Mixure Models

Evan Miller

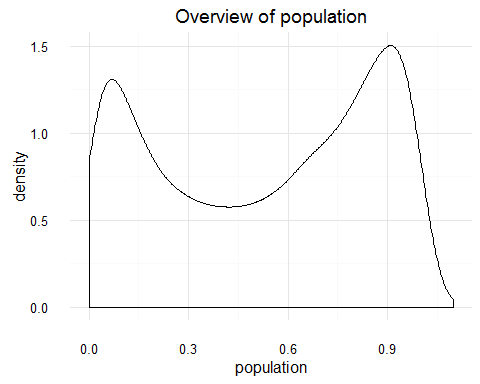
Thursday, June 16, 2016

This R markdown document will highlight the process used to estimate a mixture model from the underlying population data. The data has:

* 8000 members;
* data generated from 2 separate beta distributions

I need to:

* Find the maximum likelihood estimate (MLE) for the parameters of each component beta distribution
* Find the percentage of the data coming from each component beta distribution (5 parameters in total).



After looking at the graph there are two peaks. One at 0.1 and one at 0.95. However this doesn't give me much help as the beta distribtuion is characterised by it's alpha and beta parameters, not the mean and standard deviation.

The approach I have taken is as follows:

1. Given initial values for the beta distributions calculate the probability that a given observation would arise from either of the two beta distributions
2. Use these probabilities to calculate the class responsibilities
3. Draw randomly from a bernoulli distribution a class that is based off the observation's class responsibilities
4. Use the fitdistr function to fit a beta distribution for each class
5. Calculate the mixing proportion pi
6. Check if convergence has been achieved by differencing the initial and calculated parameters
7. If convergence hasn't been achieved repeat the process with the calculated parameters as inputs