CS510: Advanced Topics in Information Retrival **Fall 2017**

Assignment 7

Zhuo Li(zhuol2) *Handed In: November 2, 2017*

1. Priors, Posteriors, and Conjugacy

   a. The best guess of θ is $\theta = 11/98$. The statistical estimation technique we use is maximum likelihood estimation.

   b. For the following Beta Distribution

$$P(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

   We have

   i. Mean is

$$\frac{\alpha}{\alpha+\beta}$$

   ii. Variance is

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

   iii. According to the problem we have

$$\begin{cases} \frac{\alpha}{\alpha+\beta} = 0.1 \\ \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.05^2 \end{cases} \Rightarrow \begin{cases} \alpha = 3.5 \\ \beta = 31.5 \end{cases}$$

   c. After we observe the data in the samples, there are another $\theta^{11}(1-\theta)^{87}$, so we have$\alpha = 14.5$ and $\beta = 118.5$. Then we have the posterior mean

$$\frac{\alpha}{\alpha+\beta} = 0.1090$$

   d. We have

   i. The likelihood function is

$$\prod_{i=1}^{N} p(x_i|\lambda) = \frac{\lambda^{\sum_{i=1}^N x_i} e^{-N\lambda}}{\prod_{i=1}^N x_i!}$$

   ii. We have

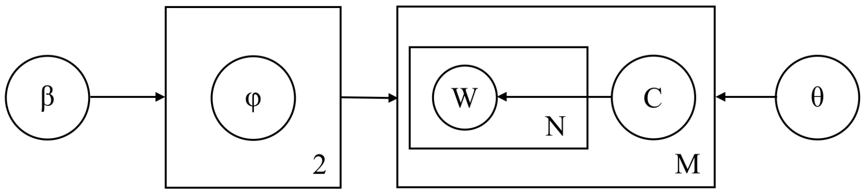
$$\begin{aligned} P(\lambda|X,\alpha,\beta) &\propto P(X|\lambda)P(\lambda|\alpha,\beta) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\frac{\lambda^{\sum_{i=1}^N x_i} e^{-N\lambda}}{\prod_{i=1}^N x_i!} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^N x_i!}\lambda^{\alpha+\sum_{i=1}^N x_i - 1}e^{-(\beta+N)\lambda} \end{aligned}$$

   So this posterior distribution is of the same probability family of the prior. i.e. it is also a Gamma distribution. And we have the new parameters $\alpha' = \alpha + \sum_{i=1}^N x_i$ and $\beta' = \beta + N$.

2. Graphical Models and Plate Notation

   a. We have

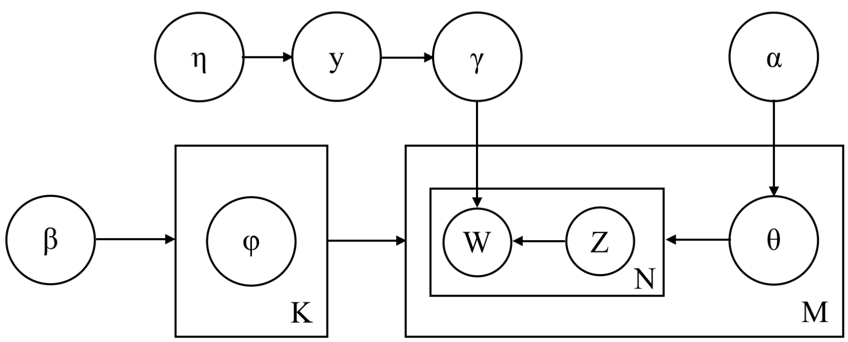
      i. The plate notation is as followings



      ii. The random variables are W, C, ϕ_0 and ϕ_1. Parameters are θ and β.

   b. i. We may have another hidden variable η to decide the distribution of the average rating y_d as $f(y_d|\eta)$. And we have another hidden parameter γ donating the distribution of word in d_i corresponding to y_d as $f(w_{d_i}|\gamma)$. Then we can first generate words for document d_j with original LDA sampling. i.e. Sample a topic assignment $z_{j,t}$ from topic proportions θ_j. Then we can sample the average rating $y_{j,t}$ from rating proportion η_j. And finally sample a word from the topic $\phi_{z_{j,t}}$ and $\gamma_{y_{j,t}}$.

      ii. The plate notation is as following



      iii. We can have the joint distribution as

$$
\begin{aligned}
&P(W, Y, Z, \Theta, \Phi, \Gamma|\alpha, \beta, \eta) \\
= &\prod P(\phi_k|\beta) \int f(y_r|\eta) f(\gamma_{j,r}|y_r) \prod P(\theta_j|\alpha) P(z_{j,t}|\theta) P(w_{r,j,t}|\phi_{z_{j,t},\gamma_{yr,j}})
\end{aligned}
$$

3. Latent Dirichlet Allocation

   a. Because we can not directly compute the distribution of

$$
P(\theta_d, Z_d|w_d, \alpha, \phi) = \int_{\theta_d} \sum_{Z_d} P(\theta_d|\alpha) P(Z_d|\theta_d) P(w_d|Z_d, \phi) d\theta_d
$$

b. Variational inference is to use a factorized distribution to approximate the posterior distribution of interest. And we need to solve

$$q(\theta_d|\gamma_d)\prod_{i=1}^{N} q(z_{d,i}|\pi_{d,i})$$

c. The Markov chain Monte Carlo algorithm is to sample from a random distribution based on Markov chain to get close to the target distribution in mean and variance.

d. For the all k samples in each iteration, we update sample $i = 1, 2...k$ for the next iteration with previous updated sample (i.e. $x_1, x_2...x_{i-1}$) and the sample at current iteration after it (i.e. $x_{i+1}, x_{i+2}...x_k$).

e. Variantional inference is faster than Gibbs sampling but leads to a distribution that is not same to the target.
Gibbs sampling can approximate to the real distribution with any small tolerance.
In general We prefer variantional inference when we need a fast but not sensitive to accuracy model, while choose Gibbs sampling when we want an accurate model but do not care about time.