MI602: Métodos Computacionais em Estatística, Atividade II

Guilherme Ludwig
gvludwig@ime.unicamp.br

17 de Novembro de 2022

- A atividade é individual. Referencie o trabalho citado, mas desenvolva as soluções na atividade. Explique resultados com suas próprias palavras.
- A solução deve ser desenvolvida em detalhe. Todas as contas devem estar explicadas e desenvolvidas. Não é preciso fazer operações elementares manualmente mas as contas devem estar claramente indicadas.
- Cada aluno deverá ser responsável por garantir que não há cópia do seu trabalho. Indícios de plágio podem levar à anulação da atividade de todos os envolvidos. Use seu bom senso para guiar as conversas com seus colegas.
- A atividade deve ser entregue através do Moodle, em um único documento em formato `PDF`.
  - As páginas da sua solução devem estar numeradas, e as linhas também devem estar numeradas. Use o pacote no LaTeX `\usepackage{lineno}` e `\linenumbers`. Isso é importante pois me ajudará a dar *feedback* à sua solução.
  - Se você usar o `Word` ou `LibreOffice`, numere linhas e páginas e imprima o documento em formato `PDF` antes de enviá-lo.
- Use a linguagem que você preferir, de preferência com o pacote `listings` do LaTeX para exibir código: `R`, `C`, `python`, `julia`, ou outra qualquer. Todos os códigos devem estar listados no final da atividade, com instruções de execução (à parte das linguages listadas).
- Se você quiser, pode usar `Rmarkdown`, `Jupyter notebooks` ou o que preferir para mesclar código e texto.
- **Prazo de entrega:** 24/11/2022, às 23:59, via Moodle. Sem tolerância para atraso.

Esse exercício está adaptada da questão 2.6 do livro de Givens and Hoeting (2012), página 56. A Tabela 1 mostra as contagens de uma população do chamado besouro-da-farinha (*tribolium confusum*) em vários instantes do tempo. Os besouros em todos os estágios de desenvolvimento foram contabilizados, e a comida disponível foi controlada pelos pesquisadores. Um modelo elementar de crescimento populacional é o modelo logístico, em que

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\left(1 - \frac{N}{K}\right),$$

onde N é o tamanho da população, t é o tempo, r é um parâmetro de taxa de crescimento, e K é um parâmetro que determina o potencial populacional do meio ambiente. A solução desta equação diferencial é dada por

$$N_t = f(t) = \frac{KN_0}{N_0 + (K - N_0)\exp\{-rt\}},$$

onde N_t é o tamanho da população no tempo t, e N_0 é o tamanho inicial.

Tabela 1: Contagem de besouros-da-farinha em todos os estágios de desenvolvimento, por dia

Dias	0	8	28	41	63	79	97	117	135	154
Besouros	2	47	192	256	768	896	1120	896	1184	1024

Considere $N_0 = 2$, e use como função objetivo a perda quadrática

$$Q(r, K) = \sum_{i=1}^{m} (N_i - f_{r,K}(t_i))^2,$$

onde $N_0 = 2, N_1 = 47, \ldots, N_{10} = 1024$, e $t_0 = 0$, $t_1 = 8, \ldots, t_{10} = 154$.

(1) Encontre o gradiente e a matriz Hessiana para o problema de mínimos quadrados não-lineares $\hat{r}$, $\hat{K}$.

(2) Implemente o método de Newton-Raphson manualmente (sem usar `optim` ou pacote similar). Comente sobre a convergência do método, experimentando diferentes valores iniciais.

(3) Implemente *line search* no método anterior. Comente sobre o ganho computacional.

(4) Uma suposição comum em modelagem populacional é a log-normalidade, isto é, $\log(N_t)$ tem distribuição normal com média $\log(f(t))$ e variância σ^2. Neste caso, encontre os estimadores de máxima verossimilhança $\hat{r}$, $\hat{K}$ utilizando o algoritmo de escore de Fisher (justifique suas contas; implemente o método sem usar `optim`).

(5) Nota extra (20%): implemente o modelo log-normal do item (4) utilizando as restrições $r > 0, K > 0$ através do `L-BFGS-B`, utilizando `optim`.

(6) Nota extra 2 (40%): faça o item (5) implementando funções objetivo e gradiente utilizando o `Rcpp/Eigen`.

Referências

G. H. Givens and J. A. Hoeting. *Computational Statistics*. John Wiley & Sons, 2012.