

# Homework 4

## Stat 215A, Fall 2019

**Due:** provide a hard copy at the beginning of the lab on Friday November 22nd or push a `homework4.pdf` file to the `lab4/` folder of your `stat-215-a` GitHub repo by Thursday November 21 11:59pm

### 1 EM Algorithm

Suppose  $X_1, \dots, X_n$  are i.i.d. observations from a mixture of two Poisson distributions,  $\text{Pois}(\mu_0)$  and  $\text{Pois}(\mu_1)$ , with mixing probabilities of  $\pi$  and  $1 - \pi$ , respectively. That is, there is an initial probability  $\pi$  that an observation  $X_i$  is drawn from  $\text{Pois}(\mu_0)$  and probability  $1 - \pi$  from  $\text{Pois}(\mu_1)$ .

Recall that if  $X \sim \text{Pois}(\mu)$ , the probability density function is given by

$$p_\mu(x) = \frac{\mu^x e^{-\mu}}{x!}.$$

1. Define the observed data vector, the latent variable vector, and the distribution of the latent variable.
2. Write down the E step for estimating  $\mu_0, \mu_1, \pi$ .
3. Write down the M step for estimating  $\mu_0, \mu_1, \pi$ .
4. Give an initial estimator to start the EM algorithm.
5. Write R code to implement the E and M steps.
6. Simulate data from a mixture of two Poisson distributions, where you know the true parameters, and run your EM algorithm on the simulated data. Show the accuracy of EM clustering as you vary the values of  $\mu_0$  and  $\mu_1$ .
7. Now simulate data from a mixture of two binomial distributions,  $\text{Binom}(100, \pi_0)$  and  $\text{Binom}(100, \pi_1)$ . Show the accuracy of your Poisson EM clustering algorithm as you vary the values of  $\pi_0$  and  $\pi_1$  in this misspecified model.