| Name | SOLUTION |
|---|---|
| PID # | |
| Instructor | Sergey Kirshner |

STAT 598G Spring 2011
Quiz #3
April 21, 2011

You are not allowed to use books or notes. Please read the directions carefully. The quiz is graded out of 3 points. You have 20 minutes to complete it. Please show all your work. Use the back of the page if you need more space.

A pdf for a mixture of two exponentials is given by

$$f(x|\pi,\lambda_0,\lambda_1) = (1-\pi)\lambda_0 e^{-\lambda_0 x} + \pi\lambda_1 e^{-\lambda_1 x},\ x > 0$$

with parameters $\pi \in [0,1]$, $\lambda_0, \lambda_1 > 0$. A data set $\{(c_1,x_1),\ldots,(c_n,x_n)\}$ is obtained by first drawing $C_i = c_i \overset{i.i.d}{\sim} \mathsf{Bern}(\pi),\ i = 1,\ldots,n$, and then drawing $X_i = x_i | C_i = c_i \sim \mathsf{Exp}(\lambda_{c_i}),\ i = 1,\ldots,n$.

1. (1.5pts) Assume that one observes a complete data set $\mathcal{D}_c = \{(c_1,x_1),\ldots,(c_n,x_n)\}$. Write down the complete data log-likelihood, and derive a closed form MLE estimate for $(\pi,\lambda_0,\lambda_1)$ from the complete data $\mathcal{D}_c$.

**Solution:**

$$\begin{aligned}
l_{\mathcal{D}_c}(\pi,\lambda_0,\lambda_1) &= \ln p(\mathcal{D}_c|\pi,\lambda_0,\lambda_1) = \ln\prod_{i=1}^{n}\left(\pi^{c_i}(1-\pi)^{1-c_i}\lambda_{c_i}e^{-\lambda_{c_i}x_i}\right) \\
&= \sum_{i=1}^{n}\left(c_i\ln\pi + (1-c_i)\ln(1-\pi) + \ln\lambda_{c_i} - \lambda_{c_i}x_i\right) \\
&= \#1\ln\pi + \#0\ln(1-\pi) + \#1\ln\lambda_1 + \#0\ln\lambda_0 - \lambda_1\sum_{\substack{i=1\\c_i=1}}^{n}x_i - \lambda_0\sum_{\substack{i=1\\c_i=0}}^{n}x_i \qquad (1)
\end{aligned}$$

where $\#1 = \sum_{i=1}^{n} c_i$ (number of 1s among $c_i$s), and $\#0 = n - \#1$ (number of 0s among $c_i$s). We will find the MLE by setting the gradient to $\mathbf{0}$ (joint distribution of $C$ and $X$ falls within the exponential family, so the parameter values corresponding to the $\mathbf{0}$ of the gradient corresponds to MLE).

$$\frac{\partial l_{\mathcal{D}_c}}{\partial \lambda_1} = \frac{\#1}{\lambda_1} - \sum_{\substack{i=1\\c_i=1}}^{n} = 0 \Longrightarrow \hat{\lambda_1} = \boxed{\#1\left(\sum_{\substack{i=1\\c_i=1}}^{n} x_i\right)^{-1}};$$

$$\frac{\partial l_{\mathcal{D}_c}}{\partial \lambda_0} = \frac{\#0}{\lambda_0} - \sum_{\substack{i=1\\c_i=0}}^{n} = 0 \Longrightarrow \hat{\lambda_0} = \boxed{\#0\left(\sum_{\substack{i=1\\c_i=0}}^{n} x_i\right)^{-1}};$$

$$\frac{\partial l_{\mathcal{D}_c}}{\partial \pi} = \frac{\#1}{\pi} - \frac{\#0}{1-\pi} = 0 \Longrightarrow \hat{\pi} = \boxed{\frac{\#1}{n}}.$$

2. (1.5pts) Now, suppose that the mixture memberships $c_1, \ldots, c_n$ are not observed. Describe in detail an algorithm for estimating of MLE for $(\pi, \lambda_0, \lambda_1)$ from incomplete data $\mathcal{D} = \{x_1, \ldots, x_n\}$. What are the properties of the obtained solution?

**Solution:** Let $\boldsymbol{\theta} = (\pi, \lambda_1, \lambda_0)$. The corresponding log-likelihood is

$$l_{\mathcal{D}}(\boldsymbol{\theta}) = \ln p(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln\left((1-\pi)\lambda_0 e^{-\lambda_0 x_i} + \pi\lambda_1 e^{-\lambda_1 x_i}\right).$$

Closed form solution is not known in the general case. This incomplete data log-likelihood is not concave, and may contain several local maxima. We will an iterative approach to estimate the maximum of $l_{\mathcal{D}}$. There many general approach to maximizing $l_{\mathcal{D}}$ (gradient ascent, conjugate gradients, etc), because the distribution over the complete data falls within the exponential family, we will apply the Expectation-Maximization (EM) approach. Denote by $\boldsymbol{\theta}^{(t)} = \left(\pi^{(t)}, \lambda_1^{(t)}, \lambda_0^{(t)}\right)$ the set of parameters at iteration $t$. In the E-step, we will estimate

$$\gamma_{ic} = P\left(C_i = c|x_i, \boldsymbol{\theta}^{(t)}\right) = \frac{\left(\pi^{(t)}\right)^c\left(1-\pi^{(t)}\right)^{1-c}\lambda_c^{(t)} e^{-\lambda_c^{(t)} x_i}}{\pi^{(t)}\lambda_1^{(t)} e^{-\lambda_1^{(t)} x_i} + \left(1-\pi^{(t)}\right)\lambda_0^{(t)} e^{-\lambda_0^{(t)} x_i}}. \tag{2}$$

In the M-step, we find $\boldsymbol{\theta}^{(t+1)}$ maximizing the expected log-likelihood

$$\begin{aligned}
\boldsymbol{\theta}^{(t+1)} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}\right) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, E_{P\left(\boldsymbol{C}|\boldsymbol{X},\boldsymbol{\theta}^{(t)}\right)} \ln P\left(\boldsymbol{C}, \boldsymbol{X}|\boldsymbol{\theta}^{(t+1)}\right)\\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n}\sum_{c=0}^{1} P\left(C_i = c|x_i, \boldsymbol{\theta}^{(t)}\right) \ln P\left(C_i = c, x_i|\boldsymbol{\theta}^{(t+1)}\right)\\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\left(\ln\pi\sum_{i=1}^{n}\gamma_{i1} + \ln(1-\pi)\sum_{i=1}^{n}\gamma_{i0} + \ln\lambda_1\sum_{i=1}^{n}\gamma_{i1} + \ln\lambda_0\sum_{i=1}^{n}\gamma_{i0} - \lambda_1\sum_{i=1}^{n}\gamma_{i1}x_i - \lambda_0\sum_{i=1}^{n}\gamma_{i0}x_i\right)
\end{aligned}$$

which is maximized similar to (1), resulting in

$$\pi^{(t+1)} = \boxed{\frac{\sum_{i=1}^{n}\gamma_{i1}}{n}}, \quad \lambda_1^{(t+1)} = \boxed{\left(\sum_{i=1}^{n}\gamma_{i1}\right)\left(\sum_{i=1}^{n}\gamma_{i1}x_i\right)^{-1}}, \quad \lambda_0^{(t+1)} = \boxed{\left(\sum_{i=1}^{n}\gamma_{i0}\right)\left(\sum_{i=1}^{n}\gamma_{i0}x_i\right)^{-1}}.$$

The algorithm is summarized in Algorithm 1.

**Algorithm 1** Expectation-Maximization

$t = 0$, initialize $\boldsymbol{\theta}^{(0)}$ ▷ e.g., $\pi^{(0)} \sim \mathsf{Unif}\,(0,1)$, $\lambda_0^{(0)}, \lambda_1^{(0)} \overset{iid}{\sim} \mathsf{Exp}\,(1)$

**repeat** ▷ iterating, iteration $t$

$\quad \gamma_{ic} = \frac{\left(\pi^{(t)}\right)^c \left(1-\pi^{(t)}\right)^{1-c} \lambda_c^{(t)} e^{-\lambda_c^{(t)} x_i}}{\pi^{(t)} \lambda_1^{(t)} e^{-\lambda_1^{(t)} x_i} + \left(1-\pi^{(t)}\right) \lambda_0^{(t)} e^{-\lambda_0^{(t)} x_i}}, \; c = 0,1, \;\; i = 1, \ldots, n$ ▷ E-step:

$\quad \pi^{(t+1)} = \frac{\sum_{i=1}^n \gamma_{i1}}{n}$ ▷ M-step: updating $\pi$

$\quad \lambda_1^{(t+1)} = \sum_{i=1}^n \gamma_{i1} \left(\sum_{i=1}^n \gamma_{i1} x_i\right)^{-1}$ ▷ M-step: updating $\lambda_1$

$\quad \lambda_0^{(t+1)} = \sum_{i=1}^n \gamma_{i0} \left(\sum_{i=1}^n \gamma_{i0} x_i\right)^{-1}$ ▷ M-step: updating $\lambda_0$

$\quad t = t + 1$

**until** convergence ▷ e.g., $l\left(\boldsymbol{\theta}^{(t+1)}\right) - l\left(\boldsymbol{\theta}^{(t)}\right) \leq$ threshold