

## Aplicação de HMMs utilizando dados sobre COVID-19 para estimar o estado de subnotificação de casos

Lucas Rebouças Lima<sup>1</sup>

UFSJ, Ouro Branco, MG

Telles Timóteo da Silva<sup>2</sup>

DEFIM/UFSJ, Ouro Branco, MG

Em 26 de fevereiro de 2020, o primeiro caso de SARS-CoV-2 (novo Coronavírus) chegava ao Brasil. Em 11 de março do mesmo ano, a Organização Mundial da Saúde (OMS) reconheceu as características de dispersão da COVID-19 como se tratando de uma pandemia, e o Brasil viria a ser um dos epicentros globais para a disseminação desta doença.

Os dados que servem como base para o monitoramento da doença foram frequentemente questionados devido à possibilidade de subnotificação por conta da baixa testagem em território nacional [2]. Neste sentido, o trabalho aqui proposto procura fornecer um método para estimar a subnotificação dos casos. A técnica se baseia na aplicação de “Hidden Markov Models” (Cadeias Ocultas de Markov ou HMM’s), um modelo estocástico capaz de analisar cenários cujo próximo estado possível depende apenas do estado atual [3].

Um HMM constitui-se de (i) uma cadeia de Markov, cujos estados não são diretamente observados e de (ii) uma seqüência de observações associada a cada estado [4,5]. O objetivo da pesquisa consiste em se estimar o estado em que se encontra o processo a partir dos dados observados para a doença. A hipótese subjacente é que haja incerteza nas notificações, sendo que a real dimensão da expansão desta doença esteja oculta devido à subnotificação.

O ponto de partida do trabalho é a realização de um estudo exploratório dos dados. Nas Figuras 1 e 2 encontram-se, respectivamente, gráficos das notificações de infecções e óbitos associados à doença na cidade de São Paulo - SP no período de 25 de fevereiro de 2020 a 14 de março de 2021.

Juntamente com a análise da variação temporal do processo, a análise das frequências do número de notificações e do número de óbitos permite definir as classes em que se encontram as observações para estipulação do HMM adequado aos dados. O padrão que se percebe para o período de 25 de fevereiro de 2020 a 14 de março de 2021 é de alta frequência para maior volume de notificações diárias e poucos óbitos, e uma baixa frequência para menor volume de notificações diárias e muitos óbitos. No caso de notificações, temos em torno de 14 classes relevantes no período:  $N_i = ]\frac{2750}{14}(i-1), \frac{2750}{14}i]$ , para  $i = 1, \dots, 14$ . No caso de óbitos, 7 classes:  $O_i = ]25(i-1), 25i]$ ,  $i = 1, \dots, 7$ .

Tendo em vista os estados ocultos a serem especificados, utilizaremos os dados observados para otimizar as matrizes que compõem o HMM. Com o HMM otimizado, poderemos fazer uma verificação inicial para validar o modelo, investigando se simulações computacionais reproduzem o padrão observado nas frequências absolutas de médias móveis. Este será um indicador importante da qualidade do modelo e da utilidade que ele terá para produzir inferências. Outro indicador importante é poder reproduzir o padrão de distribuição dos pares ordenados  $(o_t, n_{t-d})$ , onde  $o_t$  representa óbitos e  $n_t$ , notificações,  $t = 1, \dots, 384$ . O valor  $d$  representa o tempo médio decorrido

<sup>1</sup>lucasreboucas01@aluno.ufsj.edu.br, bolsista de iniciação científica.

<sup>2</sup>timoteo@ufsj.edu.br

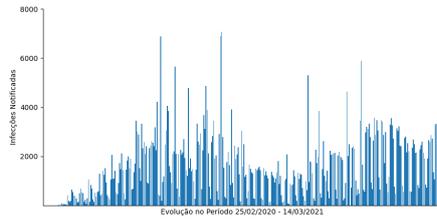


Figura 1: Casos Diários

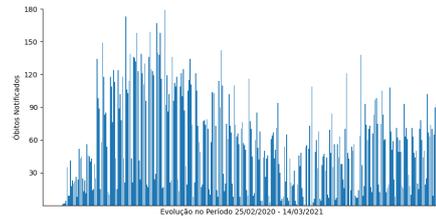


Figura 2: Óbitos Diários

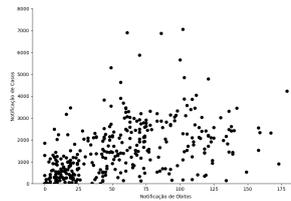


Figura 3: Comparativo Óbitos x Casos

para uma certa quantidade de infectados evoluírem a óbito. Neste trabalho vamos considerar  $d$  variando entre 7 e 21 dias [1]. A dispersão dos pontos  $(o_t, n_{t-d})$  para  $d = 7$  está ilustrada na Figura 3. Adicionalmente, poderemos simular o processo temporal das observações, comparando-o com gráficos semelhantes aos das Figuras 1 e 2.

## Agradecimentos

À Universidade Federal de São João del-Rei, que disponibiliza o auxílio financeiro ao bolsista de iniciação científica por meio de políticas de incentivo à pesquisa e desenvolvimento.

## Referências

- [1] ARRUDA, E.; ALEXANDRE, R.; FRAGOSO, M.; DO VAL, J.; THOMAS, S. A Novel Stochastic Epidemic Model with Application to COVID-19, arXiv:2102.08213v1 [q-bio.QM], 16 Feb 2021, 2021.
- [2] MAGNO, L.; ROSSI, T. A.; MENDONÇA-LIMA, F. W.; SANTOS, C. C.; CAMPOS, G. B.; MARQUES, L. M.; PEREIRA, M.; PRADO, B. L.; DOURADO, I. Desafios e propostas para ampliação da testagem e diagnóstico para COVID-19 no Brasil, *Ciência e Saúde Coletiva*, **25** [9], pp.3355-3364, 2020. DOI: 10.1590/1413-81232020259.17812020.
- [3] STAMP, M. A revealing introduction to hidden Markov models, *Science*, pp.1-20, 2004.
- [4] RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, **77** [2], pp.257-286, 1989. DOI: 10.1109/5.18626.
- [5] RABINER, L. e JUANG, B. *Fundamentals of speech recognition*. PTR Prentice Hall, New Jersey, 1993.