USP – ICMC – SME0806 - Estatística Computacional
5º trabalho – 1º/2015

1. Uma amostra de 197 objetos foi observada e distribuída em quatro categorias com frequências $(125, 18, 20, 34)$. Suponha que as probabilidades das categorias são dadas por
$$\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right).$$
   (a) Obtenha a estimativa de máxima verossimilhança de θ utilizando o algoritmo EM com aproximação de Monte Carlo no passo E.
   (b) Compare a solução do item anterior com a solução exata.

2. Os dados da tabela abaixo referem-se ao número diário de óbtidos de pessoas com 80 anos de idade ou mais em uma certa região. A coluna de frequências n_i registra o número de dias com um dado número i de óbitos. Estudos mostraram que uma distribuição de Poisson não fornece um bom ajuste a estes dados, possivelmente por diferentes padrões no inverno e no verão. Uma mistrura de duas distribuições de Poisson fornece um ajuste bem melhor.

Número de óbtidos (i)	Frequência (n_i)	Número de óbtidos (i)	Frequência (n_i)
0	162	5	61
1	267	6	27
2	271	7	8
3	185	8	3
4	111	9	1

O número de óbitos no k-ésimo dia é denotado por Y_k, $k = 1, \ldots, n$, em que $n = \sum_{i=0}^{9} n_i = 1096$. Para i óbtios, a função de probabilidade é dada por
$$P(Y_k = i; \boldsymbol{\theta}) = f(i; \boldsymbol{\theta}) = \alpha e^{-\mu_1} \frac{\mu_1^i}{i!} + (1-\alpha) e^{-\mu_2} \frac{\mu_2^i}{i!},$$
$i = 0, 1, 2, \ldots$, em que $\boldsymbol{\theta} = (\mu_1, \mu_2, \alpha)$, (μ_1, μ_2) é o vetor de médias e α é a probabilidade de que uma observação pertença à população 1. Logo, a função de verossimilhança de $\boldsymbol{\theta}$ é dada por
$$L(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=0}^{9} \left\{ \alpha e^{-\mu_1} \frac{\mu_1^i}{i!} + (1-\alpha) e^{-\mu_2} \frac{\mu_2^i}{i!} \right\}^{n_i}.$$

Seja Z_k a variável indicadora de que o k-ésimo dia pertença à população 1, sendo que $Z_k \sim \text{Bernoulli}(\alpha)$, independentes, $k = 1, \ldots, n$. Se $\widehat{\boldsymbol{\theta}}^{(j)}$ denota a estimativa de $\boldsymbol{\theta}$ na j-ésima iteração do algoritmo EM, pode ser provado que

$$\mathrm{E}[Z_k | Y_k = i, \widehat{\boldsymbol{\theta}}^{(j)}] = \frac{\widehat{\alpha}^{(j)} e^{-\widehat{\mu}_1^{(j)}} {\widehat{\mu}_1^{(j)}}^i}{\widehat{\alpha}^{(j)} e^{-\widehat{\mu}_1^{(j)}} {\widehat{\mu}_1^{(j)}}^i + (1 - \widehat{\alpha}^{(j)}) e^{-\widehat{\mu}_2^{(j)}} {\widehat{\mu}_2^{(j)}}^i},$$

$k = 1, \ldots, n$, que só depende de i, para todo $k = 1, \ldots, n$, e é denotada por $z_i^{(j)}$, $i = 0, \ldots, 9$.

O passo M do algoritmo EM é formado pelas equações

$$\widehat{\alpha}^{(j+1)} = \frac{\sum_{i=0}^{9} n_i z_i^{(j)}}{\sum_{i=0}^{9} n_i},$$

$$\widehat{\mu}_1^{(j+1)} = \frac{\sum_{i=0}^{9} i n_i z_i^{(j)}}{\sum_{i=0}^{9} n_i z_i^{(j)}} \quad \text{e} \quad \widehat{\mu}_2^{(j+1)} = \frac{\sum_{i=0}^{9} i n_i (1 - z_i^{(j)})}{\sum_{i=0}^{9} n_i (1 - z_i^{(j)})}. \tag{1}$$

(a) Prove os resultados apresentados nas equações (1).

(b) Obtenha a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ utilizando o algoritmo EM sem e com aproximação de Monte Carlo no passo E.

(c) Ajuste a distribuição de Poisson aos dados e justifique a afirmação de que o ajuste não é bom.

(d) Justifique a afirmação de que o ajuste pela mistura de distribuições de Poisson é melhor do que o ajuste pela distribuição de Poisson.