STAT 600 Statistical Computing

HW 3: EM and Extensions

Spring 2024, **Due Feb. 29th**

Homework format: Homework should be submitted as a pdf generated by LaTeX or Rmarkdown. All functions should be coded in `Rcpp/RcppArmadillo`. Please provide explanations of your solutions and appropriate graphics (labeled well).

1. Suppose that you observe $y_1, \ldots, y_n$, where $Y \sim p_\theta(y)$. Here $p_\theta(y)$ is defined

$$p_\theta(y) = \left\{ p\lambda \exp^{-\lambda y} + (1-p)\mu \exp^{-\mu y} \right\} 1_{(0,\infty)}(y),$$

and $\theta = (p, \lambda, \mu)$ with $p \in (0,1)$ and $\lambda, \mu > 0$. This is a mixture of exponentials.

It is too difficult to compute the MLE for θ analytically, so we can use the EM algorithm. The natural complete data for this problem is $X = (Y, \delta) \sim p_\theta(x)$ where

$$p_\theta(x) \propto (p\lambda \exp(-\lambda y))^{\delta} ((1-p)\mu \exp(-\mu y))^{1-\delta}$$

   (a) Derive the E-step in the EM algorithm. Hint:

$$E(\delta|Y) = \hat{\delta} = \frac{p\lambda \exp(-\lambda y)}{p\lambda \exp(-\lambda y) + (1-p)\mu \exp(-\mu y)}$$

   (b) Derive the M-step.

2. Write a function in `Rcpp` that implements the EM-algorithm above.

3. Simulate 100 data sets with $n = 100$ from the true distribution with $\theta = (p, \lambda, \mu) = (0.25, 1, 2)$, respectively.

4. Estimate the parameters of the model using the simulated data.

5. Estimate the standard errors of the parameter estimates using Louis's, bootstrap, OR SEM method. Do more than one for bonus.

6. Compare the average estimates, bias, standard errors (the square root of the average of the variances estimated with each of the methods), and coverage probability (using 95% CI) for each of the parameters in the model.