

Homework 3

● Graded

Student

AREEB AHMAD

Total Points

120 / 120 pts

Question 1

Problem 1

15 / 15 pts

✓ **+ 15 pts** Fully correct answer

+ 5 pts Part 1 (unbiased expectation)
correct

+ 4 pts Part 1 mostly correct with some
minor errors

+ 3 pts Part 1 somewhat correct with
some major errors but the basic
approach is correct

+ 10 pts Part 2 (variance of the
approximation) correct

+ 8 pts Part 2 mostly correct with some
minor errors

+ 6 pts Part 2 somewhat correct with
some major errors but the basic
approach is correct

+ 0 pts Incorrect or not attempted

Question 2

Problem 2

30 / 30 pts

✓ + 30 pts Fully correct answer

+ 8 pts Basic steps correct

+ 6 pts Correct mean-field posterior for w

+ 6 pts Correct mean-field posterior for β

+ 6 pts Correct mean-field posterior for α

+ 4 pts Overall approach (the VI algorithm steps) correctly shown with appropriate expectations correctly specified

- 4 pts Didn't have this: Overall approach (the VI algorithm steps) correctly shown with appropriate expectations correctly specified. At least the expectations required in the mean-field distributions should be specified clearly.

+ 0 pts Incorrect or not attempted

Question 3

Problem 3

20 / 20 pts

✓ **+ 20 pts** Fully correct answer for all parts

+ 7 pts Correct CP for each λ

+ 4 pts Somewhat correct CP for λ but
some mistakes

+ 7 pts Correct CP for α

+ 4 pts Somewhat correct CP for α but
some mistakes

+ 6 pts Correct CP for β

+ 4 pts Somewhat correct CP for β but
some mistakes

+ 0 pts Incorrect or not attempted

Question 4

Problem 4

15 / 15 pts

✓ **+ 15 pts** Fully correct answer for all parts

+ 5 pts Correct answer for the mean

+ 4 pts Mostly correct answer for the mean but with some small errors

+ 10 pts Correct answer for the variance

+ 8 pts Mostly correct answer for the variance but with some small errors

+ 0 pts Incorrect or not attempted

Question 5

Problem 5

40 / 40 pts

✓ **+ 20 pts** Correct implementation of the rejection sampler

+ 15 pts Mostly correct implementation of the rejection sampler with some minor mistakes

✓ **+ 20 pts** Correct implementation of the MH sampler, correct plots, and correct answer to the questions about σ^2 and rejection rate

+ 15 pts Mostly correct implementation of the MH sampler with some minor mistakes in implementation or not answering the questions about σ^2 and rejection rate correctly

+ 0 pts Did not attempt

Question assigned to the following page: [1](#)

Student Name: Areeb Ahmad
 Roll Number: 180135
 Date: November 15, 2022

We are approximating $\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})d\mathbf{z}$ with Monte-Carlo's approximation $\hat{f} = \frac{1}{S} \sum_{s=1}^S f(\mathbf{z}^{(s)})$, where S samples are $\mathbf{z}^{(s)} \sim p(\mathbf{z})$ as iid.

Proof1 :

$$\mathbb{E}[f] = \mathbb{E}[\hat{f}] \quad (1)$$

$$\begin{aligned} \mathbb{E}[\hat{f}] &= \mathbb{E}\left[\frac{1}{S} \sum_{s=1}^S f(\mathbf{z}^{(s)})\right] \\ &= \frac{1}{S} \mathbb{E}\left[\sum_{s=1}^S f(\mathbf{z}^{(s)})\right], \text{ by Linearity property} \\ &= \frac{1}{S} \sum_{s=1}^S \mathbb{E}[f] \\ &= \mathbb{E}[f] \end{aligned} \quad (2)$$

Proof2 :

$$\text{var}[\hat{f}] = \frac{1}{S} \mathbb{E}[(f - \mathbb{E}[f])^2] \quad (3)$$

$$\begin{aligned} \text{var}[\hat{f}] &= \mathbb{E}[\hat{f}^2] - \mathbb{E}[\hat{f}]^2 \\ &= \mathbb{E}\left[\left(\frac{1}{S} \sum_{s=1}^S f(\mathbf{z}^{(s)})\right)^2\right] - \mathbb{E}\left[\frac{1}{S} \sum_{s=1}^S f(\mathbf{z}^{(s)})\right]^2 \\ &= \mathbb{E}\left[\frac{1}{S^2} \sum_{s=1}^S f^2(\mathbf{z}^{(s)}) + \frac{1}{S^2} \sum_{s=1}^S \sum_{k=1}^S f(\mathbf{z}^{(s)})f(\mathbf{z}^{(k)})\right] - \mathbb{E}\left[\frac{1}{S} \sum_{s=1}^S f(\mathbf{z}^{(s)})\right]^2 \\ &= \frac{1}{S^2} \sum_{s=1}^S \mathbb{E}[f^2(\mathbf{z}^{(s)})] + \frac{1}{S^2} \sum_{s=1}^S \sum_{k=1}^S \mathbb{E}[f(\mathbf{z}^{(s)})f(\mathbf{z}^{(k)})] - \mathbb{E}\left[\frac{1}{S} \sum_{s=1}^S f(\mathbf{z})\right]^2 \\ &= \frac{1}{S^2} \sum_{s=1}^S \mathbb{E}[f^2(\mathbf{z})] + \frac{1}{S^2} \sum_{s=1}^S \sum_{k=1}^S \mathbb{E}[f(\mathbf{z})]^2 - \mathbb{E}\left[\frac{1}{S} \sum_{s=1}^S f(\mathbf{z})\right]^2 \\ &= \frac{1}{S^2} S \mathbb{E}[f^2(\mathbf{z})] + \frac{1}{S^2} S(S-1) \mathbb{E}[f(\mathbf{z})]^2 - \mathbb{E}[f(\mathbf{z})]^2 \\ &= \frac{1}{S} \mathbb{E}[f^2(\mathbf{z})] - \frac{1}{S} \mathbb{E}[f(\mathbf{z})]^2 \\ &= \frac{1}{S} \text{var}[f] \end{aligned} \quad (4)$$

Question assigned to the following page: [1](#)

hence,

$$\text{var}[\hat{f}] = \frac{1}{S}(\mathbb{E}[f^2] - \mathbb{E}[f]^2) \tag{5}$$

Question assigned to the following page: [2](#)

Student Name: Areeb Ahmad
 Roll Number: 180135
 Date: November 15, 2022

My solution to problem 2
 The joint distribution is given by:

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

Mean field VI updates:

- For β keeping other fixed

$$\begin{aligned}
 \log q_{\beta}^*(\beta) &= E_{\mathbf{w}, \alpha} \left[\frac{N}{2} \log \beta - \frac{\beta}{2} \sum_{n=1}^N (y_n - \mathbf{w}^T \mathbf{x}_n)^2 + (a_0 - 1) \log \beta - b_0 \beta \right] \\
 &= \left(\frac{N}{2} + a_0 - 1 \right) \log \beta - \beta \left[\frac{1}{2} \sum_{n=1}^N \mathbb{E} \left[(y_n - \mathbf{w}^T \mathbf{x}_n)^2 \right] + b_0 \right] \\
 \beta &\sim \text{Gamma} \left(\frac{N}{2} + a_0, \frac{1}{2} \sum_{n=1}^N \mathbb{E} \left[(y_n - \mathbf{w}^T \mathbf{x}_n)^2 \right] + b_0 \right)
 \end{aligned} \tag{7}$$

- For \mathbf{w} keeping others fixed

Question assigned to the following page: [2](#)

$$\begin{aligned}
\log q_{\mathbf{w}}^*(\mathbf{w}) &= E_{\alpha, \beta} [p(\mathbf{y}, \mathbf{w}, \alpha, \beta | \mathbf{X})] \\
&= -\frac{1}{2} \left[\mathbf{w}^T \left(\mathbb{E} [\beta] \sum_{n=1}^N x_n x_n^T + \text{diag}(\mathbb{E} [\alpha_1] \dots \mathbb{E} [\alpha_d]) \right) \mathbf{w} - 2\mathbf{w}^T \mathbb{E} [\beta] \sum_{n=1}^N y_n x_n \right] \\
\mathbf{w} &\sim \mathcal{N} \left(\mathbf{U}^{-1} \mathbb{E} [\beta] \sum_{n=1}^N y_n x_n, \mathbf{U}^{-1} \right) \\
\mathbf{U} &= \left(\mathbb{E} [\beta] \sum_{n=1}^N x_n x_n^T + \text{diag}(\mathbb{E} [\alpha_1] \dots \mathbb{E} [\alpha_d]) \right)
\end{aligned} \tag{8}$$

- For $\alpha_d \forall d$ keeping others fixed

$$\begin{aligned}
\log q_{\alpha_d}^* &= E_{\mathbf{w}, \alpha_{-d}, \beta} \left[\left(\frac{1}{2} + e_0 - 1 \right) \log \alpha_d - \alpha_d \left(f_0 + \frac{\mathbf{w}_d^2}{2} \right) \right] \\
&= \left(\frac{1}{2} + e_0 - 1 \right) \log \alpha_d - \alpha_d \left(f_0 + \frac{\mathbb{E} [\mathbf{w}_d^2]}{2} \right) \\
\alpha_d &\sim \text{Gamma} \left(\frac{1}{2} + e_0, f_0 + \frac{\mathbb{E} [\mathbf{w}_d^2]}{2} \right)
\end{aligned} \tag{9}$$

We have,

$$\begin{aligned}
\mathbf{w} &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \\
\boldsymbol{\mu} &= \Sigma \mathbb{E} [\beta] \sum_{n=1}^N y_n x_n \\
\Sigma &= \left(\mathbb{E} [\beta] \sum_{n=1}^N x_n x_n^T + \text{diag}(\mathbb{E} [\alpha_1] \dots \mathbb{E} [\alpha_d]) \right)^{-1}
\end{aligned} \tag{10}$$

$$\beta \sim \text{Gamma}(a_\beta, b_\beta)$$

$$\begin{aligned}
a_\beta &= \frac{a}{2} + a_0 \\
b_\beta &= \frac{1}{2} \sum_{n=1}^N \mathbb{E} [(y_n - \mathbf{w}^T x_n)^2] + b_0
\end{aligned} \tag{11}$$

$$\alpha_d \sim \text{Gamma}(c_{\alpha_d}, g_{\alpha_d})$$

$$\begin{aligned}
c_{\alpha_d} &= \frac{1}{2} + e_0 \\
g_{\alpha_d} &= f_0 + \frac{\mathbb{E} [\mathbf{w}_d^2]}{2}
\end{aligned} \tag{12}$$

Question assigned to the following page: [2](#)

Therefore we have:

$$\begin{aligned}
\mathbb{E}[\mathbf{w}] &= \boldsymbol{\mu} \\
\mathbb{E}[\mathbf{w}\mathbf{w}^T] &= \boldsymbol{\mu}\boldsymbol{\mu}^T + \Sigma \\
\mathbb{E}[\beta] &= \frac{a_\beta}{b_\beta} \\
\mathbb{E}[\alpha_d] &= \frac{c_{\alpha_d}}{g_{\alpha_d}} \\
\mathbb{E}[w_d^2] &= \Sigma_{dd} + \mu_d^2
\end{aligned} \tag{13}$$

Mean Field Inference VI algorithm

- Let $c_{\alpha_d} = \frac{1}{2} + \epsilon_0$ and $a_\beta = \frac{N}{2} + a_0$ and initialize $c_{\alpha_d} \forall d$ and b_β .
- Compute $\mathbb{E}[\alpha_d]^{(0)} \forall d$ and $\mathbb{E}[\beta]^{(0)}$
- For $t=1,2,\dots,T$ or until convergence

$$\begin{aligned}
\Sigma^{(t)} &= \left(\mathbb{E}[\beta]^{(t-1)} \sum_{n=1}^N x_n x_n^T + \text{diag}(\mathbb{E}[\alpha_1]^{(t-1)} \dots \mathbb{E}[\alpha_d]^{(t-1)}) \right)^{-1} \\
\boldsymbol{\mu}^{(t)} &= \Sigma^{(t)} \mathbb{E}[\beta]^{(t-1)} \sum_{n=1}^N y_n x_n \\
b_\beta^{(t)} &= \frac{1}{2} \mathbb{E}[(y_n - \mathbf{w}^T x_n)^2] + b_0 \\
\mathbb{E}[\mathbf{w}] &= \boldsymbol{\mu}^{(t)} \\
\mathbb{E}[\mathbf{w}\mathbf{w}^T] &= \boldsymbol{\mu}^{(t)} \boldsymbol{\mu}^{(t)T} + \Sigma^{(t)} \\
\mathbb{E}[\beta]^{(t)} &= \frac{a_\beta}{b_\beta^{(t)}} \\
\mathbb{E}[\alpha_d] &= \frac{c_{\alpha_d}}{g_{\alpha_d}^{(t)}} \forall d \\
g_{\alpha_d}^{(t)} &= f_0 + \frac{\mathbb{E}[\mathbf{w}_d^{2(t)}]}{2}
\end{aligned} \tag{14}$$

Question assigned to the following page: [3](#)

Conditional posterior of α

$$\begin{aligned}
 p(\alpha|\mathbf{X}, \boldsymbol{\lambda}, \beta) &\propto p(\boldsymbol{\lambda}|\alpha, \beta)p(\alpha) \\
 &\propto \prod_{n=1}^N \text{Gamma}(\lambda_n|\alpha, \beta) * \text{Gamma}(\alpha|a, b) \\
 &\propto \left[\frac{\beta^{N\alpha} (\prod_{n=1}^N \lambda_n)^{\alpha-1}}{\Gamma(\alpha)^N} \right] * \alpha^{a-1} \exp(-\alpha b)
 \end{aligned} \tag{15}$$

We don't have closed form expression. so we will need a sampling method to draw samples from this distribution.

Conditional posterior of β

$$\begin{aligned}
 p(\beta|\mathbf{X}, \boldsymbol{\lambda}, \alpha) &\propto p(\boldsymbol{\lambda}|\alpha, \beta)p(\beta) \\
 &\propto \prod_{n=1}^N \text{Gamma}(\lambda_n|\alpha, \beta) * \text{Gamma}(\beta|c, d) \\
 &\propto \beta^{N\alpha+c-1} * \alpha^{a-1} \exp(-\beta(d + \sum_{n=1}^N \lambda_n)) \\
 \beta &\sim \text{Gamma}(N\alpha + c, d + \sum_{n=1}^N \lambda_n)
 \end{aligned} \tag{16}$$

Conditional posterior of λ_n

$$\begin{aligned}
 p(\lambda_n|\mathbf{X}, \lambda_{-n}, \alpha, \beta) &\propto p(x_n|\lambda_n)p(\lambda_n|\alpha, \beta) \\
 &\propto \text{Poisson}(x_n|\lambda_n) * \text{Gamma}(\lambda_n|\alpha, \beta) \\
 &\propto \lambda_n^{\alpha+x_n-1} \exp(-(\beta+1)\lambda_n) \\
 \lambda_n &\sim \text{Gamma}(\alpha + x_n, \beta + 1)
 \end{aligned} \tag{17}$$

Gibbs Sampling:

- $\alpha^{(0)}, \beta^{(0)}$
- for $t = 1 \dots T$:

$$\begin{aligned}
 \lambda_n^{(t)} &\sim \text{Gamma}(\alpha^{(t-1)} + x_n, \beta^{(t-1)} + 1) \\
 \beta^{(t)} &\sim \text{Gamma}(N\alpha^{(t-1)} + c, d + \sum_{n=1}^N \lambda_n^{(t)}) \\
 \alpha^{(t)} &\sim p(\alpha|\mathbf{X}, \boldsymbol{\lambda}^{(t)}, \beta^{(t)}) \text{ Using Sampling}
 \end{aligned} \tag{18}$$
- Return $(\alpha^{(t)}, \beta^{(t)}, \boldsymbol{\lambda}^{(t)})_{t=0}^T$

Question assigned to the following page: [4](#)

- For Mean $r_{ij} = \mathbf{u}_i^T \mathbf{v}_j + \epsilon_{ij}$ where $\epsilon_{ij} \sim \mathcal{N}(0, \beta^{-1})$

$$\begin{aligned} \mathbb{E}[r_{ij}] &= \mathbb{E}[\mathbf{u}_i^T \mathbf{v}_j + \epsilon_{ij}] \\ &= \mathbb{E}[\mathbf{u}_i^T \mathbf{v}_j] + \mathbb{E}[\epsilon_{ij}] \\ &= \frac{1}{S} \sum_{s=1}^S \mathbf{u}_i^{(s)T} \mathbf{v}_j^{(s)} + 0 \end{aligned} \tag{19}$$

- For variance:

$$\begin{aligned} \mathbb{E}[r_{ij}^2] &= \mathbb{E}[(\mathbf{u}_i^T \mathbf{v}_j)^2 + 2\epsilon_{ij}(\mathbf{u}_i^T \mathbf{v}_j) + \epsilon_{ij}^2] \\ &= \mathbb{E}[(\mathbf{u}_i^T \mathbf{v}_j)^2] + \mathbb{E}[\epsilon_{ij}^2] \\ &= \frac{1}{S} \sum_{s=1}^S (\mathbf{u}_i^{(s)T} \mathbf{v}_j^{(s)})^2 + \beta^{-1} \\ \text{var}[r_{ij}] &= \mathbb{E}[r_{ij}^2] - \mathbb{E}[r_{ij}]^2 \\ \text{var}[r_{ij}] &= \frac{1}{S} \sum_{s=1}^S (\mathbf{u}_i^{(s)T} \mathbf{v}_j^{(s)})^2 + \beta^{-1} - \left(\frac{1}{S} \sum_{s=1}^S \mathbf{u}_i^{(s)T} \mathbf{v}_j^{(s)} \right)^2 \end{aligned} \tag{20}$$

Question assigned to the following page: [5](#)

Part I:

- Maximizing over pre-specified range, we will get $M \approx 6.5$
- Acceptance rate ≈ 0.458 .
- We know $Z_p = p(\text{accept}) * M$, using $p(\text{accept}) = 0.458$, we will get $Z_p \approx 3$. Hence the histogram and plot become similar.

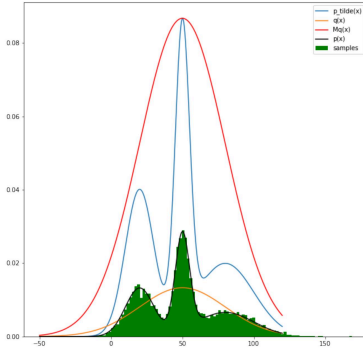


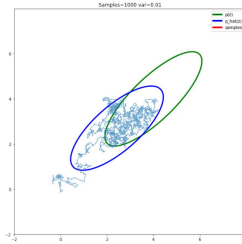
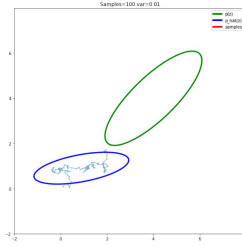
Figure 1: $\tilde{p}(x)$, $q(x)$, $Mq(x)$, $p(x)$ and samples

Part II:

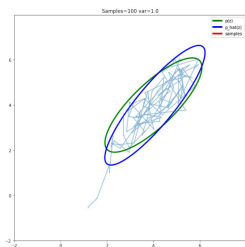
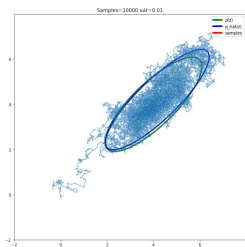
- For $\sigma^2 = 0.01$, Rejecting rate ≈ 0.088
- For $\sigma^2 = 1$, Rejecting rate ≈ 0.600
- For $\sigma^2 = 100$, Rejecting rate ≈ 0.988

Question assigned to the following page: [5](#)

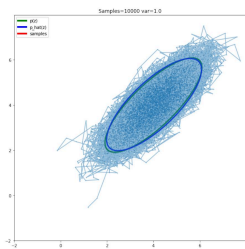
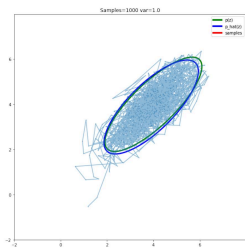
- Given 10k samples we can see from the plots that $\sigma^2 = 1$ works fairly well in terms of convergence. It is able to traverse the entire space relatively fast. We see that for $\sigma^2 = 0.01$ rejection rate is less but the traversal of space is poor. For $\sigma^2 = 100$ traversal of space is good but the convergence rate is very slow. It is for $\sigma^2 = 1$ that traversal is good and the convergence rate is not bad either. Therefore, $\sigma^2 = 1$ is the preferred choice.



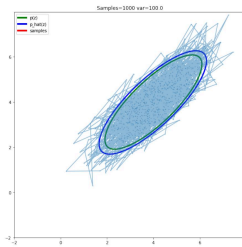
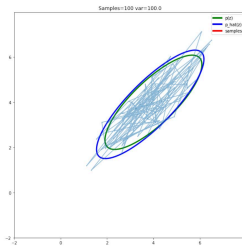
Question assigned to the following page: [5](#)



Question assigned to the following page: [5](#)



Question assigned to the following page: [5](#)



Question assigned to the following page: [5](#)

