Trabalho 1

Modelos Compartimentais em Epidemiologia e Inferência Bayesiana

18 de janeiro de 2022

Contexto

Sabe-se que o número de indivíduos infectados em função do tempo no estágio inicial de uma epidemia segue um comportamento exponencial crescente. Seja t o tempo (em dias) e $I(t)$ a média móvel do número diário de novos casos. Os dados mostrados na Fig. 1 retratam os 40 primeiros dias da epidemia em determinada localidade, desde que o primeiro caso foi diagnosticado.

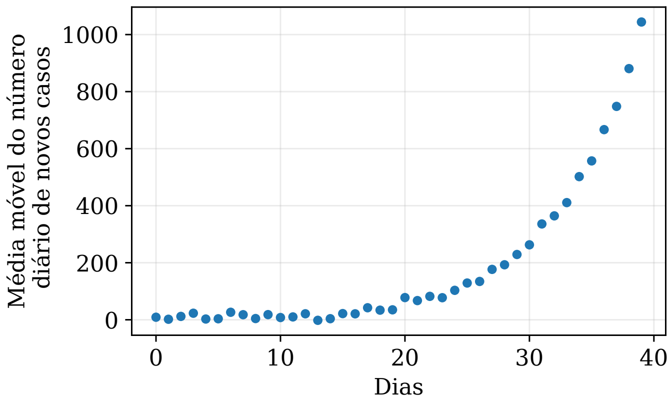

Figura 1: Média móvel do número diário de novos casos para os 40 primeiros dias da epidemia.

Modelagem matemática

Considere o teorema de Bayes, dado por

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{y})}\ , \tag{1}$$

onde $\boldsymbol{\theta}$ e $\boldsymbol{y}$ denotam os vetores de parâmetros e de dados, respectivamente. Por sua vez, o modelo capaz de descrever o comportamento dos dados disponíveis é dado por

$$\mathcal{M}_\theta(t) = I(t_1)\exp(\theta_1 t)\ .$$

Queremos conhecer os valores de $\boldsymbol{\theta}$ em que a saída do modelo $\mathcal{M}_\theta$ melhor represente o comportamento dos dados $y_1 = I(t_1), y_2 = I(t_2), \ldots$ Para isso, precisamos inferir os valores de $\boldsymbol{\theta}$, isto é, obter uma aproximação de $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ na Eq. (1).

Assuma que a função de verossimilhança é uma distribuição normal expressa por

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}) = \mathcal{N}\left(\mathcal{M}_\theta, \sigma^2\right)\ .$$

Neste caso, o vetor de parâmetros na inferência Bayesiana é dado por $\boldsymbol{\theta} = \left(\theta_1, \sigma^2\right)$. Note que, como não conhecemos a variância da função de verossimilhança, σ^2 é um hiperparâmetro do modelo, que também deve ser inferido. Assumiremos as seguintes distribuições *a priori*:

$$p\left(\theta_1\right) = \mathcal{U}\left(0{,}1,\ 0{,}2\right) \text{dia}^{-1}$$
$$p\left(\sigma\right) = \mathcal{U}\left(10,\ 15\right)\ \text{indivíduos}^2\ ,$$

onde $\mathcal{U}\left(a, b\right) = 1/(b-a)$ representa a distribuição uniforme.

Tarefa

Use um método de amostragem da sua preferência para aproximar a probabilidade *a posteriori* conjunta $p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)$, onde $\boldsymbol{\theta} = \left(\theta_1, \sigma^2\right)$ e $\boldsymbol{y} = \left(I\left(t_1\right), I\left(t_2\right), \ldots\right)$. Você deve obter as distribuições *a posteriori* marginais de cada parâmetro, assim como inferir o valor mais provável de cada distribuição aproximada. Mostre que o valor mais provável de cada parâmetro, quando usado para avaliar o modelo exponencial, fornece o melhor ajuste de $I\left(t\right)$ no intervalo em que os dados foram disponibilizados. Além disso, considere os seguintes aspectos na sua análise numérica:

- Como a variação dos parâmetros do método de amostragem adotado afetam a convergência e o custo computacional?
- Seria possível adotar outra distribuição *a priori* para algum dos parâmetros? Como isso afetaria o procedimento de inferência?
- O diagnóstico das cadeias nos leva a crer que houve convergência?
- Qual é o nível de correlação entre os parâmetros e como isso eventualmente afetou a aproximação das distribuições?

Regras

- O trabalho é composto por (1) código-fonte, (2) relatório e (3) apresentação.
- O relatório deve ser elaborado em formato de artigo, com *template* livre e sem limite de páginas.
- O código-fonte pode ser implementado em qualquer linguagem de programação.
- Não é permitido o uso de bibliotecas que contenham implementações do método de amostragem adotado.
- Qualquer método de amostragem pode ser empregado, desde que devidamente descrito no relatório, incluindo as referências utilizadas.
- O relatório pode ser redigido em português ou inglês.
- Os trabalhos serão em dupla.
- As apresentações terão duração máxima de 12 minutos, com mais 3–5 minutos para perguntas.
- A nota final será definida por média aritmética simples entre código-fonte, relatório e apresentação. Cada aluno da dupla receberá a mesma nota.

Prazo para envio dos relatórios e código-fonte: 28 de janeiro de 2022.

Data das apresentações: 31 de janeiro de 2022.