Homework 4

1. Design an EM algorithm to estimate the parameters in the zero-inflated Poisson model we studied in class. Be sure to specify the complete data likelihood, the conditional likelihood $Q(\psi, \psi^{(m)})$, the E step, and the M step. Remember that the observed data density for observation $i$ had the form:

$$f(x_i) = (1 - \pi) \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} + \pi I(x_i = 0), \ i = 1, \ldots, n$$

Write a program in R for your algorithm. The user should be able to input a data vector, an initial estimate of $\pi$, and an initial estimate of $\lambda$. The output should include the number of iterations and the parameter estimates. Check your algorithm against a simulated data set generated by:

```
zind = rbinom(500, 1, .7); zip = zind * rpois(500, 3)
```

Compare your answer for the data set you generate against both methods we studied in the computer session.

2. Modify the EM algorithm for a mixture of finite normals so that mixing occurs only over the sample means. A common value for $\sigma^2$ will need to be derived. Generate a gradient trace; the range for $\mu$ could cover $x_{(1)}$ through $x_{(n)}$ and you should output a graph with $\mu$ on the x-axis and $D_{Q_0}(\mu)$ on the y-axis. Refer to the `eruptions` variable for the `faithful` dataset; starting with a two-point mixture, how many mixture points need to be added before the trace lies entirely below 0?