

Abordagem Bayesiana adaptativa para Vigilância online de Clusters Espaciais

Max Sousa de Lima ¹
Luiz Henrique Duczmal ²

Resumo

Neste trabalho, nós desenvolvemos um método Bayesiano para modelagem, detecção e vigilância estatística de mudanças no padrão de um processo estocástico no espaço-tempo. O método é baseado em um processo a *posteriori* adaptativo e o foco das aplicações é a vigilância online de *cluster* espaciais de doenças onde a classe dos possíveis candidatos a *cluster* é excessivamente grande o que pode dificultar a performance de qualquer método desenvolvido com este propósito. O método é ilustrado usando a dados de câncer da tireóide em homens que ocorreram no Novo México durante o período de tempo 1973-1992.

palavra chave: cluster espaciais, posteriori adaptativa, vigilância estatística

1 Introdução

Métodos estatísticos para vigilância epidemiológica são importantes para a detecção online de clusters espaciais de eventos, os quais podem indicar uma rápida mudança no padrão espaço-temporal dos casos observados da doença. Estatisticamente, este problema é descrito da seguinte forma (Sonesson, 2007): Denote por \mathcal{S} uma região de estudo particionado em L -áreas A_l contíguas identificadas por um ponto s_l . Considere sobre \mathcal{S} a vigilância estatística de um processo estocástico $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots$ 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 ξ um conjunto conexo de localizações representando um candidato a cluster espacial 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 vigilância é

¹Dep. de Estatística -UFAM

²Dep. de Estatística -UFMG

detectar um *cluster* que emergiu em um tempo desconhecido $k \leq n$ e que está presente até o estágio atual n . O *cluster* detectado no espaço-tempo é expresso em função de uma mudança no padrão da distribuição do sub-processo $\{X_t(s_l) \in \mathbf{X} : s_l \in \xi, t \geq k\}$. Neste paper, abordamos o problema de vigilância online de *clusters* espaciais do ponto de vista Bayesiano, 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 espaciais ξ no tempo $\nu = k$. Ou seja, avaliamos $\Pi_{n,k}(j) = \mathbb{P}(\{\xi = \xi_j, \nu = k | \mathcal{H}_n\})$.

2 Modelagem Bayesiana do Cluster

Para a detecção do *cluster* e estimação do par (ν, ξ) considere uma abordagem bayesiana hierárquica construída da seguinte forma:

1. $(\mathbf{X}_t | \boldsymbol{\theta}_t)$ 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 $\{\boldsymbol{\theta}_t\}_{t \geq 1}$ os componentes de \mathbf{X}_t são condicionalmente independentes.
2. $(\boldsymbol{\theta}_t | \nu, \xi)$ 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. $G_0(\boldsymbol{\theta}_t)$ é uma distribuição de referência e neste modelo depois da mudança no tempo $\nu = k$, os componentes de $\boldsymbol{\theta}_t$ tais que $s_l \notin \xi_j$ possuem distribuição 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 .
3. 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 $j = 1, 2, \dots, J$. A distribuição a priori para ξ é não informativa no espaço dos candidatos a clusters Ξ , $\pi_\xi(j) = 1/J$. A distribuição para ν é geométrica(0.5) truncada em n_0 , $\pi_\nu(k) = (1 - 0.5^{1/n-n_0+1}) 0.5^{k-n_0/(n-n_0+1)}$. n_0 é o menor momento de interesse na mudança do padrão geográfico da doença e $p = 0.5$ implica que no tempo n a probabilidade a *priori* que não tenha ocorrido uma mudança é a mesma que a mudança tenha ocorrido.
4. A *priori* para $\theta_t(s_l)$ é $G_0(\theta_t(s_l)) \sim Weibull(2, 1.414)$ que fornece *Moda* = 1 e pouca variação. Os $\{\theta_t(s_l)\}$ são assumidos independentes com distribuição comum, pois é razoável supor que, se não existem mudanças

no processo sob vigilância, os $\{\theta_t(s_l)\}$ surgem de forma aleatória a partir de uma distribuição comum. Então, quando o processo esta sob-controle, pode-se interpretar a $\text{Moda}(\theta_t(s_l)) = 1$ como sendo o valor mais provável para o risco relativo de ocorrência de um caso da doença em uma área escolhida ao acaso no mapa.

Desta forma a distribuição a *posteriori* conjunta de (ν, ξ) é dada por

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

onde,

$$FB_t(j) = \prod_{s_l \in \xi_j} \frac{\int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]^{m_{t,j}}}{\int f(X_t(s_l) | \theta_t(s_l)) d[G_0(\theta_t(s_l))]} , t \geq k \quad (1)$$

depende somente do sub-processo $\mathbf{X}_t(j) = \{X_t(s_l) : t \geq k, s_l \in \xi_j\}$.

3 Vigilância online de clusters espaciais usando processos a posteriori adaptativo

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

$$\Pi_n(0) = \mathbb{P}(\nu > n | \mathcal{H}_n) \quad \text{e} \quad \Pi_n(j) = \mathbb{P}(n_0 \leq \nu \leq n, \xi = \xi_j | \mathcal{H}_n).$$

$\Pi_n(0)$ é a probabilidade de que a mudança não tenha ocorrido antes ou até o estágio atual n e $\Pi_n(j)$ 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

$$\Pi_n(j) = \sum_{k=n_0}^n \pi(\nu = k, \xi = \xi_j | \mathcal{H}_n) \quad \text{e} \quad \Pi_n(0) = \sum_{j=1}^J \sum_{k=n+1}^\infty \pi(\nu = k, \xi = \xi_j | \mathcal{H}_n). \quad (2)$$

Definimos como estratégia para detecção online de *cluster* espaciais o procedimento com tempo de alarme baseado na Máxima Probabilidade a Posteriori, $\hat{\tau} = \min\{n \geq n_0 : \max_{j \in \mathcal{J}} [\Pi_n(j)] > 0.5\}$. A parte espacial estimada

do cluster é $\hat{\xi} = \xi_{\hat{j}}$. Onde $\hat{j} = \arg \left(\max_{j \in \mathcal{J}} [\Pi_n(j)] \right)$. O tempo k estimado é $\hat{k} = \arg \left[\max_{n_0 \leq k \leq \hat{\tau}} \pi(\nu = k, \xi = \xi_{\hat{j}} | \mathcal{H}_{\hat{\tau}}) \right]$ pois,

$$\Pi_{\hat{\tau}}(\hat{j}) = \sum_{k=n_0}^{\hat{\tau}} \pi(\nu = k, \xi = \xi_{\hat{j}} | \mathcal{H}_{\hat{\tau}})$$

e grandes valores de $\pi(\nu = k, \xi = \xi_{\hat{j}} | \mathcal{H}_{\hat{\tau}})$ possuem maior contribuição para que o alarme dispare. O processo adaptativo é construído substituindo $\Xi = \{\xi_j, j = 1, 2, \dots, J\}$, por $\Xi^{(n)} = \{\xi_j^{(n-1)}, j = 1, 2, \dots, J^{(n)}\}$ que representa a classe mais plausível de candidatos a cluster no tempo anterior $n - 1$.

4 Aplicação: Detecção online de clusters espaciais de doenças

Para ilustrar o método proposto avaliamos o conjunto de casos de câncer da tireóide em homens que ocorreram no Novo México no período de tempo 1973-1992 (Kulldorff, 2001; Lima e Duczmal, 2011) Assumimos que o número de casos $(X_t(s_l) | \theta_t(s_l)) \sim Poisson(e_t(s_l) \times \theta_t(s_l))$ com $e_t(s_l)$ sendo a contagem esperada da doença no tempo t em s_l . A quantidade $\theta_t(s_l)$ é uma medida de risco relativo de ocorrência de um caso de câncer da tireóide no tempo t em A_l . Os resultados na figura 1 apresenta a função de alarme proposta para $m=2, 4, m=8$ e também o tempo, a região e as observações que provocaram a mudança no padrão espaço-temporal do processo. 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 $k = 18$.

5 Considerações Finais

Neste trabalho desenvolvemos um novo método para detecção e monitoramento de *clusters* emergentes. Os Resultados obtidos em nossa aplicação mostrou que a abordagem proposta é muito efetiva em uma vigilância epidemiológica simultânea no espaço-tempo. Analisando este mesmo conjunto

de dados, a estatística scan espaço-tempo de Kulldorff (2001) detectou um cluster emergente com valor- $p=0.016$. Lima e Duczmal (2013) detectam um *cluster* emergente 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.

Referências

- [1] Kulldorff, M., (2001) Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of Royal Statistical Society, A.*, **164**, 61-72.
- [2] Lima, M.S, Duczmal, L.H, (2013) Adaptive Likelihood Ratio Approaches for the Detection of Space-Time Disease Clusters *submitted*
- [3] Sonesson, C., (2007). A CUSUM framework for detection of space time disease clusters using scan statistic. *Statistics in Medicine.*, **26**, 4770-4789.

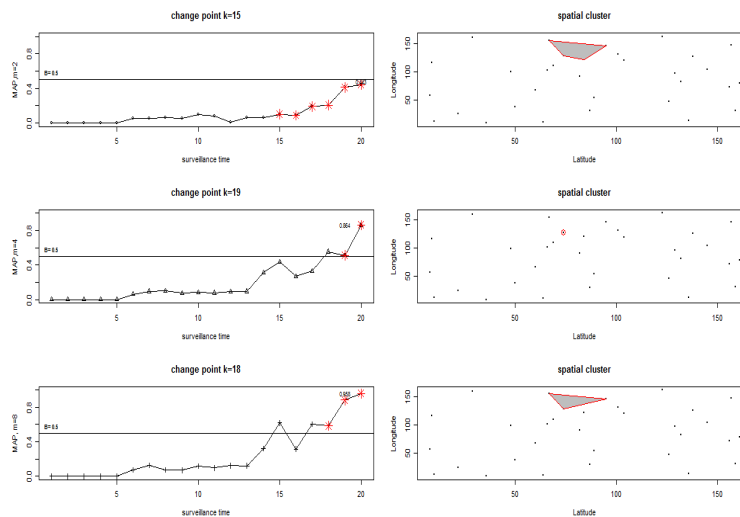


Figura 1: A esquerda: Máxima probabilidade a posteriori adaptativa (MAP) para $m=2,4,8$ e ponto de mudança; A direita: região da mudança.