# CS777 Assignment6 Task1

Group members: Zhaopu Teng, Xiaoyang Wang, Lin Chen

Our MCMC algorithm is following:

In there, k is the document clusters, n is the number of the whole documents for us to generate process, means the number of words in the document i.

Thus, we can get the PDF there:

Then, updating based on the PDF we got above, so we can get:

Where is a vector with , the total number of word in document I produced by Category C.

Then updating for a given j based on the PDF we got above and we have:

Where is a vector with the total number of times category C produced by word x.

Updating category Assignment for in n vector we will have then:

So, we have

Pr|=c|, all (

This is to compute the following for each c:

Then, normalize and update by sampling c from resultingCategorical Distribution