# Assignment 7

**# 1 Priors, Posteriors, and Conjugacy**

1. Without the study, I guess with maximum likelihood estimation.
2. i) The mean of Beta distribution

ii) The variance of Beta distribution

iii) Using the study data, we have ,

1. Applying Bayes theorem, the posterior distribution is Using the sample data, the posterior mean of :

where

which implies that

Thus, the posterior mean ,

1. i) Likelihood:

ii) The Likelihood of Poisson distribution

Also, the conjugate prior is a distribution which has a probability density function proportional to

Therefore, the posterior distribution

This is proportional to the probability density function of a distribution .

**# 2 Graphical Models and Plate Notation**

1. i) We have

ii) The random variables are the corpus , the set of classes and the class mixture .

The parameters are

1. i) Assuming there are documents in such a review dataset with a unique vocabulary , and it has topics shared among all documents and each document has its mixture of topics :

For

,

For ,

For

To generate ratings for document , we do two samplings. First, sampling a topic assignment from the topics proportions ; Second, sampling a rating from the corresponding topic . is an indicator variable that denotes which topic from was selected for .

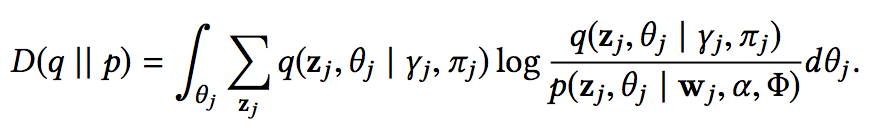
To generate words for document , we sample word from corresponding topic rating .

ii)

iii)

**# 3 Latent Dirichlet Allocation**

1. The LDA posterior is computationally intractable because there are too many possible latent. Specifically, the posterior distribution of latent variables given by , and observed data , there are many possible couplings between and over the latent topic assignment.
2. Variational inference, as an alternative strategy to Markov chain Monte Carlo, is to approximate probability density through optimization.



1. The goal of Markov chain Monte Carlo is to generate fair samples from a probability in high-dimensional space.
2. In each iteration, every variable (or block of variables) will be swept to sample from its conditional distribution with the remaining variables fixed to their current values in order to generate posterior samples. This process will continue until “convergence”.
3. The advantage of variational inference are 1) maximizing an explicit objective; 2) being faster in most cases.

Gibbs sampling has advantage of being nonparametric and asymptotically exact.