 4, 3, 2)$. Isto implica que o centróide s_5 é o segundo mais próximo de s_1 , s_4 é o terceiro centróide mais próximo de s_1 e assim sucessivamente. Nesta representação está implícito que "todo centróide é o mais próximo de si mesmo". Centrado em s_l (que corresponde a l -ésima linha de $I_{l,j}$), identifique um *cluster* por um vetor $\xi_{l_i} = (l_{[i,1]}, l_{[i,2]}, \dots, l_{[i,L]})$ construído da seguinte forma:

- 1) Defina $l_{[i,j]} = 1$ se $l = j$, $i = 1, 2, \dots, L$ e $j = 1, 2, \dots, L$.
- 2) Defina $l_{[i,I_{l,j}]} = 1$, se $s_{I_{l,j}}$ é um dos j centróides mais próximos de s_l e $j \leq i$. Caso contrário faça $l_{[i,I_{l,j}]} = 0$.

Por exemplo, suponha que $S = \{s_1, s_2, s_3, s_4, s_5\}$ e que sua matriz de adjacência é tal que na linha 1 temos o vetor de adjacências $(s_1, s_5, s_4, s_3, s_2)$. Então, usando a representação para ξ_{l_i} , temos $\xi_{1_1} = (1, 0, 0, 0, 0)$, $\xi_{1_2} = (1, 0, 0, 0, 1)$, $\xi_{1_3} = (1, 0, 0, 1, 1)$, $\xi_{1_4} = (1, 0, 1, 1, 1)$ e $\xi_{1_5} = (1, 1, 1, 1, 1)$. Observe que a cada valor de j , ξ_{l_i} recebe o valor 1 no índice do vizinho mais próximo de s_l em sua posição original no espaço. Esta representação é única a menos do cluster ξ_{l_L} que surge L vezes diferenciado apenas pelo seu centróide. Para verificarmos que de fato esta é a representação dos *clusters*, exemplificamos a formação da representação de ξ_{1_2} . Neste caso $l = 1, i = 2$ e $j = 1, 2, \dots, L$.

- 1) Quando $j = 1, l = j$ e portanto $1_{[1,1]} = 1$
- 2) Quando $j = 2, I_{1,2}=5$, e s_5 é o segundo ($i=2$) centróide mais próximo de s_l e também, $j = 2 \leq 2 = i$. Portanto, $1_{[1,5]} = 1$. Agora note que para todo $j \geq 3$ temos que $j > i$. Assim não satisfaz a condição $j \leq i$ implicando que as outras coordenadas de ξ_{1_2} são nulas. Portanto $\xi_{1_2} = (1, 0, 0, 0, 1)$.

Para formar a representação do *cluster*, este procedimento é repetido para todo $\xi_{l_i}, (l, i) = 1, 2, \dots, L$. Depois deste procedimento, a coleção de *clusters* no espaço pode ser construída em dois passos: primeiro definimos uma classe auxiliar $\tilde{\Xi} = \{\xi_{l_i} : l, i = 1, 2, \dots, L\}$ O número total de cluster em $\tilde{\Xi}$ é L^2 ; em seguida definimos

$$\Xi = \{\xi_{l_i} \in \tilde{\Xi} : (\langle \xi_{l_i}, \xi_{l_i} \rangle) \leq a\} \quad (3.17)$$

onde $(\langle . \rangle)$ denota o produto interno entre dois vetores e a é um valor fixo, que representa uma restrição na quantidade (tamanho do *cluster*) de localizações espaciais em ξ_{l_i} . Esta restrição é imposta porque muitas vezes não há interesse em detectar *clusters* geográficos com um número de áreas maior que a . Por exemplo, se $a = L/2$, então o maior número de localizações espaciais

em ξ_{l_i} é exatamente $L/2$ e $J = |\Xi| = L^2/2$. Como a distância euclidiana é usada como medida de proximidade, então a representação geométrica de ξ_{l_i} é um círculo cujo o raio é igual $r_{l_i} = \max_{s_i \in \xi_{l_i}} [d(s_l, s_i)]$ onde $d(\cdot)$ é a distância euclidiana.

3.2.6 Algoritmo para detecção de clusters emergentes no espaço tempo

Após a formulação da estrutura dos espaço dos *cluster*, podemos usar o seguinte algoritmo para detecção e monitoramento de *cluster* no espaço tempo:

- **Algoritmo**

1. no tempo $n = 0$, Escolher um centróide no mapa.
2. Calcular as distâncias para os outros centróides. Ordene estas distâncias em ordem crescente e depois guarde-as em um vetor de distâncias (este vetor de distâncias é um vetor de adjacências).
3. Para cada centróide do mapa repetir os passos 1 e 2 e então construa a matriz de adjacências de acordo com a explicação dada na seção 3.2.5.
4. Escolher novamente um centróide s_l na região. Centrado em s_l construa $\xi_{l_i}, i = 1, 2, \dots, L$ como explicado na seção 3.2.5.
5. Escolha um valor para a (por exemplo, $a = L/4$) e construa a classe

$$\Xi = \{\xi_{l_i} : (\langle \xi_{l_i}, \xi_{l_i} \rangle) \leq a, (l, i) = 1, 2, \dots, L\}$$

6. Para cada $n \geq 1$ especifique um valor para α e calcule o par $(f(\text{ALR}_n), B)$, onde f é qualquer função de monitoramento descrita na seção 3.2.4.
7. Se $f(\text{ALR}_n) \geq B$, declare que existe um *cluster* emergente no mapa e o estime pelo $\text{argmax}[f(\text{ALR}_n)]$ como dado na seção 3.2.3 . Caso contrário, incremente uma unidade no tempo e volte ao passo 6.

3.3 Aplicação

Para um melhor entendimento do funcionamento dos métodos propostos eles são aplicados a um conjunto de dados reais. Os dados são referentes ao número de casos de câncer da tireóide em homens que ocorreram no Novo México durante o período de tempo 1973-1992. Os dados usados nesta análise (Kulldorff, 2001) foram agregados em 32 áreas. A incidência de casos é de 2.4 para cada grupo de 100000 pessoas-ano no Novo México, porém variando entre as áreas de acordo com idade e raça. O objetivo desta análise é a detecção de *clusters* emergentes de câncer da tireóide na região do Novo México cujo a partição é apresentado na figura 3.1.

O modelo de cluster emergente

Assuma que $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_L))'$ é um processo de Poisson. Em que $X_t(s_l)$ é o número observado de casos e $\mu_t(s_l)$ o número de casos esperados, no tempo t na l -ésima área A_l identificada por seu centróide s_l representada espacialmente na figura 3.1 com $l = 1, 2, \dots, L = 32$. Nesta aplicação, usamos Ξ como o espaço dos *cluster* dado em (3.17) com $a = L/4 = 32/4 = 8$. De modo que $J = L^2/4 = 256$.

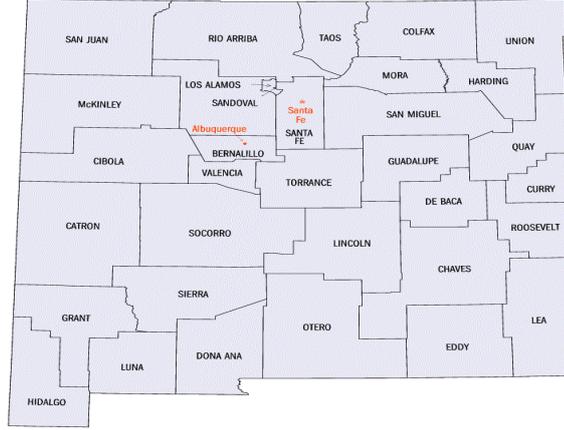


Figura 3.1: Mapa do Novo México. Fonte: US Census.

Sob H_0 temos

$$f_0(\mathbf{X}_t) = \prod_{s_l \in \mathcal{S}} \mu_t(s_l)^{X_t(s_l)} \exp\{-\mu_t(s_l)\} / X_t(s_l)!, \quad t \geq 1$$

e sob $H_{k,j}$

$$f_{k,j}(\mathbf{X}_t) = \prod_{s_l \in \mathcal{S}} \left(\mu_t(s_l) \delta_{t,j}^{I_{\{t \geq k, s_l \in \xi_j\}}} \right)^{X_t(s_l)} \exp\left\{ -\mu_t(s_l) \delta_{t,j}^{I_{\{t \geq k, s_l \in \xi_j\}}} \right\} / X_t(s_l)!.$$

Onde $I_{\{\mathbb{A}\}}$ é a função indicadora do evento \mathbb{A} . Ou seja, sob a ausência de *cluster* $X_t(s_l)$ tem valor esperado $\mu_t(s_l)$ para todo $s_l \in \mathcal{S}$ e $t \geq 1$, e depois de um tempo desconhecido $\nu = k$, $1 \leq k \leq n$, o processo muda para um novo estado com nível médio $\delta_{t,j} \mu_t(s_l)$ em um conjunto ξ_j . A quantidade $\delta_{t,j} > 1$ representa o risco relativo ou a intensidade da mudança na região do *cluster* ξ_j . Observe, ainda, que $\theta_{0,l}^t = \mu_t(s_l)$ e $\theta_{1,l}^t = \delta_{t,j} \mu_t(s_l)$. Usando (3.14) a LR adaptativa entre os tempos k e n para processo é dada por

$$ALR_{n,k}^j = \prod_{t=k}^n \left(\delta_j^{[k,t-1]} \right)^{X_t(j)} \exp\{-\mu_j^{[k,t-1]} (\delta_j^{[k,t-1]} - 1)\}. \quad (3.18)$$

onde,

$$X_t(j) = \sum_{s_l \in \xi_j} X_t(s_l) \quad \mu_t(j) = \sum_{s_l \in \xi_j} \mu_t(s_l).$$

e $(\mu_j^{[k,t-1]}, \delta_j^{[k,t-1]})$ representam as estimativas para $(\mu_t(j), \delta_{t,j})$ que dependem dos dados através

de $\{\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_{t-1} = \mathbf{x}_{t-1}\}$.

Escolha das Estimativas

Observa-se que $\mu_t(j) = \mathbb{E}_0[X_t(j)]$ representa o valor esperado para o número de casos que ocorrem no tempo t em ξ_j sob a suposição que não existe *cluster* emergente, ou seja, não depende de k . Então, para $t > 1$, uma estimativa razoável para $\mu_t(j)$ é dada por,

$$\mu_j^{[k,t-1]} = \mu_j^{[t-1]} = \frac{1}{t-1} \left(\sum_{m=1}^{t-1} N_m(j) \lambda_m \right) \triangleq \frac{1}{t-1} \sum_{m=1}^{t-1} \mu_j^{[m]}. \quad (3.19)$$

Onde, $N_m(j)$ representa a população em risco em ξ_j no tempo m e λ_m é uma estimativa no tempo m para taxa global da doença.

Para ajustarmos o número de casos esperados por idade e sexo, defina $x_{m,i,r}(s_l)$ como sendo o número de casos que ocorreram no tempo m ($1 \leq m \leq t-1$) entre as pessoas na classe de idade i pertencente a raça r na l -ésima área, e $N_{m,i,r}(s_l)$ a respectiva população em risco. Então, a estimativa para taxa global de casos em todo o mapa referente a classe de idade i na raça r no tempo m é dada por $\lambda_{m,i,r} = \sum_{s_l \in \mathcal{S}} x_{m,i,r}(s_l) / \sum_{s_l \in \mathcal{S}} N_{m,i,r}(s_l)$. Logo $\mu_{m,i,r}(s_l) = N_{m,i,r}(s_l) \lambda_{m,i,r}$ representa uma estimativa para o número de casos esperados no tempo m na classe de idade i e raça r sob a hipótese de que não existe cluster no espaço-tempo. Portanto, uma estimativa para o número total de casos esperados na l -ésima área no tempo m é dado por $\mu^{[m]}(s_l) = \sum_i \sum_r \mu_{m,i,r}(s_l)$ e conseqüentemente $\mu_j^{[m]} = \sum_{s_l \in \xi_j} \mu^{[m]}(s_l)$.

Para obtermos uma estimativa para $\delta_{t,j}$, notamos que sob $H_{k,j}$, o estimador de máxima verossimilhança de $\delta_{t,j}$ é dado por $X_t(j) / \mu_t(j)$ e quando $\delta_{t,j} = \delta_j$ este estimador é dado por $\sum_{m=k}^t X_m(j) / \sum_{m=k}^t \mu_m(j)$. Logo, a estimativa de máxima verossimilhança para $\delta_{t,j}$ baseada em $\{\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_{t-1} = \mathbf{x}_{t-1}\}$ é

$$\delta_j^{[k,t-1]} = \max \left\{ 1, x_{t-1}(j) / \mu_j^{[t-1]} \right\} \quad \text{ou} \quad \delta_j^{[k,t-1]} = \max \left\{ 1, \sum_{m=k}^{t-1} x_m(j) / \sum_{m=k}^{t-1} \mu_j^{[m]} \right\},$$

quando $\delta_{t,j} = \delta_j$. No entanto, consideramos como estimativa para $\delta_{t,j}$ uma versão suavizada da estimativa de máxima verossimilhança dada por

$$\delta_j^{[k,t-1]} = \max \left\{ 1, \tilde{x}_j^{[k,t-1]} / \tilde{\mu}_j^{[k,t-1]} \right\}.$$

Onde para $t > k$

$$\tilde{x}_j^{[k,t-1]} = x_{t-1}(j) + \rho \tilde{x}_j^{[k,t-2]}, \quad \tilde{\mu}_j^{[k,t-1]} = \mu_j^{[t-1]} + \rho \tilde{\mu}_j^{[k,t-2]}.$$

Caso contrário, $\tilde{x}_j^{[k,t-1]} = \tilde{\mu}_j^{[k,t-1]} = 0$ e $\delta_j^{[k,k-1]} = 1$. A quantidade $0 \leq \rho \leq 1$ é um parâmetro de

suavização. Note que se $\rho = 0$,

$$\delta_j^{[k,t-1]} = \max \left\{ 1, x_{t-1}(j) / \mu_j^{[t-1]} \right\}$$

e quando $\rho = 1$,

$$\delta_j^{[k,t-1]} = \max \left\{ 1, \frac{\sum_{m=k}^{t-1} x_m(j)}{\sum_{m=k}^{t-1} \mu_j^{[m]}} \right\}.$$

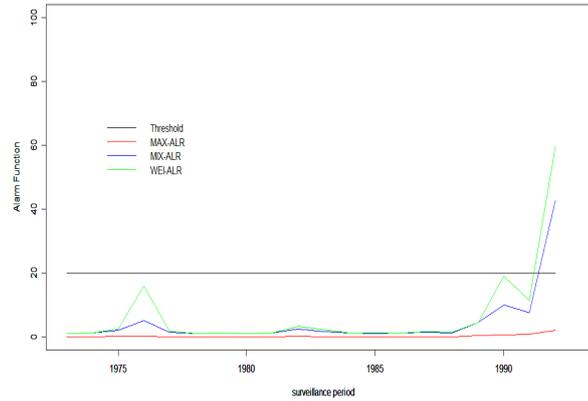
A suavização é introduzida para reduzir o efeito das observações de um passado muito distante na estimativa atual de $\delta_{t,j}$. Quando ρ é pequeno, poucas observações contribuirão para a estimativa de $\delta_{t,j}$ de modo que usamos $\rho = 0.8$.

Resultados Obtidos

Ao aplicarmos os métodos propostos para detecção e vigilância de clusters de câncer, obtemos o gráfico de monitoramento apresentado na figura (3.2) com valor de B igual a 20 e $\epsilon = 0.8$. Neste caso, a taxa de alarmes falsos para qualquer um dos métodos é no máximo 0.05. No gráfico, o MAX-ALR não ultrapassa o valor de B em nenhum momento e por isso essa estatística não dispara um alarme. Os procedimentos MIX-ALR e WEIGHT-ALR soam um alarme no tempo $n = 1992$. O WEIGHT-ALR é mais sensível a mudanças. O efeito do valor ϵ pode ser observado nos pontos de picos da função de alarme WEIGHT-ALR. Observa-se, ainda, que no tempo $t=1990$, o valor do MIX-ALR é igual 9.92 e para o WEIGHT-ALR o valor obtido foi 18.99. Estes valores não ultrapassam o limite de controle representado por $B=20$. Usando o MIX-ALR, o cluster mais provável é $\hat{\xi}^{[1980,1990]}$. onde $\hat{\xi} = \{LosAlamos, RioArriba, SantaFe, Taos\}$ representa a localização espacial do cluster (ver figura 3.1) e o intervalo $[1980, 1990]$ representa a parte temporal. O valor da razão de verossilhança adaptativa foi 308.29 com um risco relativo estimado adaptativamente por $\delta_{\hat{\xi}_j}^{[1980,1990]} = 2.27$. Para o WEIGHT-ALR, a parte espacial do cluster é estimada por $\tilde{\xi} = \{LosAlamos, Mora, Sandoval, SantaFe\}$ com uma parte temporal $[1986, 1990]$, $ALR = 48.01$ e $\delta_{\tilde{\xi}_j}^{[1986,1990]} = 2.89$. Observe que existe uma intersecção entre esses dois cluster que é representada por $\xi^{[1986,1990]} = \{LosAlamos, SantaFe\}^{[1986,1980]}$. Na prática, o MIX-ALR e o WEIGHT-ALR representam uma soma sobre vários candidatos a clusters que são possivelmente correlacionados. Uma síntese dos resultados obtidos para os últimos cinco períodos de vigilância é apresentada na tabela 3.2. O importante são os resultados referentes ao tempo atual denotado por $n = 1992$. O valor obtido para o MIX-ALR (WEIGHT-ALR) no tempo atual é 42.672 (59.529) que é um grande valor quando $B=20$. Neste caso a região que dispara o alarme é estimada por $\xi = \{LosAlamos\}$. O tempo estimado de início do cluster é $\hat{k}_{\{LosAlamos\}} = 1987$. Então o cluster no espaço-tempo estimado é representado por $\{LosAlamos\}^{[1992,1987]}$. O cluster estimado pela Estatística Scan em Kulldorff(2001) é $\{LosAlamos\}^{[1992,1989]}$. Note que embora a localização espacial seja a mesma, o tempo de duração do cluster é diferente.

Tabela 3.1: Detecção de clusters emergentes de casos de câncer da tireóide em homens que ocorreram no Novo México, usando o MIX-ALR e WEIGHT-ALR nos últimos cinco anos.

Alarm Function	Most likely cluster(ξ)	years([k,n])	$\delta_{\xi}^{[k,n-1]}$	ARL_n
MIX-ALR	SantaFe	[1980,1988]	2.34	7.41
	LosAlamos, Mora, Sandoval, SantaFe	[1986,1989]	2.29	159.42
	LosAlamos, RioArriba, SantaFe, Taos	[1980,1990]	2.27	308.29
	LosAlamos, RioArriba, SantaFe, Taos	[1980,1991]	2.23	331.15
	LosAlamos	[1987,1992]	6.98	866.88
WEIGHT-ALR	Mora, SantaFe, Taos	[1979,1988]	2.16	5.87
	LosAlamos, Mora, Sandoval, SantaFe	[1986,1989]	2.29	159.42
	LosAlamos, Mora, Sandoval, SantaFe	[1986,1990]	2.89	48.01
	LosAlamos, RioArriba, SantaFe, Taos	[1980,1991]	2.23	331.15
	LosAlamos	[1987,1992]	6.98	866.88

Figura 3.2: Monitoramento on-line usando MAX, MIX and WEIGHT-ALR com $B(\text{threshold})=20$

3.4 Extensões para abordagem Adaptativa no espaço dos clusters

Uma grande dificuldade na utilização dos procedimentos MAX e MIX-ALR é a cardinalidade J do conjunto Ξ . Quanto mais flexível em termos de forma geométrica e tamanho for a classe Ξ , melhor será a estimativa do cluster mas, maior será sua cardinalidade. Isto implica que os valores obtidos para o MAX e MIX-ALR serão pequenos de modo que o alarme pode não soar mesmo na presença de um cluster emergente no mapa. Para solucionar este problema, nós recomendamos o uso de uma abordagem adaptativa também no espaço dos clusters Ξ .

Para elaborarmos esta abordagem considere uma classe inicial de distintos candidatos a cluster $\Xi^0 = \cup \Xi_l, l = 1, 2, \dots, L$ com cardinalidade J^0 . Onde $\Xi_l = \{\xi_{l_1}, \dots, \xi_{l_L}\}$. Por exemplo, Ξ_l pode representar o conjunto de todos os círculos centrados em s_l cujo raio cresce continuamente

até que contenha 50% das localizações espaciais do mapa. Para todo $n > 1$, defina

$$\Xi_l^{[n-1]} = \{\xi_{l_j} : ALR_{n-1}^{l_j} \geq C, \quad j = 1, 2, \dots, L\} \quad \text{e} \quad \Xi^{[n-1]} = \cup \Xi_l^{[n-1]}, l = 1, 2, \dots, L.$$

para algum valor $C > 1$, então no tempo atual n , a varredura espacial para detecção de cluster emergente é realizada sobre os candidatos pertencente a classe $\Xi^{[n-1]} \subset \Xi^0$ que possui cardinalidade $J^{[n-1]}$. O fato é que se no tempo $n - 1$ a razão de verossimilhança adaptativa

$$ALR_{n-1}^{l_j} = \max_{1 \leq k \leq n-1} \frac{\mathbf{P}(\mathcal{H}_{n-1} | H_{k,l_j})}{\mathbf{P}(\mathcal{H}_{n-1} | H_0)}$$

possui valor menor ou igual a C , então ela deve ser retirada da análise no tempo n e $\Xi^{[n-1]}$ funciona como uma estimativa para o espaço dos clusters no tempo $n - 1$. Quando $\Xi_l^{[n-1]} = \emptyset$ para todo l , defina arbitrariamente $\Xi^{[n-1]} = \Xi^0$. Ou seja, quando no tempo $n - 1$ nenhum ξ_{l_j} possui $ALR_{n-1}^{l_j}$ maior que C , o processo de varredura espacial deve ser inicializada no tempo n com todos os candidatos. Note que a cada n , $J^{[n-1]} \leq J^0$. Em particular, quando

$$\Xi_l^{[n-1]} = \{\tilde{\xi}_{l_j} : \tilde{\xi}_{l_j} = \operatorname{argmax}[ALR_{n-1}^{l_j}]\}$$

temos que $J^{[n-1]} \leq L$ e $\Xi^{[n-1]}$ representa a estimativa de máxima verossimilhança adaptativa para Ξ com os dados acumulados até o tempo $n - 1$ de forma que no tempo n precisamos avaliar no máximo L candidatos.

Agora como estratégia para detecção de cluster emergente use as versões generalizadas do MAX, MIX e WEIGHT-ALR com tempo de alarme dado por:

- GMAX-ALR.

$$\tilde{\tau}_1(B) = \min \left\{ n \geq 1 : \frac{1}{J^{[n-1]}} \max_{\{\xi_{l_j} \in \Xi^{[n-1]}\}} [ALR_n^{l_j}] \geq B \right\}. \quad (3.20)$$

- GMIX-ALR.

$$\tilde{\tau}_2(B) = \min \left\{ n \geq 1 : \frac{1}{J^{[n-1]}} \sum_{\{\xi_{l_j} \in \Xi^{[n-1]}\}} ALR_n^{l_j} \geq B \right\}. \quad (3.21)$$

- : GWEIGHT-ALR

$$\tilde{\tau}_3(B) = \min \left\{ n \geq 1 : \sum_{\{\xi_{l_j} \in \Xi^{[n-1]}\}} w_{l_j} ALR_n^{l_j} \geq B \right\}. \quad (3.22)$$

Como $\Xi^{[n-1]}$ é conhecido no tempo n , a relação expressa na equação (3.10) é preservada por estes procedimentos. Para usarmos estas novas funções de monitoramento, nós usamos o seguinte algoritmo:

- **Algoritmo**

1. no tempo $n = 0$, Escolher um centróide no mapa.
2. Calcular as distâncias para os outros centróides. Ordene estas distâncias em ordem crescente e depois memoriza-las em um vetor de distâncias (este vetor de distâncias é um vetor de adjacências).
3. Para cada centróide do mapa repetir os passos 1 e 2 e então construa a matriz de adjacências de acordo com a explicação dada na seção 4.6.
4. Escolher novamente um centróide s_l na região. Centrado em s_l construa $\xi_{l_i}, i = 1, 2, \dots, L$ como explicado na seção 4.6.
5. Escolha um valor para a (por exemplo, $a = L/4$) e construa a classe

$$\Xi^0 = \{\xi_{l_i} : (\langle \xi_{l_i}, \xi_{l_i} \rangle) \leq a, (l, i) = 1, 2, \dots, L\}$$

6. Para cada $n \geq 1$, atualizamos a classe Ξ^{n-1} como explicado nesta seção. Especificamos um valor para α e calcule o par $(\tilde{f}(\text{ALR}_n), B)$. Onde \tilde{f} é qualquer função de monitoramento descrita nesta seção: GMAX-ALR, GMIX-ALR ou GWEIGHT-ALR.
7. Se $\tilde{f}(\text{ALR}_n) \geq B$, declare que existe um *cluster* emergente no mapa e o estime pelo $\text{argmax}[f(\text{ALR}_n)]$ como dado na seção 3.2.3 . Caso contrário, incremente uma unidade no tempo e volte ao passo 6.

Para compararmos as funções de alarme GMAX e GMIX-ALR com MAX e MIX-ALR, respectivamente, nós usamos estas funções na aplicação descrita na seção 3.3. Usamos $\Xi_l^{[n-1]} = \{\tilde{\xi}_{l_j} : \tilde{\xi}_{l_j} = \text{argmax}[ALR_{n-1}^{l_j}]\}$. Nota-se na figura 3.3(a) que neste caso, o GMAX-ALR dispara um alarme no tempo atual enquanto que o MAX-ALR não dispara. Quando comparamos o GMIX com o MIX, observamos que o alarme disparado pelo GMIX é mais pronunciado de modo que podemos concluir que a abordagem adaptativa no espaço do cluster torna as funções de alarme mais sensíveis na detecção do cluster emergente.

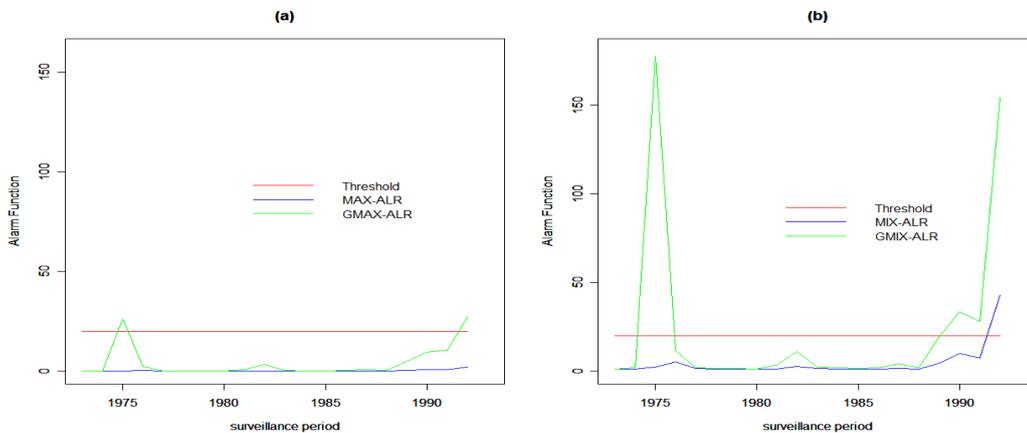


Figura 3.3: Comparação do monitoramento On-line usando o MAX-ALR com GMAX-ALR em (a) e MIX-ALR com GMIX-ALR em (b) com $B(\text{threshold})=20$.

3.5 Estudo com dados simulados

Neste estudo com dados simulados, usamos o modelo de cluster emergente utilizado na seção 3.3. Consideramos como população em risco, $N_t(s_l)$, a população de homens residentes no ano de 1992 nas 32 ($L=32$) áreas do Novo México como descrito em Kulldorff(2001). Sobre esta população em risco nós simulamos um processo de Poisson com casos simulados para um período de monitoramento $n = 20$ e 1000 réplicas do processo foram geradas. A parte espacial do cluster alocado no mapa é formado por 4 áreas, $\xi_j = \{Chaves, Eddy, Lea, Otero\}$. Este é um cluster com população mediana em relação ao mapa. Quando simulamos um processo sem *cluster* assumimos que $\mu_t(s_l) = \lambda_0 N_t(s_l)$, $t = 1, 2, \dots, 10$. Onde λ_0 é uma taxa de referência global para a doença. Para o processo com um *cluster* emergente, quando $t = 11, \dots, 20$, supomos um cluster em ξ_j com valor esperado $\theta_{t,j} \mu_t(s_l)$ para todo s_l em ξ_j . Observe que o início do *cluster* é o ponto $k = 11$. Em nossas simulações usamos:

- Sob a hipótese H_0 , $\mathbb{E}_0[X_t(s_l)] = \lambda_0 N_t(s_l)$ para todo $s_l \in \mathcal{S}$ e $t \geq 1$, de modo que o número de casos esperados na q -ésima simulação para a área A_l é proporcional a população em risco em A_l .
- Sob a hipótese $H_{k,j}$, consideramos os seguintes casos:
 - ⊗ O parâmetro $\theta_{t,j}$ assume os valores $\{1.50, 2.00\}$ para todo $s_l \in \xi_j$ e $11 \leq t \leq 20$. Ou seja, sob a hipótese de que existe um *cluster* emergente no mapa, a intensidade de casos em ξ_j é 50% e 100% maior em relação ao restante do mapa. Usamos também $\delta_{t,j} = 1.3^{t-11+1}$. Neste caso o risco cresce monotonicamente.
 - ⊗ Usamos $\lambda_0 = 0.0001$ (taxa de um caso para cada 10000 habitantes) e $\alpha = 0.05, 0.01$.

3.5.1 Performance dos métodos

Dado qualquer mapa e forma geométrica para ξ , quando $H_0 : \nu = \infty$ é verdadeira a taxa de alarmes falsos $\mathbf{FAR}[\tau^*(B)]$ dos procedimentos propostos são controlados por (3.11).

Quando a hipótese $H_{k,j} : \nu = k, \xi = \xi_j$ é verdadeira e o tempo de alarme é $\tau \geq k$ com a localização estimada do *cluster* dada por $\hat{\xi}$ (parte espacial de $\xi^{n,\hat{k}}$), pode ocorrer que $\hat{\xi}$ contenha localizações que não pertençam a ξ_j ou deixe de incluir localizações que pertencem a ξ_j de modo que o evento $\{\hat{\xi} = \xi_j\}$ pode ocorrer raramente. Então, definimos duas medidas para avaliarmos a precisão do *cluster* detectado:

- Sensitividade Condicional de um alarme $\mathbf{SSC}_{k,j}[\tau]$ - representada pela proporção de localizações espaciais pertencentes a ξ_j que foram corretamente localizadas pelo *cluster* detectado

dado que $\tau \geq k$. Esta medida é estimada por,

$$\frac{\sum_{q=1}^Q \left(\frac{\#\{\hat{\xi}^{(q)} \cap \xi_j\}}{\#\{\xi_j\}} \right) I_{\{\tau \geq k\}}}{\sum_{q=1}^Q I_{\{\tau \geq k\}}}$$

- Valor Predito Positivo Condicional de um alarme $\mathbf{VPP}_{k,j}[\tau]$ - a proporção de localizações espaciais presentes em $\hat{\xi}$ que pertencem a ξ_j dado que $\tau \geq k$. Estime essa medida por,

$$\frac{\sum_{q=1}^Q \left(\frac{\#\{\hat{\xi}^{(q)} \cap \xi_j\}}{\#\{\hat{\xi}^{(q)}\}} \right) I_{\{\tau \geq k\}}}{\sum_{q=1}^Q I_{\{\tau \geq k\}}}$$

Em que $\hat{\xi}^{(q)}$ é o estimador de ξ_j obtido na q-ésima simulação, I_A é a função indicadora do evento A e $\#\{A\}$ representa a cardinalidade do evento A .

As medidas, $\mathbf{SSC}_{k,j}[\tau]$ e $\mathbf{VPP}_{k,j}[\tau]$, avaliam o desempenho do método em relação a localização espacial estimada dado que o alarme é verdadeiro. Se $\hat{\xi} = \xi_j$ temos que $\mathbf{SSC}_{k,j}[\tau] = \mathbf{VPP}_{k,j}[\tau] = 1$. Quando $\hat{\xi} \supset \xi_j$ temos $\mathbf{SSC}_{k,j}[\tau] \geq \mathbf{VPP}_{k,j}[\tau]$ e se $\hat{\xi} \subset \xi_j$ implica em $\mathbf{SSC}_{k,j}[\tau] \leq \mathbf{VPP}_{k,j}[\tau]$. O ideal é que estas medidas sejam próximas de 1 indicando que o *cluster* localizado é próximo do verdadeiro.

A rapidez com que o método detecta o *cluster* pode ser descrita pelo atraso localizado que é representado pelo tempo médio de atraso até que o alarme soe dado que realmente existe um *cluster* emergente no mapa (condicional average detection delay). Esta medida é expressa por

$$\mathbf{ADD}_{k,j}[\tau] = \mathbb{E}(\tau - k | H_{k,j}, \tau \geq k).$$

que pode ser estimada por,

$$\hat{\mathbf{ADD}}_{k,j}[\tau] = \frac{\sum_{q=1}^Q (\tau - k) I_{\{\xi^q = \xi_j, \tau \geq k\}}}{\sum_{q=1}^Q I_{\{\xi^q = \xi_j, \tau \geq k\}}}.$$

Independente da localização estimada para o *cluster*, podemos usar o atraso médio, que pode ser estimado por

$$\hat{\mathbf{ADD}}[\tau] = \frac{\sum_{q=1}^Q (\tau - k) I_{\{\tau \geq k\}}}{\sum_{q=1}^Q I_{\{\tau \geq k\}}}.$$

3.5.2 Resultados obtidos na simulação

Os resultados para as medidas de performance estão descritos na tabela 3.2. e são apresentados apenas para os métodos GMIX e GWEIGHT-ALR. Pois estes obtiveram melhor desempenho nas simulações. Como esperado, em ambos os métodos o atraso ($\hat{\mathbf{ADD}}$) na detecção decresce e a precisão na localização cresce quando aumenta o valor de $\theta_{t,j}$. Por exemplo, quando $\theta_{t,j} = 2$, $\alpha = 0.05$ o atraso estimado para o GWEIGHT-ALR (GMIX) é igual a 2.366 (2.490)

que pode ser considerado pequeno, pois o cluster foi detectado logo no seu início. A precisão na localização medida pela **SSC** é 0.791 (0.871) enquanto que o **VPP**=0.719 (0.718). Observe na tabela que a **SSC** > 0.7 e o **VPP** > 0.6 . A **SSC** é sempre maior que o **VPP** implicando que o *cluster* localizado tende a ser maior que o verdadeiro *cluster*, porém com uma boa precisão.

De uma forma geral, o GWEIGHT-ALR é mais rápido (possui menor atraso) porém é menos preciso (possui menor **SSC**). O fato é que se o GMIX demora mais tempo para soar o alarme e então, ele realmente deveria ser mais preciso pois com o passar do tempo o cluster vai ficando mais evidente. Uma visualização da relação atraso versus precisão é mostrada através de histogramas bivariados nas figuras 3.5 e 3.6. Nestas figuras pode-se observar que mesmo quando o atraso é baixo a sensibilidade e o valor predito positivo são elevados. Uma amostra do funcionamento on-line do GMIX e GWEIGHT-ALR com dados simulados é apresentado na figura (3.4).

Tabela 3.2: Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) com diferentes valores para $\theta_{t,j}$, diferentes taxas aceitáveis de alarmes falsos (α) usando o GMIX e GWEIGHT-ALR.

		Medidas de Perfomance		
	$\theta_{t,j}$	SSC	VPP	ADD
GMIX-ALR	1.50	0.791	0.644	4.914
	$\alpha = 0.05$ 2.00	0.871	0.718	2.490
	1.3^{t-11+1}	0.861	0.750	3.198
	1.50	0.864	0.722	6.049
	$\alpha = 0.01$ 2.00	0.910	0.765	3.076
	1.3^{t-11+1}	0.921	0.814	3.604
GWEIGHT-ALR	1.50	0.729	0.616	4.197
	$\alpha = 0.05$ 2.00	0.791	0.664	2.366
	1.3^{t-11+1}	0.805	0.681	3.017
	1.50	0.811	0.687	5.289
	$\alpha = 0.01$ 2.00	0.861	0.719	2.830
	1.3^{t-11+1}	0.885	0.755	3.406

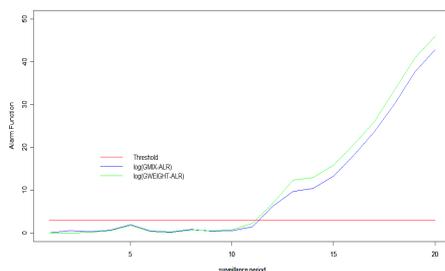


Figura 3.4: Amostra de uma simulação do processo para GMIX and GWEIGHT-ALR na escala logartímica com $\log(B)=(\text{threshold})=\log(20)$.

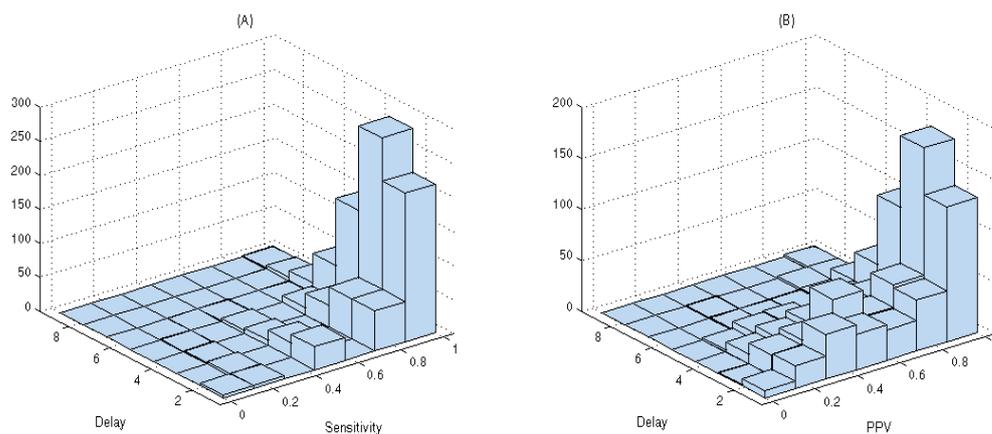


Figura 3.5: Histograma Bivariado para o GMIX-ALR: Sensitividade (Sensitivity) versus atraso (Delay) em (A) e Valor Predito Positivo (PPV-Positive Predicted Value) versus Delay(atraso) em (B).

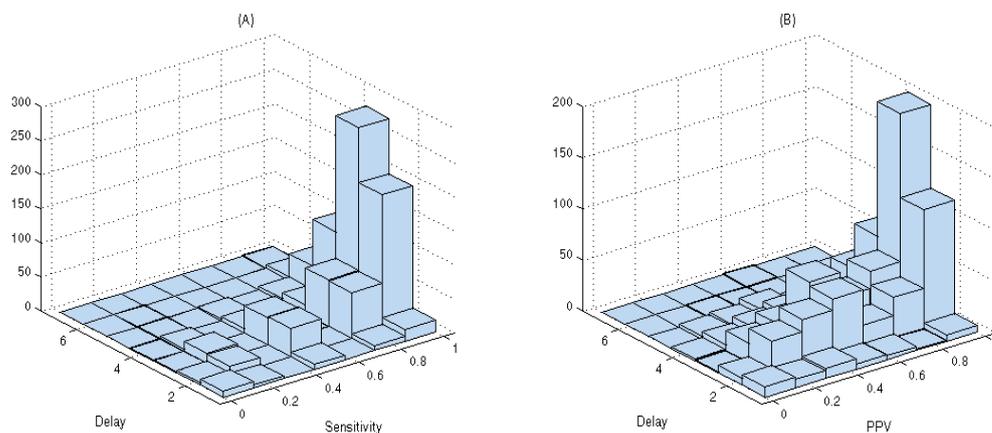


Figura 3.6: Histograma Bivariado para o WEIGHT-ALR: Sensitividade (Sensitivity) versus atraso (Delay) em (A) e Valor Predito Positivo (PPV-Positive Predicted Value) versus Delay(atraso) em (B).

3.6 Discussão

Neste capítulo desenvolvemos novos métodos para a vigilância estatística em mapas de doenças em que o foco principal é a detecção de clusters emergentes no espaço-tempo. Ao combinarmos técnicas de vigilância na escala temporal com estimação adaptativa e algoritmos

para a detecção de *clusters* espaciais de doenças, nós propusemos uma vigilância no espaço-tempo baseada no uso da razão de verossimilhança adaptativa (ALR). Em essência a ALR é uma razão de verossimilhança genuína e por isso preserva a propriedade martingal de modo que os métodos desenvolvidos neste capítulo não requerem o uso de simulações de Monte Carlo para validar a significância estatística do procedimento. Fato esse que é comum aos procedimentos existentes na literatura e pode tornar a análise muito demorada dependendo do mapa em estudo.

Na estimação adaptativa dos parâmetros nós usamos uma versão suavizada no tempo do estimador de máxima verossimilhança, isto permite que possamos dar mais pesos aos dados mais recentes diminuindo a influência de possíveis clusters históricos nos dados atuais. Uma característica importante desta abordagem adaptativa é que apesar de assumirmos independência mútua entre os vetores aleatórios, podemos capturar a dependência espacial e temporal através da estimativa. Para detectar cluster emergentes nós usamos três funções de alarme: O máximo sobre todas as ALR's, mistura de ALR's e ALR ponderada. Um exemplo onde os parâmetros são desconhecidos é usado para ilustrar os métodos propostos. Os resultados obtidos nesta aplicação foram similares aos obtidos pela usual Estatística Scan (Kulldorff 2001) o que reforça a boa habilidade dos métodos propostos em detectar cluster emergentes.

A medida que aumentamos a classe de candidatos a cluster nós observamos algumas dificuldades na detecção. Por isso aplicamos uma estimação adaptativa também no espaço dos clusters. Esta nova abordagem melhorou os métodos propostos em dois aspectos. Primeiro aumentou a velocidade na detecção e segundo diminui o espaço dos cluster tornando o algoritmo mais eficiente. A vantagem da abordagem adaptativa no espaço dos clusters é que a cada tempo precisamos avaliar no máximo L candidatos e, ainda assim, controlamos a taxa de falsos alarmes.

Embora analiticamente seja virtualmente impossível obter propriedades estatísticas destes métodos no que tange ao atraso e precisão na detecção. Empiricamente, através de um simples estudo com dados simulados, obtivemos resultados altamente satisfatórios que podem ser ratificados ou não através de estudos comparativos com outros métodos. Estes estudos comparativos não foram abordados aqui devido à dificuldade técnica de obtermos a taxa de alarmes falsos e um limite de controle adequado para os métodos competitivos, quando o valor esperado sobre a hipótese nula é desconhecida e a intensidade da mudança variam com o tempo. Também não foram considerados neste trabalho outros aspectos de um sistema de vigilância que são de vital importância para profissionais que o utilizam, por exemplo: qualidade e processamento dos dados, custo de um sistema de vigilância e possíveis perdas financeiras ocasionadas por uma ação desnecessária devido a um falso alarme.

4 *Fator de Bayes Adaptativo para detecção de cluster no espaço-tempo*

4.1 Introdução

No contexto não-bayesiano, a detecção de mudança no espaço-tempo, tem sido abordada na literatura através de teste de hipóteses. Uma suposição implícita nestes testes é que o processo observado é composto por componentes mutuamente independentes no espaço-tempo. Porém, esses componentes podem ser correlacionados, não-estacionários ou ambos. Isto implica que as propriedades obtidas para estes testes podem estar sendo violadas. Neste caso, uma alternativa é usar procedimentos bayesianos que possam acomodar esses problemas.

Do ponto de vista bayesiano os testes de hipóteses são encarados como um problema de comparação de dois modelos M_0 e M_1 , onde cada modelo pode representar, por exemplo, uma hipótese estatística de mudança ou não-mudança no padrão dos dados observados. Uma forma conveniente de resumir a evidência nos dados em favor de um dos modelos é através do uso do fator de Bayes. Este tem sido usado por Gangnon e Klayton (2007) para detecção de mudanças no espaço e por West(1986) em monitoramento na escala temporal. Neste capítulo, nós propomos o uso de um Fator de Bayes seqüencial local para detecção de mudanças no espaço-tempo. Estas mudanças são detectadas através da comparação entre as densidades preditivas para o modelo padrão, contra as preditivas para um conjunto de modelos alternativos. Estes modelos são atualizados no tempo. Essencialmente, os modelos alternativos possuem forma similar ao modelo padrão porém, permitem mudanças nos valores dos parâmetros que caracterizam o modelo padrão. A abordagem apresentada neste capítulo, pode ser entendida como uma extensão da abordagem de West(1986) discutida resumidamente na seção 2.2.5.

4.2 Análise Sequencial de Modelos

4.2.1 Distribuições preditivas e Fator de Bayes

Seja \mathcal{S} uma região em estudo particionada em L -áreas contíguas indexadas pelo ponto s_l . Usualmente, \mathcal{S} é um mapa (de municípios, bairros, etc...) e s_l é o centróide do polígono que

delimita a área. Considere sob \mathcal{S} o processo estocástico no espaço-tempo $\mathbf{X}_t = \{X_t(s_l) : t \in \{1, 2, \dots\}, l = 1, 2, \dots, L\}$ com valor observado \mathbf{x}_t , onde t representa um período ou um ponto no tempo. Sejam $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$ um conjunto conexo de localizações espaciais com o forma e tamanho conhecidos e $\Xi = \{\xi_1, \xi_2, \dots, \xi_J\}$ uma classe destes conjuntos onde os ξ'_j 's são todos distintos. Após observar o processo n vezes, denote o histórico do processo até n por \mathcal{H}_{n-1} , $\mathcal{H}_n = \{\mathcal{H}_{n-1}, \mathbf{x}_n, \mathcal{I}_n\}$ e informação inicial \mathcal{H}_0 . Onde, \mathcal{I}_n contém toda informação adicional além daquela obtida com as observações do processo. Por exemplo, \mathcal{I}_n pode incluir valores de possíveis cováriaveis conhecidas. Caso a única informação adicional obtida em cada tempo n seja \mathbf{x}_n , então $\mathcal{H}_n = \{\mathcal{H}_{n-1}, \mathbf{x}_n\}$.

Assuma que sob condições normais de funcionamento, o processo \mathbf{X}_t é representado estatisticamente por um modelo padrão (ou de referência) para \mathbf{X}_t dado por:

$$M_{t,0} = \{f(\mathbf{X}_t|\boldsymbol{\theta}_t), g_0(\boldsymbol{\theta}_t|\mathcal{H}_{t-1}), p_0(\mathbf{X}_t|\mathcal{H}_{t-1})\}.$$

Este modelo envolve: (a) A distribuição de \mathbf{X}_t com densidade condicional

$$f(\mathbf{X}_t|\boldsymbol{\theta}_t) = \prod_{s_l \in \mathcal{S}} f(X_t(s_l)|\theta_t(s_l))$$

dependendo de um parâmetro $\boldsymbol{\theta}_t = (\theta_t(s_1), \dots, \theta_t(s_L))$, ou seja, os componentes de \mathbf{X}_t são condicionalmente independentes; (b) Uma distribuição *a priori* para $\boldsymbol{\theta}_t$,

$$g_0(\boldsymbol{\theta}_t|\mathcal{H}_{t-1}) = \prod_{s_l \in \mathcal{S}} g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)).$$

Dado o histórico da série $\mathcal{H}_{t-1} = \cup_{s_l \in \mathcal{S}} \mathcal{H}_{t-1}(l)$ os componentes de $\boldsymbol{\theta}_t$ são condicionalmente independentes; (c) Uma distribuição (ou densidade) preditiva $p_0(\mathbf{X}_t|\mathcal{H}_{t-1})$ para \mathbf{X}_t dada por

$$p_0(\mathbf{X}_t|\mathcal{H}_{t-1}) = \int f(\mathbf{X}_t|\boldsymbol{\theta}_t)g_0(\boldsymbol{\theta}_t|\mathcal{H}_{t-1})d\boldsymbol{\theta}_t. \quad (4.1)$$

Onde a integral é realizada sobre o espaço paramétrico para $\boldsymbol{\theta}_t$.

Após observar $\mathbf{X}_t = \mathbf{x}_t$, a quantidade $p_0(\mathbf{x}_t|\mathcal{H}_{t-1})$ é o valor da verossimilhança do modelo padrão e representa sua habilidade preditiva. Suponha que no tempo t fixamos um ξ e $M_{t,0}$ é comparado com um modelo alternativo

$$M_{t,\xi} = \{f(\mathbf{X}_t|\boldsymbol{\theta}_t), g_\xi(\boldsymbol{\theta}_t|\mathcal{H}_{t-1}), p_\xi(\mathbf{X}_t|\mathcal{H}_{t-1})\}.$$

pertencente a uma classe $\mathcal{M} = \{M_{t,\xi}, \xi \in \Xi\}$, e que possui densidade preditiva $p_\xi(\mathbf{x}_t|\mathcal{H}_{t-1})$. O modelo $M_{t,\xi}$ é construído de modo a representar uma possível mudança na distribuição do sub-processo $\{X_t(s_l) \in \mathbf{X}_t : t \geq k, s_l \in \xi_j\}$ para algum par (k, ξ) desconhecido. Nós chamamos esta mudança de cluster no espaço-tempo (ou cluster emergente) e o denotamos por $(\hat{k}, \hat{\xi})$ indicando que a mudança ocorreu no conjunto (ou grupo) de localizações espaciais $\hat{\xi} = \{s_l \in \mathcal{S} : s_l \in \hat{\xi}\}$

emergindo a partir do tempo \tilde{k} .

Para verificar o ajuste preditivo dos modelos $M_{t,0}$ e $M_{t,\xi}$, usamos a razão de verossimilhanças preditivas ou Fator de Bayes (FB)

$$FB_t(\xi) = \frac{p_\xi(\mathbf{x}_t|\mathcal{H}_{t-1})}{p_0(\mathbf{x}_t|\mathcal{H}_{t-1})}. \quad (4.2)$$

Um valor $FB_t(\xi) > 1$ implica uma melhor performance preditiva de $M_{t,\xi}$ indicando uma possível mudança no processo.

4.2.2 Fator de Bayes Local

Como o interesse é detectar mudanças localizadas no espaço-tempo após um ponto de mudança $k \leq n$, é necessário considerar um grupo formado pelas mais recentes $n - k + 1$ observações consecutivas em algumas localizações específicas. Para isso, assuma que sob $M_{t,\xi}$ com ponto de mudança em um tempo k , a distribuição a priori para $\boldsymbol{\theta}_t$ é do tipo lei de potência expressa por

$$g_\xi(\boldsymbol{\theta}_t|\mathcal{H}_{t-1}) \propto g_0(\boldsymbol{\theta}_t|\mathcal{H}_{t-1})^{\gamma_{t,\xi}^{I(\{t \geq k, s_l \in \xi\})}}. \quad (4.3)$$

Onde $I(\{A\})$ é a função indicadora do evento A e $\gamma_{t,\xi}$ é um valor positivo. Logo, se $t < k$ temos $FB_t(\xi) = 1$ e para $t \geq k$,

$$FB_t(\xi) = \frac{p_\xi(\mathbf{x}_t|\mathcal{H}_{t-1})}{p_0(\mathbf{x}_t|\mathcal{H}_{t-1})} = \frac{\int f(\mathbf{X}_t|\boldsymbol{\theta}_t)g_\xi(\boldsymbol{\theta}_t|\mathcal{H}_{t-1})d\boldsymbol{\theta}_t}{\int f(\mathbf{X}_t|\boldsymbol{\theta}_t)g_0(\boldsymbol{\theta}_t|\mathcal{H}_{t-1})d\boldsymbol{\theta}_t}$$

e pela independência condicional dos componentes de $\boldsymbol{\theta}_t$ obtemos

$$\begin{aligned} FB_t(\xi) &= \frac{\int \prod_{s_l \in \xi} c_{t,\xi}^{(l)} f(X_t(s_l)|\theta_t(s_l)) g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l))^{\gamma_{t,\xi}} d\theta_t(s_l)}{\int \prod_{s_l \in \xi} f(X_t(s_l)|\theta_t(s_l)) g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) d\theta_t(s_l)}, \quad t \geq k. \\ &= \frac{\prod_{s_l \in \xi} p_{t,\xi}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l))}{\prod_{s_l \in \xi} p_{t,0}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l))}. \end{aligned} \quad (4.4)$$

Onde,

$$p_{t,\xi}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l)) = \int c_{t,\xi}^{(l)} f(X_t(s_l)|\theta_t(s_l)) g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l))^{\gamma_{t,\xi}} d\theta_t(s_l) \quad (4.5)$$

$$p_{t,0}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l)) = \int f(X_t(s_l)|\theta_t(s_l)) g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) d\theta_t(s_l) \quad (4.6)$$

e $c_{t,\xi}^{(l)}$ é uma constante de normalização.

Observe que quando $t \geq k$, $FB_t(\xi)$ depende somente dos dados observados em ξ . O valor $\gamma_{t,\xi}$ é explicado na seção 4.4.2. Para cada ξ , o fator de Bayes $FB_t(\xi)$ é similar ao descrito em West (1986). Agora, para uma sequencia de observações $\mathbf{x}_1, \dots, \mathbf{x}_n$, $n \geq 1$, denote por P_0 (\mathbb{P}_0) a densidade (distribuição) preditiva conjunta sob $\{M_{n,0}\}_{n \geq 1}$ e $P_{k,\xi}$ ($\mathbb{P}_{k,\xi}$) densidade(distribuição)

preditiva conjunta para $\{M_{n,\xi}\}_{n \geq 1}$ com ponto de mudança em um tempo $k \leq n$. Então a razão de verossimilhança preditiva conjunta é dada por

$$\begin{aligned}
\text{FB}_{n,k}(\xi) &= \frac{P_{k,\xi}(\mathbf{x}_n, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{P_0(\mathbf{x}_n, \dots, \mathbf{x}_1 | \mathcal{H}_0)} \\
&= \frac{P_{k,\xi}(\mathbf{x}_n | \mathbf{x}_{n-1}, \dots, \mathbf{x}_1, \mathcal{H}_0) P_{k,\xi}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{P_0(\mathbf{x}_n | \mathbf{x}_{n-1}, \dots, \mathbf{x}_1, \mathcal{H}_0) P_0(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)} \\
&= \frac{p_\xi(\mathbf{x}_n | \mathcal{H}_{n-1}) P_{k,\xi}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{p_0(\mathbf{x}_n | \mathcal{H}_{n-1}) P_0(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)} \\
&= \text{FB}_n(\xi) \times \text{FB}_{n-1,k}(\xi) = \prod_{t=k}^n \text{FB}_t(\xi). \tag{4.7}
\end{aligned}$$

$\text{FB}_{n,k}(\xi)$ é uma função que pode ser usada para localizar possíveis mudanças nas mais recentes $n - k + 1$ observações no conjunto ξ de localizações espaciais. Uma propriedade importante de $\text{FB}_{n,k}(\xi)$, é que para quaisquer ξ fixo, sob a distribuição preditiva \mathbb{P}_0 a sequência $\{\text{FB}_{n,k}(\xi), n \geq k\}$ é um martingal de valor esperado igual a 1. Para mostrar essa propriedade, observamos que se \mathbb{E}_0 é um valor esperado sob \mathbb{P}_0 , então

$$\begin{aligned}
\mathbb{E}_0[\text{FB}_{n,k}(\xi) | \mathcal{H}_{n-1}] &= \mathbb{E}_0[\text{FB}_n(\xi) \times \text{FB}_{n-1,k}(\xi) | \mathcal{H}_{n-1}] \\
&= \text{FB}_{n-1,k}(\xi) \mathbb{E}_0[\text{FB}_n(\xi) | \mathcal{H}_{n-1}]. \\
&= \text{FB}_{n-1,k}(\xi) \int \left(\frac{p_\xi(\mathbf{x}_n | \mathcal{H}_{n-1})}{p_0(\mathbf{x}_n | \mathcal{H}_{n-1})} \right) \times p_0(\mathbf{x}_n | \mathcal{H}_{n-1}) d\mathbf{x}_n. \\
&= \text{FB}_{n-1,k}(\xi) \int p_\xi(\mathbf{x}_n | \mathcal{H}_{n-1}) d\mathbf{x}_n = \text{FB}_{n-1,k}(\xi). \tag{4.8}
\end{aligned}$$

e ainda,

$$\begin{aligned}
\mathbb{E}_0[\text{FB}_{n,k}(\xi)] &= \int \cdots \int \left(\frac{P_{k,\xi}(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathcal{H}_0)}{P_0(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathcal{H}_0)} \right) P_0(\mathbf{x}_n, \dots, \mathbf{x}_1 | \mathcal{H}_0) d\mathbf{x}_1 \cdots d\mathbf{x}_n \\
&= \int \cdots \int P_{k,\xi}(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathcal{H}_0) d\mathbf{x}_1 \cdots d\mathbf{x}_n = 1. \tag{4.9}
\end{aligned}$$

de modo que $\{\text{FB}_{n,k}(\xi), n \geq k\}$ é um martingal de média 1.

4.3 Detecção de cluster no espaço-tempo via Fator de Bayes Adaptativo

4.3.1 Tempo de alarme para detecção de cluster no espaço-tempo

Para construirmos uma regra para detecção e localização do *cluster* no espaço-tempo, primeiramente assumimos que o interesse é a detecção de mudanças em uma janela temporal de comprimento fixo $w \leq n$ e definimos

$$\text{FB}_n = \max_{\xi \in \Xi} \max_{n-w+1 \leq k \leq n} \text{FB}_{n,k}(\xi).$$

Como estratégia para a detecção considere o tempo de alarme dado por

$$\tau^* = \tau(B^*) = \min \{n \geq 1 : \text{FB}_n \geq B^*\}. \quad (4.10)$$

Ou seja, o tempo de alarme, é o primeiro momento no tempo tal que a estatística FB_n excede o valor B^* que controla a probabilidade α (ou a taxa) de alarmes falsos. Na linguagem de análise sequencial $n - w + 1$ pode ser interpretado como o maior valor permitido para o atraso na detecção da mudança. Um exemplo para esta situação é a detecção de clusters emergentes de doenças infecciosas onde uma epidemia pode ser evitada se o cluster for detectado em um intervalo de tempo com comprimento menor ou igual a $w - 1$. Note que o limite de controle B^* é assumido igual para todos os candidatos a *clusters* emergentes. O fato é que o valor de B^* depende da probabilidade (ou taxa) de alarmes falsos e um valor B^* comum para todos os candidatos garante que esta taxa seja igual para todos. Se existe um n tal que o evento $\left\{ \max_{\xi \in \Xi} \max_{n-w+1 \leq k \leq n} \text{FB}_{n,k}(\xi) \geq B^* \right\}$ ocorre, estime o *cluster* no espaço-tempo por

$$\xi^{n,k} = \arg \left(\max_{\xi \in \Xi} \max_{n-w \leq k \leq n} \text{FB}_{n,k}(\xi) \right).$$

Caso contrário, defina $\tau^* = \infty$ para $\min\{\emptyset\}$. A escolha $\xi^{n,k}$ é baseada no fato de que, se o *cluster* começou no tempo \hat{k} e terminou em $n' < n$ ele deveria ter sido detectado pelo método em n' . Desde de que τ^* é baseado no valor máximo de $\text{FB}_{n,k}(\xi)$, ele é denominado MAX-FB.

4.3.2 Probabilidade de um falso alarme para o MAX-FB

O procedimento para detecção e monitoramento de *cluster* no espaço-tempo descrito em (4.10), representa um tempo aleatório de alarme que depende de \mathcal{H}_n , e de uma janela temporal de comprimento w . Um bom procedimento deve ter uma pequena probabilidade de alarmes falsos e um pequeno atraso na detecção do *cluster*. Para o método MAX-FB nós usamos a seguinte medida de um falso alarme restrito, (Lai,1995):

$$\mathbb{P}_0(\nu \leq \tau^* \leq \nu + w - 1) \leq \alpha \quad (4.11)$$

Esta medida significa que se a sequência de modelos $\{M_{t,0}, t \geq \nu\}$ é adequada para os dados, então a probabilidade de um falso alarme em qualquer janela temporal de comprimento w é menor ou igual a um nível pré-especificado $0 < \alpha < 1$.

Theorem 4 Para cada n seja J a cardinalidade de Ξ e $\Delta^* = \{\tau : \mathbb{P}_0(\nu \leq \tau \leq \nu + w - 1) \leq \alpha\}$, para algum valor fixo $0 < \alpha < 1$. Se $B^* = 2Jw/\alpha$. Então $\tau^* \in \Delta^*$.

Este teorema nos diz que se α é um nível aceitável para a probabilidade de um falso alarme e se o interesse é detectar mudanças emergentes na sequência $\{M_{t,0}\}$ usando o procedimento MAX-FB

com um nível de alarme falso inferior a 0.05, faça $B^* = 2Jw/0.05$. Para provar o teorema 4, observamos que $\Xi = \{\xi_1, \xi_2, \dots, \xi_J\}$ e o evento

$$\left\{ \max_{\xi_j \in \Xi} \max_{n-w+1 \leq k \leq n} \text{FB}_{n,k}(\xi_j) \geq B^* \right\}$$

ocorre, se $A_{n,k}(j) \triangleq \{\text{FB}_{n,k}(\xi_j) \geq B^*\}$ ocorre para algum $n \geq k$ e $j = 1, 2, \dots, J$. Como sob a distribuição preditiva \mathbb{P}_0 a sequência $\{\text{FB}_{n,k}(\xi_j), n \geq k\}$ é um martingal de valor esperado igual a 1 para todo $j = 1, 2, \dots, J$. Então,

$$\mathbb{P}_0(A_{n,k}(j)) \leq 1/B^*$$

e

$$\begin{aligned} \mathbb{P}_0 \left\{ \max_{\xi_j \in \Xi} \max_{n-w+1 \leq k \leq n} \text{FB}_{n,k}(\xi_j) \geq B^* \right\} &= \mathbb{P}_0 \left\{ \bigcup_{j=1}^J \bigcup_{k=n-w+1}^n A_{n,k}(j) \right\} \\ &\leq \sum_{j=1}^J \sum_{k=n-w+1}^n \mathbb{P}_0(A_{n,k}(j)). \end{aligned}$$

Conseqüentemente, para algum $\nu \leq n \leq \nu + w - 1$ e $j = 1, 2, \dots, J$ temos

$$\mathbb{P}_0(\nu \leq \tau^* \leq \nu + w - 1) \leq \sum_{j=1}^J \sum_{k=\nu-w+1}^{\nu+w-1} \mathbb{P}_0(A_{n,k}(j)) \leq 2Jw/B^*.$$

Agora se $B^* = 2Jw/\alpha$ obtemos,

$$\mathbb{P}_0(\nu \leq \tau^* \leq \nu + w - 1) \leq \alpha.$$

4.3.3 Fator de Bayes Adaptativo

A cada n , uma grande dificuldade que pode tornar inviável a utilização do procedimentos MAX-FB é a cardinalidade J da classe $\mathcal{M} = \{M_{n,\xi}, \xi \in \Xi\}$, pois esta é igual a cardinalidade de Ξ . Quanto mais flexível em termos de forma geométrica e tamanho for a classe Ξ , melhor será a estimativa da localização espacial do cluster porém, maior será o valor de J . Se α é pequeno e J é razoavelmente grande, então pode ocorrer $B^* = (2Jw/\alpha) \rightarrow \infty$ de modo que mesmo na presença de um cluster no espaço-tempo o alarme pode não soar, ou soar com grande atraso. Com o objetivo de solucionar este problema, nós recomendamos o uso de uma abordagem adaptativa na classe \mathcal{M} .

Para elaborarmos a abordagem adaptativa, considere uma classe inicial $\Xi^{(0)} = \cup \Xi_l, l = 1, 2, \dots, L$ com cardinalidade $J^{(0)}$, onde $\Xi_l = \{\xi_{1l}, \dots, \xi_{j_l}\}$. Neste caso, ξ_{j_l} é um conjunto de localizações espaciais com centro em s_l . Por exemplo, ξ_{j_l} pode representar um caminho aleatório

em \mathcal{S} partindo de s_l em qualquer direção. Para $n > 1$, denote

$$\xi_l^{(n-1)} = \arg \left\{ \max_{\xi_{j_l} \in \Xi_{l, n-w} \leq k \leq n-1} \max [\text{FB}_{n-1, k}(\xi_{j_l})] \right\}, \quad (4.12)$$

e

$$\Xi^{(n-1)} = \{\xi_1^{(n-1)}, \dots, \xi_L^{(n-1)}\}.$$

Seja

$$\Xi^{(n)} = \left\{ \xi_l^{(n-1)} \in \Xi^{(n-1)} : \text{FB}_{n-1}(\xi_l^{(n-1)}) > C^* \right\}$$

para alguma constante $C^* \geq 1$. Se não existe um $\xi_l^{(n-1)}$ tal que $\text{FB}_{n-1}(\xi_l^{(n-1)}) > C^*$, defina $\Xi^{(n)} = \Xi^{(n-1)}$. Finalmente, construa a classe de modelos alternativos por

$$\mathcal{M}^{(n)} = \{M_{n, \xi_l}, \xi_l \in \Xi^{(n)}\}.$$

com cardinalidade $J^{(n)} \leq L$. Note que neste caso, se $\xi_l = \xi_l^{n-1} \in \Xi^n$. Então usando (4.7) e (4.12),

$$\begin{aligned} \text{FB}_{n, k}(\xi_l) &= \frac{p_{\xi_l}(\mathbf{x}_n | \mathcal{H}_{n-1}) P_{k, \xi_l}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{p_0(\mathbf{x}_n | \mathcal{H}_{n-1}) P_0(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)} \\ &= \text{FB}_n(\xi_l^{n-1}) \times \max_{\xi_l \in \Xi^{(n-1)}_{n-w} \leq k \leq n-1} \max \left(\frac{P_{k, \xi_l}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{P_0(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)} \right) \\ &= \text{FB}_n(\xi_l^{n-1}) \text{FB}_{n-1, k}(\xi_l^{n-1}). \end{aligned} \quad (4.13)$$

Ou seja, a classe \mathcal{M}^n é formada por todos os modelos M_{n, ξ_l} com maior densidade preditiva no tempo $n-1$, dentre todos os modelos que são construídos a partir do ponto s_l , ($l = 1, 2, \dots, L$) e ainda, a análise sequencial sobre os dados acumulados ao longo do tempo é repetida somente sobre estes modelos. Após observar $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ considere *FBA* como o Fator de Bayes Adaptativo para comparar a performance preditiva de M_{n, ξ_l} versus $M_{n, 0}$ dado por

$$\text{FBA}_{n, k}(\xi_l^{n-1}) = \frac{p_{\xi_l^{n-1}}(\mathbf{x}_n | \mathcal{H}_{n-1}) P_{k, \xi_l^{n-1}}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}{p_0(\mathbf{x}_n | \mathcal{H}_{n-1}) P_0(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1 | \mathcal{H}_0)}.$$

como $\xi^{n-1} = \xi^{n-1}(\mathbf{x}_{n-1}, \dots, \mathbf{x}_1)$ é mensurável a $(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$ e portanto conhecido no tempo n , temos que $\text{FBA}_{n, k}$ preserva a propriedade martingal de $\text{FB}_{n, k}$ para qualquer conjunto ξ . Agora, definimos

$$\text{FBA}_n = \max_{\xi \in \Xi^n_{n-w+1} \leq k \leq n} \max \text{FBA}_{n, k}(\xi).$$

e como estratégia para detecção de cluster emergente considere o tempo de alarme dado por

$$\tilde{\tau} = \tau(B^{(n)}) = \min \left\{ n \geq 1 : \text{FBA}_n \geq B^{(n)} \right\}. \quad (4.14)$$

Se $B^{(n)} = 2J^{(n)}w/\alpha$ então $\tilde{\tau} \in \Delta^*$. Para mostra que $\tilde{\tau} \in \Delta^*$, no teorema 4 substitua J por $J^{(n)}$ e use o fato que $\text{FBA}_{n, k}$ preserva a propriedade martingal de $\text{FB}_{n, k}$.

Esta abordagem é adaptativa no sentido que a classe $\mathcal{M}^{(n)}$ e o limite de controle $B^{(n)}$ são

construídos com dados obtidos até o tempo $n - 1$. A vantagem desta abordagem é que o uso do Fator de Bayes Adaptativo previne que a cada n , avaliações desnecessárias sobre modelos com pobre performance preditiva no tempo anterior, tornando rápido o algoritmo de busca sobre cluster emergentes e diminuindo o atraso na detecção uma vez que diminui a cardinalidade de Ξ . Um guia para a escolha de C^* é usar a escala para o Fator de Bayes descrita em Kass e Raftery (1995). Em uma análise não sequencial, ou seja a tempo fixo, se $(J + 1)$ modelos, M_0, M_1, \dots, M_J , estão sendo considerados e M_1, \dots, M_J são comparadas com M_0 , Kass e Raftery (1995) sugeriram a seguinte interpretação:

Tabela 4.1: *Interpretação do Fator de Bayes (Kass e Raftery, 1995)*

Interpretação do Fator de Bayes		
FB	$\log(FB)$	Evidência contra M_0
$(0, 1]$	$(-\infty, 0]$	Negativa
$(1, 3]$	$(0, 1]$	Fraca
$(3, 20]$	$(1, 3]$	Positiva
$(20, 150]$	$(3, 5]$	Forte
$(150, \infty)$	$(5, \infty)$	Muito Forte

Usando esta escala nós usamos $C^* = 3$ que parece um valor razoável para uma análise preliminar no espaço Ξ . Em nossas simulações apresentadas no final deste capítulo este valor é considerado adequado para nossa aplicação.

Por fim, após a análise dos dados, se um alarme de *cluster* for disparado, estime o *cluster* emergente pelo par $(\hat{\xi}^n, k(\hat{\xi}^n))$, onde:

$$\hat{\xi}^n = \max_{\xi_l \in \Xi^{(n)}_{n-w+1 \leq k \leq n}} \max_{\xi_l} \text{FBA}_n(\xi_l) \quad (4.15)$$

e

$$k(\hat{\xi}^n) = \arg \left(\max_{n-w+1 \leq k \leq n} \text{FBA}_{n,k}(\hat{\xi}^n) \right) \quad (4.16)$$

4.4 Estrutura Geral dos Modelos

4.4.1 Modelo Padrão

Para construirmos o modelo padrão $M_{t,0}$, assumimos que no tempo t , o processo é observado sob condições normais de funcionamento. Seja $X_t(s_l)$ a variável aleatória de interesse medida em s_l e suponha que a sua densidade condicional a $\theta_t(s_l)$ é pertencente a família expo-

nencial dada por:

$$\begin{aligned} f(X_t(s_l)|\theta_t(s_l)) &= u(X_t(s_l), \phi_t(s_l)) \\ &\times \exp\{\phi_t(s_l)[X_t(s_l)\theta_t(s_l) - q(\theta_t(s_l))]\}. \end{aligned} \quad (4.17)$$

com parametro natural desconhecido $\theta_t(s_l)$ e parâmetro de precisão conhecido $\phi_t(s_l)$. Esta família de densidades tem a seguinte propriedade:

$$\mathbb{E}_0[X_t(s_l)|\theta_t(s_l)] = \frac{\partial q(\theta_t(s_l))}{\partial \theta_t(s_l)} = q'(\theta_t(s_l))$$

e

$$\mathbb{V}_0[X_t(s_l)|\theta_t(s_l)] = \frac{1}{\phi_t(s_l)} \frac{\partial^2 q(\theta_t(s_l))}{\partial \theta_t(s_l)^2} = \frac{1}{\phi_t(s_l)} q''(\theta_t(s_l)) \quad (4.18)$$

onde \mathbb{E}_0 e \mathbb{V}_0 representam, respectivamente, o valor esperado e a variância de uma variável aleatória sob o modelo $M_{t,0}$. Considere uma função monotona

$$z(\theta_t(s_l)) = q'(\theta_t(s_l)) \quad (4.19)$$

Esta define uma transformação 1 a 1 $\mu_t(s_l) = z(\theta_t(s_l))$ (com inversa $\theta_t(s_l) = z^{-1}(\mu_t(s_l))$) do espaço Θ do parâmetro $\theta_t(s_l)$ em um outro conjunto $\mathcal{E} \subset \mathcal{R}$ que representa o espaço do valor médio para a distribuição condicional de $[X_t(s_l)|\theta_t(s_l)]$. Introduzindo $\mu_t(s_l)$ em (4.18), expressamos a variância da distribuição de $[X_t(s_l)|\theta_t(s_l)]$ como uma função do valor médio, também chamada função de variância, através de

$$V(\mu_t(s_l)) = q''(z^{-1}[\mu_t(s_l)]). \quad (4.20)$$

Uma família na distribuição exponencial com parâmetro natural é unicamente determinada por sua função de variância e vice-versa (Morris, 1982). Isto é, dada uma função de variância tem-se uma classe de distribuições correspondentes. Para entender a importância da função de variância em nosso modelo padrão, suponha que

$$\begin{aligned} g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) &= v(a_t(s_l), b_t(s_l)) \\ &\times \exp\{b_t(s_l)[a_t(s_l)\theta_t(s_l)/b_t(s_l) - q(\theta_t(s_l))]\}. \end{aligned} \quad (4.21)$$

Ou seja, $g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l))$ é a densidade a *priori* de uma distribuição conjugada (Migon e Gamermam, 1999) para $\theta_t(s_l)$. Esta densidade a *priori* será denotada por $(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) \sim PC[a_t(s_l), b_t(s_l)]$ e, é completamente especificada pois o par $[a_t(s_l), b_t(s_l)]$ é assumido conhecido e função de $\mathcal{H}_{t-1}(l)$. Assuma neste trabalho que $V(\mu_t(s_l))$ é uma função no máximo quadrática, distribuições que satisfazem esta suposição são: Normal, Gama, Poisson, Binomial e Binomial negativa. Então, o par $[a_t(s_l), b_t(s_l)]$ tem uma simples interpretação em termos de $\mu_t(s_l) = z(\theta_t(s_l))$ (Gutières e Smith, 1995):

$$\mathbb{E}_0[\mu_t(s_l)|\mathcal{H}_{t-1}(l)] = a_t(s_l)/b_t(s_l),$$

e

$$b_t^{-1}(s_l) = \frac{V_0(\mu_t(s_l)|\mathcal{H}_{t-1}(l))}{\mathbb{E}_0[V_0(\mu_t(s_l))|\mathcal{H}_{t-1}(l)]} \quad (4.22)$$

com $\mu_t(s_l) = \mathbb{E}_0[X_t(s_l)|\theta_t(s_l)]$ (veja refer). O parâmetro $b_t^{-1}(s_l)$ é conhecido na área de engenharia de qualidade como relação sinal/ruído.

Agora, combinando (4.17) e (4.21) temos que a densidade preditiva de $X_t(s_l)$, veja a expressão (4.6), é

$$\begin{aligned} p_{t,0}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l)) &= \int f(X_t(s_l)|\theta_t(s_l))g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l))d\theta_t(s_l) \\ &= u(X_t(s_l), \phi_t(s_l))v(a_t(s_l), b_t(s_l)) \\ &\times \int \exp\{[a_t(s_l) + \phi_t(s_l)X_t(s_l)]\theta_t(s_l) - [b_t(s_l) + \phi_t(s_l)]q(\theta_t(s_l))\} d\theta_t(s_l) \\ &= \frac{u(X_t(s_l), \phi_t(s_l)) \times v(a_t(s_l), b_t(s_l))}{v([a_t(s_l) + \phi_t(s_l)X_t(s_l)], [b_t(s_l) + \phi_t(s_l)])}. \end{aligned} \quad (4.23)$$

Esta é a forma da densidade preditiva que serve de base para avaliar o ajuste da distribuição $[X_t(s_l)|\mathcal{H}_{t-1}(l)]$ sob o modelo padrão $M_{t,0}$. Temos, ainda,

$$\mathbb{E}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] = \mathbb{E}_0([\mathbb{E}_0[X_t(s_l)|\theta_t(s_l)]|\mathcal{H}_{t-1}(l)] = \mathbb{E}_0[\mu_t(s_l)|\mathcal{H}_{t-1}(l)] = a_t(s_l)/b_t(s_l).$$

e

$$\begin{aligned} \mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] &= \mathbb{E}_0(\mathbb{V}_0[X_t(s_l)|\theta_t(s_l)]|\mathcal{H}_{t-1}(l)) \\ &+ \mathbb{V}_0(\mathbb{E}_0[X_t(s_l)|\theta_t(s_l)]|\mathcal{H}_{t-1}(l)) \end{aligned}$$

Combinando (4.18) com (4.20) obtemos em termos de $\mu_t(s_l)$,

$$\begin{aligned} \mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] &= \mathbb{E}_0\left(\frac{1}{\phi_t(s_l)}V_0(\mu_t(s_l)|\mathcal{H}_{t-1}(l))\right) + V_0(\mu_t(s_l)|\mathcal{H}_{t-1}(l)) \\ &= \mathbb{E}_0\left(\frac{1}{\phi_t(s_l)}V_0(\mu_t(s_l)|\mathcal{H}_{t-1}(l))\right) + b_t^{-1}(s_l)\mathbb{E}_0[V_0(\mu_t(s_l))|\mathcal{H}_{t-1}(l)] \\ &= \left(\frac{1}{\phi_t(s_l)} + \frac{1}{b_t(s_l)}\right)\mathbb{E}_0[V_0(\mu_t(s_l))|\mathcal{H}_{t-1}(l)]. \end{aligned}$$

a qual é equivalente em termos de $\theta_t(s_l)$

$$\mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] = \left(\frac{1}{\phi_t(s_l)} + \frac{1}{b_t(s_l)}\right)\mathbb{E}_0[q''(\theta_t(s_l))|\mathcal{H}_{t-1}(l)].$$

4.4.2 Modelo Alternativo

Observa-se na expressão para a variância da distribuição de $[X_t(s_l)|\mathcal{H}_{t-1}(l)]$, que esta depende da distribuição a priori através de $b_t(s_l)$; a quantidade $\phi_t^{-1}(s_l) + b_t^{-1}(s_l)$ representa uma

medida de precisão da densidade preditiva que tem valor esperado $a_t(s_l)/b_t(s_l)$. Logo fixando $m_t = a_t(s_l)/b_t(s_l)$ e decrescendo $b_t(s_l)$ obtemos uma densidade preditiva mais difusa e crescendo $b_t(s_l)$ o efeito é reverso. Então, uma densidade preditiva apropriada para detecção de mudanças em $M_{t,0}$ depois de um tempo k em algum conjunto ξ , pode ser construída por substituir em (4.23) $b_t(s_l)$ por outro valor menor que $b_t(s_l)$ para todo $t \geq k$ e $s_l \in \xi$. Esta substituição é feita diretamente por usarmos a distribuição a priori do tipo lei de potência dada em (4.3) com $0 < \gamma_{t,\xi} < 1$. Para esta distribuição a priori em (4.3) obtemos a densidade preditiva para $[X_t(s_l)|\mathcal{H}_{t-1}(l)]$ sob o modelo $M_{t,\xi}$ por $p_{t,\xi}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l))$ em (4.5). Neste caso a distribuição a priori em (4.21) é substituída por

$$\begin{aligned} g_{t,\xi}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) &= v(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l)) \\ &\times \exp\{\gamma_{t,\xi}b_t(s_l)[a_t(s_l)\theta_t(s_l)/b_t(s_l) - q(\theta_t(s_l))]\} \end{aligned} \quad (4.24)$$

que representa a distribuição a priori para $\sim \theta_t(s_l)|\mathcal{H}_{t-1}(l)$ sob $M_{t,\xi}$. Combinando (4.17) com (4.24) obtenha que,

$$\begin{aligned} p_{t,\xi}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l)) &= \int f(X_t(s_l)|\theta_t(s_l))g_{t,\xi}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l))d\theta_t(s_l) \\ &= u(X_t(s_l), \phi_t(s_l))v(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l)) \\ &\times \int \exp\{[\gamma_{t,\xi}a_t(s_l) + \phi_t(s_l)X_t(s_l)]\theta_t(s_l) - [\gamma_{t,\xi}b_t(s_l) + \phi_t(s_l)]q(\theta_t(s_l))\} \\ &= \frac{u(X_t(s_l), \phi_t(s_l)) \times v(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l))}{v([\gamma_{t,\xi}a_t(s_l) + \phi_t(s_l)X_t(s_l)], [\gamma_{t,\xi}b_t(s_l) + \phi_t(s_l)])}. \end{aligned} \quad (4.25)$$

Neste caso a densidade (4.25) é obtida pela substituição do par $[a_t(s_l), b_t(s_l)]$ por $[\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l)]$ em (4.23). Por usar este fato, obtenha diretamente que,

$$\mathbb{E}_{k,\xi}[X_t(s_l)|\mathcal{H}_{t-1}(l)] = \mathbb{E}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] = a_t(s_l)/b_t(s_l).$$

e

$$\begin{aligned} \mathbb{V}_{k,j}[X_t(s_l)|\mathcal{H}_{t-1}(l)] &= \left(\frac{1}{\phi_t(s_l)} + \frac{1}{\gamma_{t,\xi}b_t(s_l)} \right) \mathbb{E}_0[q''(\theta_t(s_l))|\mathcal{H}_{t-1}(l)] \\ &> \mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)], \quad \text{para } 0 < \gamma_{t,\xi} < 1. \end{aligned} \quad (4.26)$$

Onde $\mathbb{E}_{k,\xi}$ e $\mathbb{V}_{k,j}$ representam, respectivamente, o valor esperado e a variância de uma variável aleatória sobre o modelo $M_{t,\xi}$ para $t \geq k$. Nota-se então que: 1) o valor esperado para as distribuições preditivas é o mesmo sob os modelos $M_{t,0}$ e $M_{t,\xi}$. 2) O spread $\left(\frac{1}{\phi_t(s_l)} + \frac{1}{\gamma_{t,\xi}b_t(s_l)} \right)$ da densidade preditiva sob $M_{t,\xi}$ é maior que a precisão $\left(\frac{1}{\phi_t(s_l)} + \frac{1}{b_t(s_l)} \right)$ da distribuição preditiva sob $M_{t,0}$. Isto significa que $p_{t,\xi}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l))$ pode ter a forma de uma versão suavizada de $p_{t,0}^{(l)}(X_t(s_l)|\mathcal{H}_{t-1}(l))$ (ver figura 4.1). Deste modo, a densidade preditiva do modelo alternativo $M_{t,\xi}$ dá maior probabilidade a regiões não favorecidas pela densidade do modelo padrão $M_{t,0}$ e será mais difusa em regiões com alta densidade preditiva sob $M_{t,0}$. Para que a distribuição a

priori $\theta_t(s_l)|\mathcal{H}_{t-1}(l)$ seja completamente especificada sobre o modelo $M_{t,\xi}$ precisamos especificar o valor de $\gamma_{t,\xi}$. Para isso suponha inicialmente que para cada $s_l \in \xi$ determinamos um valor $\gamma_{t,l}$. Então usamos

$$\gamma_{t,\xi} = \frac{1}{|\xi|} \sum_{s_l \in \xi} \gamma_{t,l}$$

onde $|\xi|$ representa a cardinalidade de ξ . A quantidade $\gamma_{t,l}$ é calculada como em West (1986) e dada por,

$$\gamma_{t,l} = \rho/[1 + (1 - \rho)(b_t(s_l)/\phi_t(s_l))]$$

para algum $0 < \rho < 1$ fixo. Esta escolha implica que a precisão relativa entre as densidades preditivas sob os modelos padrão e alternativo é constante no tempo e expressa por

$$\frac{\left(\frac{1}{\phi_t(s_l)} + \frac{1}{\gamma_{t,l}b_t(s_l)}\right)}{\left(\frac{1}{\phi_t(s_l)} + \frac{1}{b_t(s_l)}\right)} = \frac{1}{\rho}. \quad (4.27)$$

Notando que neste caso,

$$\frac{\mathbb{V}_{k,j}[X_t(s_l)|\mathcal{H}_{t-1}(l)]}{\mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)]} = \frac{\left(\frac{1}{\phi_t(s_l)} + \frac{1}{\gamma_{t,l}b_t(s_l)}\right) \mathbb{E}_0[q''(\theta_t(s_l))|\mathcal{H}_{t-1}(l)]}{\left(\frac{1}{\phi_t(s_l)} + \frac{1}{b_t(s_l)}\right) \mathbb{E}_0[q''(\theta_t(s_l))|\mathcal{H}_{t-1}(l)]} = \frac{1}{\rho}. \quad (4.28)$$

Então podemos dizer que o desvio padrão preditivo em $M_{t,\xi}$ é um fator $1/\sqrt{\rho}$ vezes maior que o desvio padrão preditivo em $M_{t,0}$. Para $\rho \in (0.11, 0.44)$ este fator pertence ao intervalo $(1.5, 3)$ West(1986) tem recomendado valores no intervalo $(0.1, 0.3)$ e nós usamos $\rho = 0.2$ de modo que $1/\sqrt{\rho} \approx 2$. Para finalizar esta seção, combinamos (4.25) com (4.23) e (4.4), obtemos que para qualquer ξ com estrutura de modelo de cluster dado nesta seção o Fator de Bayes é expresso por:

$$\begin{aligned} \text{FB}_t(\xi) &= \prod_{s_l \in \xi} \frac{u(X_t(s_l), \phi_t(s_l)) \times v(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l))}{v([\gamma_{t,j}a_t(s_l) + \phi_t(s_l)X_t(s_l)], [\gamma_{t,\xi}b_t(s_l) + \phi_t(s_l)])} \\ &\times \frac{v([a_t(s_l) + \phi_t(s_l)X_t(s_l)], [b_t(s_l) + \phi_t(s_l)])}{u(X_t(s_l), \phi_t(s_l)) \times v(a_t(s_l), b_t(s_l))} \\ &= \prod_{s_l \in \xi} \frac{v(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l))}{v(a_t(s_l), b_t(s_l))} \\ &\times \frac{v([a_t(s_l) + \phi_t(s_l)X_t(s_l)], [b_t(s_l) + \phi_t(s_l)])}{v([\gamma_{t,j}a_t(s_l) + \phi_t(s_l)X_t(s_l)], [\gamma_{t,\xi}b_t(s_l) + \phi_t(s_l)])} \quad t \geq k. \end{aligned} \quad (4.29)$$

4.5 Representação Espacial do cluster

Para contruímos um algoritmo para a detecção e localização do *cluster*, primeiramente reconhecemos que a noção de um conjunto de localizações espaciais ξ é inerentemente vaga e carece de qualquer definição a priori. Pode não haver verdadeiro ξ no sentido de um verdadeiro valor do parâmetro. Por exemplo, Ξ pode ser definido como um conjunto de zonas circulares

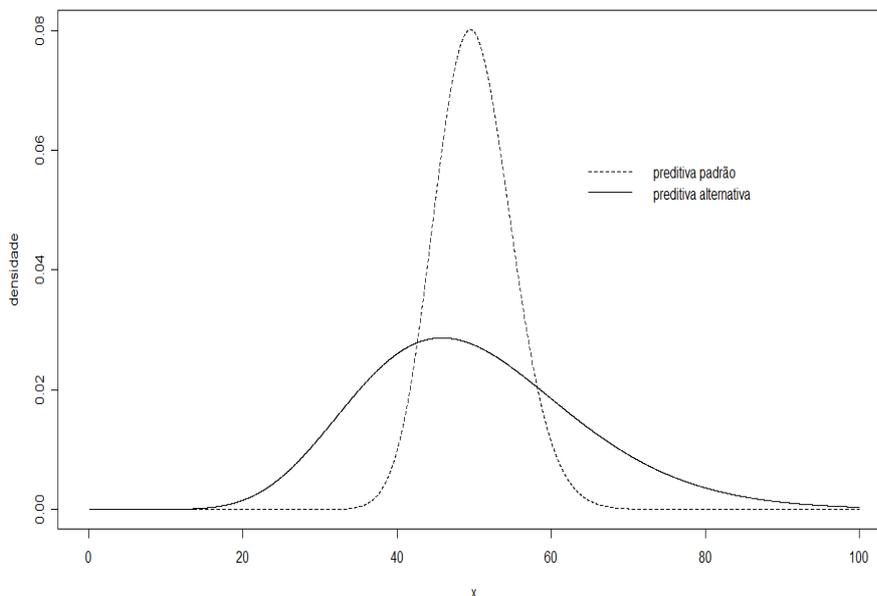


Figura 4.1: Uma possível representação para as densidades preditivas em relação aos modelos padrão e alternativo.

como as definidas na seção 2.1 (Openshaw et al., 1988; Besag e Newel, 1991; Kulldorff, 1997). No entanto, a mudança pode ter ocorrido em um conjunto de localizações espaciais com formato irregular igual ao descrito em Duczmal and Assunção (2004). Na prática, um conjunto ξ é determinado por uma estimativa ou aproximação fornecida pelos dados. Quanto mais flexível, em termos de formato geométrico for a classe Ξ , melhor será essa aproximação. Essa flexibilidade pode torna a cardinalidade de Ξ excessivamente grande, mas ao usarmos a abordagem adaptativa na classe Ξ nós podemos controlar o valor de sua cardinalidade o que permite o uso de outros algoritmos para a representação espacial do *cluster*. Por isso neste capítulo nós usamos uma representação elíptica (Kulldorff et al., 2006) para a localização espacial do *cluster*

Para construção de Ξ considere que ξ possui forma, tamanho e direção desconhecida. Uma classe de representações geométricas Ξ pode ser definida através de elipses. Primeiramente observe a figura 4.2. Esta figura apresenta uma elipse com orientação arbitrária centrada em um ponto específico s , que representa o centróide de uma área do mapa. Nesta elipse temos os seguintes elementos:

- a e b são os semi-eixos da elipse. " a " é o semi-eixo maior e " b " é o menor.
- " e " é a excentricidade.
- F_1 e F_2 são os focos dessa elipse.

Uma elipse centrada em um ponto s pode ser unicamente definida por três parâmetros: excentricidade, " e ", parâmetro de forma; ângulo= ϱ , define a direção da elipse; o tamanho é

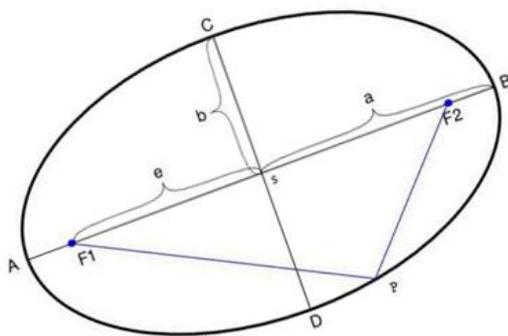


Figura 4.2: Exemplo de uma Elipse

mensurado pelo semi-eixo maior $a = \varsigma$. Em uma elipse, temos que $e^2 = 1 - (a/b)^{-2}$ de modo que o parâmetro $\varphi = a/b$ define a forma da elipse. Quando φ é grande a elipse é longa e estreita, quando $\varphi = 1$ temos um círculo. O parâmetro ϱ é o ângulo entre a linha horizontal e o semi-eixo maior da elipse. Em um mapa nós podemos construir infinitas elipses. No entanto, nós consideramos apenas um conjunto finito de elipses centradas em cada centróide s_l . Em cada ponto s_l nós podemos construir várias elipses com forma (φ), tamanho (ς) e direção (ϱ) diferente. Então, seja ξ_{l_j} uma elipse centrada em s_l com parâmetros $(\varphi_j, \varsigma_j, \varrho_j)$, desde que conheçamos as coordenadas geográficas de s_l e os pontos extremos do mapa, os valores $(\varphi_j, \varsigma_j, \varrho_j)$ estão todos bem definidos de modo que a representação de ξ_{l_j} é única. Na figura figura 4.3 apresentamos algumas formas geométricas de *cluster* que podem ser detectadas: C1 e C3 são elipses com orientações diferentes e C2 tem um formato aproximadamente circular.

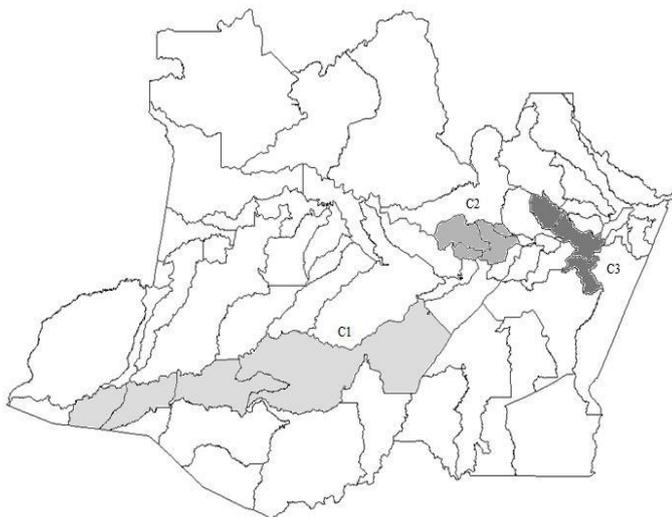


Figura 4.3: Formato de Clusters que podem ser detectados

Uma observação importante, é que para cada forma elíptica podemos construir um grande número de diferentes ângulos. Para uma elipse com forma mais compacta, tal que $\varphi = 1.5$, os centróides incluídos em uma elipse não mudam com uma suave mudança no valor do ângulo ϱ , ou seja várias elipses representarão o mesmo *cluster*. No extremo, os centróides das regiões incluídas na elipse com $\varphi = 1$ (elipse=círculo) não dependem do ângulo. Para uma longa e

estreita elipse (ς grande e φ pequeno), uma pequena mudança no ângulo resutará em muitos conjuntos com centróides diferentes. Desta maneira, parece lógico que devemos usar uma grande coleção de ângulos quando ς é grande e φ é pequeno. Neste caso podemos fazer um refinamento *a priori* no espaço dos *cluster* para avaliar suaves valores para o parâmetro ϱ . Podemos calcular, por exemplo, o percentual de intersecção entre duas elipses centradas no mesmo ponto s_l , com os mesmos parâmetros de forma φ e tamanho ς porém, com ângulos e quantidades de ângulos diferentes. A partir deste refinamento, podemos eliminar as elipses que possuem valores de ϱ que não são de interesse científico, ou não fazem sentido serem avaliados. Na tabela 4.2, por exemplo é apresentado o percentual de área de uma elipse que é também parte de qualquer outra elipse com o mesmo centro (s_l), forma (φ) e tamanho (ς), porém com quantidades (#) diferentes de ângulos. Segundo a sugestão apresentada em Kulldorff(2006), usaremos o número de ângulos escolhidos de forma que duas elipses de mesma forma, centro e tamanho, com ângulos consecutivos tenham intersecção de no mínimo 70% de suas áreas. Assim (ver tabela 4.2) tomamos dois ângulos para elipses com $\varphi = 1.5$, seis ângulos para elipses com $\varphi = 2.0$, etc...

Tabela 4.2: *Percentual de área de uma elipse que é também parte de qualquer outra elipse com o mesmo centro, forma e tamanho, porém com quantidades (#) diferentes de ângulos*

# de ângulo ϱ	parâmetro de forma φ				
	1.5	2.0	4.0	6.0	8.0
2 (90°)	75	59	31	21	16
4 (45°)	82	67	41	29	22
6 (30°)	87	77	52	38	30
9 (20°)	91	84	64	50	41
12 (15°)	93	88	71	59	49
15 (12°)	94	90	76	65	56
18 (10°)	95	92	80	70	62
24 (7.5°)	95	94	85	77	70
30 (6°)	97	95	88	81	75
36 (5°)	98	95	90	84	79
45 (4°)	98	97	92	87	83
60 (3°)	99	98	94	90	87
90 (2°)	99	98	96	94	91
180(1°)	100	99	98	97	96
360(0.5°)	100	100	99	98	98

4.6 Exemplo

Para entendermos melhor a estrutura dos modelos discutidos na seção 4.4.1, suponha que monitoramos o processo $\{Y_t(s_l) : l = 1, 2, \dots, L \text{ e } t = 1, 2, \dots\}$ com realização $\mathbf{y}_t = (y_t(s_1), y_t(s_2), \dots, y_t(s_L))'$. Em que, $y_t(s_l)$ representa o número de casos observados de uma determinada doença no tempo t em s_l . Para a distribuição do número de casos, assumimos que,

$$(Y_t(s_l)|\vartheta_t(s_l)) \sim \text{Poisson}(e_t(s_l) \times \vartheta_t(s_l)) \quad (4.30)$$

com,

$$f(y_t(s_l)|\vartheta_t(s_l)) = (y_t(s_l)!)^{-1} e_t(s_l)^{y_t(s_l)} \exp \left\{ e_t(s_l) \left(\frac{y_t(s_l)}{e_t(s_l)} \log \vartheta_t(s_l) - \vartheta_t(s_l) \right) \right\}.$$

identificando os termos em (4.17) temos que

$$\phi_t(s_l) = e_t(s_l), \quad \theta_t(s_l) = \log \vartheta_t(s_l), \quad q(\theta_t(s_l)) = \exp\{\theta_t(s_l)\} = \vartheta_t(s_l).$$

$$X_t(s_l) = Y_t(s_l)/e_t(s_l), \quad u(X_t(s_l), e_t(s_l)) = \left((Y_t(s_l)!)^{-1}, e_t(s_l)^{Y_t(s_l)} \right).$$

Notamos ainda que pelas definições em (4.19) e (4.20) obtemos que a média de $(X_t(s_l)|\theta_t(s_l))$

$$\mu_t(s_l) = z(\theta_t(s_l)) = q'(\theta_t(s_l)) = \exp\{\theta_t(s_l)\} = \vartheta_t(s_l)$$

de forma similar obtemo que a função de variância

$$V(\mu_t(s_l)) = q''(z^{-1}[\mu_t(s_l)]) = \vartheta_t(s_l).$$

Neste modelo, $e_t(s_l)$ é assumido conhecido e representa a contagem esperada da doença no tempo t em s_l e, $\boldsymbol{\vartheta}_t = (\vartheta_t(s_1), \vartheta_t(s_2), \dots, \vartheta_t(s_L))'$ é um vetor de riscos relativos no tempo t . Condicional ao processo $\{\boldsymbol{\vartheta}_t\}$, $\{Y_t\}$ é um processo de Poisson no espaço-tempo. A idéia deste modelo, é que em cada área as observações flutuam em torno de uma média $e_t(s_l)$. Porém, essa média está sujeita a variações que são quantificadas por $\vartheta_t(s_l)$. Se a doença está sob-controle sobre toda a região em estudo \mathcal{S} , esperamos que seu risco relativo seja próximo de 1, tal que a distribuição a *a priori* para $(\vartheta_t(s_l)|\mathcal{D}_{t-1})$ deve estar concentrada no intervalo $(1 - \epsilon, 1 + \epsilon)$ para algum pequeno valor ϵ apresentando poucas variações. Caso contrário, para um processo realizado sobre um possível surto ou epidemia, estas variações são maiores e uma distribuição a *priori* mais dispersa é apropriada para a detecção de possíveis mudanças.

A distribuição a *priori* padrão para $\theta_t(s_l) = \log \vartheta_t(s_l)$ é, de acordo com a expressão (4.21)

$$\begin{aligned}
g_{t,0}^{(l)}(\theta_t(s_l)|\mathcal{H}_{t-1}(l)) &= v(a_t(s_l), b_t(s_l)) \\
&\times \exp\{b_t(s_l)[a_t(s_l)\log\vartheta_t(s_l)/b_t(s_l) - \vartheta_t(s_l)]\}. \quad (4.31)
\end{aligned}$$

Usando para $(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l))$ uma distribuição a *priori* padrão na família conjugada $\sim Gama(a_t(s_l) + 1, b_t(s_l))$ com média $[a_t(s_l) + 1]/b_t(s_l)$ e variância $[a_t(s_l) + 1]/b_t^2(s_l)$ obtemos que,

$$\begin{aligned}
g_{t,0}^{(l)}(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) &= v(a_t(s_l) + 1, b_t(s_l)) \\
&\times \exp\{b_t(s_l)[a_t(s_l)\log\vartheta_t(s_l)/b_t(s_l) - \vartheta_t(s_l)]\}. \quad (4.32)
\end{aligned}$$

que tem a mesma estrutura de (4.31). Então, a *priori* para o modelo alternativo descrita em 4.24 pode ser aplicada usando uma distribuição a *priori* para $\theta_t(s_l)$ ou $\vartheta_t(s_l)$. No primeiro, caso os parâmetros $(a_t(s_l), b_t(s_l))$ da distribuição a *priori* padrão são substituídos por $(\gamma_{t,\xi}a_t(s_l), \gamma_{t,\xi}b_t(s_l))$. No segundo caso, a substituição é feita por $(\gamma_{t,\xi}(a_t(s_l) + 1), \gamma_{t,\xi}b_t(s_l))$. O segundo caso é usado até o final desta seção. Usando, então, a distribuição a *priori* para $(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l))$, obtemos que o valor esperado e a variância para distribuição preditiva de $X_t(s_l) = Y_t(s_l)/e_t(s_l)$ sob o modelo padrão são:

$$\begin{aligned}
\mathbb{E}_0(X_t(s_l)|\mathcal{H}_{t-1}(l)) &= \mathbb{E}_0 \left[\mathbb{E}_0 \left(\frac{Y_t(s_l)}{e_t(s_l)} | \vartheta_t(s_l) \right) | \mathcal{D}_{t-1} \right] \\
&= \frac{1}{e_t(s_l)} \mathbb{E}_0(e_t(s_l)\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) = \mathbb{E}_0(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) = \frac{[a_t(s_l) + 1]}{b_t(s_l)}.
\end{aligned}$$

e,

$$\begin{aligned}
\mathbb{V}_0(X_t(s_l)|\mathcal{H}_{t-1}(l)) &= \mathbb{E}_0 \left[\mathbb{V}_0 \left(\frac{Y_t(s_l)}{e_t(s_l)} | \vartheta_t(s_l) \right) | \mathcal{D}_{t-1} \right] + \mathbb{V}_0 \left[\mathbb{E}_0 \left(\frac{Y_t(s_l)}{e_t(s_l)} | \vartheta_t(s_l) \right) | \mathcal{D}_{t-1} \right] \\
&= \frac{1}{e_t(s_l)} \mathbb{E}_0[\vartheta_t(s_l)|\mathcal{D}_{t-1}] + \mathbb{V}_0[\vartheta_t(s_l)|\mathcal{D}_{t-1}] \\
&= \frac{[a_t(s_l) + 1]}{e_t(s_l)b_t(s_l)} + \frac{[a_t(s_l) + 1]}{b_t^2(s_l)} = \left(\frac{1}{e_t(s_l)} + \frac{1}{b_t(s_l)} \right) \frac{[a_t(s_l) + 1]}{b_t(s_l)}. \quad (4.33)
\end{aligned}$$

Substituindo os parâmetros $(a_t(s_l) + 1, b_t(s_l))$ da distribuição a *priori* padrão por $(\gamma_{t,\xi}(a_t(s_l) + 1), \gamma_{t,\xi}b_t(s_l))$ na distribuição a *priori* alternativa, obtenha diretamente que:

$$\mathbb{E}_{k,\xi}[X_t(s_l)|\mathcal{H}_{t-1}(l)] = \mathbb{E}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)] = \frac{[a_t(s_l) + 1]}{b_t(s_l)}.$$

e

$$\begin{aligned}
\mathbb{V}_{k,j}[X_t(s_l)|\mathcal{H}_{t-1}(l)] &= \left(\frac{1}{e_t(s_l)} + \frac{1}{\gamma_{t,\xi}b_t(s_l)} \right) \frac{[a_t(s_l) + 1]}{b_t(s_l)} \\
&> \mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)], \quad \text{para } 0 < \gamma_{t,\xi} < 1. \quad (4.34)
\end{aligned}$$

Note agora que,

$$\frac{\mathbb{V}_{k,j}[X_t(s_l)|\mathcal{H}_{t-1}(l)]}{\mathbb{V}_0[X_t(s_l)|\mathcal{H}_{t-1}(l)]} = \frac{\left(\frac{1}{e_t(s_l)} + \frac{1}{\gamma_{t,l}b_t(s_l)}\right)}{\left(\frac{1}{e_t(s_l)} + \frac{1}{b_t(s_l)}\right)} = \frac{1}{\rho}. \quad (4.35)$$

Então, comparando 4.27, 4.28 e 4.35, conclua que usando a distribuição a *priori* para $\theta_t(s_l)$ ou $\vartheta_t(s_l)$ o resultado da análise não se altera.

Para calcularmos o Fator de Bayes, usamos o fato que $(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) \sim \text{gamma}(a_t(s_l) + 1, b_t(s_l))$. Então

$$\begin{aligned} g_{t,0}^{(l)}(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) &= \frac{b_t(s_l)^{[a_t(s_l)+1]}}{\Gamma([a_t(s_l) + 1])} \\ &\times \exp\{b_t(s_l)[a_t(s_l)\log\vartheta_t(s_l)/b_t(s_l) - \vartheta_t(s_l)]\}. \end{aligned} \quad (4.36)$$

identifique que a função $v(a_t(s_l) + 1, b_t(s_l))$ dada em 4.36 é igual a

$$v(a_t(s_l) + 1, b_t(s_l)) = \frac{b_t(s_l)^{[a_t(s_l)+1]}}{\Gamma([a_t(s_l) + 1])}. \quad (4.37)$$

Substituindo 4.37 em 4.29 com $\phi_t(s_l) = e_t(s_l)$ e $X_t(s_l) = Y_t(s_l)/e_t(s_l)$. Obtemos que Fator de Bayes em 4.29 é igual a

$$\text{FB}_t(\xi) = \prod_{s_l \in \xi} \frac{PG_\xi(\gamma_{t,\xi}b_t(s_l), \gamma_{t,\xi}a_t(s_l))}{PG_0(b_t(s_l), a_t(s_l))}. \quad (4.38)$$

onde $PG(.,.)$, denota a distribuição Poisson-Gama com densidade abaixo.

$$\begin{aligned} PG(\gamma_{t,\xi}b_t(s_l), \gamma_{t,\xi}a_t(s_l)) &= \frac{\Gamma(\gamma_{t,\xi}[a_t(s_l) + 1] + Y_t(s_l))}{\Gamma(\gamma_{t,\xi}[a_t(s_l) + 1])Y_t(s_l)!} \left(\frac{\gamma_{t,\xi}b_t(s_l)}{\gamma_{t,\xi}b_t(s_l) + e_t(s_l)}\right)^{\gamma_{t,\xi}[a_t(s_l)+1]} \\ &\times \left(\frac{e_t(s_l)}{\gamma_{t,\xi}b_t(s_l) + e_t(s_l)}\right)^{Y_t(s_l)} \end{aligned}$$

e,

$$\begin{aligned} PG(b_t(s_l), a_t(s_l)) &= \frac{\Gamma([a_t(s_l) + 1] + Y_t(s_l))}{\Gamma([a_t(s_l) + 1])Y_t(s_l)!} \left(\frac{b_t(s_l)}{b_t(s_l) + e_t(s_l)}\right)^{[a_t(s_l)+1]} \\ &\times \left(\frac{e_t(s_l)}{b_t(s_l) + e_t(s_l)}\right)^{Y_t(s_l)} \end{aligned}$$

Então, neste caso, após observar uma sequência de vetores $(\mathbf{y}_1, \dots, \mathbf{y}_n)$, comparar a sequência de modelos $\{M_{n,0}\}_{n \geq 1}$ com a sequência $\{M_{n,\xi}\}_{n \geq 1}$ é equivalente a comparar uma sequência de razões de densidades preditivas na família de densidades Poisson-Gama.

Agora assumamos que no tempo inicial a classe de conjuntos de localizações espaciais é igual a Ξ^0 . Então, após observar o processo n vezes e atualizar Ξ^n como explicado na seção 4.3.3. O

Fator de Bayes Adaptativo dado pela expressão 4.13 é igual a

$$\text{FBA}_{n,k}(\xi_l^{n-1}) = \text{FB}_n(\xi_l^{n-1}) \prod_{t=k}^{n-1} \prod_{s_l \in \xi_l^{n-1}} \frac{PG_\xi(\gamma_{t,\xi_l^{n-1}} b_t(s_l), \gamma_{t,\xi_l^{n-1}} a_t(s_l))}{PG_0(b_t(s_l), a_t(s_l))}.$$

com função de alarme,

$$\text{FBA}_n = \max_{\xi_l^{n-1} \in \Xi^{n-n-w+1}} \max_{1 \leq k \leq n} \text{FBA}_{n,k}(\xi_l^{n-1}). \quad (4.39)$$

e, tempo de alarme

$$\tilde{\tau} = \tau(B^{(n)}) = \min \left\{ n \geq w : \text{FBA}_n \geq B^{(n)} \right\}. \quad (4.40)$$

com $B^{(n)} = 2J^{(n)}w/\alpha$, para algum par (w, α) fixo. Após o cômputo destas quantidades, use o algoritmo seguinte para a detecção de mudanças no espaço-tempo, usando o fato que Ξ tem representação elíptica.

• **Algoritmo para a detecção de clusters emergentes no espaço tempo**

1. no tempo $n = 0$, Escolher um centróide no mapa.
2. Calcular as distâncias para os outros centróides. Ordene estas distâncias em ordem crescente e depois memoriza-las em um vetor de distâncias (este vetor de distâncias é um vetor de adjacências).
3. Para cada centróide do mapa repetir os passos 1 e 2 e então construa a matriz de adjacências de acordo com a explicação dada na seção 3.2.5.
4. Escolher novamente um centróide s_l na região. Centrado em s_l construa elipses (ξ_{l_j}) da seguinte forma:
 - ★ O semi-eixo maior (ς_j) da c -ésima elipse ξ_{l_j} é igual a distância $d(s_l, s_c)$ entre os centróides s_l e s_c . Essas distâncias já estão armazenadas na matriz de adjacências definida no passo 3.
 - ★ defina $0 \leq e_j < 1$ um valor de excentricidade para elipse ξ_{l_j} e encontre φ_j através da relação $e_j^2 = 1 - \varphi_j^{-2}$. Quando $e_j = 0$, temos que ξ_{l_j} é um círculo. O semi-eixo menor da elipse é $\varphi_j^{-1}\varsigma_j$.
 - ★ o ângulo ϱ_j da elipse ξ_{l_j} é definido como o ângulo entre a linha horizontal leste-oeste e o eixo maior da elipse.
5. Escolha um valor para c (por exemplo, $c = L/4$) e construa

$$\Xi^0 = \{ \xi_{l_i} : (\langle \xi_{l_i}, \xi_{l_i} \rangle) \leq c, l = 1, 2, \dots, L; i = 1, 2, \dots, L \times \text{nang} \times \text{nex} \}.$$

Onde nang e nex representam, respectivamente, o número de ângulos e de excêntridade utilizados. Neste caso, c é o maior número de localizações espaciais em ξ_{l_i} que possui semi-eixo maior (ς) igual a $d(s_l, s_c)$.

6. Para cada $n \geq 1$ atualize a classe de subconjuntos de localizações espaciais Ξ^n como explicado na seção 4.3.3 e calcule o par $(FBA_n, B^{(n)})$.
7. Se $FBA_n \geq B^{(n)}$, declare que existe um *cluster* emergente no mapa e o estime pelo par $(\hat{\xi}^n, k(\hat{\xi}^n))$ dados respectivamente em 4.15 e 4.16. Caso contrário incremente uma unidade no tempo e volte ao passo 6.

4.7 Uma aplicação a dados reais: Detecção de cluster de Hanseníase no Amazonas-Brasil

Para ilustrarmos a utilização do método proposto para detecção de *cluster* emergentes em problemas reais, nós realizamos uma aplicação usando os dados de registros de novos casos de Hanseníase em menores de 15 anos no Estado do Amazonas no período de janeiro de 2008 a dezembro de 2009. Embora, esses casos de hanseníase tenham ocorrido a tempo fixado no passado e uma análise retrospectiva seja uma escolha mais adequada, nós usaremos estes dados em uma análise prospectiva considerando dezembro de 2009 como tempo atual de vigilância. O total de novos casos observados no período em estudo foram 1287 com uma taxa mensal de 0.16 casos para cada 10 mil habitantes. Os dados contém informações dos novos casos de Hanseníase para os 62 municípios do estado. A hanseníase é uma doença infecto-contagiosa, por isso é passível de ser transmitida de pessoa para pessoa. É uma doença endêmica no estado, porém pode evoluir para uma epidemia. Esta doença está relacionada com as condições de vida da população e seu acompanhamento epidemiológico é realizado por meio do coeficiente de detecção de novos casos. Este coeficiente é obtido dividindo-se o total de novos casos pela população em risco e depois multiplicando o resultado por 10000 habitantes. O coeficiente de detecção em menores de 15 anos é um indicador que reflete a gravidade do nível endêmico da hanseníase e a exposição precoce ao *Mycobacterium leprae*, por isso há um interesse maior em sua vigilância.

4.7.1 Análise exploratória dos dados

No acompanhamento epidemiológico, parâmetros usados pela secretária estadual de saúde (Imbiriba, 2009), estabelecem como hiperendêmico um coeficiente de detecção acima de 4 por 10000 habitantes; muito alto, um coeficiente entre (4, 2); alto: (2, 1); médio ou normal: (1, 0.2) e baixo: menor que 0.2. Analisando a Figura 4.4, observamos que de uma forma geral, o Estado do Amazonas encontra-se em um período não endêmico. No entanto, este padrão não é replicado para os seus municípios. Por exemplo, ao observarmos a Figura 4.4 notamos uma grande variação no coeficiente de detecção entre as cidades de Manaus, Humaitá e Itacoatiara. Manaus aparentemente, possui uma grande variação no padrão de casos durante este período e o mesmo ocorre na cidade de Itacoatiara. Na cidade de Humaitá parece ter surgido um *cluster* emergente no mês 12. Este tipo de doença (como a Hanseníase) é muito comum na maioria dos

municípios uma vez que ela está associada a condições de vida da população, de modo que a nível local os componentes de uma série temporal, como tendência e sazonalidade podem ser ignorados na análise.

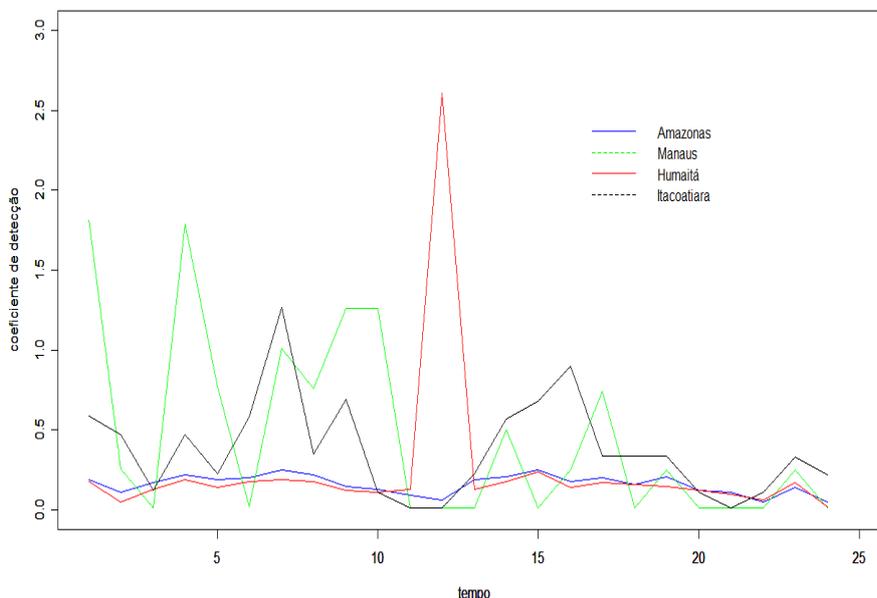


Figura 4.4: *Coeficiente de Detecção de Hanseníase (por 10000 ml hab) no Estado do Amazonas(em azul), segundo o mês, nas cidades: Manaus(em verde), Humaitá (em vermelho) e Itacoatiara (em preto)*

4.7.2 Aplicação do Fator de Bayes Adaptativo para os dados de Hanseníase

Seja $X_t(s_l)$ o número de novos casos da doença, Hanseníase, no tempo $t = 1, 2, \dots, n = 24$ (jan/2008 a dez/2009) na l -ésima área A_l identificada por seu centróide s_l . A_l representa um município do Estado do Amazonas, com $l = 1, 2, \dots, L = 62$. Cada um destes municípios é representado espacialmente pelas coordenadas geográficas de seu centróide, apresentados na figura 4.5. Os dados de população e coordenadas geográficas foram obtidos na Secretaria de Planejamento do Estado do Amazonas-SEPLAN.

Assuma que $\mathbf{Y}_t = (Y_t(s_1), \dots, Y_t(s_L))'$ é um processo de Poisson de contagem em dados de área no espaço-tempo tal que $Y_t(s_l) \sim \text{Poisson}(e_t(s_l) \times \vartheta_t(s_l))$. Onde, $e_t(s_l)$ é assumido conhecido e representa a contagem esperada da doença no tempo t em s_l e, $\boldsymbol{\vartheta}_t = (\vartheta_t(s_1), \vartheta_t(s_2), \dots, \vartheta_t(s_L))'$ é um vetor de riscos relativos no tempo t para cada município. Então o Fator de Bayes $FB_t(\boldsymbol{\xi})$ é dado pela expressão 4.38 onde $\boldsymbol{\xi}$ tem representação espacial dada na seção 4.5. Para aplicar o método de monitoramento no espaço-tempo, primeiro precisamos determinar o número de casos esperados em cada município do mapa. Em mapeamento de doença, o número de casos esperados pode ser calculado através de uma padronização externa (ajustando uma taxa padrão conhecida de casos pela população em risco) ou padronização interna (usando os dados observados para

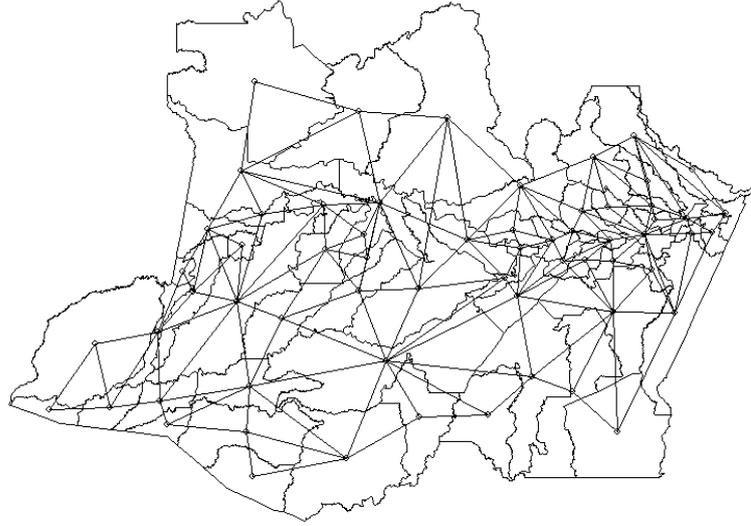


Figura 4.5: Mapa do Estado do Amazonas com os municípios espacialmente representados pelas coordenadas de seus centróides.

estimar o valor esperado), veja por exemplo, Banerjee et al, (2004). Geralmente a abordagem via padronização externa é a melhor (Zhou e Lawson, 2008) pois evita o uso dos dados corrente para estimarmos quantidades que em teoria deveriam ser conhecidas. Uma forma de usar a padronização interna sem utilizar os dados correntes é usar métodos (como o Fator de Bayes usado aqui) com janela temporal w que permitem que o valor esperado $e_t(s_l)$, quando desconhecido, possa ser estimado com os dados observados no intervalo de tempo $t = 1, 2, \dots, n - w$. Este tipo de abordagem é muito importante principalmente quando a série de dados observada possui, por exemplo, tendência ou sazonalidade o que não é o caso dos dados de Hanseníase e portanto, usamos então uma padronização interna tal que $e_t(s_l) = N_t(s_l) \times \lambda_{t-w}, t \geq w$. Ou seja, o número de casos esperados $e_t(s_l)$ é proporcional a população em risco e ajustado por um fator λ_{t-w} o qual representa uma taxa de referência global no mapa obtida com os dados acumulados até o tempo $t - w$. Para todo usamos

$$\lambda_t = \frac{\sum_{T=1}^t \sum_{s_l \in \mathcal{S}} Y_T(s_l)}{\sum_{T=1}^t \sum_{s_l \in \mathcal{S}} N_T(s_l)}.$$

No tempo t e município $l = 1, 2, \dots, 62$, uma distribuição a priori $g_{t,0}^{(l)}(\vartheta_t(s_l) | \mathcal{H}_{t-1}(l))$ e um mecanismo de evolução do tempo $t - 1$ para o tempo t faz-se necessário para o início do monitoramento via Fator de Bayes. A distribuição para $(\vartheta_t(s_l) | \mathcal{H}_{t-1}(l))$ na família conjugada (ver seção 4.6) é uma distribuição Gama. A suposição de uma distribuição Gama para os riscos relativos, tem sido considerada por vários autores como uma suposição adequada (Martuzzi e Hills, 1995), pois além de unimodal ela pode tomar uma grande variedade de formas regulares podendo, por exemplo, acomodar à assimetria da distribuição. Para o mecanismo de evolução temporal, assumamos mais especificamente que a relação entre $\vartheta_t(s_l)$ e $\vartheta_{t-1}(s_l)$ é da forma multiplicativa

$$\vartheta_t(s_l) = \frac{\zeta_t(s_l)}{m} \vartheta_{t-1}(s_l), \quad 0 < m \leq 1, \quad \zeta \sim \text{Beta}(m a_{t-1}(s_l), (1 - m) b_{t-1}(s_l)). \quad (4.41)$$

Assumindo que o processo $\zeta_t(s_l)$ é independente de $\vartheta_{t-1}(s_l)$ obtemos diretamente que $\mathbb{E}(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) = \vartheta_{t-1}(s_l)$ e ainda (west e Harrisson, 1997) $(\vartheta_{t-1}(s_l)|\mathcal{H}_{t-1}(l)) \sim \text{Gama}(a_{t-1}(s_l), b_{t-1}(s_l))$ implica que $(\vartheta_t(s_l)|\mathcal{H}_{t-1}(l)) \sim \text{Gama}(a_t(s_l), b_t(s_l))$ com

$$a_t(s_l) = ma_{t-1}(s_l) \quad \text{e} \quad b_t(s_l) = mb_{t-1}(s_l)$$

forneendo a estrutura conjugada necessária para o cômputo do Fator de Bayes a cada tempo t . Essencialmente, m controla a passagem de informação dos dados do tempo $t - 1$ para o tempo t . Grandes valores de m (próximo de 1) garantem uma maior passagem de informação entre os períodos de tempo tal que para uma abordagem adaptativa com janela temporal w pequena é importante que quase toda informação contida na janela seja avaliada de modo que usamos $m = 0.9$ e também, $a_0(s_l) = b_0(s_l) = 100$ implicando em uma distribuição *priori* inicial com média 1 pouca variabilidade.

4.8 Resultados da Aplicação

Ao aplicarmos o método proposto, Fator de Bayes Adaptativo (FBA) para a detecção e vigilância de clusters de casos de hanseníase em menores de 15 anos no estado do Amazonas no período 2008-2009, obtemos o gráfico de monitoramento na escala logarítmica apresentado na figura ?? com janela temporal $w = 3$ e probabilidade aceitável de alarmes falsos $\alpha = 0.05$. Neste gráfico, no tempo $t = 16$ o valor do logaritmo do FBA é igual a 14.220 e ultrapassa o limite de controle $\log(B_{16}) = 7.67$. Neste momento o método soa um alarme na localização espacial mostrado na figura 4.7. Este *cluster* persiste até o tempo $t=17$ com $\log(\text{FBA}) = 27.567$ e tempo de mudança estimado em 15, de modo que o cluster emergente estimado com os dados até o tempo $t=17$ é igual a $\hat{\xi} \times [15, 17]$ onde $\hat{\xi}$ é representado pelo conjunto de regiões marcadas na figura 4.7. Observe o formato geométrico do conjunto de localizações espaciais que indica claramente o caráter elíptico do algoritmo. Prosseguindo a análise no tempo observamos a indicação de um novo tempo de alarme $t = 18$. Neste caso a localização espacial do *cluster* é mostrada na figura 4.8. Observe novamente o caráter elíptico do conjunto de localizações espaciais que pertencem ao *cluster*. Existe neste *cluster* uma certa desconexidade isto deve-se ao fato que o semi-eixo menor desta elipse não tem o tamanho suficiente para incorporar o centróide da região vizinha mostrada no mapa e, isto cria uma certa desconexidade no *cluster*. Este fato cria um problema apenas de representação do *cluster* e não altera a análise dos dados. Estes *clusters* identificados no passado são chamados de *clusters* históricos com tempo final $t = 18$. Após este tempo estes *clusters* não mais existem e os dados sucessivos não trazem mais informações relevantes. Após continuarmos o monitoramento no tempo observamos que no momento atual $n = 24$ não existe indicação de qualquer mudança na distribuição do processo espaço-temporal.

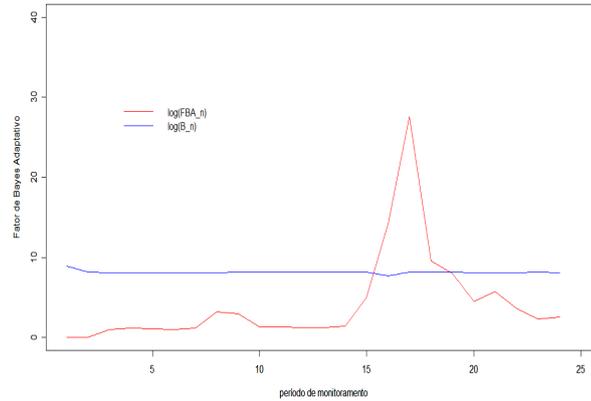


Figura 4.6: Monitoramento on-line usando Fator de Bayes Sequencial Adaptativo com janela $w = 3$, $\alpha = 0,05$

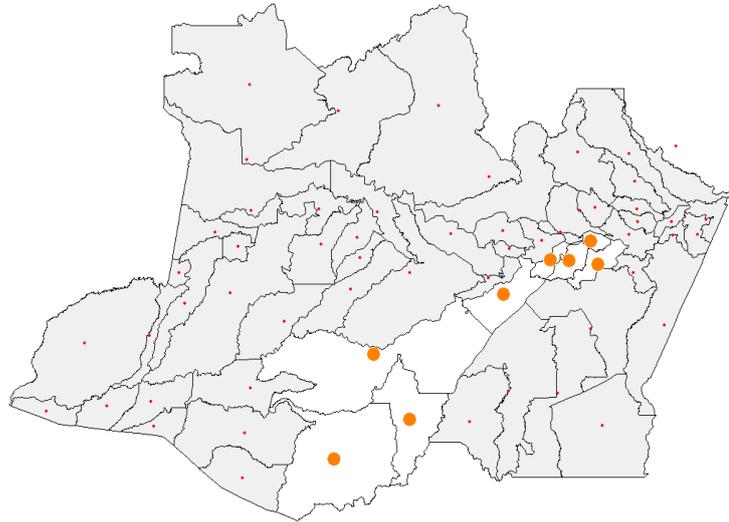


Figura 4.7: *cluster detectado usando o Fator de Bayes Sequencial Elíptico* $t \in [15, 17]$

4.9 Estudo com dados simulados

4.9.1 Dados Simulados

Para ilustrar a aplicação e avaliar a performance do método proposto para a detecção de *cluster* no espaço-tempo, nós realizamos um estudo com dados simulados com o objetivo de detectar mudanças na variação do risco de ocorrências de casos de uma doença sobre o espaço-tempo. O estudo é conduzido de forma similar ao descrito em 3.5 com poucas modificações. Consideramos como população em risco, $N_t(s_l)$, a população de homens residentes no ano de 1992 nas 32 ($L=32$) áreas do Novo México como descrito em Kulldorff(2001). Sobre esta população em risco nós simulamos um processo de Poisson

$$\mathbf{Y}_t = (Y_t(s_1), \dots, Y_t(s_L))'$$

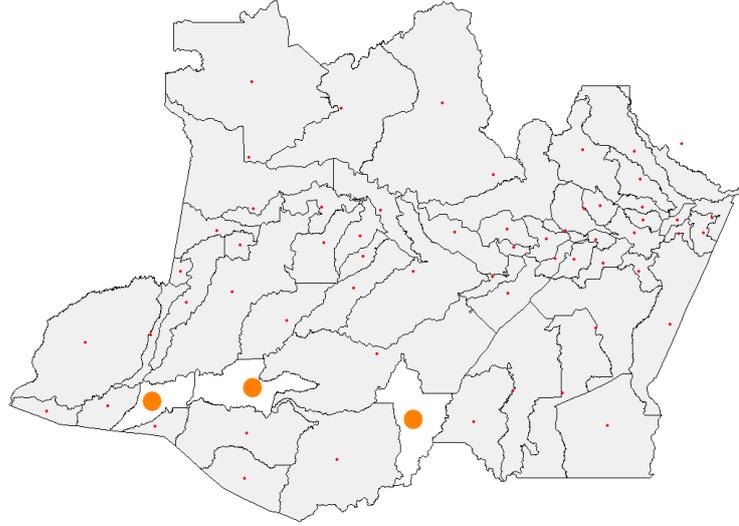


Figura 4.8: *cluster detectado usando o Fator de Bayes Sequencial Elíptico* $t \in [16, 18]$

condicional a um outro processo de risco relativo

$$\boldsymbol{\vartheta}_t = (\vartheta_t(s_1), \vartheta_t(s_2), \dots, \vartheta_t(s_L))'$$

tal que $\mathbb{E}(Y_t(s_l) | \vartheta_t(s_l)) = e_t(s_l) \times \vartheta_t(s_l)$ com $e_t(s_l)$ representando o número de casos esperados em s_l no tempo t , e dado por $e_t(s_l) = \lambda_0 N_t(s_l)$. λ_0 é uma taxa de referência global para a doença.

Os casos são simulados para um período de monitoramento $t = 1, 2, \dots, n = 20$ e 1000 réplicas do processo foram geradas. A parte espacial do cluster alocado no mapa é formado por 4 áreas, $\xi = \{Chaves, Eddy, Lea, Otero\}$ que podem ser identificadas na figura 5.2. Este é um cluster com população mediana em relação ao mapa. Quando simulamos um processo sem *cluster* assumimos que sob o modelo padrão $\{\boldsymbol{\vartheta}_t\}$ é realização de um processo Gama tal que,

$$\vartheta_t(s_l) \sim \text{Gama}(100, 100), t = 1, 2, \dots, 20 \quad \text{e} \quad l = 1, 2, \dots, 32.$$

Esta distribuição a *priori*, produz uma pequena variação no risco sobre todo o mapa tal que o valor esperado de $Y_t(s_l)$ esta próximo de $e_t(s_l)$. Para o modelo de *cluster* emergente definimos um risco alternativo por $\tilde{\vartheta}_t(s_l) = \vartheta_t(s_l) \delta^{I_{\{t \geq 11, s_l \in \xi\}}}$. para valores de $\delta = 1.5, 2, 3$. Note que neste caso o ponto de mudança é $k = 12$. Usamos, ainda, $\lambda_0 = 0.0001$ (taxa de um caso para cada 10000 habitantes); probabilidade de falso alarme $\alpha = 0.05, 0.01$; uma janela temporal de tamanho $w = 3$ e $\rho = 0.2, 0.4$. Isto gera um total de 11 cenários diferentes sobre o modelo de *cluster* e 2 cenários diferentes ($\rho = 0.2, 0.4$) sobre o modelo padrão. Obviamente este total de cenários é muito pequeno em relação a classe de todos possíveis. No entanto representam parâmetros(δ, α, λ) comumente utilizados em outros estudos de simulação de alguns trabalhos já existentes na literatura (Zhou e Lawson, 2008; Tango et al, 2010). Após gerarmos estes cenários aplicamos o Fator de Bayes usando o algoritmo descrito na seção 4.6 com mecanismo de evolução temporal dado em 4.41.

O algoritmo descrito em na seção 4.6 é usado da seguinte forma: usamos quatro valores (1,1.5,2,4) para o parâmetro de forma φ da elipse e então escolhemos os ângulos de modo que duas elipses de mesma forma, centro e tamanho, com ângulos consecutivos tenham intersecção de no mínimo 70% de suas áreas. Assim (ver tabela 4.2), quando $\varphi = 1.5$ usamos 2 ângulos de 90 graus, $\varphi = 2$ usamos 6 ângulos de 60 graus, $\varphi = 4$ usamos 12 ângulos de 15 graus. Para $\varphi = 1$ a representação geométrica do *cluster* é simplesmente um círculo, ou seja usamos apenas um ângulo. Como descrito na seção 4.5 usamos o fato que a excentricidade, "e" da elipse satisfaz a relação $e^2 = 1 - (a/b)^{-2} = 1 - \varphi^{-2}$ de modo que desta relação podemos obter o valor das quatro excentricidades usadas. Neste caso, para cada s_l e parâmetro de forma fixo, a cardinalidade de Ξ_l^0 no passo 5 do algoritmo dado em 4.6 é igual a

$$\begin{aligned} J_l^0 &= L \times \text{número de ângulos} \times \text{número de excentricidades} \\ &= L \times \text{número de ângulos} \times \# \text{ de parâmetros de forma usados.} \end{aligned}$$

Então, o número total de *clusterS* avaliados inicialmente é

$$32 \times 1 + 32 \times 2 + 32 \times 6 + 32 \times 12 = 672$$

isto significa que, neste caso com $L = 32$, em cada s_l podemos construir 672 conjuntos de localizações espaciais com formato elíptico. Logo o número total de subconjuntos elípticos é igual $32 \times 672 = 21504$. Usamos um valor de c dado no passo 5 do algoritmo proposto igual a $c = L/4$, teremos, então, que o número total de subconjunto avaliados no passo inicial é 5376. No entanto a partir da primeira atualização da classe Ξ^n o número de subconjuntos avaliados reduz a 32. E isso é o ganho ao se utilizar a abordagem adaptativa.

Para finalizar, em cada réplica do processo gerado, avaliamos as seguintes medidas de performance: Sensitividade Condicional de um alarme **SSC**; Valor Predito Positivo Condicional de um alarme **VPP**; atraso na detecção **ADD**. Todas estas medidas de avaliação foram propostas por Lima e Duczmal (2011) e estão descritas na seção 3.5.1.

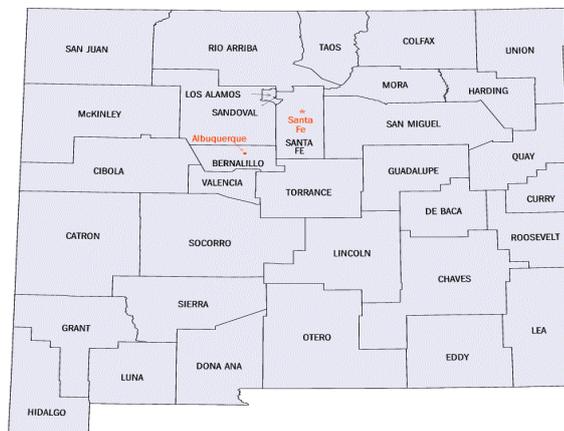


Figura 4.9: Mapa do Novo México. Fonte: US Census.

4.9.2 Análise dos resultados obtidos com dados simulados

Nós apresentamos nesta seção o resultado da aplicação do Fator de Bayes para os cenários simulados descrito na seção 4.9. Na figura 4.10 apresentamos os resultados para a distribuição do máximo do Fator de Bayes adaptativo, e da cardinalidade J^n de Ξ^n para valores de $\rho = 0.2, 0.4$. Um valor de $\rho = 0.2(0.4)$ significa um desvio padrão preditivo do modelo alternativo 2,23 (1.58) vezes maior em relação ao modelo padrão. Observando a figura 4.10 nota-se claramente que praticamente não há mudança na distribuição do Máximo do Fator de Bayes quando modificamos o valor de ρ . O percentil de ordem 85, em ambas as distribuições do Máximo é, aproximadamente, representado pelo valor 3 que nós usamos como um valor razoável para C^* , dado na seção 4.3.3.

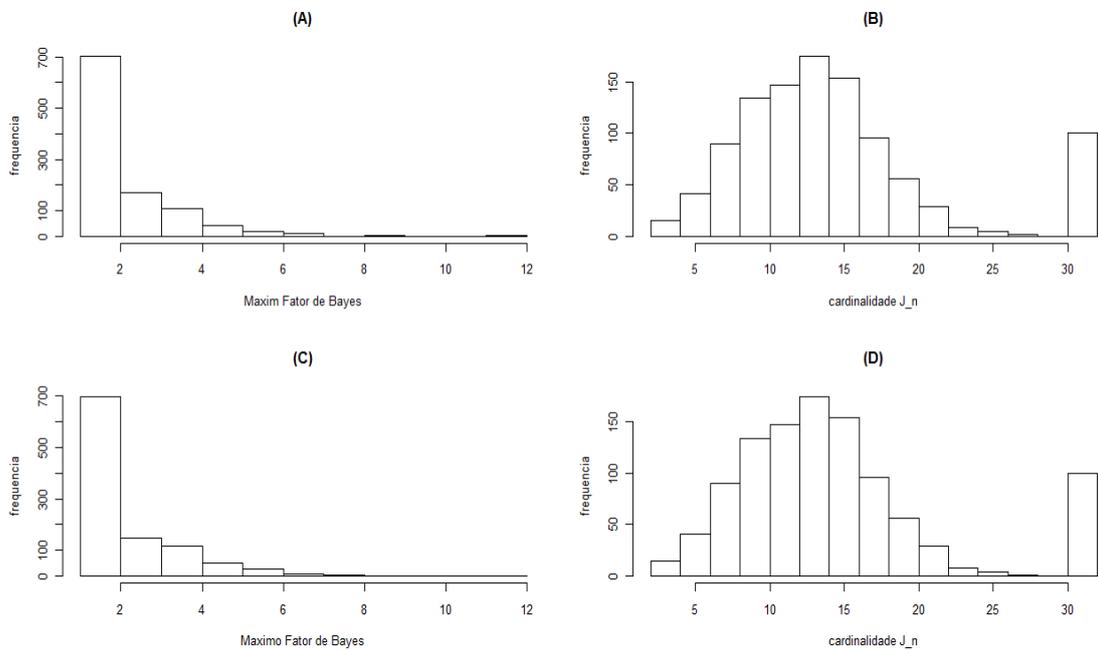


Figura 4.10: Máximo do Fator de Bayes Adaptativo em (A), Cardinalidade J^n de Ξ^n em (B) para 1000 simulações do processo sobre o modelo padrão com $\rho = 0.2$; Máximo do Fator de Bayes Adaptativo em (C), Cardinalidade J^n de Ξ^n em (D) para 1000 simulações do processo sobre o modelo padrão com $\rho = 0.4$

Os resultados para as medidas de performance estão descritos na tabela 4.3. Nota-se, na tabela que o atraso (**ADD**) na detecção decresce a Sensitividade (**SSC**) na localização cresce e o Valor predito positivo (**PPV**) diminui quando aumentamos o valor de δ . A **SSC** é sempre maior que o **VPP** implicando que o *cluster* localizado tende a ser maior que o verdadeiro *cluster*, porém com uma boa precisão. Por exemplo, quando $\delta = 2$, $\alpha = 0.05$ e $\rho = 0.2$ o atraso estimado é igual a 1.429, que pode ser considerado muito pequeno, pois o cluster foi detectado logo no seu início. A precisão na localização medida pela **SSC** é 0.746 enquanto que o **VPP**=0.724.

De uma forma geral, o método é bastante rápido (possui pequeno atraso) e possui uma boa precisão. Uma outra observação interessante é que, quando usamos um valor de $\rho = 0.4$ os resultados quase não se alteram. Uma visualização da relação atraso versus precisão é mostrada

através de histogramas bivariados nas figuras 4.11 e 4.12. Nestas figuras pode-se observar que o atraso é baixo a Sensitividade e o Valor predito positivo são relativamente bons. Uma amostra das densidades preditivas na região do *cluster* alocado artificialmente no mapa é mostrado na figura 4.13. Nesta figura é possível observar que a preditiva alternativa é uma versão suavizada da preditiva padrão.

Tabela 4.3: Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) com diferentes valores para δ , diferentes probabilidades de alarmes falsos (α) usando o Fator de Bayes Adaptativo Eliptico com janela $w=3$.

		Medidas de Performance			
		δ	SSC	VPP	A\hat{D}D
	$\alpha = 0.05$	1.50	0.608	0.781	1.989
		2.00	0.746	0.724	1.429
		3.0	0.838	0.681	1.195
$\rho = 0.2$					
	$\alpha = 0.01$	1.50	0.598	0.778	2.159
		2.00	0.759	0.733	1.508
		3.0	0.850	0.689	1.206
	$\alpha = 0.05$	1.50	0.605	0.751	1.889
		2.00	0.740	0.723	1.410
		3.0	0.800	0.678	1.190
$\rho = 0.4$					
	$\alpha = 0.01$	1.50	0.590	0.770	2.140
		2.00	0.750	0.732	1.500
		3.0	0.850	0.680	1.202

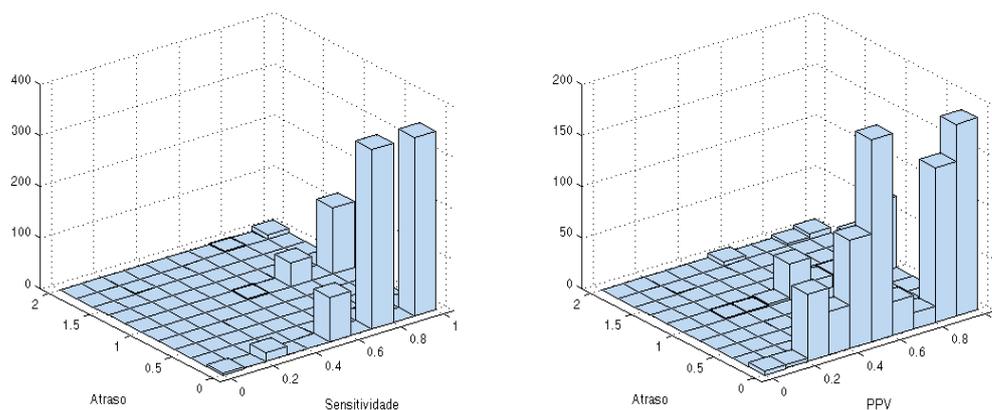


Figura 4.11: Histograma Bivariado para o Fator de Bayes Adaptativo com $\rho = 0.2$: Sensitividade versus atraso e Valor Predito Positivo versus atraso

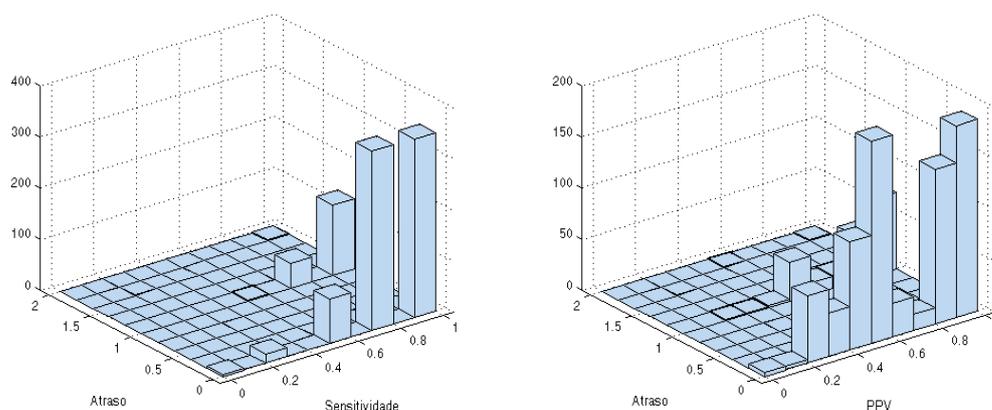


Figura 4.12: Histograma Bivariado para o Fator de Bayes Adaptativo com $\rho = 0.4$: Sensitividade versus atraso e Valor Predito Positivo versus atraso

4.10 Discussão

A proposta deste capítulo é o desenvolvimento de um novo método para detecção e monitoramento de *clusters* emergentes. Nós desenvolvemos e descrevemos uma abordagem Bayesiana, baseada em um Fator Bayes Adaptativo com janela temporal limitada, e usamos uma estrutura de *cluster* onde o espaço é representado por subconjuntos de localizações espaciais com formato elíptico. Três importantes características do método proposto são: 1) A utilização de uma janela temporal w que permite a estimação, de qualquer parâmetro de interesse, usando dados históricos pertencentes ao conjunto de tempos $t = 1, 2, \dots, t - w$; 2) Uma abordagem adaptativa no espaço dos *cluster* que permite tornar o algoritmo de varredura no espaço-tempo "relativamente" rápido. 3) ajusta automaticamente o limite de controle para a função de alarme, mantendo constante a probabilidade de um falso alarme e, ainda, tornando o atraso na detecção muito pequeno. A forma explícita da distribuição preditiva também é muito importante neste caso, pois permite analiticamente a obtenção de todas as quantidades de interesse para cada tempo de monitoramento e qualquer área do mapa em estudo.

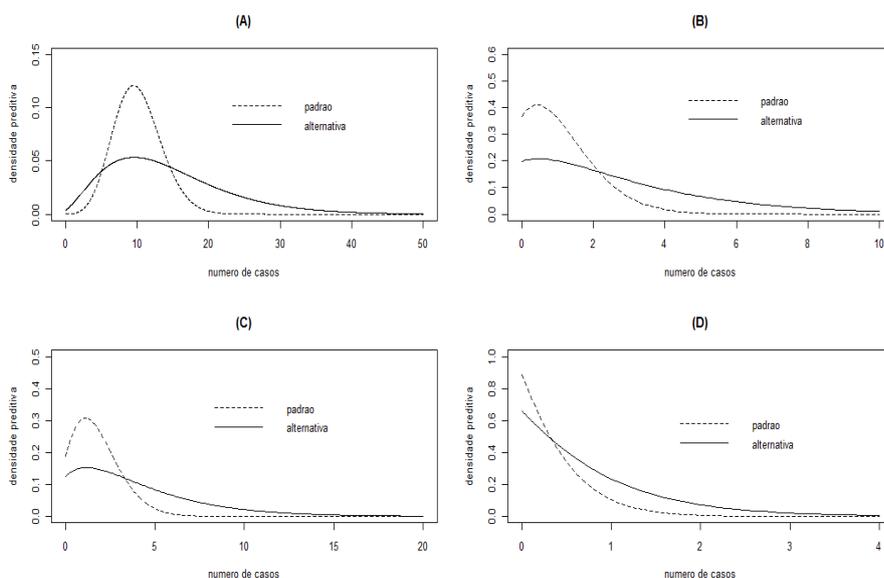


Figura 4.13: Distribuições Preditivas e Padrão após o ponto de mudança nas áreas que compõe o *cluster*: Chaves em (A), Eddy em (B), Lea em (C) e Otero em (D).

Este método desenvolvido, pode ser entendido como uma versão espaço-temporal do Fator de Bayes Sequencial descrito em West(1986) e por isso também pode ser estendido para os modelos analisados em West e Harrison (1986). Além da abordagem adaptativa, outras diferenças desse método de monitoramento espaço-temporal em relação aos existentes na literatura são: uma possível incorporação na análise de alguns componentes de uma série temporal como tendência e sazonalidade; controle adequado da probabilidade de um falso alarme; não requer o uso de testes baseados em simulação de Monte Carlo para validar a significância estatística dos resultados. Uma desvantagem é utilização de distribuições *a priori* somente na família conjugada de distribuições.

5 *Detecção de clusters emergentes via análise a posteriori adaptativa*

5.1 Introdução

Os algoritmos desenvolvidos anteriormente para a detecção de *clusters* emergentes não levam em consideração a incerteza sobre o formato, tamanho e tempo de início do mesmo. De uma maneira implícita, a busca realizada por estes algoritmos supõe que todos os candidatos a *cluster* emergente são equiprováveis. Por isso, neste capítulo abordamos o problema da detecção de *clusters* emergentes usando um processo a *posteriori* adaptativo. Este processo é construído a partir da distribuição conjunta a *posteriori* de que tenha ocorrido uma mudança no conjunto de localizações ξ no tempo $\nu = k$. Ou seja, avaliamos diretamente o evento $\{\xi = \xi_j, \nu = k\}$. Nós apresentamos uma estrutura geral de distribuição a *priori* para parte espacial (região da mudança) e temporal (ponto de mudança) do *cluster* emergente e mostramos também, que esta abordagem minimiza a perda esperada a *posteriori* baseada em uma combinação linear da probabilidade de um falso alarme com a probabilidade de uma falsa localização.

5.2 Construção hierárquica do modelo

Considere, novamente, \mathcal{S} como a região em estudo. Seja $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots, L\}$ o processo sob vigilância. A cada tempo discreto $t \geq 1$ observamos um vetor L -variado $\mathbf{X}_t = (X_t(s_1), X_t(s_2), \dots, X_t(s_L))'$, em que $X_t(s_l)$ representa algum valor observado em um ponto $s_l \in \mathcal{S}$. Sejam $\xi = \{s_l \in \mathcal{S} : s_l \in \xi\}$ um conjunto conexo de localizações espaciais em \mathcal{S} , $\mathcal{H}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ o conjunto de dados acumulados até o instante n e $f(\mathbf{X}_t | \boldsymbol{\theta}_t)$ uma função de densidade conhecida para \mathbf{X}_t , dado um valor selecionado para o parâmetro $\boldsymbol{\theta}_t$, para cada t , tal que $\boldsymbol{\theta}_t = (\theta_t(s_1), \theta_t(s_2), \dots, \theta_t(s_L))'$ pertence a algum conjunto aberto Θ do espaço finito dimensional. O objetivo da vigilância é verificar se existe um *cluster* emergente no mapa. O interesse aqui, é detectar um *cluster* que começou em um tempo desconhecido $\nu \leq n$ e que está presente até o estágio atual n . O *cluster* no espaço-tempo é expresso em função de uma mudança no padrão espaço-temporal de eventos em ξ que emerge depois de um tempo desconhecido ν .

Para a detecção do *cluster* e estimação do par (ν, ξ) considere uma abordagem bayesiana

hierárquica construída da seguinte forma:

1. Assuma que $(\mathbf{X}_t|\boldsymbol{\theta}_t)$ tem densidade

$$f(\mathbf{X}_t|\boldsymbol{\theta}_t) = \prod_{s_l \in \mathcal{S}} f(X_t(s_l)|\theta_t(s_l)).$$

ou seja, dado o processo $\{\boldsymbol{\theta}_t\}$ os componentes de \mathbf{X}_t são condicionalmente independentes.

2. Suponha que $(\boldsymbol{\theta}_t|\nu, \xi)$ tem função de distribuição $G(\boldsymbol{\theta}_t|\nu, \xi) = [G_0(\boldsymbol{\theta}_t)]^{m_{t,\xi} I_{\{t \geq \nu, s_l \in \xi\}}}$, para alguma função de distribuição $G_0(\boldsymbol{\theta}_t)$ e $m_{t,\xi} > 1$ inteiro, conhecidos.
3. Considere que ν e ξ são a priori independentes tal que a densidade a priori conjunta para o par (ν, ξ) é dada por $\pi(\nu = k, \xi = \xi_j) = \pi_\nu(k)\pi_\xi(j)$, $k = 1, 2, \dots$ e $j = 1, 2, \dots, J$.

Neste trabalho, $G_0(\boldsymbol{\theta}_t)$ representa uma distribuição de referência no seguinte sentido:

- Se não há mudanças no padrão espaço-temporal do processo $\{\mathbf{X}_t\}$, então o processo $\{\boldsymbol{\theta}_t\}$ tem função de distribuição $G_0(\boldsymbol{\theta}_t)$.

Para a interpretação de $m_{t,\xi}$, assumamos que

$$G_0(\boldsymbol{\theta}_t) = \prod_{s_l \in \mathcal{S}} G_0(\theta_t(s_l)).$$

Então,

$$G(\boldsymbol{\theta}_t|\nu = k, \xi = \xi_j) = \begin{cases} \prod_{s_l \in \mathcal{S}} G_0(\theta_t(s_l)) & \text{se } t \leq k - 1 \\ \prod_{s_l \notin \xi_j} G_0(\theta_t(s_l)) \prod_{s_l \in \xi_j} [G_0(\theta_t(s_l))]^{m_{t,j}} & \text{se } t \geq k, s_l \in \xi_j \end{cases}$$

ou seja, depois da mudança no tempo $\nu = k$, os componentes de $\boldsymbol{\theta}_t$ tais que $s_l \notin \xi_j$ possuem distribuição de referência G_0 . Enquanto que os componentes tais que $s_l \in \xi_j$ possuem a distribuição do máximo de $m_{t,j}$ -variáveis aleatórias com distribuição G_0 , isto implica que para todo $s_l \in \xi_j$, $\theta_t(s_l)$ possui distribuição estocasticamente maior que o $\theta_t(s_l)$ tal qual $s_l \notin \xi_j$. Para termos uma idéia de quanto é esta grandeza estocástica, notamos que se Z_1 é uma variável aleatória com função de distribuição $F(z_1)$ e Z_2 tem função de distribuição $F^m(z_2)$, então

$$\mathbb{P}(Z_1 < Z_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} d[F(z_1)]d[F^m(z_2)]. \quad (5.1)$$

Fazendo a mudança de variável $Y_i = F(z_i) \sim U(0, 1)$, $i = 1, 2$, obtemos,

$$\mathbb{P}(Z_1 < Z_2) = \int_0^1 \left(\int_0^{y_2} d[y_1] \right) d[y_2^m] = \int_0^1 y_2 d[y_2^m] = \frac{m}{m+1}$$

de modo que podemos escolher m de acordo com a grandeza estocástica desejada.

5.3 Distribuições a posteriori

Para detecção do par (ν, ξ) consider uma análise a posteriori conjunta. Primeiramente pela independência condicional, denote para $t \leq k - 1$,

$$\begin{aligned}
P_0(\mathbf{X}_t) &= P_0(\mathbf{X}_t | \nu = k, \xi = \xi_j) = \int f(\mathbf{X}_t | \boldsymbol{\theta}_t) dG(\boldsymbol{\theta}_t | \nu = k, \xi = \xi_j), \quad t \leq k - 1 \\
&= \int f(\mathbf{X}_t | \boldsymbol{\theta}_t) dG_0(\boldsymbol{\theta}_t) \\
&= \int \prod_{s_l \in \mathcal{S}} f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)) \\
&= \prod_{s_l \in \mathcal{S}} \int f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)). \tag{5.2}
\end{aligned}$$

e usando o fato que

$$\prod_{s_l \in \mathcal{S}} f(X_t(s_l) | \theta_t(s_l)) = \prod_{s_l \notin \xi_j} f(X_t(s_l) | \theta_t(s_l)) \times \prod_{s_l \in \xi_j} f(X_t(s_l) | \theta_t(s_l)).$$

denote, quando $t \geq k$ e $\xi = \xi_j$

$$\begin{aligned}
P_j(\mathbf{X}_t) &= P_j(\mathbf{X}_t | \nu = k, \xi = \xi_j) = \int f(\mathbf{X}_t | \boldsymbol{\theta}_t) dG(\boldsymbol{\theta}_t | \nu = k, \xi = \xi_j), \quad t \geq k, s_l \in \xi_j \\
&= \int \prod_{s_l \notin \xi_j} f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)) \prod_{s_l \in \xi_j} f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]^{m_{t,j}} \\
&= \prod_{s_l \notin \xi_j} \int f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)) \\
&\times \prod_{s_l \in \xi_j} \int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]^{m_{t,j}}. \tag{5.3}
\end{aligned}$$

Após observar o processo n -vezes e combinando (5.2) com (5.3) temos que,

$$f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n | \nu = k, \xi = \xi_j) = \prod_{t=1}^{(k-1) \wedge n} P_0(\mathbf{X}_t) \prod_{t=k \vee 1}^n P_j(\mathbf{X}_t) \tag{5.4}$$

onde definimos, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$ e $\prod_{t>n}^n P_j(\mathbf{X}_t) = 1$. Note que se a mudança ocorreu no tempo $\nu = k \leq n$, temos que $\min\{(k-1), n\} = k-1$ e o $\max\{k, 1\} = k$ de modo que, (5.4) é igual a,

$$f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n | \nu = k, \xi = \xi_j) = \prod_{t=1}^{(k-1)} P_0(\mathbf{X}_t) \prod_{t=k}^n P_j(\mathbf{X}_t) \tag{5.5}$$

caso contrário, se a mudança ainda não ocorreu tal que $k > n$ temos,

$$f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n | \nu = k, \xi = \xi_j) = \prod_{t=1}^n P_0(\mathbf{X}_t) \tag{5.6}$$

Usando a distribuição a *priori* para o par (ν, ξ) dada na seção 5.2, obtenha que a distribuição a *posteriori* conjunta de (ν, ξ) é representada por

$$\pi(\nu = k, \xi = \xi_j | \mathcal{H}_n) = \begin{cases} \frac{\pi_\nu(k)\pi_\xi(j)\prod_{t=1}^n P_0(\mathbf{X}_t)}{f(\mathcal{H}_n)} & \text{se } k > n \\ \frac{\pi_\nu(k)\pi_\xi(j)\prod_{t=1}^{(k-1)} P_0(\mathbf{X}_t)\prod_{t=k}^n P_j(\mathbf{X}_t)}{f(\mathcal{H}_n)} & \text{se } k \leq n, \xi = \xi_j \end{cases} \quad (5.7)$$

onde

$$f(\mathcal{H}_n) = \sum_{i=1}^J \sum_{k=1}^n \pi_\nu(k)\pi_\xi(j) \prod_{t=1}^{(k-1)} P_0(\mathbf{X}_t) \prod_{t=k}^n P_j(\mathbf{X}_t) + \sum_{k=n+1}^{\infty} \pi_\nu(k) \prod_{t=1}^n P_0(\mathbf{X}_t).$$

Agora, multiplique o denominador e o numerador de (5.7) por

$$\frac{1}{\prod_{t=1}^n P_0(\mathbf{X}_t)}$$

e obtenha que,

$$\pi(\nu = k, \xi = \xi_j | \mathcal{H}_n) = \begin{cases} \frac{\pi_\nu(k)\pi_\xi(j)}{\sum_{i=1}^J \sum_{k=1}^n \pi_\nu(k)\pi_\xi(i) \prod_{t=k}^n FB_t(i) + \sum_{k=n+1}^{\infty} \pi_\nu(k)} & \text{se } k > n \\ \frac{\pi_\nu(k)\pi_\xi(j) \prod_{t=k}^n FB_t(j)}{\sum_{i=1}^J \sum_{k=1}^n \pi_\nu(k)\pi_\xi(i) \prod_{t=k}^n FB_t(i) + \sum_{k=n+1}^{\infty} \pi_\nu(k)} & \text{se } k \leq n, \xi = \xi_j \end{cases} \quad (5.8)$$

onde

$$FB_t(j) = \frac{P_j(\mathbf{X}_t)}{P_0(\mathbf{X}_t)}, t \geq k. \quad (5.9)$$

Notando em (5.2) que

$$\begin{aligned} \prod_{s_l \in \mathcal{S}} \int f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)) &= \prod_{s_l \notin \xi_j} \int f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)) \\ &\times \prod_{s_l \in \xi_j} \int f(X_t(s_l) | \theta_t(s_l)) dG_0(\theta_t(s_l)). \end{aligned}$$

Obtenha que

$$FB_t(j) = \frac{P_j(\mathbf{X}_t)}{P_0(\mathbf{X}_t)} = \prod_{s_l \in \xi_j} \frac{\int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]^{m_{t,j}}}{\int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]} , t \geq k \quad (5.10)$$

que depende somente do sub-processo $\mathbf{X}_t(j) = \{X_t(s_l) : t \geq k, s_l \in \xi_j\}$. Em uma estrutura geral computamos $FB_t(j)$ da seguinte forma:

1. Para computarmos a integral $\int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]^{m_{t,j}}$, assumimos que $G_0(\theta_t(s_l))$ é contínua e fazemos a mudança de variável

$$Y = G_0(\theta_t(s_l)) \sim U(0, 1) \Rightarrow \theta_t(s_l) = G_0^{-1}(y)$$

de modo que

$$\begin{aligned} \int f(X_t(s_l)|\theta_t(s_l))d[G_0(\theta_t(s_l))]^{m_{t,j}} &= \int_0^1 f(X_t(s_l)|G_0^{-1}(y))d[y^{m_{t,j}}] \\ &= \int_0^1 f(X_t(s_l)|G_0^{-1}(y))m_{t,j}y^{m_{t,j}-1}dy \\ &= \mathbb{E}_Z [f(X_t(s_l)|G_0^{-1}(z))], \end{aligned} \quad (5.11)$$

onde, $Z \sim \text{Beta}(m_{t,j}, 1)$, $s_l \in \xi_j$.

2. Se a expressão de $f(X_t(s_l)|G_0^{-1}(z))$ é analiticamente complicada, duas abordagens podem ser usadas para determinar $\mathbb{E}_Z [f(X_t(s_l)|G_0^{-1}(z))]$.

(a) Gere Q valores z_1, z_2, \dots, z_Q de uma distribuição $\text{Beta}(m_{t,j}, 1)$ e use a aproximação

$$\mathbb{E}_Z [f(X_t(s_l)|G_0^{-1}(z))] \approx \frac{1}{Q} \sum_{q=1}^Q f(X_t(s_l)|G_0^{-1}(z_q)). \quad (5.12)$$

Quanto maior o valor de Q , mais precisa será esta abordagem. Mesmo se L for razoavelmente grande esta abordagem não torna o algoritmo de busca por *cluster* emergente muito demorado, pois exige apenas a geração de uma matriz $L \times Q$ de valores.

(b) Aproximar $\mathbb{E}_Z [f(X_t(s_l)|G_0^{-1}(z))]$ usando o método Delta, de modo que

$$\begin{aligned} \mathbb{E}_Z [f(X_t(s_l)|G_0^{-1}(z))] &\approx [f(X_t(s_l)|G_0^{-1}(z))]_{z=\mathbb{E}(Z)} \\ &= [f(X_t(s_l)|G_0^{-1}(z))]_{z=m_{t,j}/(m_{t,j}+1)}. \end{aligned} \quad (5.13)$$

3. Obviamente a quantidade $\int f(X_t(s_l)|\theta_t(s_l))d[G_0(\theta_t(s_l))]$ é obtida com $m_{t,j} = 1$.

Para detectarmos possíveis mudanças considere o processo a *posteriori* $\mathbf{\Pi}_n = \{\Pi_n(0), \Pi_n(j)\}_{n \geq 1}$ com $j = 1, 2, \dots, J$. Onde,

$$\Pi_n(0) = \mathbb{P}(\nu > n | \mathcal{H}_n) \quad \text{e} \quad \Pi_n(j) = \mathbb{P}(\nu \leq n, \xi = \xi_j | \mathcal{H}_n).$$

Ou seja, depois de observar o processo n vezes, $\Pi_n(0)$ é a probabilidade de que a mudança não tenha ocorrido antes ou até o estágio atual n . Enquanto que $\Pi_n(j)$ é a probabilidade a *posteriori* conjunta de que tenha ocorrido uma mudança depois de ν e que está presente até o tempo atual no conjunto $\xi = \xi_j$ que representa a região da mudança. Observe que

$$\Pi_n(j) = \sum_{k=1}^n \pi(\nu = k, \xi = \xi_j | \mathcal{H}_n) \quad \text{e} \quad \Pi_n(0) = \sum_{j=1}^J \sum_{k=n+1}^{\infty} \pi(\nu = k, \xi = \xi_j | \mathcal{H}_n). \quad (5.14)$$

Se a distribuição a *priori* para ν é truncada em algum valor $k_0 = k_0(n, w) = n - w + 1$, usamos para detecção de mudanças o processo a *posteriori* com janela temporal limitada w , denotado por

$$\mathbf{\Pi}_{n,w} = \{\Pi_n(0), \Pi_{n,w}(j)\}_{n \geq k_0(n,w)} \quad \text{com } j = 1, 2, \dots, J.$$

Onde

$$\Pi_{n,w}(j) = \mathbb{P}(n - w + 1 \leq \nu \leq n, \xi = \xi_j | \mathcal{H}_n) = \sum_{k=n-w+1}^n \pi(\nu = k, \xi = \xi_j | \mathcal{H}_n).$$

O valor $k_0(n, w) = n - w + 1$ reflete o momento a partir do qual o pesquisador teria interesse em detectar uma mudança no padrão do processo no espaço-tempo. Notamos que $k_0(n, n) = 1$ e $k_0(n, 1) = n$. Por exemplo, se o interesse é detectar um *cluster* que tenha emergido a pelo menos 7 períodos de tempo (7 dias, meses, etc,...), então $k_0(n, 6) = n - 7$ e $w=6$. Claramente, quando $w = n$ temos o processo a *posteriori* dado em 5.14.

5.4 Tempo de alarme baseado no processo a posteriori

Um teste sequencial usado para a detecção de *cluster* emergente (*cluster* no espaço-tempo) pode ser representado por um par $\delta = (\tau, d)$ onde τ é um tempo de alarme (ou tempo de parada) e $d = d(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\tau)$ é uma função de decisão assumindo valores no conjunto $\mathcal{J} = \{1, 2, \dots, J\}$, tal que $d = j$ implica tomar uma decisão em favor de ξ_j de modo que

$$\{\tau < \infty, d = j\} = \{\delta \text{ existe um cluster emergente no conjunto } \xi_j\}.$$

Considere um algoritmo para a detecção e localização de um *cluster* no espaço-tempo para computar o par $\{\tau, d\}$ baseado em \mathcal{H}_n e que tenta minimizar os seguintes erros:

1. **falso alarme:** Se o tempo de alarme é $\tau < \nu$. Então teremos uma decisão final incorreta com $d = j$. Denote por a_{0d} a perda incorrida no evento $\{\tau < \nu, d = j\}$. Então a perda esperada por um falso alarme é $\mathbb{E}[a_{0d} I_{\{\tau < \nu\}}]$.
2. **falsa localização:** O *cluster* é detectado e localizado depois de ν de modo que o tempo de alarme $\tau \geq \nu$ é verdadeiro e $d = j$ porém, $\xi = \xi_i$ com $\xi_j \cap \xi_i = \emptyset$. Denote por $a_{\xi d}$ a perda incorrida no evento $\{\tau \leq \nu, d = j, \xi = i\}$. Então a perda esperada por uma falsa localização é $\mathbb{E}[a_{\xi d} I_{\{\tau \leq \nu\}}]$.

Após observar o processo n -vezes, $n \geq 1$, seja τ o tempo de alarme (ou tempo de parada) e $\mathbb{R}(\delta | \mathcal{H}_\tau)$ a função risco de Bayes a *posteriori* associada a τ . Então

$$\mathbb{R}(\delta | \mathcal{H}_\tau) = \mathbb{E}[a_{0d} I_{\{\tau < \nu\}} | \mathcal{H}_\tau] + \mathbb{E}[a_{\xi d} I_{\{\tau \geq \nu\}} | \mathcal{H}_\tau]. \quad (5.15)$$

onde o valor esperado é realizado sobre a distribuição a *posteriori* para (ν, ξ) . Agora, considere

uma classe de estratégias de detecção sequencial dada por:

$$\Delta = \{\delta = (\tau, d) : \tau \text{ é mesurável a } \sigma(\mathbf{X}_1, \dots, \mathbf{X}_\tau) \text{ e } d = d(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\tau) \in \mathcal{J}\}.$$

onde σ denota a sigma-álgebra. Então, para todo $\delta \in \Delta$, temos que

$$\mathbb{R}(\delta|\mathcal{H}_\tau) = \sum_{j=1}^J I_{\{\tau < \infty, d=j\}} \sum_{i=0}^J a_{ij} \Pi_\tau(i). \quad (5.16)$$

Equanto a formulação do risco de Bayes dada em (5.15) é expressa em termos dos valores para as variáveis aleatórias não observáveis (ν, ξ) . A expressão (5.16) da uma versão para o risco de Bayes baseado na distribuição a *posteriori* conjunta de (ν, ξ) . Para provar (5.16), usamos $I_{\{A \cap B\}} = I_{\{A\}} I_{\{B\}}$ e observamos que o evento $\{\tau = n, d = j\}$ é mensurável a $\sigma(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Agora,

$$\mathbb{E}[a_{0d} I_{\{\tau < \nu\}} | \mathcal{H}_\tau] = \sum_{j=1}^J \mathbb{E}[a_{0d} I_{\{\tau < \nu, d=j\}} | \mathcal{H}_\tau] = \sum_{j=1}^J a_{0j} \mathbb{E}[I_{\{\tau < \nu, d=j\}} | \mathcal{H}_\tau]$$

mas

$$\begin{aligned} \mathbb{E}[I_{\{\tau < \nu, d=j\}} | \mathcal{H}_\tau] &= \sum_{n=1}^{\infty} \mathbb{E}[I_{\{\tau=n, d=j, \nu > n\}} | \mathcal{H}_n] = \sum_{n=1}^{\infty} \mathbb{E}[I_{\{\tau=n, d=j\}} I_{\{\nu > n\}} | \mathcal{H}_n] \\ &= \sum_{n=1}^{\infty} I_{\{\tau=n, d=j\}} \mathbb{E}[I_{\{\nu > n\}} | \mathcal{H}_n] = \sum_{n=1}^{\infty} I_{\{\tau=n, d=j\}} \Pi_n(0) \\ &= \lim_{M \rightarrow \infty} \sum_{n=1}^M I_{\{\tau=n, d=j\}} \Pi_n(0) = \lim_{M \rightarrow \infty} I_{\{\tau \leq M, d=j\}} \Pi_\tau(0) \\ &= I_{\{\tau < \infty, d=j\}} \Pi_\tau(0). \end{aligned} \quad (5.17)$$

e portanto,

$$\mathbb{E}[a_{0d} I_{\{\tau < \nu\}} | \mathcal{H}_\tau] = \sum_{j=1}^J a_{0j} I_{\{\tau < \infty, d=j\}} \Pi_\tau(0) \quad (5.18)$$

Onde usamos também, $\lim_{M \rightarrow \infty} \{\tau \leq M\} = \bigcup_{M=1}^{\infty} \{\tau \leq M\} = \{\tau < \infty\}$ (Ross, 1983). Temos ainda, que

$$\mathbb{E}[a_{\xi d} I_{\{\tau \geq \nu\}} | \mathcal{H}_\tau] = \sum_{j=1}^J \sum_{i=1}^J \mathbb{E}[a_{\xi d} I_{\{\tau \geq \nu, d=j, \xi=i\}} | \mathcal{H}_\tau] = \sum_{j=1}^J \sum_{i=1}^J a_{ij} \mathbb{E}[I_{\{\tau \geq \nu, d=j, \xi=i\}} | \mathcal{H}_\tau]$$

mas,

$$\begin{aligned}
\mathbb{E}[I_{\{\tau \geq \nu, d=j, \xi=i\}} | \mathcal{H}_\tau] &= \sum_{n=1}^{\infty} \mathbb{E}[I_{\{\tau=n, d=j, \nu \leq n, \xi=i\}} | \mathcal{H}_n] = \sum_{n=1}^{\infty} \mathbb{E}[I_{\{\tau=n, d=j\}} I_{\{\nu \leq n, \xi=i\}} | \mathcal{H}_n] \\
&= \sum_{n=1}^{\infty} I_{\{\tau=n, d=j\}} \mathbb{E}[I_{\{\nu \leq n, \xi=i\}} | \mathcal{H}_n] = \sum_{n=1}^{\infty} I_{\{\tau=n, d=j\}} \Pi_n(i) \\
&= \lim_{M \rightarrow \infty} \sum_{n=1}^M I_{\{\tau=n, d=j\}} \Pi_\tau(i) = \lim_{M \rightarrow \infty} I_{\{\tau \leq M, d=j\}} \Pi_\tau(i) \\
&= I_{\{\tau < \infty, d=j\}} \Pi_\tau(i).
\end{aligned} \tag{5.19}$$

Logo,

$$\mathbb{E}[a_{\xi d} I_{\{\tau \geq \nu\}} | \mathcal{H}_\tau] = \sum_{j=1}^J \sum_{i=1}^J a_{ij} I_{\{\tau < \infty, d=j\}} \Pi_\tau(i). \tag{5.20}$$

Substituindo (5.18) e (5.20) em (5.15) obtenha que

$$\begin{aligned}
\mathbb{R}(\delta | \mathcal{H}_\tau) &= \sum_{j=1}^J a_{0j} I_{\{\tau < \infty, d=j\}} \Pi_\tau(0) + \sum_{j=1}^J \sum_{i=1}^J a_{ij} I_{\{\tau < \infty, d=j\}} \Pi_\tau(i) \\
&= \sum_{j=1}^J I_{\{\tau < \infty, d=j\}} \left[a_{0j} \Pi_\tau(0) + \sum_{i=1}^J a_{ij} \Pi_\tau(i) \right] = \sum_{j=1}^J I_{\{\tau < \infty, d=j\}} \sum_{i=0}^J a_{ij} \Pi_\tau(i).
\end{aligned} \tag{5.21}$$

Para construirmos o tempo de alarme, definimos as funções,

$$u(y) = \min_{j \in \mathcal{J}} [u_j(y)] \quad \text{onde} \quad u_j(y) = \sum_{i=0}^J y(i) a_{ij}, \quad j \in \mathcal{J}.$$

Então, para todo $\delta = (\tau, d) \in \Delta$,

$$\begin{aligned}
\mathbb{R}(\delta | \mathcal{H}_\tau) &= \mathbb{R}((\tau, d) | \mathcal{H}_\tau) = \sum_{j=1}^J I_{\{\tau < \infty, d=j\}} \sum_{i=0}^J a_{ij} \Pi_\tau(i) \\
&= I_{\{\tau < \infty\}} \sum_{j=1}^J I_{\{d=j\}} u_j(\mathbf{\Pi}_\tau) \\
&\geq I_{\{\tau < \infty\}} u(\mathbf{\Pi}_\tau) = \mathbb{R}(\tau, \tilde{d}).
\end{aligned} \tag{5.22}$$

Onde o evento $I_{\{\tau < \infty\}}$ define uma regra de decisão \tilde{d} e, \tilde{d} é qualquer índice $\tilde{d} = \tilde{j}$ tal $u_{\tilde{j}}(\mathbf{\Pi}_\tau) = u(\mathbf{\Pi}_\tau)$. Ou seja, minimizar o risco de Bayes a *posteriori* consiste em escolher $\tilde{j} = \arg \left(\min_{j \in \mathcal{J}} [u_j(\mathbf{\Pi}_\tau)] \right)$. Desta forma a decisão ótima depende do valor do processo $\mathbf{\Pi}$, somente no estágio em que o alarme é disparado. Note que,

$$u_j(\mathbf{\Pi}_\tau) = \sum_{i=0}^J a_{ij} \Pi_\tau(i) = a_{0j} \Pi_\tau(0) + \sum_{i=1}^J a_{ij} \Pi_\tau(i)$$

Assuma, que $a_{0j} = 1$ e $a_{ij} = 1 - I_{\{0\}}(j - i)$ é a perda 0-1, então

$$\begin{aligned} u_j(\mathbf{\Pi}_\tau) &= \Pi_\tau(0) + \sum_{i=1}^J \Pi_\tau(i) - \Pi_\tau(j) \\ &= \Pi_\tau(0) + (1 - \Pi_\tau(0)) + \Pi_\tau(j) = 1 - \Pi_\tau(j). \end{aligned}$$

De modo que sob a perda 0-1, a decisão ótima é escolher o conjunto ξ_j com a maior probabilidade a *posteriori*. Então, seja $0 < B < 1$ e

$$\tilde{\Delta} = \{\delta \in \Delta : \mathbb{R}(\delta | \mathcal{H}_n) \leq 1 - B\}$$

Defina como estratégia para detecção de *cluster* emergente um procedimento $\tilde{\delta} = (\tilde{\tau}, \tilde{d})$ com tempo de alarme,

$$\tilde{\tau} = \min\{n \geq 1 : \max_{j \in \mathcal{J}} [\Pi_n(j)] > B\} \quad (5.23)$$

e $\tilde{d} = \tilde{j} = \arg \left(\max_{j \in \mathcal{J}} [\Pi_n(j)] \right)$. então $\tilde{\delta} \in \tilde{\Delta}$. Note nesta estratégia que $[\Pi_{\tilde{\tau}}(\tilde{j})] > B$ e obtenha em 5.23 que $u(\mathbf{\Pi}_{\tilde{\tau}}) = 1 - (\mathbf{\Pi}_{\tilde{\tau}}) \leq 1 - B$ e ainda, a localização espacial do *cluster* é obtida diretamente por $\xi_{\tilde{j}} = \tilde{j}$. Para estimarmos o tempo de início do *cluster* (a parte temporal), notamos por (5.14) que

$$\begin{aligned} \Pi_{\tilde{\tau}}(\tilde{j}) &= \sum_{k=1}^{\tilde{\tau}} \pi(\nu = k, \xi = \xi_{\tilde{j}} | \mathcal{H}_{\tilde{\tau}}) \\ &= \pi(\nu = 1, \xi = \xi_{\tilde{j}} | \mathcal{H}_1) + \pi(\nu = 2, \xi = \xi_{\tilde{j}} | \mathcal{H}_2) + \cdots + \pi(\nu = \tilde{\tau}, \xi = \xi_{\tilde{j}} | \mathcal{H}_{\tilde{\tau}}). \end{aligned}$$

Grandes valores de $\pi(\nu = k, \xi = \xi_{\tilde{j}} | \mathcal{H}_{\tilde{\tau}})$ possuem maior contribuição para que o alarme dispare, de forma que

$$\tilde{k} = \arg \left[\max_{1 \leq k \leq \tilde{\tau}} \pi(\nu = k, \xi = \xi_{\tilde{j}} | \mathcal{H}_{\tilde{\tau}}) \right] \quad (5.24)$$

é um estimador para o início do *cluster*.

Notando que para todo w fixo ($w \leq n$), que o evento $\{\tau = k_0(n, w), d = j\}$ é mensurável a $\sigma(\mathbf{X}_1, \dots, \mathbf{X}_n)$ e usando o fato que $\sum_{n=w}^{\infty} I_{\{\tau=k_0(n,w), d=j\}} \Pi_{n,w}(j) = I_{\{\tau < \infty, d=j\}} \Pi_{\tau,w}(j)$. Definindo como estratégia para detecção de *cluster* emergente um procedimento $\tilde{\delta}_w = (\tilde{\tau}_w, \tilde{d}_w)$ com tempo de alarme,

$$\tilde{\tau}_w = \min\{n \geq w : \max_{j \in \mathcal{J}} [\Pi_{n,w}(j)] > B\} \quad (5.25)$$

e $\tilde{d}_w = \tilde{j} = \arg \left(\max_{j \in \mathcal{J}} [\Pi_{n,w}(j)] \right)$. então $\tilde{\delta}_w \in \tilde{\Delta}$. O fato é que, com w fixado, o raciocínio para o processo $\mathbf{\Pi}_{n,w}$ é análogo ao desenvolvido para o processo $\mathbf{\Pi}_n$.

5.5 Detecção de cluster emergente usando processo a posteriori adaptativo

Nesta seção usamos a equivalência $\{\xi = \xi_j\} = \{\xi = j\}$ e discutimos a abordagem adaptativa para o processo Π_n que pode ser estendida diretamente para o processo $\Pi_{n,w}$ pela simples substituição de $\Pi_n(j)$ por $\Pi_{n,w}(j)$. Para elaborarmos um processo a *posteriori* adaptativo, considere uma classe inicial de candidatos a parte espacial do *cluster* emergente representada por $\Xi^{(0)} = \cup \Xi_l, l = 1, 2, \dots, L$ com cardinalidade $J^{(0)}$. Onde Ξ_l é um processo estocástico com espaço de estados $\{\xi_{l_1}, \dots, \xi_{l_j}\}$. Para $n > 1$, denote $\xi_l^{(n-1)}$ como o estado do processo Ξ_l no tempo $n-1$, onde

$$\xi_{l_j}^{(n-1)} = \arg \left\{ \max_{\xi_{l_j} \in \Xi_l} [\Pi_{n-1}(\xi_{l_j})], j = 1, 2, \dots, J_l \right\}, \quad (5.26)$$

onde l identifica s_l e \tilde{j} identifica o \tilde{j} -ésimo candidato centrado em s_l . Agora, defina

$$\Xi^{(n-1)} = \{\xi_{1_{\tilde{j}}}^{(n-1)}, \dots, \xi_{L_{\tilde{j}}}^{(n-1)}\}$$

e, no tempo n construa um novo processo $\Xi^{(n)}$ com espaço de estados

$$\Xi^{(n)} = \left\{ \xi_{l_j}^{(n-1)} \in \Xi^{(n-1)} : \Pi_{n-1}(\xi_{l_j}^{(n-1)}) > B^*, l = 1, 2, \dots, L \right\}$$

para alguma constante $0 < B^* < 1$, obviamente $B^* < B$. Se não existe um $\xi_{l_j}^{(n-1)}$ tal que $\Pi_{n-1}(\xi_{l_j}^{(n-1)}) > B^*$ defina $\Xi^{(n)} = \Xi^{(n-1)}$. Finalmente, construa o processo a *posteriori* adaptativo por

$$\tilde{\Pi} = \{\tilde{\Pi}_n(0), \tilde{\Pi}_n(\xi_{l_j}^{(n-1)})\}_{n \geq 1} \quad \text{com} \quad l_j = 1, 2, \dots, |\Xi^{(n)}|. \quad (5.27)$$

onde $|\Xi^{(n)}|$ denota a cardinalidade de $\Xi^{(n)}$. Para detecção de *cluster* emergente usamos o tempo de alarme

$$\hat{\tau} = \min\{n \geq 1 : \max_{\xi_{l_j}^{(n-1)} \in \Xi^{(n)}} [\tilde{\Pi}_n(\xi_{l_j}^{(n-1)})] > B\} \quad (5.28)$$

e função de decisão $d = \tilde{l}_{\tilde{j}}$, onde $\tilde{l}_{\tilde{j}}$ identifica $\xi_{\tilde{l}_{\tilde{j}}}^{(n-1)}$ tal que,

$$\xi_{\tilde{l}_{\tilde{j}}}^{(n-1)} = \arg \left(\max_{\xi_{l_j}^{(n-1)} \in \Xi^{(n)}} [\tilde{\Pi}_n(\xi_{l_j}^{(n-1)})] \right).$$

Para uma análise "objetiva", nós usamos $B^* = 1/(J^0)$ este valor significa que a *priori*, qualquer subconjunto de localizações espaciais ξ tem a mesma probabilidade de representar a parte espacial do *cluster* emergente.

5.6 Escolha da distribuição a priori para região e ponto de mudança

Desde que o interesse em nosso trabalho é encontrar *clusters* localizados e não uma possível mudança global na região em estudo, é razoável supor, por exemplo, que pequenos *clusters* (conjuntos com poucas localizações espaciais ou poucas áreas) são mais prováveis que um *cluster* grande (conjuntos com muitas localizações espaciais ou muitas áreas). Outra questão de interesse, pode ser a detecção de pequenos *cluster* com uma certa tendência espacial. Baseado neste pensamento, nós usamos a idéia de modelo de *cluster* descrita em Gangnon e Clayton (2000) (veja novamente, seção 2.1.5) e construímos uma classe $\Xi = \{\xi_1, \dots, \xi_J\}$ de candidatos com uma estrutura probabilística associada que depende somente das características geométricas (forma, tamanho e direção) de ξ . Especificamente, seja s_l uma localização espacial na área em estudo e ξ_{l_j} um candidato a parte espacial do *cluster* emergente com parâmetros de forma (φ_{l_j}), tamanho (ς_{l_j}) e direção (ϱ_{l_j}) conhecidos. Então, para a distribuição de ξ_{l_j} definimos

$$\pi(\xi_{l_j}) = c \times \exp\{-p_1(l_j)E(\varphi_{l_j}) - p_2(l_j)E(\varsigma_{l_j}) - p_3(l_j)E(\varrho_{l_j})\} \quad (5.29)$$

em que $E(\varphi_{l_j})$, $E(\varsigma_{l_j})$, $E(\varrho_{l_j})$ são, respectivamente, escores de tamanho, forma e direção do *cluster*, a terna $(p_1(l_j), p_2(l_j), p_3(l_j))$ são valores não negativos representando pesos de importância e

$$c = \sum_{\xi_{l_j} \in \Xi} \exp\{-p_1(l_j)E(\varphi_{l_j}) - p_2(l_j)E(\varsigma_{l_j}) - p_3(l_j)E(\varrho_{l_j})\},$$

é uma constante de normalização. Estas distribuições a *priori* são importantes para refletir a incerteza sobre a estrutura do *cluster*. e podemos defini-las baseado em alguns critérios, como os listados a seguir (estes são apenas algumas sugestões):

1. um *cluster* pequeno é mais provável que um *cluster* grande. Isto preserva a similaridade espacial.
2. um *cluster* em formato circular é mais provável que um irregular. Podemos supor que não existe tendência espacial.
3. um *cluster* elíptico é mais provável desde que o interesse seja detectar uma tendência espacial.
4. uma área com pequena população tem menos chance de pertencer ao *cluster*, pois suas taxas são muito instáveis.

Por exemplo a figura 5.1 apresenta uma classe Ξ de candidatos a parte espacial *clusters* com formato geométrico circular, porém com tamanho variável. Em cada ponto s_l arbitrário da região em estudo podemos centrar um círculo e aumentar seu raio. Para esta classe nós podemos

assumir que $E(\varphi_{l_j})$ é igual ao raio de ξ_{l_j} , r_{l_j} . Observe que r_{l_j} grande $\pi(\xi_{l_j})$ diminui, tornando um *cluster* pequeno mais provável (situação 1). Desde de que a forma está fixada (formato circular) tomamos $E(\varsigma_{l_j})$, pois $E(\varsigma_{l_j}) > 0$ penaliza a forma do *clusters*. Notando que o formato circular independe da direção, tomamos $E(\varrho_{l_j})$. Para escolher $p_1(l_j)$, podemos proceder da seguinte forma: sejam ξ_{l_j} e $\xi_{l_j}^*$ dois candidatos a *cluster* tal que $r_{l_j} = 2$ e $r_{l_j}^* = 3$. Então, obtemos que

$$\frac{\pi(\xi_{l_j})}{\pi(\xi_{l_j}^*)} = e^{p_1(l_j)} \Rightarrow p_1(l_j) = \log \left(\frac{\pi(\xi_{l_j})}{\pi(\xi_{l_j}^*)} \right)$$

tal que $p_1(l_j)$ reflete pesos de preferência na escala logarítmica. Por exemplo, nós podemos supor que ao aumentarmos o raio (aumentamos o tamanho do *cluster*) em uma unidade de distância (por exemplo, 1Km) ocorra uma perda de similaridade espacial entre o centro do *cluster* e as áreas da borda do *cluster*, tal que ξ_{l_j} seja (devido a perda de similaridade espacial de $\xi_{l_j}^*$) 10% mais provável que $\xi_{l_j}^*$. Então escolhemos $p_1(l_j) = \log(1,1) \cong 0,1$. A escolha $p_1(l_j) = 1$ implica que círculos com mesmo raio possuem a mesma probabilidade de ser a verdadeira parte espacial do *cluster* emergente.

Em geral, diferentes funções escore para cada parte espacial do *cluster* emergente podem ser usadas para refletir diferentes crenças a *priori* sobre a estrutura do *cluster*. Especificações apropriadas requerem o conhecimento de um *expert* sobre o padrão dos eventos em questão. Quando $E(\varphi_{l_j}) = E(\varsigma_{l_j}) = E(\varrho_{l_j}) = 0$, teremos uma distribuição a *priori* uniforme no espaço Ξ dada por $\pi(\xi_{l_j}) = 1/|\Xi|$, $\forall \xi_{l_j} \in \Xi$, onde $|\Xi|$ é a cardinalidade de Ξ .

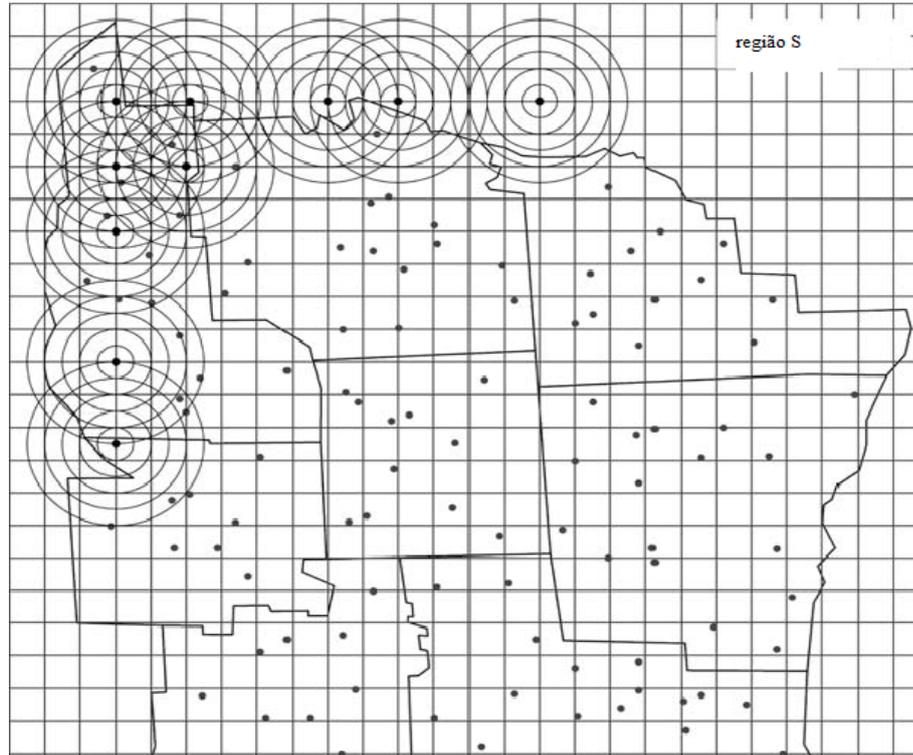


Figura 5.1: Uma Classe de clusters com formato circular e tamanho variável

A distribuição a *priori* para o ponto de início do *cluster* emergente (parte temporal) é baseada numa versão truncada da distribuição geométrica. Assumimos que

$$\pi(\nu = k) = \pi_\nu(k) = p(1-p)^{k-(n-w+1)}, \quad k = n-w+1, n-w+2, \dots \quad (5.30)$$

onde $w \leq n$ é uma janela temporal de valor pré-especificado. Obviamente quando $w = n$ temos a distribuição geométrica usual com

$$\pi_\nu(k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots \quad (5.31)$$

Em uma análise puramente temporal, a distribuição geométrica dada em (5.31) foi usada pela primeira vez por Shiryaev (1963) como distribuição a *priori* para um provável ponto de mudança. Recentemente, esta distribuição tem sido usada para monitorar mudanças no tempo de chegada em um processo de Poisson em várias categorias com taxa de chegada conhecida (Brown, 2008) e desconhecida (Brown, 2009). Uma classe mais geral de distribuições a *priori*, no qual a geométrica está inserida, usada para monitoramento de mudanças na escala temporal tem sido apresentada em Borovkov(1998). Um grande problema ao usár-mos esta abordagem é a especificação de p . Na escala temporal, Frisén and Wessman (1999), Sonesson and Bock (2003), tem recomendado o uso de pequenos valores para p (por exemplo $p = 0.20$) de modo que os resultados possam ser comparados com o método da razão de verossimilhança (veja, 2.2.2). Pequenos valores de p indicam que a probabilidade de uma mudança é pequena. Nós assumimos que no tempo n a probabilidade a *priori* que não tenha ocorrida uma mudança é p_0 . Desta forma, p satisfaz a equação,

$$\mathbf{P}(\nu > n) = p_0$$

mas,

$$\begin{aligned} \mathbf{P}(\nu > n) &= \sum_{k=n+1}^{\infty} p(1-p)^{k-(n-w+1)} = (1-p)^{-n+w} \sum_{k=n+1}^{\infty} p(1-p)^{k-1} \\ &= (1-p)^{-n+w} \times (1-p)^n = (1-p)^w. \end{aligned}$$

então obtenha $p = 1 - p_0^{1/w}$. Uma escolha natural é $p_0 = 1/2$ de modo que $p = 1 - 0.5^{1/w}$. Para esta escolha, a probabilidade a *priori* de que não tenha ocorrido mudança no ponto k é maior que 0.5.

5.7 Aplicação

Para ilustrar a utilização do método para detecção de *cluster* no espaço-tempo baseado em um processo a *posteriori*, nós avaliamos o conjunto de dados referente aos casos de câncer da tireóide em homens que ocorreram no Novo México durante o período de tempo 1973-1992. Este conjunto de dados foi descrito e utilizado na seção 3.3 desta tese e têm sido usado por alguns

pesquisadores (Kulldorff, 2001; Lima e Duczmal, 2011) como um banco de dados teste para avaliar a performance de um método proposto para a detecção de *cluster* emergente.

5.7.1 Análise dos dados de câncer da tireóide

A região de estudo é o mapa apresentado na figura 5.2 representando as 32 áreas do Novo México ($l = 1, 2, \dots, L = 32$). Para a construção do método, seja $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_L))'$ o processo sob monitoramento, em que $X_t(s_l)$ representa o número de casos câncer da tireóide no tempo t na área A_l identificada por s_l que representa o centróide da área. assumimos que condicional a $\boldsymbol{\theta}_t = (\theta_t(s_1), \theta_t(s_2), \dots, \theta_t(s_L))'$, o vetor $\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_L))'$ é um processo de Poisson tal que $(X_t(s_l) | \theta_t(s_l)) \sim \text{Poisson}(e_t(s_l) \times \theta_t(s_l))$ com $e_t(s_l)$, assumido conhecido, representando a contagem esperada da doença no tempo t em s_l , variando entre as áreas de acordo com idade e raça. A quantidade $\theta_t(s_l)$ é uma medida de risco relativo (em relação ao mapa) de ocorrência de um caso de câncer da tireóide no tempo t em A_l . Se a doença está sob-controle no mapa, esperamos que esta medida de risco seja próximo de 1 com pequenas variações.

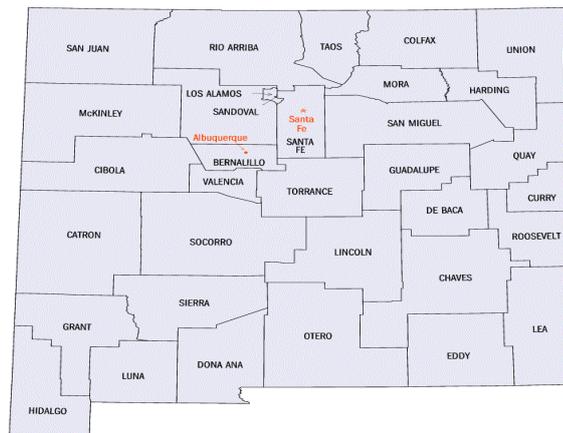


Figura 5.2: Mapa do Novo México. Fonte: US Census.

Determinação do valor de $e_t(s_l)$

Para determinarmos o valor de $e_t(s_l)$, usamos uma padronização interna ajustada por idade e raça. Especificamente, defina $x_{T,i,r}(s_l)$ como sendo o número de casos que ocorreram no tempo T ($1 \leq T \leq t$) entre as pessoas na classe de idade i pertencente a raça r na l -ésima área, e $N_{T,i,r}(s_l)$ a respectiva população em risco. Então, a estimativa para taxa global de casos em todo o mapa referente a classe de idade i na raça r no tempo T é dada por $\lambda_{T,i,r} = \sum_{s_l \in \mathcal{S}} x_{T,i,r}(s_l) / \sum_{s_l \in \mathcal{S}} N_{T,i,r}(s_l)$. Logo $e_{T,i,r}(s_l) = N_{T,i,r}(s_l) \lambda_{T,i,r}$ representa uma estimativa para o número de casos esperados no tempo T na classe de idade i e raça r sob a suposição de que não existe cluster no espaço-tempo. Portanto, uma estimativa para o número total de casos

esperados na l -ésima área no tempo T é dado por $e_T(s_l) = \sum_i \sum_r e_{T,i,r}(s_l)$ e, finalmente usamos,

$$e_t(s_l) = \frac{1}{t-w} \sum_{T=1}^{t-w} e_T(s_l), \quad t \geq w+1.$$

Escolha das distribuições a priori

1. Distribuição a priori para $\theta_t(s_l)$: Supomos que

$$G_0(\theta_t(s_l)) = 1 - \exp \left\{ - \left[\frac{\theta_t(s_l)}{b} \right]^a \right\}.$$

Obviamente, os $\{\theta_t(s_l) : l = 1, 2, \dots, L \text{ e } t \geq 1\}$ são assumidos independentes com distribuição comum $\sim Weibull(a, b)$ onde a e b são, respectivamente, os parâmetros de forma e escala dessa distribuição. A distribuição $Weibull(a, b)$ surge, neste capítulo, como uma distribuição alternativa em relação ao uso de uma distribuição na família conjugada (ver seção 4.6), pois ela pode tomar uma grande variedade de formas regulares podendo, por exemplo, acomodar à assimetria da distribuição. Esta distribuição, possui

$$\text{Moda}(\theta_t(s_l)) = b \left(\frac{a-1}{a} \right)^{1/a}, \quad a > 1 \text{ e } b > 0.$$

e

$$\text{Variância} = b^2 (\Gamma(1 + 2/a) - \Gamma((1 + 1/a))^2)$$

Em que $\Gamma(\cdot)$ denota a função gama. Em problemas de detecção de *cluster* emergente, é razoável supor que, se não existem mudanças no processo espaço-temporal em monitoramento, os $\{\theta_t(s_l) : l = 1, 2, \dots, L \text{ e } t \geq 1\}$ surgem de forma aleatória a partir de uma distribuição comum. Desta forma, quando o processo esta sob-controle, pode-se interpretar a $\text{Moda}(\theta_t(s_l))$ como sendo o valor mais provável para o risco relativo de ocorrência de um caso de câncer da tireóide em uma área escolhida ao acaso no mapa. Observe que, se

$$b = \left(\frac{a-1}{a} \right)^{-1/a}, \quad \text{então a } \text{Moda}(\theta_t(s_l)) = 1.$$

Logo, sobre esta escolha para b , podemos selecionar o valor a de modo a obtermos pequenas variações em torno de 1 para a distribuição de $\theta_t(s_l)$. Como escolha apropriada, nós usamos o par $(a, b) = (2, 1.414)$ que fornece $\text{Moda} = 1$ e pouca variação (veja figura 5.3). Agora, seja

$$y = G_0(\theta_t(s_l)) \Rightarrow G_0^{-1}(y) = b (-\log(1-y))^{1/a}, \quad 0 < y < 1.$$

e

$$f(X_t(s_l) | \theta_t(s_l)) = (X_t(s_l)!)^{-1} (e_t(s_l) \theta_t(s_l))^{X_t(s_l)} \exp\{-e_t(s_l) \theta_t(s_l)\}$$

então,

$$f(X_t(s_l)|G_0^{-1}(y)) = (X_t(s_l)!)^{-1} \left(e_t(s_l) b(-\log(1-y))^{1/a} \right)^{X_t(s_l)} \times \exp\{-e_t(s_l)[b(-\log(1-y))^{1/a}]\}. \quad (5.32)$$

Usamos como grandeza estocástica desejada $m = 2, 4, 8$, computamos a expressão (5.12) e obtemos o valor de $FB_t(j)$ em (5.10). O comportamento da distribuição *a priori*, do máximo de m -variáveis aleatórias com distribuição G_0 e $1 - G_0^m$ são mostrados na figura 5.3.

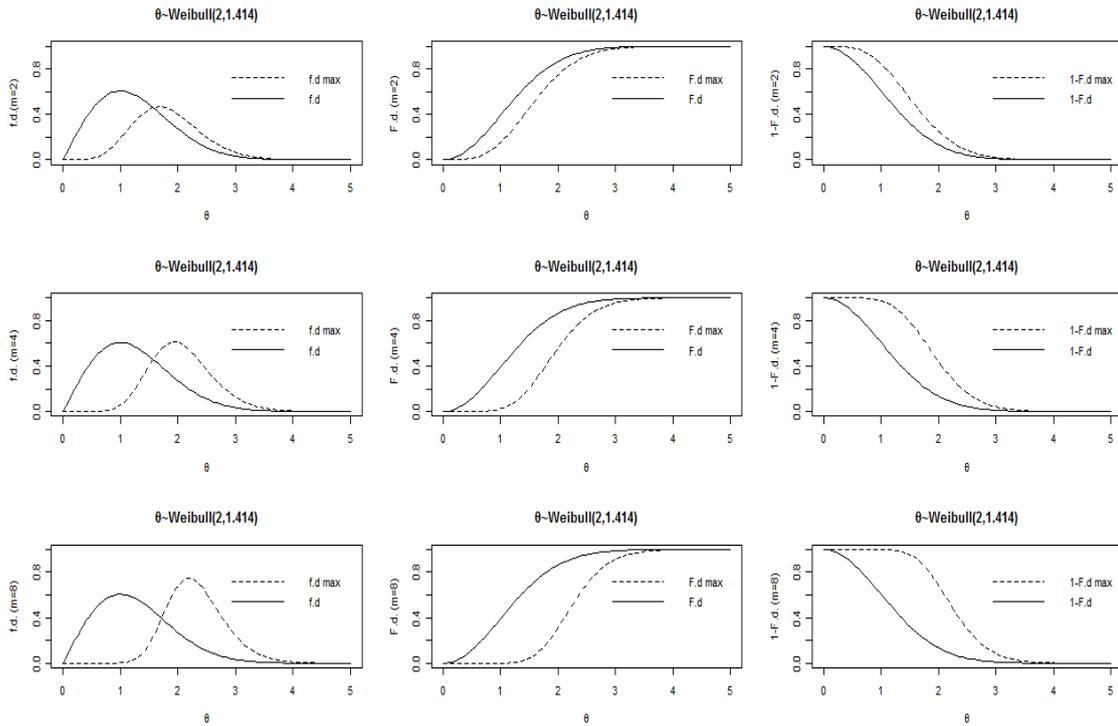


Figura 5.3: função de densidade weibull (f.d), função de densidade do máximo de m -variáveis weibull (f.d. max), Função de distribuição weibull (F.d), Função de distribuição do máximo de m -variáveis weibull (F.d. max), 1-Função de distribuição weibull (1-F.d), 1-Função de distribuição do máximo de m -variáveis weibull (1-F.d. max)

2. Distribuição *a priori* para ξ : Para compararmos nossos resultados com os obtidos em Kulldorff (2001) e Lima e Duczmal (2011). Nós supomos que ξ tem formato geométrico circular, porém com tamanho variável como mostrado na figura 5.1. Em cada ponto s_l arbitrário do mapa nós centramos um círculo e aumentamos seu raio r_{l_j} até que contenha todos os outros centróides. Nesta formulação não existe restrição ao tamanho do círculo de modo que temos um total inicial de círculos igual a $L^2 = 32^2$. Usando um simples algoritmo para eliminação de círculos com as mesmas localizações espaciais, obtemos um total de

$$J^0 = 743 \quad (5.33)$$

círculos distintos. Usando $p_1(l_j) = 1$ a expressão (5.29) simplifica em,

$$\pi(\xi_{l_j}) = c \times \exp(-r_{l_j}), \quad \text{onde} \quad c = \sum_{\xi_{l_i} \in \Xi} \exp(-r_{l_i}). \quad (5.34)$$

A escolha $p_1(l_j) = 1$ implica que círculos com mesmo raio possuem a mesma probabilidade *a priori* de ser a verdadeira parte espacial do *cluster* emergente. Para obtenção do raio r_{l_j} , realizamos um reescalonamento nas distâncias entre os centróides do mapa. Centrado em s_l , seja $d(s_l, s_{l'})$ a distância de s_l a qualquer outro ponto $s_{l'}$ do mapa e $Max(s_l)$ a maior distância observada, então definimos $\tilde{d}(s_l, s_{l'}) = d(s_l, s_{l'})/Max(s_l)$. Usando $\tilde{d}(s_l, s_{l'})$ como medida de similaridade espacial construímos Ξ de acordo com a explicação dada na seção 3.2.5 e computamos a expressão (5.34).

3. Distribuição *a priori* para ν : A distribuição para ν é baseada numa versão truncada da distribuição geométrica discutida na seção 5.6. Ou seja,

$$\pi(\nu = k) = \pi_\nu(k) = p(1 - p)^{k-(n-w+1)}, \quad k = n - w + 1, n - w + 2, \dots \quad (5.35)$$

onde $w \leq n$ é a janela temporal de valor pré-especificado. Usamos $w = 6$ que representa o mesmo valor usado em Kulldorff (2001). Como discutido na seção 5.6 p satisfaz a equação $p = 1 - p_0^{1/w}$ onde p_0 é a probabilidade *a priori* que não tenha ocorrido uma mudança até o tempo n . A tabela 5.1 apresenta alguns valores para p de acordo com o valor de w e $p_0 = 0.5$ que é o valor usado nesta aplicação.

Tabela 5.1: Valores para p de acordo com o valor de w e $p_0 = 0.5$.

$\mathbf{P}(\nu > n)$	w	p
	3	0.21
	4	0.16
$p_0 = 0.5$	5	0.13
	6	0.11
	7	0.09
	8	0.08

5.7.2 Obtenção do processo *a posteriori* adaptativo

Para obtermos a distribuição do processo *a posteriori* adaptativo, usamos (5.8) combinada com a abordagem adaptativa desenvolvida na seção 5.5. Como a distribuição *a priori* para ν é

truncada em $n - w + 1$ com $w = 6$, obtemos por (5.8) que para todo $n - 5 \leq k \leq n, \xi = \xi_j$,

$$\pi(\nu = k, \xi = \xi_j | \mathcal{H}_n) = \frac{\pi_\nu(k) \pi_\xi(l_j) \prod_{t=k}^n FB_t(l_j)}{\sum_{\xi_{l_i} \in \Xi^n} \sum_{k=n-5}^n \pi_\nu(k) \pi_\xi(l_i) \prod_{t=k}^n FB_t(l_i) + \sum_{k=n+1}^{\infty} \pi_\nu(k)}$$

e pela substituição dos termos descritos na seção 5.7.1, temos:

1. $\pi_\nu(k) = p(1-p)^{k-(n-5)}$ dada em (5.35) com $p = 1 - 0.5^{1/6}$.
2. $\sum_{k=n+1}^{\infty} \pi_\nu(k) = \mathbf{P}(\nu > n) = p_0 = 0.5$ (valor assumido nesta aplicação).
3. $\pi_n(\xi_{l_j}) = c_n \times \exp(-r_{l_j})$ (*priori* no tempo n) dada em (5.34) com $c_n = \sum_{\xi_{l_i} \in \Xi(n)} \exp(-r_{l_i})$
4. Combinando 5.10 com 5.12, aproximamos $FB_t(l_j)$ por,

$$FB_t(l_j) \approx \prod_{s_l \in \xi_{l_j}} \frac{\sum_{q=1}^Q f_m(X_t(s_l) | G_0^{-1}(z_q))}{\sum_{q=1}^Q f_1(X_t(s_l) | G_0^{-1}(z_q))}.$$

onde z_1, z_2, \dots, z_Q são $Q = 1000$ valores gerados de uma distribuição Beta($m, 1$). O índice em $f_1(\cdot)$ significa que $m = 1$ e $f_m(\cdot)$ é dada em (5.32) substituindo y por z_q .

Desta forma,

$$\Pi_{n,w}(l_j) = \mathbb{P}(n - w + 1 \leq \nu \leq n, \xi = \xi_{l_j} | \mathcal{H}_n) = \sum_{k=n-w+1}^n \pi(\nu = k, \xi = \xi_{l_j} | \mathcal{H}_n). \quad (5.36)$$

é dada pela expressão (com $w = 6$),

$$\frac{\sum_{k=n-5}^n (1 - 0.5^{1/6})(0.5^{1/6})^{k-(n-5)} \exp(-r_{l_j}) \prod_{t=k}^n \prod_{s_l \in \xi_{l_j}} \frac{\sum_{q=1}^Q f_m(X_t(s_l) | G_0^{-1}(z_q))}{\sum_{q=1}^Q f_1(X_t(s_l) | G_0^{-1}(z_q))}}{\sum_{\xi_{l_i} \in \Xi(n)} \sum_{k=n-5}^n (1 - 0.5^{1/6})(0.5^{1/6})^{k-(n-5)} \exp(-r_{l_i}) \prod_{t=k}^n \prod_{s_l \in \xi_{l_i}} \frac{\sum_{q=1}^Q f_m(X_t(s_l) | G_0^{-1}(z_q))}{\sum_{q=1}^Q f_1(X_t(s_l) | G_0^{-1}(z_q))}} + 0.5/c_n \quad (5.37)$$

Obviamente,

$$\Pi_n(0) = 1 - \sum_{\xi_{l_i} \in \Xi(n)} \Pi_{n,w}(l_j).$$

Como em (5.33) temos $J^0 = 743$ obtemos da seção 5.5 que $B^* = 1/743 \approx 0.0013$. Agora, usamos um limite de controle $B = 0.5$ e aplicamos o processo a *posteriori* adaptativo nos conjunto de dados de câncer da tireóide (Kulldorff, 2001) e os resultados obtidos são apresentados e discutidos na próxima seção.

5.7.3 Análise dos resultados da aplicação

Os resultados mais importantes envolvendo a aplicação do método de monitoramento proposto, baseado na Máxima Probabilidade a *Posteriori* (MPA) conjunta de um possível tempo de mudança em alguma região do mapa, são apresentados nas figuras 5.4 e 5.5. A figura 5.4

apresenta o funcionamento da função de alarme baseada na máxima probabilidade a posteriori adaptativa para $m=2$, $m=8$ e também o tempo, a região e as observações que provocaram a mudança no padrão espaço-temporal do processo. Para $m=2$, observa-se que quando $t = 15$ a MAP cresce e atinge o maior valor 0.443 no tempo final de monitoramento $n=20$ na sub-região $\tilde{\xi}$ do mapa formada pelas áreas (ver, figura 5.4): Los Alamos, Rio Arriba, Taos e Santa Fé. A área central do *cluster* é Rio Arriba. Todos os pontos marcados (por *) no gráfico são provenientes de $\tilde{\xi}$ de modo que o *cluster* emergente detectado é representado geometricamente por um cilindro $[15, 20] \times \tilde{\xi}$ de base $\tilde{\xi}$ e altura $[15, 20]$. Obviamente, que usando um limite de controle $B = 0.5$, este cluster não dispara o alarme e portanto não é considerado significativo. Ainda na figura 5.4, quando $m=8$, observa-se uma sequência crescente de valores a partir do ponto $t = 18$ até o tempo atual $n = 20$. Nesta sequência de pontos no tempo, a máxima probabilidade a posteriori assume, respectivamente, os valores 0.584, 0.882 e 0.958, mostrando claramente a presença de um *cluster* emergente significativo que começa no tempo de mudança $k = 18$ e áreas pertencentes ao *cluster* $\hat{\xi}$ (ver, figura 5.4): Los Alamos, Rio Arriba e Taos. Este *cluster* emergente é geometricamente representado pelo cilindro $[18, 20] \times \hat{\xi}$ de base $\hat{\xi}$ e altura $[18, 20]$. Este está contido em $[15, 20] \times \tilde{\xi}$ que foi o cilindro obtido com $m=2$. Na figura 5.5, Apresentamos conjuntamente, o comportamento da máxima probabilidade a posteriori para $m=2, 4$ e 8 em (A), e também, a razão prospectiva para os casos observados por casos esperados nas áreas que compõem o *cluster* significativo. A razão prospectiva é calculada por

$$\sum_{t=k}^n \sum_{s_l \in \hat{\xi}} \frac{X_t(s_l)}{e_t(s_l)}, k = 1, 2, \dots, n.$$

Os valores observados para esta razão, mostram que nos últimos dois períodos de tempo ($t = 19, 20$) esta razão foi superior para todas as áreas pertencentes ao *cluster* indicando um aumento no risco relativo de ocorrência de um caso da doença.

Analisando este mesmo conjunto de dados, a estatística scan espaço-tempo de Kulldorff (2001) detectou o cluster emergente $[17, 20] \times \xi^*$ ($\xi^* = \{LosAlamos\}$) com um valor- $p=0.016$. Lima e Duczmal (2011) usando uma função de alarme baseada na razão de verossimilhança adaptativa detectam o *cluster* emergente $[16, 20] \times \xi^*$ com razão de verossimilhança adaptativa em favor do *cluster* igual a 42.672. Obviamente, uma probabilidade a posteriori igual a 0.958 é mais conclusiva.

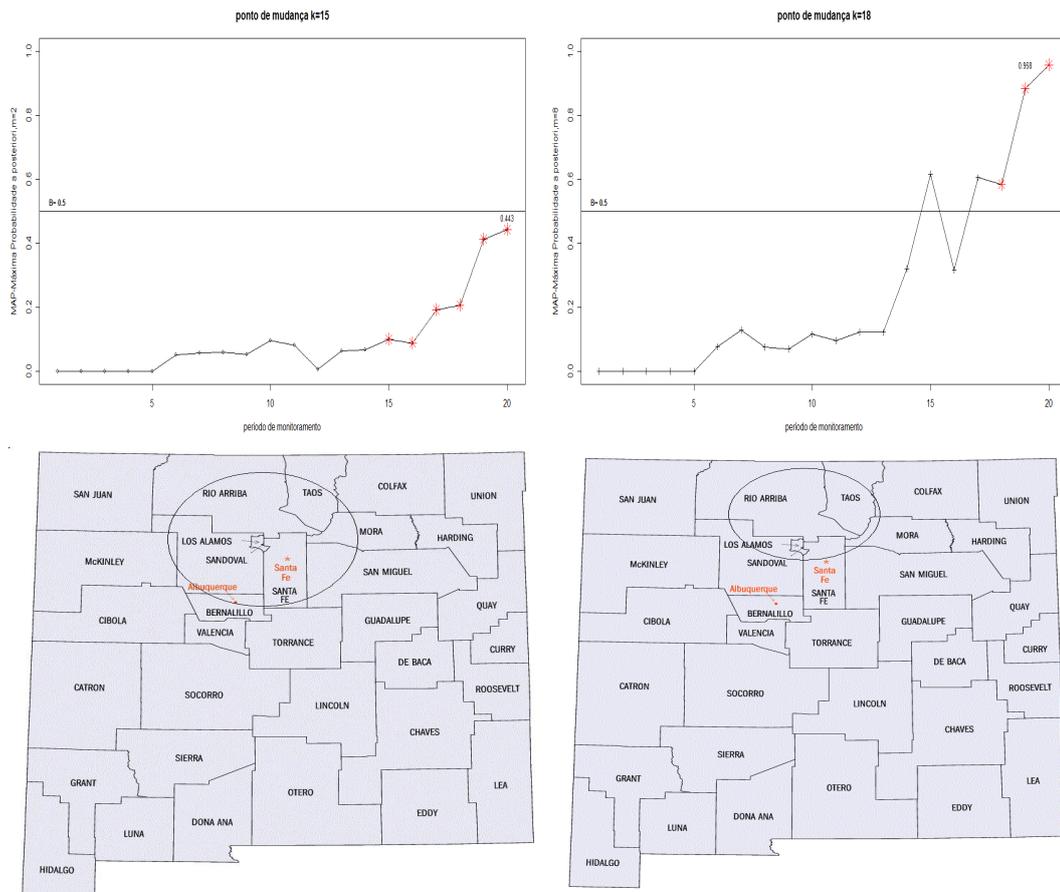


Figura 5.4: A esquerda: Máxima probabilidade a posteriori adaptativa para $m=2$, ponto $k=15$ e região da mudança; A direita: Máxima probabilidade a posteriori adaptativa para $m=8$, ponto $k=19$ e região da mudança.

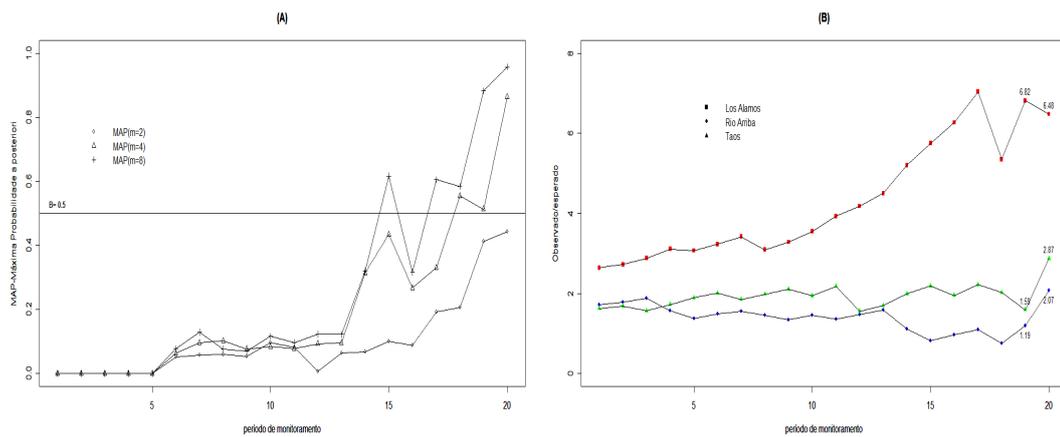


Figura 5.5: Em (A): Máxima probabilidade a posteriori adaptativa para $m=2, 4$ e 8 ; Em (B) Razão valor observado por valor esperado $x_t(s_t)/e_t(s_t), t \geq k = 1, 2, \dots, n$ para as regiões do cluster emergente detectado e considerado significativo

5.8 Estudo com dados simulados

5.8.1 Dados simulados

Para avaliar a performance do método baseado em um processo a *posteriori* para a detecção de *cluster* no espaço-tempo, realizamos um estudo com dados simulados com o objetivo de detectar mudanças na distribuição do risco de ocorrências de casos de uma doença sobre o espaço-tempo. Consideramos como população em risco, $N_t(s_l)$, a população de homens residentes no ano de 1992 nas 32 ($L=32$) áreas do Novo México como descrito em Kulldorff(2001). Sobre esta população em risco nós simulamos um processo de Poisson

$$\mathbf{X}_t = (X_t(s_1), \dots, X_t(s_L))'$$

condicional a um outro processo de risco relativo

$$\boldsymbol{\theta}_t = (\theta_t(s_1), \theta_t(s_2), \dots, \theta_t(s_L))'$$

tal que $\mathbb{E}(X_t(s_l)|\theta_t(s_l)) = e_t(s_l) \times \theta_t(s_l)$ com $e_t(s_l)$ representando o número de casos esperados em s_l no tempo t . e dado por $e_t(s_l) = \lambda_0 N_t(s_l)$. λ_0 é uma taxa de referência global para a doença.

Os casos são simulados para um período de monitoramento $t = 1, 2, \dots, n = 20$ e 1000 réplicas do processo foram geradas. A parte espacial do cluster alocado no mapa é formado por 4 áreas, $\xi = \{\text{chaves}, \text{Eddy}, \text{Lea}, \text{Otero}\}$ que podem ser identificadas na figura 5.2. Este é um cluster com população mediana em relação ao mapa. Quando simulamos um processo sem *cluster* assumimos que sob o modelo padrão $\{\boldsymbol{\theta}_t\}$, é realização de um processo Gama tal que,

$$\theta_t(s_l) \sim \text{Gama}(100, 100), t = 1, 2, \dots, 20 \quad \text{e} \quad l = 1, 2, \dots, 32.$$

Esta distribuição a *priori* produz uma pequena variação no risco sobre todo o mapa tal que o valor esperado de $X_t(s_l)$ esta próximo de $e_t(s_l)$. Para o modelo de *cluster* emergente definimos um risco alternativo por $\tilde{\theta}_t(s_l) = \theta_t(s_l) \delta^{I_{\{t \geq 11, s_l \in \xi\}}}$. onde $\theta_t(s_l)$ é o valor gerado da distribuição de referência $\text{Gama}(100, 100)$ e $\delta = 1.5, 2, 3$. Observa-se que neste caso o ponto de mudança é $k = 12$. Usamos, ainda, $\lambda_0 = 0.0001$ (taxa de um caso para cada 10000 habitantes); uma janela temporal de tamanho $w = 6$. Para detectarmos o *cluster* no espaço tempo, nós usamos $G_0 \sim \text{Weibull}(2, 1.414)$ e $G_0 \sim \text{Unif}(0, 2)$. Isto gera um total de 18 cenários diferentes sobre o modelo de *cluster* e sobre estes cenários nós aplicamos como função de monitoramento a expressão 5.37 e calculamos as medidas de performance: Sensitividade Condicional de um alarme **SSC**; Valor Predito Positivo Condicional de um alarme **VPP**; atraso na detecção **ADD**. descritas na secção 3.5.1. Uma observação importante, é que embora o processo $\{\boldsymbol{\theta}_t\}$ seja gerado de uma distribuição Gama, usamos como distribuições de referência as distribuições Uniforme e Weibull.

5.8.2 Análise dos resultados obtidos com dados simulados

Os resultados para as medidas de performance estão descritos em termos de médias na tabela 5.2. Nota-se, na tabela que o atraso (**A \hat{D} D**) na detecção decresce tornando-se muito pequeno quando aumentamos o valor de δ e m . A **SSC** é sempre menor que o **VPP** implicando que o *cluster* localizado tende a ser menor que o verdadeiro *cluster*, porém com uma boa precisão. Isto deve ocorrer como efeito da distribuição a priori para a parte espacial do *cluster* que favorece de uma maneira mais forte os *clusters* com pequenas áreas (pois as distâncias entre seus centróides são mais próximas). Por exemplo, quando $\delta = 2$, $m = 8$ e $G_0 \sim Weibull$ o atraso estimado é igual a 0.581, que pode ser considerado muito pequeno, pois o cluster foi detectado logo no seu início. A precisão na localização medida pela **SSC** é 0.907 enquanto que o **VPP**=0.946.

De uma forma geral, o método é bastante rápido (possui pequeno atraso), possui uma boa precisão e aparentemente parece ser robusto no sentido de que mesmo usando uma distribuição a priori diferente da verdadeira, os resultados obtidos foram satisfatórios. Uma amostra do funcionamento on-line do método com dados simulados é apresentado na figura 5.6. Nesta figura observamos que quando m cresce a máxima probabilidade a posteriori se aproxima de 1.

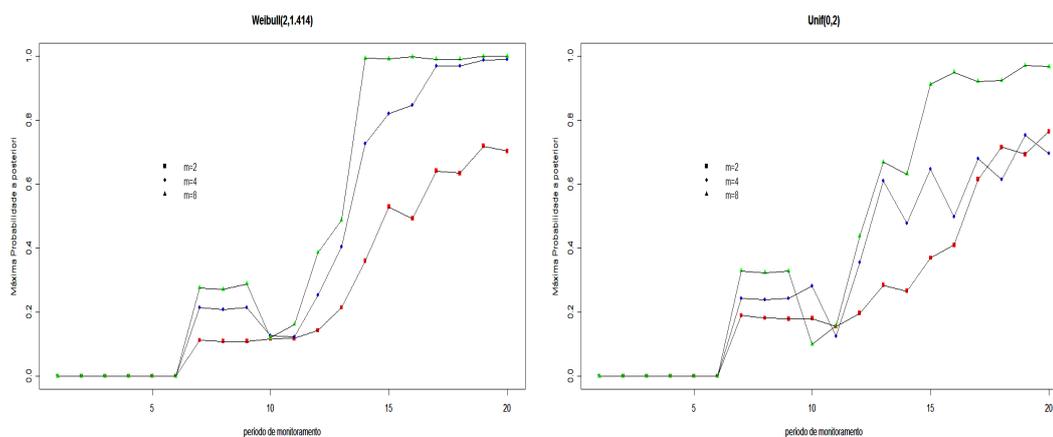


Figura 5.6: Amostra de uma simulação do processo a posteriori adaptativo para distribuições de referência Weibull e Uniforme com: $\delta = 2$, $B=0.5$, $m=2,4,8$

Tabela 5.2: Valores estimados para Sensitividade (**SSC**), Valor Predito Positivo (**VPP**) e Atraso (**ADD**) usando o processo a posteriori adaptativo com janela $w=6$, diferentes valores para δ e duas distribuições para G_0 : Weibull(2,1.414) e Uni(0,2)

		Medidas de Performance			
		δ	SSC	VPP	ADD
Weibull(2,1.414)	$m = 2$	1.50	0.571	0.634	3.785
		2.00	0.510	0.742	3.462
		3.00	0.341	0.679	2.917
	$m = 4$	1.50	0.940	0.953	2.950
		2.00	0.907	0.941	1.478
		3.00	0.805	0.928	1.224
	$m = 8$	1.50	0.933	0.958	1.845
		2.00	0.907	0.946	0.581
		3.00	0.877	0.897	0.463
Uni(0,2)	$m = 2$	1.50	0.741	0.779	3.170
		2.00	0.781	0.813	3.000
		3.00	0.639	0.823	2.602
	$m = 4$	1.50	0.846	0.887	2.884
		2.00	0.819	0.835	1.257
		3.00	0.847	0.883	0.915
	$m = 8$	1.50	0.816	0.873	2.553
		2.00	0.809	0.833	0.933
		3.00	0.800	0.807	0.384

5.9 Discussão

Neste capítulo desenvolvemos um novo método para detecção e monitoramento de *clusters* emergentes. Ao incorporarmos na análise, através de distribuições *a priori*, a incerteza na

localização espacial e no ponto de mudança no tempo, nós realizamos um monitoramento no espaço-tempo baseado na máxima probabilidade a *posteriori* conjunta de que tenha ocorrido uma mudança no padrão de eventos no espaço-tempo. Esta nova metodologia é uma combinação dos seguintes métodos descritos nas seções 2.2 e 2.1: Shiryaev-Roberts, usado para detecção de mudanças na escala temporal; e o método Bayesiano de Gangnon e Clayton (2000), usado para a detecção de mudanças no padrão espacial de eventos. Mostramos que esta abordagem minimiza uma perda esperada a *posteriori* representada por uma combinação de uma perda por um falso alarme e uma falsa localização. Desenvolvemos uma estrutura geral de distribuições a *priori* para a parte espacial do *cluster* que levam em consideração o tamanho, forma e direção do *cluster*. Para identificar possíveis mudanças na taxa de ocorrências de eventos, nós elaboramos uma abordagem alternativa em relação ao uso de distribuições na família conjugada (descritas na seção 4.4), baseada em mudanças na função de distribuição a *priori*. Essas mudanças são capturadas por uma distribuição que é estocasticamente maior do que a considera como distribuição padrão (ou distribuição de referência). Resultados mostrados em nossas simulações mostraram uma certa robustez no método na presença de distribuições a *priori* mal especificada. Em nossas aplicações obtivemos resultados mais conclusivos que os apresentados em Kulldorff (2001) e Lima e Duczmal (2011).

6 Conclusão

Nesta tese de doutorado foram desenvolvidos vários métodos para a detecção de mudanças no espaço-tempo que podem ser usados em uma variedade de sistemas de vigilância: Sistema de vigilância em crimes; Sistema de vigilância em saúde pública; Sistemas de monitoramento ambiental. O foco principal é a detecção de *clusters* no espaço-tempo.

Ao combinarmos técnicas de vigilância estatística no domínio do tempo com estimação adaptativa e algoritmo para detecção de *cluster* espacial com representação geométrica circular, nós desenvolvemos um método de monitoramento no espaço-tempo baseada no uso da Razão de Verossimilhança Adaptativa (ALR). Em essência a ALR é uma razão de verossimilhança genuína e por isso preserva a propriedade martingal de modo que os métodos baseados na ALR não requerem o uso de testes de Monte Carlo para validar a significância estatística do procedimento o que tornou bastante rápido e eficiente o algoritmo desenvolvido para detecção do tempo e localização da mudança.

Ao considerarmos a possível observação de processos espaço-temporais compostos por componentes correlacionados, não-estacionários ou com localização espacial da mudança com um formato não circular, Nós desenvolvemos, para a detecção de *clusters* emergentes nestes tipos de processos, uma abordagem Bayesiana baseada em um Fator Bayes Adaptativo com janela temporal limitada e usamos uma estrutura de *cluster* onde o espaço é representado por subconjuntos de localizações espaciais com formato elíptico. Três importantes características observadas nesse novo método são: 1) A utilização de uma janela temporal w que permite a estimação, de qualquer parâmetro de interesse, usando dados históricos pertencentes ao conjunto de tempos $t = 1, 2, \dots, t - w$; 2) Uma abordagem adaptativa no espaço dos *cluster* que permite tornar o algoritmo de varredura no espaço-tempo "relativamente" rápido. 3) ajusta automaticamente o limite de controle para a função de alarme mantendo constante a probabilidade de um falso alarme e, ainda, tornando o atraso na detecção muito pequeno.

Ao incorporarmos na análise, através de distribuições *a priori*, a incerteza na localização espacial e no ponto de mudança no tempo combinada com algoritmos para detecção de *cluster* espacial, nós realizamos um monitoramento no espaço-tempo baseada na probabilidade *a posteriori* conjunta de que tenha ocorrido uma mudança no padrão de eventos de um determinado

fenômeno em estudo em uma sub-região do mapa a um tempo desconhecido. Neste caso foi usada como função de alarme, a máxima probabilidade a *posteriori* conjunta (MAP), que torna mais simples a interpretação da abordagem adaptativa em termos de probabilidades a *posteriori*. Mostramos que esta abordagem minimiza uma perda esperada a *posteriori* representada por uma combinação de uma perda por um falso alarme e uma falsa localização. Desenvolvemos uma estrutura geral de distribuições a *priori* para a parte espacial do *cluster* que levam em consideração o tamanho, forma e direção do *cluster*.

Embora analiticamente, tenha sido impossível obter propriedades dos métodos desenvolvidos no que tange ao atraso e precisão na detecção, empiricamente, através de alguns simples estudo com dados artificiais, obtivemos resultados altamente satisfatórios, que podem ser ratificados ou não através de estudos comparativos com outros métodos. Estes estudos comparativos não foram abordados aqui devido à algumas dificuldades técnicas. Uma importante observação nos resultados e simulações é que a performance dos métodos propostos sugerem que a abordagem adaptativa na classe dos subconjuntos de localizações espaciais é muito efetiva em um monitoramento simultâneo no espaço-tempo.

Referências Bibliográficas

- Assunção, R., Corrêa, T (2009). Surveillance to detect emerging space-time clusters *Computational Statistics and Data Analysis*, **53**, 2817-2830.
- Ailamaki, A., Faloutsos, C., Fischbeck, P (2003). An environmental sensor network to determine drinking water quality and security *SIGMOD Record*, **32**, 47-52.
- Banerjee S, Carlin B.P, Gelfand A.E (2004). Hierarchical Modeling and Analysis for Spatial Data. *Chapman Hall, Boca Raton, FL*
- Besag, J. and Newell. J (1991). The detection of cluster in rare diseases. *J.R. Statist. Soc. A*, **154**, 143-155.
- Borovkov, A.A., (1998). Asymptotically optimal solutions in the change-point problem. *Theory Probab. Appl*, **43**, 539-561.
- Brown,M. (2008). Monitoring a Poisson process in several categories subject to changes in the arrival rates *Statistics and Probability Letters* **78**,2637-2643.
- Brown,M. (2008). Bayesian Detection of Changes of a Poisson Process Monitored at Discrete Time Points Where the Arrival Rates are Unknown *Sequential Analysis* **27**,68-77.
- Cuzick, J. and Edwards, R (1990). Spatial clustering for inhomogeneous populations. *Journal of Royal Statistical Society. B*, **52**, 73-104.
- Duczmal, L.H., Assunção, R.M., (2004). A simulated annealing strategy for the detection of arbitrary shaped spatial clusters. *Computational Statistics and Data Analysis*, **45**, 269-286.
- Duczmal, L., Cançado, A.L., Takahashi, R.H., Bessegato, L.F. (2007). A genetic algorithm for irregularly shaped spatial scan statistics. *Computational Statistics and Data Analysis* **52**, 43-52.
- Dwass, M., (1957). Modified randomization test for nonparametric hypotheses. *Ann. Math. Statist*, **28**, 181-187.
- Frisén, M. and Wessman, P. (1999). Evaluations of likelihood ratio methods for surveillance: differences and robustness *Communs Statist. Simuln Computn* **28**, 597-622.
- Gangnon, R.D., Clayton, M. K. (2007). Cluster detection using Bayes factors from overparameterized cluster models. *Environ Ecol Stat*, **14**, 69-82.
- Gutierrez, P. E, and Smith A.F.M, (1995). Conjugate Parametrizations for Natural Exponential Families. *Journal of the American Statistical Association* **90**, 1347-1356.
- Kass, R.E and Adrian, E.R. (1995). Bayes Factors *Journal of the American Statistical Association* **90**,430, 773-795.
- Knorr-Held, L., Raßer G., (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, **56**, 13-21.
- Kulldorff, M., (1997) A spatial scan statistic. *Communs Statist. Theory Meth.*, **26**, 1481-1496.

- Kulldorff, M., (2001) Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of Royal Statistical Society, A.*, **164**, 61-72.
- Kulldorff, M., Huang, Lan., Pickle, Linda., Duczmal, L. (2006). An Elliptic Spatial Scan Statistic. *Statistics In Medicine*, **25**, 3929-3943.
- Gorr, W., Harries, R. (2003). Introduction to crime forecasting. *International Journal of Forecasting*, **19**, 551-555.
- Imbiriba, E.N.B, Neto,A.L.S, Souza, W.V, Pedrosa, V., Cunha, M.G, Garnelo, L. (2009). Social inequality, urban growth and leprosy in Manaus: a Spatial approach. *Rev Saude Publica*, **43**,4, 656-665.
- Lai, T.L., (1995). Sequential change point detection in quality control and dynamical systems. *Journal of Royal Statistical Society, A.*, **57**, 613-658.
- Lai, T.L., (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory.*, **44**, 2917-2929.
- Lima, M.S, Duczmal, L.H, (2009) Endemic Disease Surveillance Using Bayes Factor. *International Society for Disease Surveillance Eighth Annual Conference, Miami, EUA*
- Lima, M.S, Duczmal, L.H., (2011). Surveillance and Detection of Space-Time Clusters using Adaptive Bayes Factor. *in: Cancer Clusters. Editor: Bradley D. Germanno, Series: Cancer Etiology, Diagnosis and Treatments, Nova Science Publishers, ISBN: 978-1-61209-516-5*
- Lima, M.S, Duczmal, L.H, (2011) Adaptive Likelihood Ratio Approaches for the Detection of Space-Time Disease Clusters *submitted*
- Lorden, G., Pollak M., (2005). Nonanticipating estimation applied to sequential analysis and changepoint detection. *Annals of Statistics*, **33**, 1422-1454.
- Marshall, J.B., Spitzner, D.J., Woodall, W.H., (2007). Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time. *Statistics in Medicine.*, **26**, 1576-1593.
- Migon, H.S. and Gamermam, D. (1999). Statistical Inference: An Integrated Approach. *Hodder Arnold, London*
- Morris, C.N (1982). Natural Exponential Families with Quadratic Variance Functions. *Annals of Statistics* ., **10**, 65-80.
- Neill, D.B, Moore, A.W., Sabhnani, M.R., (2005). Detecting elongated disease clusters. Morbidity and Mortality Weekly *Supplement on Syndromic Surveillance, Report*, **54**, 197.
- Neill, D.B, (2006). Detection of Spatial and Spatio-Temporal Clusters. Ph.D. Thesis, *University Pittsburgh*.
- Neill, D.B, (2009). Expectation-based scan statistics for monitoring spatial time. *International Journal of Forecasting* **25**, 498-517.
- Openshaw, S., Craft, A.W., Charlton, M., and Birch, J.M., (1988). Investigation of leukaemia cluster by use of a geographical analysis machine. *Lancet*, **1**, 272-273.
- Page, E. S., (1954). Continuous inspection schemes. *Biometrika*, **41**, 100-115.
- Patil, G.P., Taillie, C., (2003). Geographic and network surveillance via scan statistics for critical area detection. *Statistical Science*, **18**, 457-465.

- Pavlov, I.V., (2003). Sequential Procedure of Testing Composite Hypotheses with Applications to the Kiefer-Weiss. *Theory Prob. Appl.*, **35**, 280-292.
- Pollak, M., Siegmund, D., (1975) Aproximations to the expected sample size of certain sequential test. *Annals of Statistics*, **3**, 1267-1282.
- Pollak, M. (1985) Optimal detection of a change in distribution. *Annals of Statistics*, **13**, 206-227.
- Pollak, M. (1987) Average run lengths of an optimal method of detecting a change in distribution. *Annals of Statistics*, **15**, 749-779.
- Porter, D.M., (2007) Some Adaptive approaches for space-time anomaly detection. *First International Workshop in Sequential Methodologies, IWSM*.
- Roberts, S. W. (1966) A comparison of some control chart procedures. *Technometrics.*, **8** 411-430.
- Sabhnani, M., Neill, D.B., Moore, A. W., Tsui, F.C., (2005). Detecting anomalous clusters in pharmacy retail data. In *Proceedings of the KDD, workshop on data mining methods for anomaly detection 2005*, 58-61.
- Shiryayev, A.N., (1986). On optimum methods in quickest detection problems. *Theory Probab. Appl.*, **8**, 22-46.
- Shiryayev, A.N., (1978). Optimal Stopping Rules. In: *Springer-Verlag, New York*
- Sonesson, C., Bock, D., (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of Royal Statistical Society, A.*, **166**, 5-21.
- Sonesson, C., (2007). A CUSUM framework for detection of space time disease clusters using scan statistic. *Statistics in Medicine.*, **26**, 4770-4789.
- Tartakovsky, A.G., Rozovskii, L.R., Blazek, R.B., Kim, H. (2006). Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology.*, **3**, 252-293 .
- Tango, T. (1995). A class of tests for detecting general and focused clustering of rare diseases. *Statistics in Medicine.*, **14**, 2323-2334.
- Tango, T., Takahashi, K., (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geography*, **4**, 4-11.
- Tango, T., Takahashi, K., Kohriyama, Kazuaki. (2010). A Space-Time Scan Statistic for Detecting Emerging Outbreaks *Biometrics*, **4**, 1-10.
- Tuia, D., Lasaponarab, R., Telescab, L., Kanevskia, M., (2008). Emergence of spatio-temporal patterns in forest-fire sequences. *Physica A*, **387**, 3271-3280.
- Turnbull, B.W., Iwano, E.J., Burnett, W.S., HOWE, H.L, and Clark, L.L (1990). Monitoring for Cluster of disease: Application to leukemia incidence in upstate New York. *American J. epidem*, **88**, **132**, S136-S143.
- Walter, S.D. (1994). A simple test for spatial patten in regional health data. *Statistics in Medicine.*, **13**, 1037-1044.
- West, M., (1986). Bayesian Model monitoring. *Journal of Royal Statistical Society, B.*, **48**, 70-78.

West, M. and Harrison, P.J. (1986). Monitoring and Adaptation in Bayesian Forecasting Models *Journal of the American Statistical Association* **81**,395, 741-750.

West, M. and Harrison, P.J. (1997). Bayesian Forecasting and Dynamic Models *Springer-Verlag New York, Second Edition*

Zhou, H. and Lawson, A.B, (2008). EWMA smoothing and Bayesian spatial modeling for health surveillance *Statistics in Medicine* **48**

Ross,S.M (1983). Stochastic Processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. *John Wiley Sons Inc., New York, 1983. Lectures in Mathematics, 14.*