Topics in Probabilistic Modeling & Inference (CS698X), Spring 2021
Indian Institute of Technology Kanpur
Homework Assignment Number 3

QUESTION 1

Student Name: Piyush Kumar Gaurav
Roll Number: 20104442
Date: May 9, 2021

My solution to Problem 1
Given: The i.i.d. observations as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. Thus

$$Likelihood = p(\mathbf{x}|\theta) = \prod_{n=1}^{N} p(\mathbf{x}_n|\theta)$$

where θ is the model parameters. Thus the prior over θ is defined as

$$Prior = p(\theta)$$

To prove: The solution of following expression would yield posterior distribution over θ (i.e. $p(\theta|\mathbf{x})$)

$$\underset{q(\theta)}{argmin} - \sum_{n=1}^{N} \left[\int q(\theta) log(p(\mathbf{x}_n|\theta)) d\theta \right] + KL(q(\theta)||p(\theta)) \tag{1}$$

where $q(\theta)$ is some probability distribution and $KL(q(\theta||p(\theta))$ represents the K L Divergence between the probability distributions $q(\theta)$ and $p(\theta)$
Hence our goal is to approximate $p(\theta|x)$ with $q(\theta)$ such that the distance between the two distributions $q(\theta)$ and $p(\theta)$ is minimized (where distance between two probability distribution is measured in terms of K L Divergence). Hence

$$\underset{q(\theta)}{argmin} KL(p(\theta|\mathbf{x})||q(\theta)) \tag{2}$$

where the posterior distribution $p(\theta|\mathbf{x})$ is not known.
In order to achieve our goal, we consider the marginal likelihood (which is constant w.r.t. θ as θ is marginalised). log of Marginal likelihood can be written as,

$$log(p(\mathbf{x})) = \left[\underbrace{\int q(\theta) log \left\{ \frac{p(\mathbf{x}, \theta)}{q(\theta)} \right\} d\theta}_{\mathcal{L}(q)} \right] + \left[\underbrace{- \int q(\theta) log \left\{ \frac{p(\theta|\mathbf{x})}{q(\theta)} \right\}}_{KL(p(\theta|\mathbf{x})||q(\theta))} \right] \tag{3}$$

Now, as mentioned earlier, $log(p(\mathbf{x}))$ is constant and hence minimizing $KL(p(\theta|\mathbf{x})||q(\theta))$ is equivalent to maximizing $\mathcal{L}(q)$ i.e.

$$
\begin{aligned}
\underset{q(\theta)}{argmin}\, KL(p(\theta|\mathbf{x})||q(\theta)) &= \underset{q(\theta)}{argmax}\, \mathcal{L}(q) \\
&= \underset{q(\theta)}{argmax} \int q(\theta) log \left\{ \frac{p(\mathbf{x},\theta)}{q(\theta)} \right\} d\theta \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log(p(\mathbf{x},\theta)) d\theta - \int q(\theta) log(q(\theta)) d\theta \right] \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log(p(\mathbf{x}|\theta)\, p(\theta)) d\theta - \int q(\theta) log(q(\theta)) d\theta \right] \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log(\underbrace{p(\mathbf{x}|\theta)}_{likelihood}) d\theta + \int q(\theta) log(\underbrace{p(\theta)}_{prior}) d\theta - \underbrace{\int q(\theta) log(q(\theta)) d\theta}_{entropy} \right] \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log(p(\mathbf{x}|\theta)) d\theta - \int q(\theta) log \left(\frac{q(\theta)}{p(\theta)} \right) d\theta \right] \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log(p(\mathbf{x}|\theta)) d\theta \right] - KL(q(\theta)||p(\theta)) \\
&= \underset{q(\theta)}{argmax} \left[\int q(\theta) log \left(\prod_{n=1}^{N} p(\mathbf{x}|\theta) \right) d\theta \right] - KL(q(\theta)||p(\theta)) \\
&= \underset{q(\theta)}{argmax} \left[\sum_{n=1}^{N} \int q(\theta) log \left(p(\mathbf{x}|\theta) \right) d\theta \right] - KL(q(\theta)||p(\theta)) \\
&= \underset{q(\theta)}{argmin} \left[- \sum_{n=1}^{N} \int q(\theta) log \left(p(\mathbf{x}|\theta) \right) d\theta \right] + KL(q(\theta)||p(\theta))
\end{aligned} \tag{4}
$$

Hence, from (1) and (4) we can conclude,

$$
\begin{aligned}
p(\theta|\mathbf{x}) &\cong \underset{q(\theta)}{argmin} KL(p(\theta|\mathbf{x})||q(\theta)) \\
&= \underset{q(\theta)}{argmin} - \sum_{n=1}^{N} \left[\int q(\theta) log(p(\mathbf{x}_n|\theta)) d\theta \right] + KL(q(\theta)||p(\theta))
\end{aligned}
$$

Intuition: The $q(\theta)$ which minimizes the objective function in Eqn. (1) is close to the likelihood $(p(\mathbf{x}_n|\theta))$ for the given data (as it minimizes Term 1 in Eqn. (1)) at the same time $q(\theta)$ is also close to the Prior $(p(\theta))$ (as it minimizes Term 2)

Student Name: Piyush Kumar Gaurav
Roll Number: 20104442
Date: May 9, 2021

QUESTION 2

My solution to problem 2

Given: The observations from a regression model are $\{\mathbf{x}_n, t_n\}_{n=1}^{N}$. Following are the details of the regression model,

$$Likelihood \equiv y_n \sim \mathcal{N}(y_n|\mathbf{w}^T\mathbf{x}_n, \beta^{-1})$$

$$Prior(\mathbf{w}) \equiv \mathcal{N}(\mathbf{w}|0, \Sigma)$$

$$Prior(\beta) = Gamma(\beta|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}\beta^{a_0-1}exp(-b_0\beta)$$

$$Prior(\alpha_d) = Gamma(\alpha_d|e_0, f_0) = \frac{f_0^{e_0}}{\Gamma(e_0)}\alpha_d^{e_0-1}exp(-f_0\alpha_d)$$

where $\Sigma = diag(\alpha_1, \alpha_2, \dots, \alpha_d)$

The joint distribution of all the parameters can be written as

$$\begin{aligned} p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \dots, \alpha_D|\mathbf{X}) &= p(\mathbf{y}|\mathbf{w}, \beta, \mathbf{X})\, p(\mathbf{w}|\alpha_1, \dots, \alpha_d)\, p(\beta)\, p(\alpha_1, \dots, \alpha_D) \\ &= \prod_{n=1}^{N} p(y_n|\mathbf{w}, \mathbf{x}_n, \beta)\, p(\mathbf{w}|\alpha_1, \dots, \alpha_d)\, p(\beta) \prod_{d=1}^{D} p(\alpha_d) \end{aligned} \tag{5}$$

Taking log both side we get,

$$\begin{aligned} &log(p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \dots, \alpha_D|\mathbf{X})) \\ &= \sum_{n=1}^{N} log(\mathcal{N}(y_n|\mathbf{w}, \mathbf{x}_n, \beta))\, log(\mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha_1, \dots, \alpha_d))\, log(Gamma(\beta)) \sum_{d=1}^{D} log(Gamma(\alpha_d)) \\ &= \sum_{n=1}^{N} log\left(\sqrt{\frac{\beta}{2\pi}}exp\left(-\frac{\beta}{2}(y_n - \mathbf{w}^T\mathbf{x}_n)^2\right)\right) + log\left(\sqrt{\frac{\prod_{d=1}^{D}\alpha_d}{(2\pi)^D}}exp\left(\frac{-\mathbf{w}^T\Sigma\mathbf{w}}{}\right)\right) \\ &+ log\left(\frac{b_0^{a_0}}{\Gamma(a_0)}\beta^{a_0-1}exp(-b_0\beta)\right) + \sum_{d=1}^{D} log\left(\frac{f_0^{e_0}}{\Gamma(e_0)}\alpha_d^{e_0-1}exp(-f_0\alpha_d)\right) \\ &= \frac{N}{2}log(\beta) - \frac{\beta}{2}\sum_{n=1}^{N}(y_n - \mathbf{w}^T\mathbf{x}_n)^2 + \frac{1}{2}\sum_{d=1}^{D}log(\alpha_d) - \frac{1}{2}\mathbf{w}^T\Sigma\mathbf{w} \\ &+ (a_0 - 1)log(\beta) - b_0\beta + (e_0 - 1)\sum_{d=1}^{D}log(\alpha_d) - f_0\sum_{d=0}^{D}\alpha_d + constants \end{aligned} \tag{6}$$

Following the formulas and notations as mentioned in the lectures for evaluating $q^*_{\mathbf{w}}(\mathbf{w})$, $q^*_{\beta}(\beta)$ and $q^*_{\alpha_d}(\alpha_d)$ which are variational posterior for $\mathbf{w}$, β and α_d respectively. Thus we can write,

$$log(q^*_{\mathbf{w}}(\mathbf{w})) = \mathbb{E}_{q_{\beta,\alpha_1,\ldots,\alpha_D}}[log(p(\mathbf{y},\mathbf{w},\beta,\alpha_1,\ldots,\alpha_D|\mathbf{X}))] + \text{constant w.r.t } \mathbf{w} \tag{7}$$

$$log(q^*_{\beta}(\beta)) = \mathbb{E}_{q_{\mathbf{w},\alpha_1,\ldots,\alpha_D}}[log(p(\mathbf{y},\mathbf{w},\beta,\alpha_1,\ldots,\alpha_D|\mathbf{X}))] + \text{constant w.r.t } \beta \tag{8}$$

$$log(q^*_{\alpha_d}(\alpha_d)) = \mathbb{E}_{q_{\mathbf{w},\beta,\alpha_{-d}}}[log(p(\mathbf{y},\mathbf{w},\beta,\alpha_1,\ldots,\alpha_D|\mathbf{X}))] + \text{constant w.r.t } \alpha_d \tag{9}$$

Using (6) and (7) we can write

$$\begin{aligned} log(q^*_{\mathbf{w}}(\mathbf{w})) &= \mathbb{E}\left[-\frac{\beta}{2}\sum_{n=1}^{N}(y_n - \mathbf{w}^T\mathbf{x}_n)^2 - \frac{1}{2}\mathbf{w}^T\Sigma\mathbf{w}\right] + \text{constant w.r.t } \mathbf{w} \\ &= -\frac{1}{2}\left[\mathbf{w}^T\left(\mathbb{E}[\beta]\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^T + \mathbb{E}[\Sigma]\right)\mathbf{w} - 2\mathbf{w}^T\mathbb{E}[\beta]\sum_{n=1}^{N}y_n\mathbf{x}_n\right] + \text{constant w.r.t } \mathbf{w} \end{aligned} \tag{10}$$

Considering the fact that the $Prior(\mathbf{w})$ is Gaussian and the form of $q^*_{\mathbf{w}}(\mathbf{w})$ is Gaussian too, we write

$$q^*_{\mathbf{w}}(\mathbf{w}) \equiv \mathcal{N}(\mathbf{w}|\mu_{\mathbf{w},\Sigma_{\mathbf{w}}}) \tag{11}$$

where

$$\Sigma_{\mathbf{w}} = \left(\mathbb{E}[\beta]\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^T + \mathbb{E}[\Sigma]\right)$$

$$\mu_{\mathbf{w}} = \Sigma_{\mathbf{w}}^{-1}\mathbb{E}[\beta]\sum_{n=1}^{N}y_n\mathbf{x}_n$$

$$\mathbb{E}[\Sigma] = diag(\mathbb{E}[\alpha_1],\ldots,\mathbb{E}[\alpha_D])$$

Using (6) and (8) we can write

$$\begin{aligned} log(q^*_{\beta}(\beta)) &= \mathbb{E}\left[\frac{N}{2}log(\beta) - \frac{\beta}{2}\sum_{n=1}^{N}(y_n - \mathbf{w}^T\mathbf{x}_n)^2 + (a_0 - 1)log(\beta) - b_0\beta\right] + \text{constant w.r.t } \beta \\ &= \left(\frac{N}{2} + a_0 + 1\right)log(\beta) - \beta\left(\sum_{n=1}^{N}\frac{1}{2}\mathbb{E}[(y_n - \mathbf{w}^T\mathbf{x}_n)^2] + b_0\right) + \text{constant w.r.t } \beta \end{aligned} \tag{12}$$

Considering the fact that the $Prior(\beta)$ is Gamma and the form of $q^*_{\mathbf{w}}(\mathbf{w})$ is Gamma too,

$$q^*_{\beta}(\beta) \equiv Gamma(\beta|i,j) \tag{13}$$

where

$$i = \frac{N}{2} + a_0$$

$$
\begin{aligned}
j &= \left(\sum_{n=1}^{N} \frac{1}{2} \mathbb{E}[(y_n - \mathbf{w}^T \mathbf{x}_n)^2] + b_0 \right) \\
&= \left(\sum_{n=1}^{N} \frac{1}{2} \left(y_n^2 - 2 y_n \mathbb{E}[\mathbf{w}]^T \mathbf{x}_n + \mathbf{x}_n^T \mathbb{E}[\mathbf{w}^T \mathbf{w}] \mathbf{x}_n \right) + b_0 \right)
\end{aligned}
$$

Using (6) and (9) we can write

$$
\begin{aligned}
log(q^*_{\alpha_d}(\alpha_d)) &= \mathbb{E}\left[\frac{1}{2} log(\alpha_d) - \frac{1}{2} w_d^2 \alpha_d + (e_0 - 1) log(\alpha_d) - f_0 \alpha_d \right] + \text{constant w.r.t } \alpha_d \\
&= \left(\frac{1}{2} + e_0 - 1 \right) log(\alpha_d) - \alpha_d \left(f_0 + \frac{1}{2} \mathbb{E}[w_d^2] \right) + \text{constant w.r.t } \alpha_d
\end{aligned} \tag{14}
$$

Considering the fact that the $Prior(\alpha_d)$ is Gamma and the form of $q^*_{\alpha_d}(\alpha_d)$ is Gamma too,

$$
q^*_{\alpha_d}(\alpha_d) \equiv Gamma(\alpha_d | k, l) \tag{15}
$$

where

$$
\begin{aligned}
k &= \frac{1}{2} + e_0 \\
l &= \left(f_0 + \frac{1}{2} \mathbb{E}[w_d^2] \right)
\end{aligned}
$$

The Mean Field V.I. Algorithm

1. Initialize $\mu_{\mathbf{w}}$, $\Sigma_{\mathbf{w}}$, i, j, k and l. Hence, we initialize the distributions $q^*_{\mathbf{w}}(\mathbf{w})$, $q^*_{\beta}(\beta)$ and $q^*_{\alpha_d}(\alpha_d)$ for $d = 1, 2, \ldots, D$

2. Compute Evidence Lower Bound (ELBO) using the initialized value in Step 1

$$ELBO = \mathcal{L}(\mathbf{q}) = \mathbb{E}_{\mathbf{q}}[log(p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \ldots, \alpha_D|\mathbf{X}))] + \mathbb{E}_{\mathbf{q}}\mathbf{q}$$

   where $\mathbf{q} = q^*_{\mathbf{w}}(\mathbf{w})q^*_{\beta}(\beta)\prod_{d=1}^{D} q^*_{\alpha_d}(\alpha_d)$

3. For t = 1 to T (until convergence)

   a. Compute $\mathbb{E}_{q^*_{\beta}(\beta)}[\beta]$ and $\mathbb{E}_{q^*_{\alpha}(\alpha_d)}[\alpha_d]$ for $d = 1, 2, \ldots, D$

   b. Compute

$$q^*_{\mathbf{w}}(\mathbf{w}) \equiv \mathcal{N}(\mathbf{w}|\mu_{\mathbf{w},\Sigma_{\mathbf{w}}})$$

   where

$$\Sigma_{\mathbf{w}} = \left(\mathbb{E}[\beta]\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^T + \mathbb{E}[\Sigma]\right)$$

$$\mu_{\mathbf{w}} = \Sigma_{\mathbf{w}}^{-1}\mathbb{E}[\beta]\sum_{n=1}^{N} y_n\mathbf{x}_n$$

$$\mathbb{E}[\Sigma] = diag(\mathbb{E}[\alpha_1], \ldots, \mathbb{E}[\alpha_D])$$

   c. Compute $\mathbb{E}_{q^*_{\mathbf{w}}(\mathbf{w})}[\mathbf{w}]$ (and hence, $\mathbb{E}_{q^*_{w_d}(w_d)}[w_d]$) and $\mathbb{E}_{q^*_{\mathbf{w}}(\mathbf{w})}[\mathbf{w}^T\mathbf{w}]$

   d. Compute

$$q^*_{\beta}(\beta) \equiv Gamma(\beta|i, j)$$

   and

$$q^*_{\alpha_d}(\alpha_d) \equiv Gamma(\alpha_d|k, l)$$

   where

$$i = \frac{N}{2} + a_0$$

$$j = \left(\sum_{n=1}^{N}\frac{1}{2}\left(y_n^2 - 2y_n\mathbb{E}[\mathbf{w}]^T\mathbf{x}_n + \mathbf{x}_n^T\mathbb{E}[\mathbf{w}^T\mathbf{w}]\mathbf{x}_n\right) + b_0\right)$$

$$k = \frac{1}{2} + e_0$$

$$l = \left(f_0 + \frac{1}{2}\mathbb{E}[w_d^2]\right)$$

   e. Compute ELBO using the updated $q^*_{\mathbf{w}}(\mathbf{w})$, $q^*_{\beta}(\beta)$ and $q^*_{\alpha_d}(\alpha_d)$ for $d = 1, 2, \ldots, D$. Check for the convergence of ELBO, if not converged go to Step 3.a else Stop

Topics in Probabilistic Modeling & Inference (CS698X), Spring 2021
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Piyush Kumar Gaurav
Roll Number: 20104442
Date: May 9, 2021

QUESTION 3

My solution to problem 3
The count value observations are denoted as $x_1, x_2, \ldots, x_N$. Hence,

$$Likelihood = p(x_n|\lambda_n) = Poisson(x_n|\lambda_n)$$

where $n = 1, 2, \ldots, N$. Thus,

$$Total\ Likelihood = \prod_{n=1}^{N} p(x_n|\lambda_n) = \prod_{n=1}^{N} Poisson(x_n|\lambda_n)$$

The Priors for the parameters are

$$Prior(\lambda_n) = p(\lambda_n|\alpha, \beta) = Gamma(\lambda_n|\alpha, \beta)$$

$$Prior(\lambda_1, \ldots, \lambda_N) = \prod_{n=1}^{N} p(\lambda_n|\alpha, \beta) = \prod_{n=1}^{N} Gamma(\lambda_n|\alpha, \beta)$$

$$Prior(\alpha|a, b) = p(\alpha|a, b) = Gamma(\alpha|a, b)$$

$$Prior(\beta|c, d) = p(\beta|c, d) = Gamma(\beta|c, d)$$

where a, b, c, d are fixed parameters.

The Joint Distribution of $x_1, \ldots, x_N, \lambda_1, \ldots, \lambda_N, \alpha, \beta$ can be defined as

$$p(x_1, \ldots, x_N, \lambda_1, \ldots, \lambda_N, \alpha, \beta) = \prod_{n=1}^{N} p(x_n|\lambda_n) \prod_{n=1}^{N} p(\lambda_n|\alpha, \beta)\, p(\alpha|a, b)\, p(\beta|c, d) \tag{16}$$

Computation of Conditional Posteriors:
Conditional Posterior for λ_n (where $n = 1, 2, \ldots, N$) is given by,

$$\begin{aligned} p(\lambda_n|x_n, \alpha, \beta) &\propto p(x_n|\lambda_n) p(\lambda_n|\alpha, \beta) \\ &= Poisson(x_n|\lambda_n)\, Gamma(\lambda_n|\alpha, \beta) \\ &\propto \left[\lambda_n^{x_n}\, e^{-\lambda_n}\right] \left[\lambda_n^{\alpha-1}\, e^{-\beta\lambda_n}\right] \\ &= Gamma(\lambda_n|\alpha + x_n, \beta + 1) \end{aligned} \tag{17}$$

Conditional Posterior for α is given by,

$$
\begin{aligned}
p(\alpha|\lambda_1,\ldots,\lambda_N,\beta) &= \prod_{n=1}^{N} p(\lambda_n|\alpha,\beta)p(\alpha|a,b) \\
&= \prod_{n=1}^{N} Gamma(\lambda_n|\alpha,\beta)\; Gamma(\alpha|a,b) \\
&= \prod_{n=1}^{N} \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_n^{\alpha-1}exp(-\beta\lambda_n)\right]\left[\frac{b^{a}}{\Gamma(a)}\alpha^{a-1}exp(-b\alpha)\right] \\
&\propto \prod_{n=1}^{N} \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_n^{\alpha-1}\right]\left[\alpha^{a-1}exp(-b\alpha)\right]
\end{aligned}
\tag{18}
$$

Hence a closed form solution is not available for $p(\alpha|\lambda_1,\ldots,\lambda_N,\beta)$

Conditional Posterior for β is given by,

$$
\begin{aligned}
p(\beta|\lambda_1,\ldots,\lambda_N,\alpha) &= \prod_{n=1}^{N} p(\lambda_n|\beta,\alpha)p(\beta|c,d) \\
&= \prod_{n=1}^{N} Gamma(\lambda_n|\beta,\alpha)\; Gamma(\beta|c,d) \\
&= \prod_{n=1}^{N} \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_n^{\alpha-1}exp(-\beta\lambda_n)\right]\left[\frac{b^{a}}{\Gamma(a)}\beta^{a-1}exp(-b\beta)\right] \\
&\propto \beta^{N\alpha+a-1}exp(-\beta(\sum_{n=1}^{N}\lambda_n+b)) \\
&= Gamma(\beta|i,j)
\end{aligned}
\tag{19}
$$

where

$$
\begin{aligned}
i &= N\alpha + a \\
j &= \sum_{n=1}^{N}\lambda_n + b
\end{aligned}
$$

Hence a closed form solution is available for CP of λ_n and β whereas the same is not possible for CP of α

Topics in Probabilistic Modeling & Inference (CS698X), Spring 2021
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Piyush Kumar Gaurav
Roll Number: 20104442
Date: May 9, 2021

QUESTION 4

My solution to problem 4
Since the samples of $\{\mathbf{U}^{(s)}, \mathbf{V}^{(s)}\}_{s=1}^{S}$ are taken from a Gibbs sampler (Conditional posterior of $\mathbf{U}, \mathbf{V}$). Posterior Predictive Distribution (PPD) can be written as:

$$p(r_{ij}|\mathbf{R}) = \int \underbrace{p(r_{ij}|u_i, v_j)}_{Likelihood} \underbrace{p(u_i, v_j|\mathbf{R})}_{Posterior} \, du_i \, dv_j$$

Mean of the PPD can be written as:

$$\begin{aligned} Mean &= \mathbb{E}_{Posterior}[Likelihood] = \mathbb{E}_{p(u_i,v_j|\mathbf{R})}[p(r_{ij}|u_i, v_j)] \\ &= \mathbb{E}_{Posterior}[{u_i^{(s)}}^T v_j^{(s)} + \epsilon_{ij}] \\ &= \mathbb{E}_{Posterior}[{u_i^{(s)}}^T v_j^{(s)}] + \cancelto{0}{\mathbb{E}_{Posterior}[\epsilon_{ij}]} \\ &= \frac{1}{S}\sum_{s=1}^{S} {u_i^{(s)}}^T v_j^{(s)} \end{aligned} \tag{20}$$

Variance of the PPD can be written as

$$\begin{aligned} Variance &= \mathrm{Var}_{Posterior}[Likelihood] = \mathrm{Var}_{p(u_i,v_j|\mathbf{R})}[p(r_{ij}|u_i, v_j)] \\ &= \mathrm{Var}_{Posterior}[u_i^T v_j + \epsilon_{ij}] \\ &= \mathrm{Var}_{Posterior}[{u_i}^T v_j] + \mathrm{Var}_{Posterior}[\epsilon_{ij}] \\ &\text{(Considering quantities } ({u_i}^T v_j) \text{ and } \epsilon_{ij} \text{ to be independent)} \\ &= \mathrm{Var}[u_i]\,\mathrm{Var}[v_j] + \mathrm{Var}[u_i]\mathbb{E}[v_j]^2 + \mathrm{Var}[v_j]\mathbb{E}[u_i]^2 + \mathrm{Var}[\epsilon_{ij}] \\ &\text{(Representing } \mathrm{Var}_{Posterior} \; as \; \mathrm{Var} \text{ and } \mathbb{E}_{Posterior} \text{ as } \mathbb{E} \text{ for brevity and} \\ &\quad \text{considering quantities } u_i \text{ and } v_j \text{ to be independent)} \end{aligned} \tag{21}$$

where

$$\mathbb{E}[u_i] = \frac{1}{S}\sum_{s=1}^{S} u_i^{(s)}$$

$$\mathbb{E}[v_j] = \frac{1}{S}\sum_{s=1}^{S} v_j^{(s)}$$

$$\mathrm{Var}[u_i] = \frac{1}{S}\sum_{s=1}^{S} \Big[(u_i^{(s)} - \mathbb{E}[u_i])^T (u_i^{(s)} - \mathbb{E}[u_i])\Big]$$

$$\mathrm{Var}[v_j] = \frac{1}{S}\sum_{s=1}^{S} \Big[(v_j^{(s)} - \mathbb{E}[v_j])^T (v_j^{(s)} - \mathbb{E}[v_j])\Big]$$

$$\mathrm{Var}[\epsilon_{ij}] = \beta^{-1}$$

Student Name: Piyush Kumar Gaurav
Roll Number: 20104442
Date: May 9, 2021

QUESTION 5

My solution to problem 5
Given Distribution is $p(x) \propto exp(sin(x))$. Hence,

$$\tilde{p}(x) = exp(sin(x))$$

In order to find maximum when we differentiate $\tilde{p}(x)$. We get, $exp(sin(x))cos(x)$, for which Max is at $\frac{\pi}{2}$ and Min at $x = -\frac{\pi}{2}$ If $q(x)$ is proposal density and M be an real no. then the following condition must hold for Rejection sampling:

$$\begin{aligned} Mq(x) &\geq \tilde{p}(x) \\ M &\geq \frac{\tilde{p}(x)}{q(x)} \\ M &\geq \frac{exp(sin(x))}{\frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2} x^2)} \\ M &\geq \sqrt{2\pi\sigma^2}\, exp\left(sin(x) + \frac{x^2}{2\sigma^2}\right) \end{aligned} \tag{22}$$

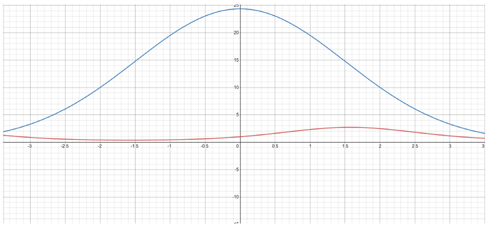
The terms inside the exponential can be put to limit by fixing x to specific values. This can be done by considering the following two points

1. The first term $sin(x)$ can be put to max of $\tilde{p}(x)$ i.e. $x = \frac{\pi}{2}$
2. The second term $\frac{x^2}{2\sigma^2}$ can be used as bias value such the $Mq(x)$ lies above $\tilde{p}(x)$ at its boundaries i.e. $-\pi$ and π, Hence using x as pi we get $\frac{x^2}{2\sigma^2} = \frac{\pi^2}{2\sigma^2}$

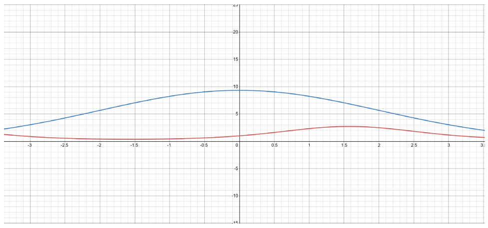
Thus finally we get,

$$M \geq \sqrt{2\pi\sigma^2}\, exp\left(1 + \frac{\pi^2}{2\sigma^2}\right) \tag{23}$$

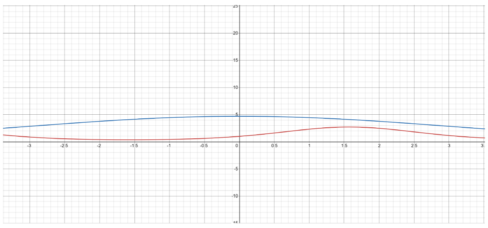
Following are some of the plots we can observe of different values of σ. Also note that $Mq(x)$ is symmetrical w.r.t. σ. So $Mq(x)$ just depends on the magnitude of σ and not on its sign. (courtesy: https://www.desmos.com/calculator)



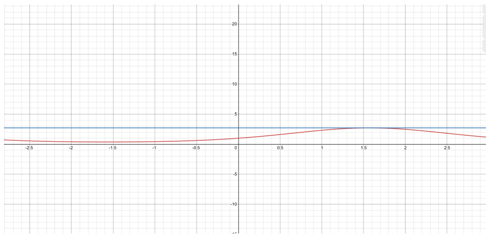
(a) $\sigma = 1.5$.



(b) $\sigma = 2$.



(c) $\sigma = 3$.



(d) $\sigma = \text{inf}$.

Figure 1: Plot showing $\tilde{p}(x)$ in red and $Mq(x)$ in blue for different σ.

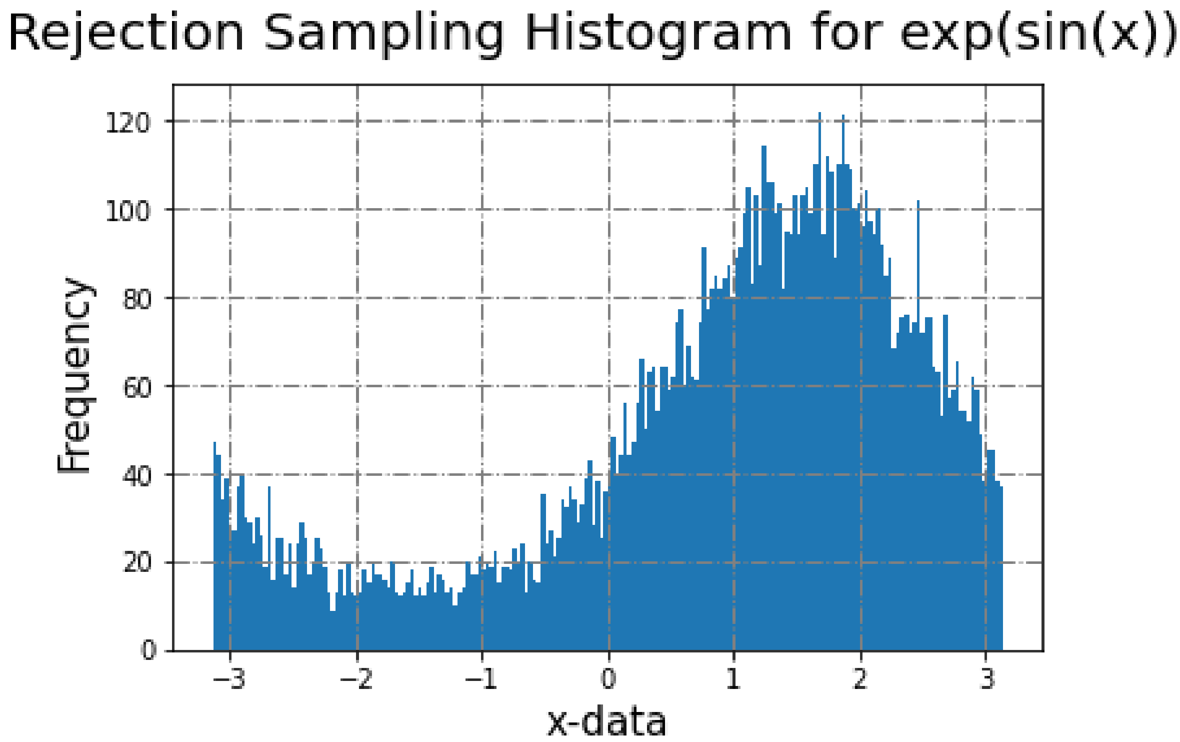


Figure 2: Rejection Sampling Histogram for exp(sin(x)) with $\sigma = 3$ (based on plots from Figure 1) and the value of $M = 40$ (tuned based on Equation (23))