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

Student Name: Bholeshwar Khurana
Roll Number: 170214
Date: June 7, 2021

QUESTION 1

The given problem is

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

Plugging in the expression of KL, the problem becomes:

$$\begin{aligned}
q^*(\theta) &= \underset{q(\theta)}{\arg\min} \quad -\sum_{n=1}^{N}\left[\int q(\theta)log(p(\mathbf{x}_n|\theta))d\theta\right] - \int q(\theta)\log\left(\frac{p(\theta)}{q(\theta)}\right)d\theta \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\sum_{n=1}^{N} log(p(\mathbf{x}_n|\theta))d\theta - \int q(\theta)\log\left(\frac{p(\theta)}{q(\theta)}\right)d\theta \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)log(p(\mathbf{X}|\theta))d\theta - \int q(\theta)\log\left(\frac{p(\theta)}{q(\theta)}\right)d\theta \quad \text{(Observations are i.i.d.)} \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\log\left(\frac{p(\mathbf{X}|\theta)p(\theta)}{q(\theta)}\right)d\theta \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\log\left(\frac{p(\theta|\mathbf{X})p(\mathbf{X})}{q(\theta)}\right)d\theta \quad \text{(Using Bayes' rule)} \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\log\left(\frac{p(\theta|\mathbf{X})}{q(\theta)}\right)d\theta - \int q(\theta)\log(p(\mathbf{X}))d\theta \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\log\left(\frac{p(\theta|\mathbf{X})}{q(\theta)}\right)d\theta - \log(p(\mathbf{X})) \quad \left(\text{since } \int q(\theta)d\theta = 1\right) \\
&= \underset{q(\theta)}{\arg\min} \quad -\int q(\theta)\log\left(\frac{p(\theta|\mathbf{X})}{q(\theta)}\right)d\theta \quad \left(\text{since } p(\mathbf{X}) \text{ doesn't depend on } q(\theta)\right) \\
&= \underset{q(\theta)}{\arg\min} \quad KL\left(q(\theta)||p(\theta|\mathbf{X})\right) \\
&= p(\theta|\mathbf{X})
\end{aligned}$$

Hence solving the above problem is equivalent to finding the posterior distribution of θ, i.e. $q^*(\theta) = p(\theta|\mathbf{X})$.

The given objective function tries to maximize $\sum_{n=1}^{N}\left[\int q(\theta)log\left(p(\mathbf{x}_n|\theta)\right)d\theta\right]$, i.e. $q(\theta)$ tries to explain the data well. Also the objective function tries to minimize $KL\left(q(\theta)||p(\theta)\right)$, keeping $q(\theta)$ close to the prior $p(\theta)$.

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

Student Name: Bholeshwar Khurana
Roll Number: 170214
Date: June 7, 2021

QUESTION 2

Given for N i.i.d. observations $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$:

$$\begin{aligned}
\text{Likelihood model:} \quad & p(y_n|\mathbf{w},\mathbf{x}_n) = \mathcal{N}\left(y_n \middle| \mathbf{w}^\top \mathbf{x}_n, \beta^{-1}\right) \\
& p(\mathbf{y}|\mathbf{w},\mathbf{X}) = \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{w}, \beta^{-1}\mathbf{I}_N) \\
\text{Prior on } \mathbf{w}: \quad & p(\mathbf{w}) = \mathcal{N}\left(\mathbf{w}|0, \operatorname{diag}(\alpha_1^{-1}, \alpha_2^{-1}, \ldots, \alpha_D^{-1})\right) \\
\text{Prior on } \beta: \quad & p(\beta) = \operatorname{Gamma}(\beta|a_0, b_0) \\
\text{Prior on } \alpha_d: \quad & p(\alpha_d) = \operatorname{Gamma}(\alpha_d|e_0, f_0), \ \forall d \\
& \operatorname{Gamma}(\eta|\tau_1,\tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} \exp(-\tau_2 \eta)
\end{aligned}$$

Firstly, we need to find the joint distribution $p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \cdots, \alpha_D|\mathbf{X})$. Using the chain rule of probability, we can write

$$\begin{aligned}
p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \ldots, \alpha_D|\mathbf{X}) &= p(\mathbf{y}|\mathbf{w}, \beta, \mathbf{X}) p(\mathbf{w}|\alpha_1, \ldots, \alpha_D) p(\beta) p(\alpha_1, \ldots, \alpha_D) \\
&= p(\mathbf{y}|\mathbf{w}, \beta, \mathbf{X}) p(\mathbf{w}|\alpha_1, \ldots, \alpha_D) p(\beta) \prod_{d=1}^{D} p(\alpha_d)
\end{aligned}$$

Taking log on both sides. Let $M = \log p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \ldots, \alpha_D|\mathbf{X})$

$$\begin{aligned}
M &= \log p(\mathbf{y}|\mathbf{w}, \beta, \mathbf{X}) + \log p(\mathbf{w}|\alpha_1, \ldots, \alpha_D) + \log p(\beta) + \sum_{d=1}^{D} \log p(\alpha_d) \\
&= \log \left(\sqrt{\frac{\beta^N}{(2\pi)^N}} \exp\left(\frac{-\beta}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) \right) \right) + \log \left(\sqrt{\frac{\alpha_1 \ldots \alpha_D}{(2\pi)^D}} \exp\left(\frac{-\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}{2} \right) \right) \\
&\quad + \sum_{d=1}^{D} \log \left(\frac{f_0^{e_0}}{\Gamma(e_0)} \alpha_d^{e_0 - 1} \exp(-f_0 \alpha_d) \right) + \log \left(\frac{b_0^{a_0}}{\Gamma(a_0)} \beta^{a_0 - 1} \exp(-b_0 \beta) \right) \\
&\propto \frac{N}{2} \log \beta - \frac{\beta}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) + \sum_{d=1}^{D} \left[\left(\frac{1}{2} + (e_0 - 1) \right) \log \alpha_d - f_0 \alpha_d \right] - \frac{1}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\
&\quad + (a_0 - 1) \log \beta - b_0 \beta \qquad \text{(Ignoring constants)}
\end{aligned}$$

where, $\boldsymbol{\Sigma} = \operatorname{diag}(\alpha_1, \ldots, \alpha_D)$

We now need to compute mean-field updates for each parameter (while keeping other parameters fixed):

1. For $\mathbf{w}$:

$$\begin{aligned}
\log q^*_{\mathbf{w}}(\mathbf{w}) &= \mathbb{E}_{q_{\beta,\alpha}}\left[\log p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \ldots, \alpha_D|\mathbf{X})\right] && \text{(Ignoring constants)} \\
&= \mathbb{E}_{q_{\beta,\alpha}}\left[-\frac{\beta}{2}\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right)^\top\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right) - \frac{1}{2}\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}\right] \\
&= -\frac{1}{2}\left[\mathbf{w}^\top\left(\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{X} + \mathrm{diag}(\mathbb{E}[\alpha_1], \ldots, \mathbb{E}[\alpha_D])\right)\mathbf{w} + 2\mathbf{w}^\top\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{y}\right]
\end{aligned}$$

This shows that $q^*(\mathbf{w})$ has a Gaussian form:

$$\begin{aligned}
q^*(\mathbf{w}) &= \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}) \\
\text{where, } \boldsymbol{\Sigma}_{\mathbf{w}} &= \left(\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{X} + \mathrm{diag}\left(\mathbb{E}[\alpha_1], \ldots, \mathbb{E}[\alpha_D]\right)\right)^{-1} \\
\boldsymbol{\mu}_{\mathbf{w}} &= \boldsymbol{\Sigma}_{\mathbf{w}}\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{y}
\end{aligned}$$

2. For β:

$$\begin{aligned}
\log q^*_\beta(\beta) &= \mathbb{E}_{q_{\mathbf{w},\alpha}}[\log p(\mathbf{y}, \mathbf{w}, \beta, \alpha_1, \ldots, \alpha_D|\mathbf{X})] && \text{(Ignoring constants)} \\
&= \mathbb{E}_{q_{\mathbf{w},\alpha}}\left[-\frac{\beta}{2}\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right)^\top\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right) + \frac{N}{2}\log\beta + (a_0-1)\log\beta - b_0\beta\right] \\
&= \left(\frac{N}{2} + a_0 - 1\right)\log\beta - \beta\left(\frac{1}{2}\mathbb{E}_{\mathbf{w}}\left[\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right)^\top\left(\mathbf{y}-\mathbf{X}\mathbf{w}\right)\right] + b_0\right) \\
&= \left(\frac{N}{2} + a_0 - 1\right)\log\beta - \beta\left(\frac{1}{2}\mathbb{E}_{\mathbf{w}}\left[\mathbf{y}^\top\mathbf{y} + \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w} - 2\mathbf{w}^\top\mathbf{X}^\top\mathbf{y}\right] + b_0\right) \\
&= \left(\frac{N}{2} + a_0 - 1\right)\log\beta - \beta\left(\frac{1}{2}\mathrm{Trace}\left(\mathbf{y}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\mathbb{E}\left[\mathbf{w}\mathbf{w}^\top\right] - 2\mathbb{E}[\mathbf{w}]^\top\mathbf{X}^\top\mathbf{y}\right) + b_0\right)
\end{aligned}$$

This shows that $q^*_\beta(\beta)$ has a Gamma form:

$$\begin{aligned}
q^*_\beta(\beta) &= \mathrm{Gamma}(\beta|a, b) \\
\text{where, } a &= \frac{N}{2} + a_0 \\
b &= \frac{1}{2}\mathrm{Trace}\left(\mathbf{y}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\mathbb{E}\left[\mathbf{w}\mathbf{w}^\top\right] - 2\mathbb{E}[\mathbf{w}]^\top\mathbf{X}^\top\mathbf{y}\right) + b_0
\end{aligned}$$

3. For α_d:

$$\begin{aligned}
\log q^*_{\alpha_d}(\alpha_d) &= \mathbb{E}_{q_{\mathbf{w},\beta,\alpha_1,\ldots,\alpha_{d-1},\alpha_{d+1},\ldots,\alpha_D}}\left[\log p\left(\mathbf{y}, \mathbf{w}, \alpha, \beta|\mathbf{X}\right)\right] && \text{(Ignoring constants)} \\
&= \mathbb{E}_{q_{\mathbf{w},\beta,\alpha_1,\ldots,\alpha_{d-1},\alpha_{d+1},\ldots,\alpha_D}}\left[\left(e_0 - \frac{1}{2}\right)\log\alpha_d - f_0\alpha_d - \frac{w_d^2\alpha_d}{2}\right] \\
&= \left(\frac{1}{2} + e_0 - 1\right)\log\alpha_d - \alpha_d\left(f_0 + \frac{\mathbb{E}[w_d^2]}{2}\right)
\end{aligned}$$

This shows that $q^*_{\alpha_d}(\alpha_d)$ has a Gamma form:

$$\begin{aligned}
q^*_{\alpha_d}(\alpha_d) &= \text{Gamma}(\alpha_d|e_d, f_d) \\
\text{where, } e_d &= \frac{1}{2} + e_0 \\
f_d &= f_0 + \frac{\mathbb{E}[w_d^2]}{2}
\end{aligned}$$

Since $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{w}), \beta \sim \text{Gamma}(\beta|a, b)$ and $\alpha_d \sim \text{Gamma}(\alpha_d|e_d, f_d)$, the required expectations are:

$$\begin{aligned}
\mathbb{E}[\mathbf{w}] &= \boldsymbol{\mu}_\mathbf{w} \\
\mathbb{E}[\mathbf{w}\mathbf{w}^\top] &= \boldsymbol{\Sigma}_\mathbf{w} + \boldsymbol{\mu}_\mathbf{w}\boldsymbol{\mu}_\mathbf{w}^\top \\
\mathbb{E}[w_d^2] &= (\boldsymbol{\Sigma}_\mathbf{w})_{dd} + (\boldsymbol{\mu}_\mathbf{w})_d^2 \\
\mathbb{E}[\beta] &= \frac{a}{b} \\
\mathbb{E}[\alpha_d] &= \frac{e_d}{f_d} \quad \forall d
\end{aligned}$$

Following is the final Mean-field VI algorithm:

Algorithm 1: Mean-field VI

Input: $\mathbf{y}$, $\mathbf{X}$ and hyper-parameters a_0, b_0, e_0, f_0

Set $e_d = e_0 + \frac{1}{2}\ \forall d$ and $a = a_0 + \frac{N}{2}$

Initialize b and $f_d \ \ \forall d$

while *(NOT CONVERGED)* **do**

$$\begin{aligned}
\mathbb{E}[\beta] &= \frac{a}{b} \\
\mathbb{E}[\alpha_d] &= \frac{e_d}{f_d} \quad \forall d \\
\boldsymbol{\Sigma}_\mathbf{w} &= \left(\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{X} + \text{diag}(\mathbb{E}[\alpha_1], \ldots, \mathbb{E}[\alpha_D])\right)^{-1} \\
\boldsymbol{\mu}_\mathbf{w} &= \boldsymbol{\Sigma}_\mathbf{w}\left(\mathbb{E}[\beta]\mathbf{X}^\top\mathbf{y}\right) \\
\mathbb{E}[\mathbf{w}] &= \boldsymbol{\mu}_\mathbf{w} \\
\mathbb{E}[\mathbf{w}\mathbf{w}^\top] &= \boldsymbol{\Sigma}_\mathbf{w} + \boldsymbol{\mu}_\mathbf{w}\boldsymbol{\mu}_\mathbf{w}^\top \\
\mathbb{E}[w_d^2] &= (\boldsymbol{\Sigma}_\mathbf{w})_{dd} + (\boldsymbol{\mu}_\mathbf{w})_d^2 \quad \forall d \\
b &= \frac{1}{2}\text{Trace}\left(\mathbf{y}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\mathbb{E}\left[\mathbf{w}\mathbf{w}^\top\right] - 2\mathbb{E}[\mathbf{w}]^\top\mathbf{X}^\top\mathbf{y}\right) + b_0 \\
f_d &= f_0 + \frac{\mathbb{E}[w_d^2]}{2} \quad \forall d
\end{aligned}$$

end

Return $\{\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{w}\}, \{a, b\}, \{e_d, f_d\}_{d=1}^D$

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

Student Name: Bholeshwar Khurana
Roll Number: 170214
Date: June 7, 2021

QUESTION 3

Given for N i.i.d. observations $\{x_n\}_{n=1}^{N}$:

$$\begin{aligned}
\text{Likelihood model:} \quad & p(x_n|\lambda_n) = \text{Poisson}(x_n|\lambda_n) \\
\text{Prior on } \lambda_n\text{:} \quad & p(\lambda_n|\alpha,\beta) = \text{Gamma}(\lambda_n|\alpha,\beta) \\
\text{Prior on } \alpha\text{:} \quad & p(\alpha|a,b) = \text{Gamma}(\alpha|a,b) \\
\text{Prior on } \beta\text{:} \quad & p(\beta|c,d) = \text{Gamma}(\beta|c,d)
\end{aligned}$$

1. Conditional Posterior of λ_n:

$$\begin{aligned}
p(\lambda_n|\mathbf{X},\lambda_{-n},\alpha,\beta) &\propto p(\mathbf{X}|\lambda_n)p(\lambda_n|\alpha,\beta) && \text{(Using Bayes' rule)} \\
&\propto p(x_n|\lambda_n)p(\lambda_n|\alpha,\beta) && \text{(Since only } x_n \text{ depends on } \lambda_n) \\
&\propto \frac{\exp(-\lambda_n)\lambda_n^{x_n}}{x_n!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_n^{\alpha-1}\exp\left(-\beta\lambda_n\right) \\
&\propto \lambda_n^{\alpha+x_n-1}\exp\left(-\left(\beta+1\right)\lambda_n\right)
\end{aligned}$$

This gives the conditional posterior of λ_n as

$$p(\lambda_n|\mathbf{X},\lambda_{-n},\alpha,\beta) = \text{Gamma}\left(\lambda_n|\alpha+x_n,\beta+1\right)$$

2. Conditional Posterior of β:

$$\begin{aligned}
p(\beta|\mathbf{X},\boldsymbol{\lambda},\alpha) &\propto p(\boldsymbol{\lambda}|\alpha,\beta)p(\beta) \\
&\propto \prod_{n=1}^{N} p(\lambda_n|\alpha,\beta)p(\beta) && \text{(Since } \lambda_n\text{s are i.i.d.)} \\
&\propto \prod_{n=1}^{N}\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_n^{\alpha-1}\exp\left(-\beta\lambda_n\right)\right]\frac{d^c}{\Gamma(c)}\beta^{c-1}\exp\left(-\beta d\right) \\
&\propto \left[\beta^{N\alpha}\exp\left(-\beta\sum_{n=1}^{N}\lambda_n\right)\right]\beta^{c-1}\exp\left(-\beta d\right) \\
&\propto \beta^{N\alpha+c-1}\exp\left(-\left(d+\sum_{n=1}^{N}\lambda_n\right)\beta\right)
\end{aligned}$$

This gives the conditional posterior of β as

$$p(\beta|\mathbf{X},\boldsymbol{\lambda},\alpha) = \text{Gamma}\left(\beta\middle|N\alpha+c, d+\sum_{n=1}^{N}\lambda_n\right)$$

3. Conditional Posterior of α:

$$
\begin{aligned}
p(\alpha|\mathbf{X}, \boldsymbol{\lambda}, \beta) &\propto p(\boldsymbol{\lambda}|\alpha, \beta)p(\alpha) \\
&= \prod_{n=1}^{N} \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda_n^{\alpha-1} \exp\left(-\beta\lambda_n\right) \right] \frac{b^a}{\Gamma(a)} \alpha^{a-1} \exp\left(-\alpha b\right) \\
&\propto \left[\frac{\beta^{n\alpha} \left(\prod_{n=1}^{N} \lambda_n \right)^{\alpha-1}}{\Gamma(\alpha)^N} \right] \alpha^{a-1} \exp\left(-\alpha b\right)
\end{aligned}
$$

This is not any standard known form. Hence, CPs for all the parameters except α are available in closed forms.

Following is the Gibbs sampling algorithm for this model:

Algorithm 2: Gibbs Sampling

Input: $\mathbf{X}$ and hyper-parameters a, b, c, d

Initialize $\alpha^{(0)}, \beta^{(0)}$

for $t = 1, 2, \ldots, T$ **do**

$$
\begin{aligned}
\lambda_n^{(t)} &\sim \text{Gamma}\left(\lambda_n \middle| \alpha^{(t-1)} + x_n, \beta^{(t-1)} + 1\right) \quad \forall n \\
\beta^{(t)} &\sim \text{Gamma}\left(\beta \middle| N\alpha^{(t-1)} + c, d + \sum_{n=1}^{N} \lambda_n^{(t)}\right) \\
\alpha^{(t)} &\sim p\left(\alpha \middle| \mathbf{X}, \boldsymbol{\lambda}^{(t)}, \beta^{(t)}\right) \qquad \text{(Use any sampling method for this)}
\end{aligned}
$$

end

Return $\left\{ \lambda_n^{(t)}, \beta^{(t)}, \alpha^{(t)} \right\}_{t=1}^{\top}$

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

QUESTION 4

Student Name: Bholeshwar Khurana
Roll Number: 170214
Date: June 7, 2021

Finding expectation of r_{ij}:

$$\begin{aligned}\mathbb{E}[r_{ij}] &= \mathbb{E}\left[\mathbf{u}_i^\top \mathbf{v}_j + \epsilon_{ij}\right] \\ &= \mathbb{E}\left[\mathbf{u}_i^\top \mathbf{v}_j\right] + \mathbb{E}\left[\epsilon_{ij}\right]\end{aligned}$$

Since $\epsilon_{ij} \sim \mathcal{N}\left(\epsilon_{ij} \middle| 0, \beta^{-1}\right)$, hence $\mathbb{E}[\epsilon_{ij}] = 0$ and using Monte-Carlo approximation on the known samples $\mathbf{U}^{(s)} = \left\{\mathbf{u}_i^{(s)}\right\}_{i=1}^N$ and $\mathbf{V}^{(s)} = \left\{\mathbf{v}_j^{(s)}\right\}_{j=1}^N$, we can write the expectation as:

$$\mathbb{E}[r_{ij}] \approx \frac{1}{S}\sum_{s=1}^{S} \mathbf{u}_i^{(s)\top} \mathbf{v}_j^{(s)}$$

Finding variance of r_{ij}:
We know that $\text{var}(r_{ij}) = \mathbb{E}\left[r_{ij}^2\right] - \mathbb{E}\left[r_{ij}\right]^2$. Finding $\mathbb{E}\left[r_{ij}^2\right]$:

$$\begin{aligned}\mathbb{E}\left[r_{ij}^2\right] &= \mathbb{E}\left[\left(\mathbf{u}_i^\top \mathbf{v}_j + \epsilon_{ij}\right)^2\right] \\ &= \mathbb{E}\left[\left(\mathbf{u}_i^\top \mathbf{v}_j\right)^2 + \epsilon_{ij}^2 + 2\epsilon_{ij}\left(\mathbf{u}_i^\top \mathbf{v}_j\right)\right] \\ &= \mathbb{E}\left[\left(\mathbf{u}_i^\top \mathbf{v}_j\right)^2\right] + \mathbb{E}\left[\epsilon_{ij}^2\right] + 2\mathbb{E}\left[\epsilon_{ij}\right]\mathbb{E}\left[\mathbf{u}_i^\top \mathbf{v}_j\right] && (\mathbf{u}_i^\top \mathbf{v}_j \text{ and } \epsilon_{ij} \text{ are independent}) \\ &= \mathbb{E}\left[\left(\mathbf{u}_i^\top \mathbf{v}_j\right)^2\right] + \beta^{-1} && \left(\mathbb{E}\left[\epsilon_{ij}\right] = 0, \mathbb{E}\left[\epsilon_{ij}^2\right] = \beta^{-1}\right) \\ &\approx \frac{1}{S}\sum_{s=1}^{S}\left(\mathbf{u}_i^{(s)\top} \mathbf{v}_j^{(s)}\right)^2 + \beta^{-1} && \text{(Using Monte-Carlo approximation)}\end{aligned}$$

Hence, we can write the variance as:

$$\begin{aligned}\text{var}(r_{ij}) &= \mathbb{E}\left[r_{ij}^2\right] - \mathbb{E}[r_{ij}]^2 \\ \text{var}(r_{ij}) &\approx \frac{1}{S}\sum_{s=1}^{S}\left(\mathbf{u}_i^{(s)\top} \mathbf{v}_j^{(s)}\right)^2 + \beta^{-1} - \frac{1}{S^2}\left[\sum_{s=1}^{S} \mathbf{u}_i^{(s)\top} \mathbf{v}_j^{(s)}\right]^2\end{aligned}$$

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

QUESTION 5

Student Name: Bholeshwar Khurana
Roll Number: 170214
Date: June 7, 2021

Given: $\tilde{p}(x) = \exp(\sin(x))$ and the proposal distribution $q(x) = \mathcal{N}\left(x \middle| 0, \sigma^2\right)$.
We need to find an optimal M such that $Mq(x) \geq \tilde{p}(x)$ for $x \in [-\pi, \pi]$:

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

Since $\sin(x) \leq 1$ and $x^2 \leq \pi^2$ (in the given domain of x), we get

$$\frac{\tilde{p}(x)}{q(x)} \geq \sqrt{2\pi\sigma^2} \exp\left(1 + \frac{\pi^2}{2\sigma^2}\right)$$

For optimal value of M, we want that $Mq(x)$ envelopes $\tilde{p}(x)$ tightly, hence

$$M = \sqrt{2\pi\sigma^2} \exp\left(1 + \frac{\pi^2}{2\sigma^2}\right)$$

The histogram of 10000 samples drawn from $p(x)$ for $\sigma = 2$ is:

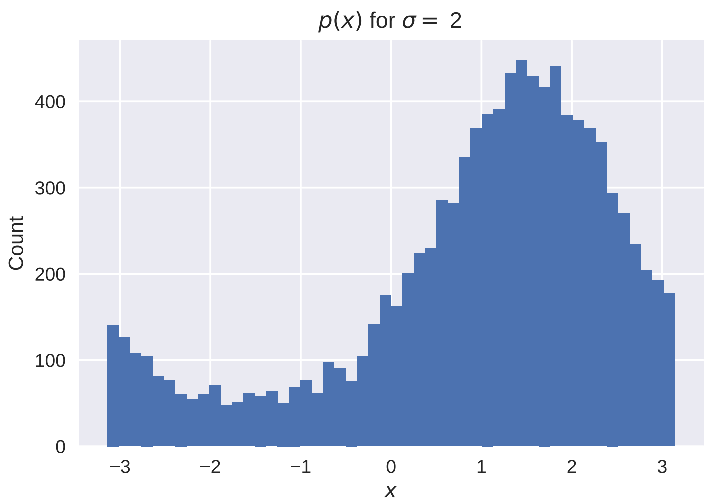

Figure 1: Histogram of 10000 samples drawn from $p(x)$ for $\sigma = 2$