

Cadeias de Markov Ocultas aplicadas a notificações de Dengue no município de Lavras-MG

Deive Ciro de Oliveira ^{1 2 3}

Rodrigo Ferreira Abreu ^{2 3}

Rafael Agostinho Ferreira ^{2 3}

João Domingos Scalon ^{2 3}

Resumo: A dengue é uma importante enfermidade e exige por parte de agentes públicos atuação para minimizar seus efeitos na população. Neste sentido, uma das ações de diagnósticos é determinar períodos temporais de maior gravidade da endemia para intensificação de prevenção, maior contingente de medicamentos, médicos etc. Considerada a observação de séries de notificação de casos de dengue (anualidade), as cadeias de markov com estados ocultos permitem fazer inferência sobre janelas de maior ocorrência de notificações. Neste trabalho aplicou-se cadeia de markov ocultas a dados de notificação de casos de dengue no município de Lavras-MG em 2015. São apresentados resultados de estimação dos parâmetros além da discussão da estimação das janelas temporais de maior intensidade.

Palavras-chave: *Cadeias de Markov Ocultas, Processos de Contagem, Endemia de Dengue*

1 Introdução

A dengue tem sido uma das doenças infecciosas com maior destaque nos dias atuais. O vírus da dengue pertencente a família *Flaviviridae* e sua principal forma de transmissão ocorre por meio da picada do mosquito *Aedes Aegypti* em humanos (1).

Dentre países endêmicos que mais sofrem com o problema da dengue está o Brasil. De acordo com o Ministério da Saúde, foram notificados, no ano de 2015, 1.688.688 casos prováveis de dengue. Em 2016, esse número foi reduzido para 1.500.535 casos (2).

De acordo com estudos na área, o padrão temporal da doença revela um comportamento sazonal durante o ano. Em geral, entre os meses de Fevereiro e Junho, a frequência de casos notificados tendem a aumentar se comparado à outras épocas do ano.

¹ICSA - UNIFAL-MG: Universidade Federal de Alfenas;

²DES; UFLA: Universidade Federal de Lavras. Email: *deive.oliveira@unifal-mg.edu.br*

³Agradecimento a FAPEMIG e a CAPES pelo suporte financeiro

Assim, é necessário que haja ferramentas de combate e prevenção desta doença para que estes números possam ser reduzidos. Além da conscientização por parte da população, métodos e técnicas estatísticas podem ser utilizadas como ferramentas para prevenção a fim de que se possa conhecer melhor o padrão da doença.

Portanto, o objetivo deste trabalho é modelar os casos notificados de dengue no município de Lavras - MG, em 2015, utilizando Cadeias de Markov Ocultas.

2 Material e Métodos

2.1 Metodologia

Modelos de Markov com estados ocultos especificam dois processos estocásticos correlacionados $\{S_t, O_t \mid t \in \Gamma\}$. O conjunto Γ identifica a indexação do processo, a qual neste trabalho esta relacionada ao tempo e é discreta. A peculiaridade é que o processo $\{S_t\}$ por definição, segue uma cadeia de markov e não é observável diretamente. A expressão de $\{S_t\}$ é dada por $\{O_t\}$. Por pressuposição do modelo, o chamado processo observável $\{O_t\}$ é condicionalmente dependente de $\{S_t\}$. Dada a especificidade do problema, deseja-se inferir sobre obter a dinâmica de dependência na parcela não observável do processo estabelecida pela Cadeia de Markov (distribuição de $\{S_t|S_{t-1}\}$ no caso de processos de indexação discreta). Ademais, interessa-se pela dependência entre os processos observável e não observável (distribuição de $\{O_t|S_t\}$). O modelo será detalhado a seguir.

2.1.1 Cadeias de Markov

O princípio elementar de uma Cadeia de Markov é que "o presente depende de um passado restrito". Aplicando este princípio na modelagem de um processo $\{S_t|t \in \Gamma\}$ e considerando Ω_{S_t} o suporte de S_t , temos:

$$\begin{aligned} \pi_{1_{s_1}} &= P(S_1 = s_1) \text{ (Distribuição Inicial),} \\ A_{[s_{t-1}, s_t]} &= P(S_t = s_t | S_{t-1} = s_{t-1}) \text{ (Matriz de transição).} \end{aligned} \tag{1}$$

Como a Cadeia tratada é estacionária, temos que $A_{[s_{t-1}, s_t]}$ independem de t e podem ser reduzidas a notação $A_{[i,j]} = P(S_t = j | S_{t-1} = i)$ com $i, j \in \Omega_{S_t}$. O Conjunto de Ω_{S_t} é chamado de conjuntos de estados da cadeia.

2.1.2 Processos de Contagem

Para modelar o processo estocástico condicionado $\{O_t|S_t\}$ utiliza-se distribuições adequadas para natureza de O_t . No caso particular da aplicação considera-se que a variável aleatória $O_t|S_t = i \sim \text{Poisson}(\lambda_i)$ com $i \in \Omega_{S_t}$. Denotaremos por $\Lambda = \lambda_i$ como o vetor de parâmetros para distribuição condicional $\{O_t|S_t = i\}$.

2.1.3 Cadeias de Markov Ocultas

Uma cadeia de markov com estados ocultos $\theta = (\pi, A, \Lambda)$ (3) é composta de 3 entidades paramétricas associadas a distribuição inicial π , a matriz de transição A e o vetor Λ associados a taxas de contagens condicionadas aos estados da cadeia oculta. Note que a observação do processo somente se dá por $\{o_t\}$.

Com base na observação do processo $\{o_t\}$, pode-se obter a função de log-verossimilhança da cadeia oculta $L(\theta|o_t)$. Via o método iterativo *EM* (4) (Expectation-Maximization Algorithm) obtém-se estimativas de máxima verossimilhança $\hat{\theta} = \arg \max_{\theta \in \Omega_\theta} L(\theta|o_t)$.

Além disso, para a escolha do melhor modelo, pode-se utilizar o Critério de Informação de Akaike (AIC), dado por $AIC = -2 \log L(\theta|o_t) + 2\theta$.

2.2 Dados

Os dados analisados englobam notificações de casos de dengue no período de 01/01/2015 a 30/11/2015 no município de Lavras (Minas Gerais), que totalizam 6326 notificações. São dados com periodicidade semanal. Assim $\{o_t\}$ denota o número de notificações da casos de dengue na t -ésima semana.

3 Resultados e Discussão

Por meio do histograma das notificações registradas no município (Figura 1), observa-se um número elevado de ocorrências entre os meses de Fevereiro e Junho de 2015. Neste período, observa-se, também, picos nos meses de março, abril e maio.

Para a modelagem, foram considerados ajustes de modelos com 2, 3, 4 e 5 estados ocultos. Pelo critério AIC, pôde-se observar, conforme a Tabela 1, que o modelo com 5 estados ocultos apresentou um melhor ajuste aos dados.

Dentre os modelos ajustados, o que melhor se adequou a partir deste critério foi o modelo utilizando 5 estados ocultos.

No entanto, vale ressaltar, para este caso, que a medida em que se aumenta o número

Figura 1: Histograma do número de notificações semanais de casos de dengue ao longo de 2015.

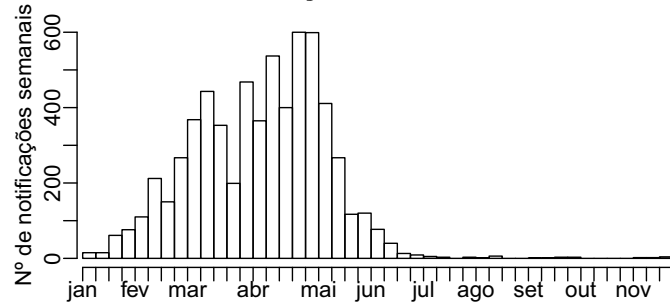


Tabela 1: Seleção de modelos

Modelos	θ	$\log L$	AIC
2 estados ocultos	2	-1225.707	2455.414
3 estados ocultos	6	-505.0046	1022.0092
4 estados ocultos	12	-338.9973	701.9946
5 estados ocultos	20	-317.9129	675.8258

de estados ocultos da cadeia, o valor do AIC tende a diminuir. Uma justificativa para este comportamento pode ser devido ao fato de que não houve uma mudança brusca na taxa λ_i (Tabela 2) ao longo das semanas, conforme pode-se notar no histograma da Figura 1.

Tabela 2: Estimativas dos parâmetros da distribuição condicional $\{O_t|S_t = i\}$.

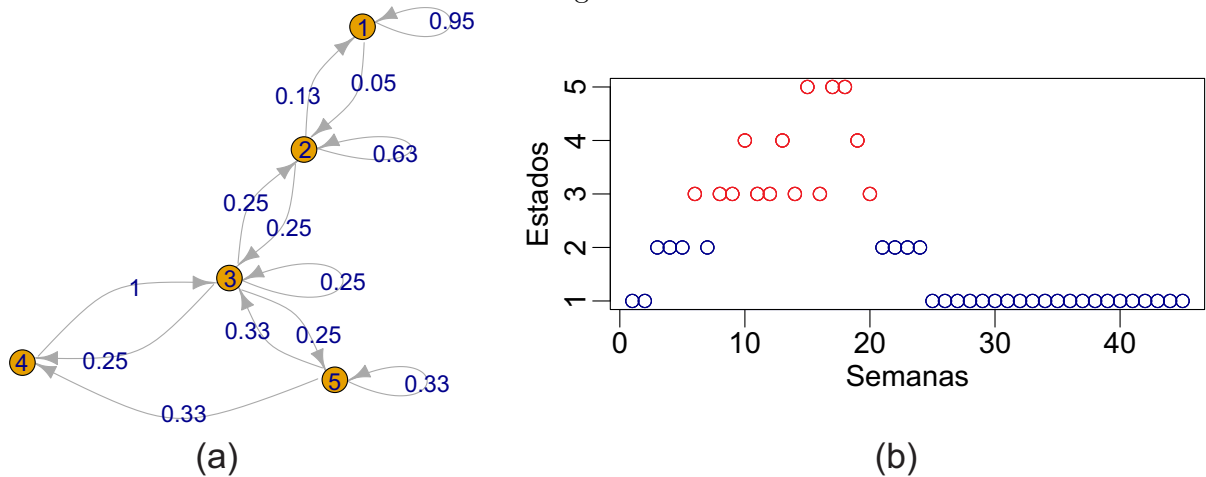
i	λ_i
1	3,74
2	93,88
3	303,87
4	440,69
5	578,68

Na estimação do modelo com 5 estados, as probabilidades de transição entre os estados da cadeia oculta estão representadas em forma de grafo na Figura 2(a).

Assim, pode-se atribuir aos estados 1 e 2 níveis mais baixos de notificações da dengue, ao passo que, os estados 3, 4 e 5 representam níveis mais elevados, o que indicam possíveis surtos da doença.

A partir do modelo $\hat{\theta}$ estimado é possível aplicar o algoritmo de Viterbi que obtém a sequência mais verossímil de estados $\{s_t^*\}$, assumindo o conjunto de observações $\{o_t\}$. ($\{s_t^*\} = \arg \max_{s_t \in \Omega_{\{s_t\}}^T} L(\{s_t\}|\{o_t\})$). A sequência de estados $\{s_t^*\}$ é apresentada na Figura 2(b).

Figura 2:



4 Conclusão

A utilização de cadeias de markov ocultas é uma ferramenta adequada na estimação de janelas endêmicas, sobretudo quando se evidencia mudanças relevantes de comportamentos no conjunto de dados em estudo. No caso específico da aplicação, considerado o modelo de 5 estados e arbitrariamente os estados 1 e 2 como moderados e 3, 4 e 5 como estados endêmicos pode-se identificar a janela endêmica entre a 6ª semana e a 20ª semana.

A estimação de janelas endêmicas podem ser informação muito importante no planejamento de ações de combate a doença que enfatizem o período de maior incidência da doenças.

A consideração de distintos padrões de taxas de contagem em estados distintos podem abrir a possibilidade da utilização de famílias de distribuições diferenciadas por estado ou mesmo processos não homogêneos intra estados.

Referências

- 1 GUBLER, D. J. Dengue and Dengue Hemorrhagic Fever. **Clinical Microbiology Reviews**, v. 11, n. 03, p. 480 – 496, jul 1998.
- 2 MINISTÉRIO DA SAÚDE. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 4, 2017. **Boletim Epidemiológico**, v. 48, n. 5, 2017.
- 3 RABINER, L. R. A tutorial on hidden markov models and selected applications in speech recognition. **Proceedings of the IEEE**, v. 77, n. 2, p. 257–286, Feb 1989. ISSN 0018-9219.
- 4 DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Wiley], v. 39, n. 1, p. 1–38, 1977. ISSN 00359246. Disponível em: <http://www.jstor.org/stable/2984875>.