Homework 3
Stat 215A, Fall 2022

Due: push a `homework3.pdf` file to your `stat-215-a` GitHub repo by **Friday, October 21 11:59pm**

1 EM Algorithm

Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a mixture of two Poisson distributions, Pois (μ_0) and Pois (μ_1), with mixing probabilities of π and $1-\pi$, respectively. That is, there is an initial probability π that an observation X_i is drawn from Pois (μ_0) and probability $1-\pi$ from Pois (μ_1). Recall that if $X \sim \text{Pois}(\mu)$, the probability density function is given by

$$p_\mu(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{1.1}$$

- Define the observed data vector, the latent variable vector, and the distribution of the latent variable.
- Write down the E step for estimating μ_0, μ_1, π.
- Write down the M step for estimating μ_0, μ_1, π.
- Give an initial estimator to start the EM algorithm.
- Write R code to implement the E and M steps.
- 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 μ_0 and μ_1.
- Now simulate data from a mixture of two binomial distributions, Binom $(100, \pi_0)$ and Binom $(100, \pi_1)$. Show the accuracy of your Poisson EM clustering algorithm as you vary the values of π_0 and π_1 in this misspecified model.

2 Miscellaneous

- What are two statistical benefits of randomized controlled studies when using Neyman-Rubin causal model?
- Why is it of interest to adjust for covariates when analyzing the results of randomized experiments?
- What was Freedman's critique of covariate adjustment using OLS and what remedy does Lin propose?