Probabilistic Machine Learning (CS772A), Fall 2022
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Subhajyoti Saha
Roll Number: 21111269
Date: November 18, 2022

QUESTION
1

1 Monte-Carlo Approximations (15 marks)

$$
\begin{aligned}
\hat{f} &= \frac{1}{S}\sum_{s=1}^{S} f(z^{(s)}) \\
\Rightarrow E[\hat{f}] &= \frac{1}{S}\sum_{s=1}^{S} E[f(z^{(s)})] \\
&= \frac{1}{S} S.E[f] \\
&= E[f] \text{ [Prooved]}
\end{aligned}
$$

It is prooved that the MCMC preserves the mean.

$$
\begin{aligned}
Var[\hat{f}] &= \frac{1}{S^2}\sum_{i=1}^{S} Var[f(z^{(s)}] \\
&= \frac{1}{S^2} S.Var[f] \\
&= \frac{1}{S} Var[f] \\
&= \frac{1}{S} E[(f - E[f])^2]
\end{aligned}
$$

From the above equation, it is proven that the variation of MCMC samplers got reduced as the no of samples got incresed.

Probabilistic Machine Learning (CS772A), Fall 2022
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Subhajyoti Saha
Roll Number: 21111269
Date: November 18, 2022

QUESTION 2

$$
\begin{aligned}
\ln q^*(w) &= E_{\beta,\alpha_1,...\alpha_D}[\ln P(y, X, w, \beta, \alpha_1, ..\alpha_D)] + const \\
&= E_{\beta,\alpha_1,...\alpha_D}[\ln P(y|X, w)P(w)P(\beta)P(\alpha_1)...P(\alpha_D)] + const \\
&= \int q(\beta) \ln P(y|X, w) d\beta + \int q(\alpha_1)...q(\alpha_D) \ln P(w) d\alpha_1...d\alpha_D + const \\
&= \sum_{i=1}^{n} \int q(\beta) \ln P(y_i|x_i, w) d\beta + \int q(\alpha_1)...q(\alpha_D) \sum_{i=1}^{D} \ln P(w_i|\alpha_i) d\alpha_1...d\alpha_D + const \\
&= \sum_{i=1}^{n} \int q(\beta) \ln P(y_i|x_i, w) d\beta + \sum_{i=1}^{D} \int q(\alpha_i) \ln P(w_i|\alpha_i) d\alpha_i + const \\
&= \sum_{i=1}^{n} \int q(\beta) \ln N(y_i|w_i^T x_i, \beta^{-1}) d\beta + \sum_{i=1}^{D} \int q(\alpha_i) \ln N(w_i|0, \alpha_i^{-1}) d\alpha_i + const \\
&= \sum_{i=1}^{n} -E_\beta[\frac{\beta}{2}(y_i - w^T x_i)^2] - \sum_{i=1}^{D} E_{\alpha_i}[\frac{\alpha_i}{2} w_i^2] + const \\
&= -\sum_{i=1}^{n} \frac{E_\beta[\beta]}{2}(y_i - w^T x_i)^2 - \sum_{i=1}^{D} \frac{E_{\alpha_i}[\alpha_i]}{2} w_i^2 + const \\
&= -\sum_{i=1}^{n} \frac{E_\beta[\beta]}{2}(w^T x_i x_i^T w - 2 y_i w^T x_i) - \sum_{i=1}^{D} \frac{E_{\alpha_i}[\alpha_i]}{2} w_i^2 + const \\
&= -\frac{1}{2} w^T [E_\beta[\beta] \sum_{i=1}^{n} x_i x_i^T + diag(E_{\alpha_i}[\alpha_i], ...E_{\alpha_i}[\alpha_i])] w + w^T (\sum_{i=1}^{n} y_i x_i) E_\beta[\beta] + const \\
\Rightarrow \ln q^*(w) &= \ln N(w|\mu, \Sigma) \\
\Rightarrow (\Sigma^{-1}) &= \Big[E_\beta[\beta] \sum_{i=1}^{n} x_i x_i^T + diag(E_{\alpha_i}[\alpha_i], ...E_{\alpha_i}[\alpha_i]) \Big] \\
\mu &= \Sigma (\sum_{i=1}^{n} y_i x_i) E_\beta[\beta]
\end{aligned}
$$

$$
\begin{aligned}
\ln q^*(\beta) &= E_{w,\alpha_1,...\alpha_D}[\ln P(y, X, w, \beta, \alpha_1, ...\alpha_D)] + const \\
&= E_w[\ln P(y|X, w) + \ln P(\beta)] + const \\
&= \sum_{i=1}^{n} E_w[\ln N(y_i|w^T x_i, \beta^{-1})] + \ln Gamma(\beta|a_o, b_o) + const \\
&= \frac{n}{2}\ln\beta - \frac{\beta}{2}\sum_{i=1}^{n} E_w[(y_i - w^T x_i)^2] + (a_o - 1)\ln\beta - b_o \ln\beta \\
&= \frac{n}{2}\ln\beta - \frac{\beta}{2}\sum_{i=1}^{n}(y_i - E_w[w]^T x_i)^2 + (a_o - 1)\beta - b_o \ln\beta + const \\
&= (\frac{n}{2} + a_o - 1)\ln\beta - (b_o + \frac{1}{2}\sum_{i=1}^{n}(y_i - E_w[w]^T x_i)^2)\beta + const
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \ln q^*(\beta) &= \ln Gamma(\beta|\hat{a}, \hat{b}) \\
\Rightarrow \hat{a} &= (\frac{n}{2} + a_o) \\
\Rightarrow \hat{b} &= (b_o + \frac{1}{2}\sum_{i=1}^{n}(y_i - E_w[w]^T x_i)^2)
\end{aligned}
$$

$$
\begin{aligned}
\ln q^*(\alpha_i) &= E_{w,\beta,\alpha_{j:j\neq i}}[\ln P(y, X, w, \beta, \alpha_{j:j\neq i})] + const \\
&= E_w[\ln P(w) + \ln P(\alpha_i)] + const \\
&= E_w[\sum_{i=1}^{D} \ln P(w_i) + \ln P(\alpha_i)] + const \\
&= E_w[\sum_{i=1}^{D} \ln N(w_i|0, \alpha_i) + \ln P(\alpha_i)] + const \\
&= E_w[\ln N(w_i|0, \alpha_i) + \ln P(\alpha_i)] + const \\
&= E_w[\frac{1}{2}\ln\alpha_i - \frac{\alpha_i}{2} w_i^2 + (e_o - 1)\ln\alpha_i - f_o\alpha_i] + const \\
&= (\frac{1}{2} + e_o - 1)\ln\alpha_i - (\frac{E_w[w_i^2]}{2} + f_o)\alpha_i + const
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \ln q^*(\alpha_i) &= \ln Gamma(\alpha_i|\hat{e}, \hat{f}) \\
\Rightarrow \hat{e} &= \frac{1}{2} + e_o \\
\Rightarrow \hat{f} &= \frac{E_w[w_i^2]}{2} + f_o
\end{aligned}
$$

The above distribution is same for all q$^*(\alpha_i) for i = 1, ...D$.

Probabilistic Machine Learning (CS772A), Fall 2022
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Subhajyoti Saha
Roll Number: 21111269
Date: November 18, 2022

QUESTION 3

The CPs of all the variable are shown below:

$$
\begin{aligned}
P(\lambda_i|\lambda_{j_{j\neq i}}, \alpha, \beta, x_1, ..x_n) &= P(\lambda_i|x_1, \alpha, \beta) \\
&= P(x_i|\lambda_i)P(\lambda_i|\alpha, \beta) \\
&= Poisson(x_i|\lambda_i)Gamma(\lambda_i|\alpha, \beta) \\
&= \frac{\lambda_i^{x_i}e^{-\lambda_i}}{x_i!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\frac{\lambda_i}{\beta}} \\
&\propto \lambda_i^{x_i+\alpha-1}e^{-\frac{_i(1+\beta)}{\beta}} \\
\Rightarrow P(\lambda_i|x_1, \alpha, \beta) &= Gamma(_i|\hat{\alpha}, \hat{\beta}) \\
\Rightarrow \hat{\alpha} &= x_i + \alpha \\
\Rightarrow \hat{\beta} &= \frac{\beta}{1+\beta} \\
P(\alpha|\lambda_1, \lambda_2, ..\lambda_N, x_1, ...x_N, \beta) &= P(\alpha|\lambda_1, ...\lambda_N, \beta) \\
&= \prod_{i=1}^{N} P(\lambda_i|\alpha, \beta)P(\alpha)P(\beta) \\
&= \prod_{i=1}^{N} Gamma(\lambda_i|\alpha, \beta)Gamma(\alpha|a, b)Gamma(\beta|c, d)
\end{aligned}
$$

The above CP is not in closed form.

We may use MH to sample from this.

$$
\begin{aligned}
P(\beta|\lambda_1, ..\lambda_N, x_1, ...x_N, \alpha) &= P(\beta|\lambda_1, ..\lambda_N, \alpha) \\
&= \prod_{i=1}^{N} P(\lambda_i|\alpha, \beta)P(\beta)P(\alpha) \\
&= \prod_{i=1}^{N} Gamma(_i|\alpha, \beta)Gamma(\alpha|a, b)Gamma(\beta|c, d)
\end{aligned}
$$

The above CP is also not in closed form.

Probabilistic Machine Learning (CS772A), Fall 2022
Indian Institute of Technology Kanpur
Homework Assignment Number 3

Student Name: Subhajyoti Saha
Roll Number: 21111269
Date: November 18, 2022

QUESTION 4

$$
\begin{aligned}
P(r_{ij}|R) &\approx P(\hat{r_{ij}}|R) \\
&= \frac{1}{S}\sum_{s=1}^{S} P(r_{ij}|u_i^{(s)}, v_j^{(s)}) \\
&= \frac{1}{S}\sum_{s=1}^{S} N(r_{ij}|u_i^{(s)T} v_j^{(s)}, \beta^{-1}) \\
\Rightarrow E[r_{ij}|R)] &= \int r_{ij} p(r_{ij}|R) dr_{ij} \\
&= \frac{1}{S}\int r_{ij} \sum_{s=1}^{S} N(r_{ij}|u_i^{(s)T} v_j^{(s)}, \beta^{-1}) dr_{ij} \\
&= \frac{1}{S}\sum_{s=1}^{S} \int r_{ij} N(r_{ij}|u_i^{(s)T} v_j^{(s)}, \beta^{-1}) dr_{ij} \\
&= \frac{1}{S}\sum_{s=1}^{S} (u_i^{(s)T} v_j^{(s)})
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow Var_{P(\hat{r_{ij}}|R}[r_{ij}|R)] &= Var[u_i^T v_j + \epsilon] \\
&= Var[\epsilon] \text{ [The other one is const.]} \\
&= \frac{1}{\beta}
\end{aligned}
$$