# Trabalho II: o algoritmo EM.

Disciplina: Inferência Estatística
Professor: Luiz Max de Carvalho

24 de Agosto de 2022

**Data de Entrega: 28 de Setembro de 2022.**

## Orientações

- Enuncie e prove (ou indique onde se pode encontrar a demonstração) de todos os resultados não triviais necessários aos argumentos apresentados;
- Lembre-se de adicionar corretamente as referências bibliográficas que utilizar e referenciá-las no texto;
- Equações e outras expressões matemáticas também recebem pontuação;
- Você pode utilizar figuras, tabelas e diagramas para melhor ilustrar suas respostas;
- Indique com precisão os números de versão para quaisquer software ou linguagem de programação que venha a utilizar para responder às questões[1];

## Introdução

O estimador de máxima verossimilhança (EMV) possui uma gama de propriedades atraentes, como consistência, invariância e normalidade asintótica. Em muitas situações práticas, no entanto, este estimador é difícil de obter, especialmente quando parte dos dados está faltando ("*missing data*"). Por exemplo, podemos estar interessados em estudar a relação entre peso e altura, mas na nossa amostra temos os pesos de alguns indivíduos e alturas de outros.

O algoritmo EM ("Expectation Maximisation") é um método iterativo para aproximar o EMV em situações com dados faltantes. Começamos com um valor inicial $\theta^{(0)}$ e depois para ir do passo $j$ para o passo $j+1$, escrevemos a *verossimilhança dos dados completos*, que é a log-verossimilhança dos dados se os tivéssemos observado completos.

O passo "E" ("esperança") do algoritmo consiste em computar a distribuição condicional das observações faltantes dadas as observações existentes se o parametro tivesse o valor $\theta^{(j)}$, e tomar a esperança destes dados faltantes tratando $\theta$ como fixo – e os dados faltantes como variáveis aleatórias. Já o passo "M"

[1]Não precisa detalhar o que foi usado para preparar o documento com a respostas. Recomendo a utilização do ambiente LaTeX, mas fique à vontade para utilizar outras ferramentas.

("marginalização") envolve escolher $\theta^{(j+1)}$ que maximize a distribuição condicional obtida no passo E.

## Questões

1. Defina claramente todos os passos do algoritmo EM (faça um glossário de termos se precisar);

2. **Um exemplo motivador:** Suponha que temos duas moedas, Moeda 1 e Moeda 2 de modo que $\Pr(\text{Cara} \mid \text{Moeda} = 1) = p_1$ e $\Pr(\text{Cara} \mid \text{Moeda} = 2) = p_2$; Suponha agora que fazemos o seguinte experimento:

   (i) Selecionamos uma moeda aleatóriamente com probabilidade 1/2;
   (ii) Lançamos a moeda selecionada $m$ vezes;
   (iii) Repetimos (i) e (ii) $n$ vezes.

   Podemos representar os dados advindos deste experimento como

$$\begin{array}{ccccc} X_{11} & \ldots & X_{1m} & & M_1 \\ X_{21} & \ldots & X_{2m} & & M_2 \\ \vdots & \ldots & \vdots & & \vdots \\ X_{n1} & \ldots & X_{nm} & & M_n \end{array}.$$

   onde os $X_{ij}$ são variáveis Bernoulli que guardam o resultados do lançamento da moeda e $M_i \in \{1, 2\}$ é a variável aleatória que guarda qual moeda foi selecionada na $i$-ésima rodada do experimento.

   **Desenvolva** um esquema EM para obter o EMV para $\theta = (p_1, p_2)$ quando desconhecemos os valores de $M_i$, isto é, quando não sabemos que moeda foi escolhida em cada uma das $n$ rodadas.

3. Mostre que a sequência $\theta^{(j)}$ é monotônica e não descrescente com respeito à verossimilhança, isto é,

$$L\left(\theta^{(j+1)}\right) \geq L\left(\theta^{(j)}\right)$$

   Dica (fortemente aconselhado): ver exercício 7.31 de Casella & Berger (2002).

4. Discuta a importância do método EM: quando ele é aplicável? Vale sempre a pena? O que o item anterior demonstra sobre o método?