TMBD
Projeto 1: Inferência Probabilística

Cadeias de Markov Discretas aplicadas à pandemia COVID-19

Gonçalo Almeida
Martim Sousa

Dezembro 2020

1 Introdução

Este pequeno documento serve de suporte para o código produzido pelo nosso grupo, que pode ser encontrado no repositório git para o efeito. Vamos demonstrar de que forma as Cadeias de Markov podem ser utilizadas para previsão, desde que tenhamos informação necessária para estimar as probabilidades de transição entre estados. Algo que, infelizmente, não é possível para o nosso exemplo por falta de dados[1] Por fim, vamos dar resposta a certas questões com base na nossa matriz de transição. Por exemplo, quantos dias, em média, demora uma pessoa a passar de suscetível a infetada? No nosso problema, dispomos de 6 estados: S, I, H, UCI, R, O, que são: suscetível, infetado, hospitalizado, internado em unidade de cuidados intensivos, recuperado e óbito, respetivamente.

[1] Os dados utilizados para obter a matriz de probabilidades de transição são artificiais.

2 Ajuste dos dados à DTMC

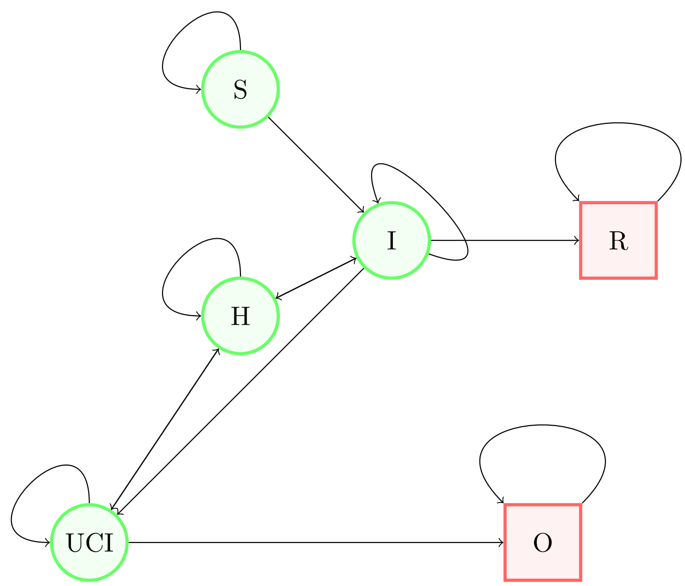
Tivemos acesso a uma base de dados, que contém uma amostra de 10000 pessoas da população portuguesa com o registo do seu estado ao longo de 1000 dias. Com recurso à função markovchainFit da linguagem R, obtivemos pelo método de máxima verosimilhança para um I.C. a 95% a seguinte matriz de transição[2]:

$$\begin{array}{c} \begin{array}{cccccc} S & I & H & UCI & O & R \end{array} \\ \begin{pmatrix} 0.99 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.80 & 0.04 & 0.01 & 0.00 & 0.15 \\ 0.00 & 0.15 & 0.80 & 0.05 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.05 & 0.80 & 0.15 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix} \end{array}$$

3 Propriedades da DTMC

A nossa Cadeia de Markov discreta pode ser visualizada através de um grafo. Os estados a verde, são transientes, i.e, existe probabilidade positiva de transitarem para outro estado, os estados a vermelho são recorrentes. O nosso problema pode ser modelado por uma Cadeia de Markov, porque o próximo estado, depende apenas do estado atual. Para além disso, é uma cadeia de Markov homogénea, porque a probabilidade de um paciente transitar do estado i para o estado j do dia n para o dia m, apenas depende do tamanho do intervalo $|m - n|$.

[2]Arredondado a 2 casas decimais



Estado	Legenda
S	Suscetível
I	Infetado
H	Hospitalizado
UCI	Internado UCI
O	Óbito
R	Recuperado

4 Previsão via DTMC

Seja P a matriz de transição, para calcularmos o número de pessoas nos diversos estados no dia k+1, sabendo o número de pessoas nos diversos estados no dia anterior k:[3]

$$X^{k+1} = P^T X^k$$

[3] X é um vetor coluna de dimensão 6 (número de estados)

Aplicando a nossa Cadeia de Markov a uma população inicial de 1.000.000 pessoas suscetíveis, ou seja:

$$X^0 = \begin{bmatrix} 1000000 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Obtemos a seguinte propagação de estados:

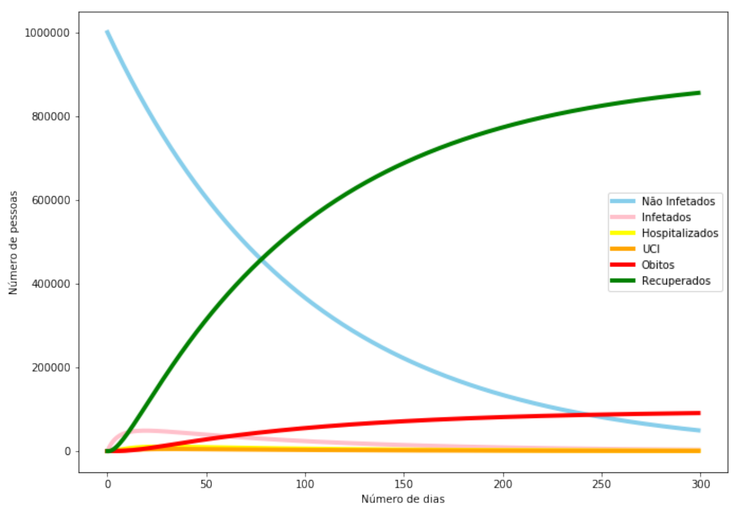


Figure 1: Evolução da Cadeia de Markov para todos os estados

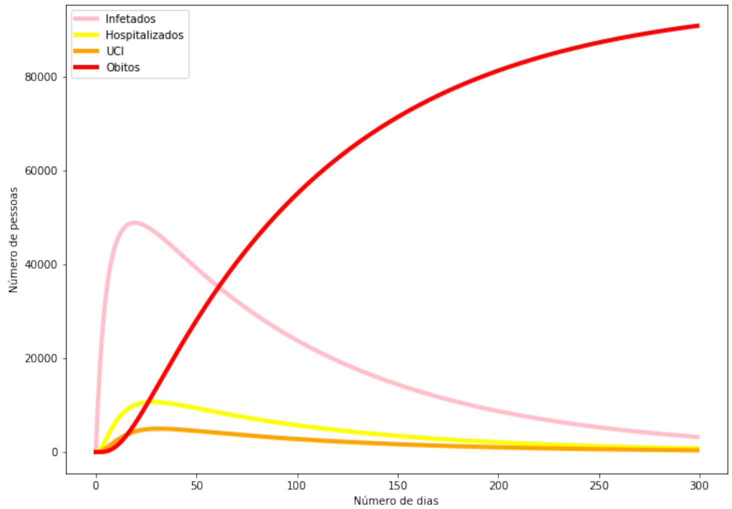


Figure 2: Evolução da Cadeia de Markov para 4 estados

Podemos constatar que ao fim de 20 dias dá-se o pico de infetados ativos, com 48.833 infetados na nossa população. O pico de hospitalizados dá-se 9 dias depois com 10.706 pessoas hospitalizadas, já o pico de internados na UCI dá-se no 33º dia com 4.983 internados em UCI. Repare-se que, embora a primeira derivada da curva de óbitos seja sempre positiva, porque a função é crescente, a segunda derivada vai decaindo progressivamente, porque os infetados vão-se dirigindo a um dos dois estados recorrentes: recuperado ou óbito.

5 Miscelânea de questões associadas à DTMC

Calculando a distribuição estacionária[4]:

$$\hat{\Pi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.096 & 0.904 \end{bmatrix}$$

Retomando o exemplo anterior de 1.000.000 suscetíveis, verificamos que ao fim de 100.000 dias temos 903.614 óbitos e 96.385 recuperados, o que verifica os valores da distribuição estacionária. Relembremos a nossa Matriz transição:

[4]Arredondado a 3 casas decimais

$$
\begin{array}{cccccc}
S & I & H & UCI & O & R \\
\end{array}
$$
$$
\begin{pmatrix}
0.99 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.80 & 0.04 & 0.01 & 0.00 & 0.15 \\
0.00 & 0.15 & 0.80 & 0.05 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.05 & 0.80 & 0.15 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00
\end{pmatrix}
$$

Estamos interessados em calcular as seguintes probabilidades:

1. P(Recuperar no primeiro dia após ser infetado)
2. P(Infetado recuperar ao fim de N dias)
3. P(um infetado recuperar em menos de N dias sem ser hospitalizado)
4. P(um infetado passar menos de N dias hospitalizado)
5. P(um infetado ou hospitalizado, passar apenas os 2 dias seguintes em UCI)
6. P(morrer em menos de N dias, sabendo que estava infetado ou hospitalizado)

Para N=10, obtivemos os seguintes resultados com recurso ao módulo Pykov da linguagem Python:

probabilidade	valor
1	0.15
2	0.02
3	0.67
4	0.17
5	0.01
6	0.03

Através do método de Monte Carlo, podemos estimar o número médio de dias de demora de ir de um estado i para um estado j. A tabela infra, contempla alguns desses resultados:

transição	número médio de dias
S$\longrightarrow I$	100
I$\longrightarrow R$	7
H $\longrightarrow I$	5
H $\longrightarrow UCI$	20
H $\longrightarrow R$	10
UCI $\longrightarrow O$	7

6 Notas Finais

O objetivo primário deste trabalho era utilizar Cadeias de Markov Discretas para modelar a progressão diária da pandemia de COVID-19 numa determinada região ou país, por exemplo, Portugal. Todavia, não dispomos de dados suficientes que nos permitam estimar a matriz de transições, à semelhança do que foi feito na secção 2 para dados artificiais. Apenas são divulgados os números diários de confirmados, casos ativos, recuperados, internados, internados uci, óbitos, sem fazer alusão a quantas pessoas passaram de um destes estados para outro. Ainda assim, julgamos que este breve texto demonstra de que forma as Cadeias de Markov Discretas podem também ser utilizadas para problemas normalmente associados à área de Machine Learning.