STAT7301 – Assignment 5
Joel Thomas 44793203

Q7.15

$$q(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{n_{\boldsymbol{x}}}, \qquad y \in \mathcal{N}(\boldsymbol{x})$$

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{\frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}, 1\right\} = \min\left\{\frac{n_{\boldsymbol{x}}}{n_{\boldsymbol{y}}}, 1\right\}$$

We wish to solve $f(\boldsymbol{x})q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = f(\boldsymbol{y})q_{MH}(\boldsymbol{x}|\boldsymbol{y})\ \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Let $\boldsymbol{y} \neq \boldsymbol{x}$ since $\boldsymbol{y} = \boldsymbol{x}$ case is trivial.

$$q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = q(\boldsymbol{y}|\boldsymbol{x})\alpha(\boldsymbol{x}, \boldsymbol{y})$$

$$q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = q(\boldsymbol{y}|\boldsymbol{x}) \min\left\{\frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}, 1\right\}$$

Let $\frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})} < 1$ so that $\min\left\{\frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}, 1\right\} = \frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}$:

$$\therefore q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = q(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}$$

$$q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{n_{\boldsymbol{x}}} \cdot \frac{n_{\boldsymbol{x}}}{n_{\boldsymbol{y}}} = \frac{1}{n_{\boldsymbol{y}}}$$

$$q_{MH}(\boldsymbol{x}|\boldsymbol{y}) = q(\boldsymbol{x}|\boldsymbol{y})\alpha(\boldsymbol{y}, \boldsymbol{x})$$

$$q_{MH}(\boldsymbol{x}|\boldsymbol{y}) = q(\boldsymbol{x}|\boldsymbol{y}) \min\left\{\frac{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}, 1\right\}$$

If $\frac{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})} < 1 \rightarrow \frac{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})} > 1$ so that $\min\left\{\frac{f(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x})}{f(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})}, 1\right\} = 1$:

$$\therefore q_{MH}(\boldsymbol{x}|\boldsymbol{y}) = q(\boldsymbol{x}|\boldsymbol{y}) \cdot 1$$

$$q_{MH}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{n_{\boldsymbol{y}}}$$

$$\therefore f(\boldsymbol{x})q_{MH}(\boldsymbol{y}|\boldsymbol{x}) = f(\boldsymbol{y})q_{MH}(\boldsymbol{x}|\boldsymbol{y})$$

$$f(\boldsymbol{x}) \cdot \frac{1}{n_{\boldsymbol{y}}} = f(\boldsymbol{y}) \cdot \frac{1}{n_{\boldsymbol{y}}}$$

$$f(\boldsymbol{x}) = f(\boldsymbol{y})\ \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$$

Since the last line holds, then it must be that $f(\boldsymbol{x}) = \frac{1}{n_x}$ i.e. each $\boldsymbol{x} \in \mathcal{X}$ follows the discrete uniform distribution.

This is sensible because we can interpret the result as follows: given another state that is a neighbour of $\boldsymbol{x}$, the probability (discrete distribution $\rightarrow f(\boldsymbol{x})$ is the pmf) of arriving in state $\boldsymbol{x}$ is equally weighted amongst all of its neighbours. Thus, this is the limiting distribution assuming the chain is irreducible and aperiodic.

Q7.18

a)

With $\boldsymbol{X} = (X, Y)^T$, we have:

$$\mathbb{E}[\boldsymbol{X}] = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{Y,X} & \sigma_X^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Since $\sigma_X^2 = \sigma_Y^2 = 1 \rightarrow \frac{\sigma_{X,Y}}{\sqrt{\sigma_X^2 \sigma_X^2}} = \sigma_{X,Y} = \rho$ is the correlation between X and Y. Note that X, Y are not independent in general since $\rho \neq 0$ in general.

The joint pdf for the bivariate normal distribution is given by:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-p^2)}[x^2 - 2\rho xy + y^2]\right)$$

$f(y|x) = \frac{f(x,y)}{f(x)} \propto f(x, y)$ as a function of y alone.

$$\therefore f(y|x) \propto \exp\left(-\frac{1}{2(1-p^2)}[-2\rho xy + y^2]\right)$$

Completing the square in for $y^2 - 2\rho xy$:

$$y^2 - 2\rho xy = y^2 - 2\rho xy + (\rho x)^2 - (\rho x)^2$$

$$y^2 - 2\rho xy = (y - \rho x)^2 - (\rho x)^2$$

$$\therefore f(y|x) \propto \exp\left(-\frac{1}{2(1-p^2)}[(y-\rho x)^2 - (\rho x)^2]\right)$$

$$f(y|x) \propto \exp\left(-\frac{1}{2}\frac{(y-\rho x)^2}{(1-p^2)}\right)$$

$$\rightarrow (Y|X=x) \sim N(\rho x, 1-\rho^2)$$

Additionally, by symmetry of the problem, $(X|Y=y) \sim N(\rho y, 1-\rho^2)$ as required.

b)

See attached code $Q7\_18.m$:

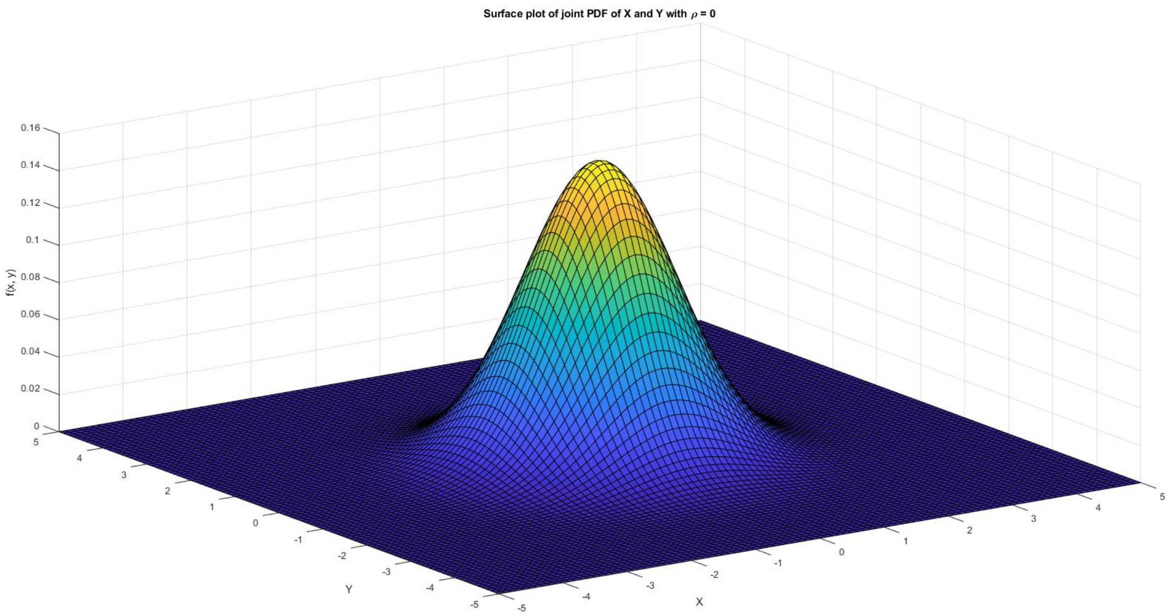

Figure 1: Surface plot of the joint pdf of X and Y denoted by $f_{X,Y}(x,y)$ with $\rho = 0$.

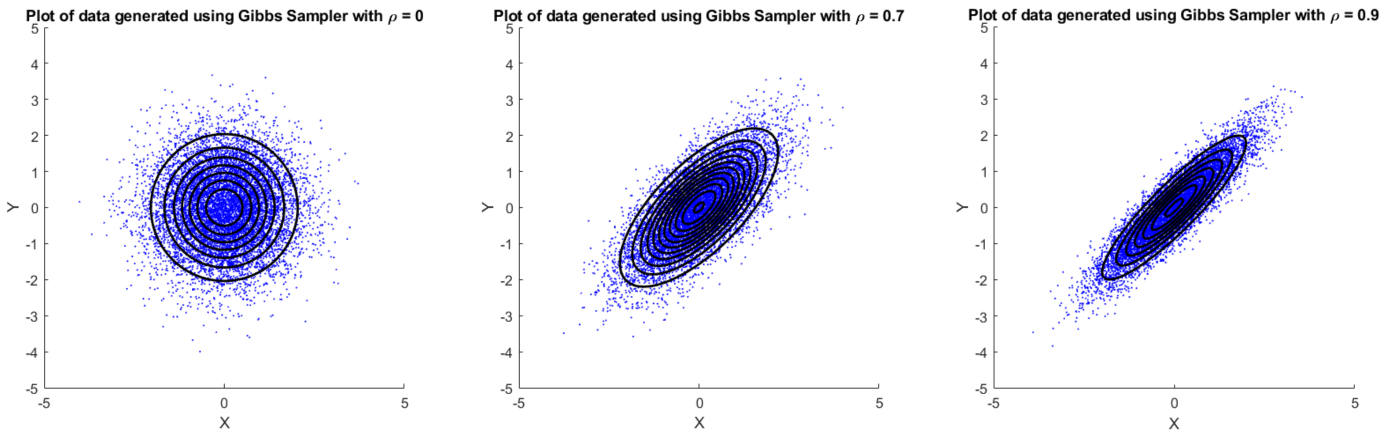

Figure 2: Plot of the generated data using Gibbs Sampler for $\rho = 0, 0.7, 0.9$.

Q8.6

By Theorem 8.2:

$$Z_i = \frac{Y_i}{\sum_{j=1}^{m+1} Y_j}, \qquad Y_j \sim Gamma(\alpha_j, 1), \qquad j = 1, \dots, m+1$$

$Z_i = (\boldsymbol{Z})_i$ where $\boldsymbol{Z} \sim Dirichlet(\alpha_1, \dots, \alpha_{m+1})$, $i = 1, \dots, m$. The Z_i can be rewritten as:

$$Z_i = \frac{Y_i}{Y_i + \sum_{j \neq i} Y_j}$$

Note that the $\{Y_j\}_{j=1}^{m+1}$ are also independent from Theorem 8.2 Now proving that $\sum_{j \neq i} Y_j \sim Gamma(\sum_{j \neq i} \alpha_j, 1)$ using the MGF for a Gamma random variable:

$$M_{Y_j}(t) = \mathbb{E}[e^{tY_j}]$$

$$M_{Y_j}(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} = (1-t)^{-\alpha_j}$$

$$M_{\sum_{j \neq i} Y_j}(t) = \mathbb{E}\left[e^{t \sum_{j \neq i} Y_j}\right]$$

$$M_{\sum_{j \neq i} Y_j}(t) = \mathbb{E}[e^{tY_1} \cdot e^{tY_2} \cdot \dots \cdot e^{tY_{m+1}}]$$

$$M_{\sum_{j \neq i} Y_j}(t) = \mathbb{E}[e^{tY_1}] \cdot \mathbb{E}[e^{tY_2}] \cdot \dots \cdot \mathbb{E}[e^{tY_{m+1}}], \qquad \left(\{Y_j\}_{j=1}^{m+1} \sim Gamma(\alpha_j, 1) \text{ independent from Theorem 8.2}\right)$$

$$M_{\sum_{j \neq i} Y_j}(t) = (1-t)^{-\alpha_1} \cdot (1-t)^{-\alpha_2} \cdot \dots \cdot (1-t)^{-\alpha_{m+1}}$$

$$M_{\sum_{j \neq i} Y_j}(t) = (1-t)^{-\sum_{j \neq i} \alpha_j}$$

Which is the corresponding MGF for a $Gamma(\sum_{j \neq i} \alpha_j, 1)$ random variable. Using the uniqueness property of MGFs $\rightarrow \sum_{j \neq i} Y_j \sim Gamma(\sum_{j \neq i} \alpha_j, 1)$. Finally, using the well-known result that if $A \sim Gamma(\lambda_1, \theta)$ and $B \sim Gamma(\lambda_2, \theta)$ are independent then $\frac{A}{A+B} \sim Beta(\lambda_1, \lambda_2)$, we have:

$$Z_i = \frac{Y_i}{Y_i + \sum_{j \neq i} Y_j} \sim Beta\left(\alpha_i, \sum_{j \neq i} \alpha_j\right)$$

Q8.10

a)

$$\sum_{i=1}^{n}(x_i-\mu)^2$$

$$=\sum_{i=1}^{n}\left((x_i-\bar{x})+(\bar{x}-\mu)\right)^2$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+2(x_i-\bar{x})(\bar{x}-\mu)+(\bar{x}-\mu)^2$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\mu)+\sum_{i=1}^{n}(\bar{x}-\mu)^2$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x})+n(\bar{x}-\mu)^2$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+2(\bar{x}-\mu)\times 0+n(\bar{x}-\mu)^2, \qquad \left(\sum_{i=1}^{n}(x_i-\bar{x})=0\right)$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2$$

b)

$$f(\sigma^2|\boldsymbol{x})$$

$$=\int f(\mu,\sigma^2|\boldsymbol{x})\,d\mu$$

$$\propto\int(\sigma^2)^{-\frac{n}{2}-1}\exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right)\,d\mu$$

$$=\int(\sigma^2)^{-\frac{n}{2}-1}\exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2}{\sigma^2}\right)\,d\mu$$

$$=(\sigma^2)^{-\frac{n}{2}-1}\exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}\right)\int\exp\left(-\frac{1}{2}\frac{n(\bar{x}-\mu)^2}{\sigma^2}\right)\,d\mu$$

$$= (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}\right) \int \exp\left(-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\frac{\sigma^2}{n}}\right) d\mu$$

We know that:

$$\int \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left(-\frac{1}{2}\frac{(\mu-\bar{x})^2}{\frac{\sigma^2}{n}}\right) d\mu = 1$$

$$\therefore \int \exp\left(-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\frac{\sigma^2}{n}}\right) d\mu = \sqrt{2\pi}\frac{\sigma}{\sqrt{n}}$$

$$\therefore f(\sigma^2|\boldsymbol{x})$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}\right) \cdot \sqrt{2\pi}\frac{\sigma}{\sqrt{n}}$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1+\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}\right)$$

$$\therefore f(\sigma^2|\boldsymbol{x}) \propto (\sigma^2)^{-\frac{n}{2}-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}\right)$$

c)

If $Z \sim InvGamma(\alpha, \beta)$, then:

$$f(z) \propto z^{-\alpha-1} \exp\left(-\frac{\beta}{z}\right)$$

Since $f(\sigma^2|\boldsymbol{x}) \propto (\sigma^2)^{-\frac{n}{2}-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}\right) = (\sigma^2)^{-\frac{n-1}{2}-1} \exp\left(-\frac{\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}\right)$

$$\therefore (\sigma^2|\boldsymbol{x}) \sim InvGamma\left(\alpha = \frac{n-1}{2}, \beta = \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

Observe that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)s_x^2$ where s_x^2 is the classical sample variance of the $\{x_i\}$.

$$\therefore (\sigma^2|\boldsymbol{x}) \sim InvGamma\left(\frac{n-1}{2}, \frac{1}{2}(n-1)s_x^2\right)$$

d)

The $1-\gamma$ credible interval (q_1, q_2), where q_1 and q_2 are the $\frac{\gamma}{2}$ and $1-\frac{\gamma}{2}$ quantiles of the

$InvGamma\left(\frac{n-1}{2}, \frac{1}{2}(n-1)s_x^2\right)$, is given by:

$\mathbb{P}(q_1 < (\sigma^2|\boldsymbol{x}) < q_2) = 1-\gamma$

Observe that:

$$\left((n-1)s_x^2\right)^{\frac{n-1}{2}-1} \cdot f(\sigma^2|\boldsymbol{x}) \propto \left(\frac{\sigma^2}{(n-1)s_x^2}\right)^{-\frac{n-1}{2}-1} \exp\left(-\frac{\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2}{\sigma^2}\right)$$

$$\therefore \left(\frac{\sigma^2}{(n-1)s_x^2}\middle|\boldsymbol{x}\right) \sim InvGamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

$$\rightarrow \left(\frac{(n-1)s_x^2}{\sigma^2}\middle|\boldsymbol{x}\right) \sim Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

Using the well-known result that a $Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$ is equivalent to a χ_ν^2 with ν degrees of freedom:

$$\left(\frac{(n-1)s_x^2}{\sigma^2}\middle|\boldsymbol{x}\right) \sim Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right) \sim \chi_{n-1}^2$$

Let $\chi_{n-1;\gamma}^2$ be the γ-quantile for the χ_{n-1}^2 distribution. The classic $1-\gamma$ confidence interval for $(\sigma^2|\boldsymbol{x})$ is given by:

$$\mathbb{P}\left(\chi_{n-1;\frac{\gamma}{2}}^2 < \left(\frac{(n-1)s_x^2}{\sigma^2}\middle|\boldsymbol{x}\right) < \chi_{n-1;1-\frac{\gamma}{2}}^2\right) = 1-\gamma$$

$$\mathbb{P}\left(\chi_{n-1;1-\frac{\gamma}{2}}^2 < \left(\frac{\sigma^2}{(n-1)s_x^2}\middle|\boldsymbol{x}\right) < \chi_{n-1;\frac{\gamma}{2}}^2\right) = 1-\gamma$$

$$\mathbb{P}\left(\frac{(n-1)s_x^2}{\chi_{n-1;1-\frac{\gamma}{2}}^2} < (\sigma^2|\boldsymbol{x}) < \frac{(n-1)s_x^2}{\chi_{n-1;\frac{\gamma}{2}}^2}\right) = 1-\gamma$$

$$\therefore \mathbb{P}(q_1 < (\sigma^2|\boldsymbol{x}) < q_2) = \mathbb{P}\left(\frac{(n-1)s_x^2}{\chi_{n-1;1-\frac{\gamma}{2}}^2} < (\sigma^2|\boldsymbol{x}) < \frac{(n-1)s_x^2}{\chi_{n-1;\frac{\gamma}{2}}^2}\right) = 1-\gamma$$

$$\rightarrow (q_1, q_2) = \left(\frac{(n-1)s_x^2}{\chi_{n-1;1-\frac{\gamma}{2}}^2}, \frac{(n-1)s_x^2}{\chi_{n-1;\frac{\gamma}{2}}^2}\right)$$

Hence, the $1-\gamma$ credible interval for $(\sigma^2|\boldsymbol{x})$ is identical to the classic confidence interval for $(\sigma^2|\boldsymbol{x})$.

e)

$f(\mu|\boldsymbol{x})$

$$= \int f(\mu, \sigma^2|\boldsymbol{x})\, d\sigma^2$$

$$\propto \int (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right) d\sigma^2$$

$$= \Gamma\left(\frac{n}{2}\right)\cdot\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)^{-\frac{n}{2}} \int \frac{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}(\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right) d\sigma^2$$

We know that the above expression involving the integral is the integral of the exact pdf of an Inverse-Gamma random variable with $\alpha = \frac{n}{2}$ and $\beta = \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$. Thus:

$$\int \frac{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}(\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right) d\sigma^2 = 1$$

$\therefore f(\mu|\boldsymbol{x})$

$$\propto \Gamma\left(\frac{n}{2}\right)\cdot\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)^{-\frac{n}{2}}\cdot 1$$

$$\propto \frac{1}{2^{\frac{n}{2}}}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)^{-\frac{n}{2}}$$

$$\propto \left(\sum_{i=1}^{n}(x_i-\mu)^2\right)^{-\frac{n}{2}}$$

Now using result from part **a)**:

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

$\therefore f(\mu|\boldsymbol{x})$

$$\propto \left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right)^{-\frac{n}{2}}$$

$$= \left((n-1)s_x^2 + n(\bar{x}-\mu)^2\right)^{-\frac{n}{2}}$$

$$= \left((n-1)s_x^2\right)^{-\frac{n}{2}} \left(1 + \frac{n(\mu-\bar{x})^2}{(n-1)s_x^2}\right)^{-\frac{n}{2}}$$

$$\propto \left(1 + \frac{n(\mu-\bar{x})^2}{(n-1)s_x^2}\right)^{-\frac{n}{2}}$$

Let $\nu = n - 1 \rightarrow n = \nu + 1$

$$\therefore f(\mu|\boldsymbol{x}) \propto \left(1 + \frac{n(\mu-\bar{x})^2}{\nu s_x^2}\right)^{-\frac{\nu+1}{2}}$$

Using 2.23, a random variable $Y \sim t_\nu$ has pdf given by:

$$f(y) \propto \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Let $y^2 = \frac{n(\mu-\bar{x})^2}{s_x^2} = \frac{(\mu-\bar{x})^2}{\frac{s_x^2}{n}}$

$$\rightarrow y = \sqrt{\frac{(\mu-\bar{x})^2}{\frac{s_x^2}{n}}} = \frac{\mu-\bar{x}}{\frac{s_x}{\sqrt{n}}}, \qquad \text{(want positive square root only)}$$

Since $f(\mu|\boldsymbol{x})$ has the same form as $f(y)$:

$$\therefore \left(\frac{\mu-\bar{x}}{\frac{s_x}{\sqrt{n}}}\middle|\boldsymbol{x}\right) \sim t_{n-1}$$

f)

The $1-\gamma$ credible interval (q_1, q_2), where q_1 and q_2 are the $\frac{\gamma}{2}$ and $1-\frac{\gamma}{2}$ quantiles of $f(\mu|\boldsymbol{x})$ is given by:

$$\mathbb{P}(q_1 < (\mu|\boldsymbol{x}) < q_2) = 1-\gamma$$

From part **e)**, we know that $\left(\frac{\mu-\bar{x}}{\frac{s_x}{\sqrt{n}}}\middle|\boldsymbol{x}\right) \sim t_{n-1}$ (we wish to use this as our pivot for constructing the classic confidence interval). Then let $t_{n-1;1-\frac{\gamma}{2}}$ represent the $1-\frac{\gamma}{2}$ quantile of the t_{n-1} distribution. Constructing the classic confidence interval in the same manner as 5.18:

$$\mathbb{P}\left(\left|\frac{\mu-\bar{x}}{\frac{s_x}{\sqrt{n}}}\right| < t_{n-1;1-\frac{\gamma}{2}}\right) = 1-\gamma$$

$$\mathbb{P}\left(-t_{n-1;1-\frac{\gamma}{2}} < \left(\frac{\mu-\bar{x}}{\frac{s_x}{\sqrt{n}}}\middle|\boldsymbol{x}\right) < t_{n-1;1-\frac{\gamma}{2}}\right) = 1-\gamma$$

$$\mathbb{P}\left(\bar{x} - t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}} < (\mu|\boldsymbol{x}) < \bar{x} + t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}}\right) = 1-\gamma$$

$$\therefore \mathbb{P}(q_1 < (\mu|\boldsymbol{x}) < q_2) = \mathbb{P}\left(\bar{x} - t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}} < (\mu|\boldsymbol{x}) < \bar{x} + t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}}\right) = 1-\gamma$$

$$\rightarrow (q_1, q_2) = \left(\bar{x} - t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}}, \bar{x} + t_{n-1;1-\frac{\gamma}{2}}\frac{s_x}{\sqrt{n}}\right)$$

Hence, the $1-\gamma$ credible interval for $(\mu|\boldsymbol{x})$ is identical to the classic confidence interval for $(\mu|\boldsymbol{x})$.

Q8.15

$$X_1, \dots, X_n \sim^{iid} Exp(\lambda)$$

$$\rightarrow f(x_k|\lambda) = \lambda e^{-\lambda x_k}, \qquad k = 1, \dots, n$$

The likelihood is given by:

$$f(\boldsymbol{x}|\lambda) = \prod_{i=k}^{n} f(x_k|\lambda), \qquad (X_k \sim^{iid} Exp(\lambda), k = 1, \dots, n)$$

$$f(\boldsymbol{x}|\lambda) = \prod_{i=k}^{n} \lambda e^{-\lambda x_k}$$

$$f(\boldsymbol{x}|\lambda) = \lambda^n e^{-\lambda \sum_{k=1}^{n} x_k}$$

This is of the form:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = c(\boldsymbol{\theta})^n \exp\left(\sum_{i=1}^{m} \eta_i(\boldsymbol{\theta}) \sum_{k=1}^{n} t_i(x_k)\right) \prod_{k=1}^{n} h(x_k)$$

With:

$$\boldsymbol{\theta} = [\lambda], \qquad m = 1, \qquad \eta(\lambda) = -\lambda, \qquad t(x_k) = x_k, \qquad c(\lambda) = \lambda, \qquad h(x_k) = 1$$

Hence, by theorem 8.4, a conjugate prior for the exponential likelihood is given by:

$$f(\lambda)$$

$$\propto c(\lambda)^b \exp\left(\sum_{i=1}^{m} \eta_i(\lambda) a_i\right)$$

$$= \lambda^b \exp(\eta(\lambda) \cdot a)$$

$$= \lambda^{(b+1)-1} e^{-a\lambda}$$

$$\therefore \lambda \sim Gamma(b+1, a)$$

Hence, an appropriate conjugate family for the $Exp(\lambda)$ distribution is given by the Gamma distribution with the posterior being $\lambda \sim Gamma(b+1, a)$.

Q8.16

From **Q8.15** above, we know that if $X_1, \dots, X_n \sim^{iid} Exp\left(\frac{1}{\theta}\right)$ then:

$$f(\boldsymbol{x}|\theta) = \theta^{-n} e^{-\frac{1}{\theta}\sum_{k=1}^{n} x_k}$$

If $\theta \sim InvGamma(\alpha_0, \lambda_0)$ is a conjugate prior for the $Exp\left(\frac{1}{\theta}\right)$ distribution, then we expect the posterior $(\theta|\boldsymbol{x})$ to also have an Inverse Gamma distribution.

$f(\theta|\boldsymbol{x})$

$\propto f(\boldsymbol{x}|\theta)f(\theta)$

$= \theta^{-n}e^{-\frac{1}{\theta}\sum_{k=1}^{n}x_k} \times \theta^{-\alpha_0-1}e^{-\frac{\lambda_0}{\theta}}$

$$= \theta^{-(n+\alpha_0)-1}\exp\left(-\frac{1}{\theta}\sum_{k=1}^{n}x_k - \frac{\lambda_0}{\theta}\right)$$

$$= \theta^{-(n+\alpha_0)-1}\exp\left(-\frac{n\bar{x}+\lambda_0}{\theta}\right)$$

$\therefore (\theta|\boldsymbol{x}) \sim InvGamma(n+\alpha_0, n\bar{x}+\lambda_0)$

Hence, $\theta \sim InvGamma(\alpha_0, \lambda_0)$ is indeed a conjugate prior for the $Exp\left(\frac{1}{\theta}\right)$ distribution.

Q6. 2020 Final Exam

a)

Want to show that $(p_1, p_2|\boldsymbol{x}) \sim Dirichlet$ if the prior $(p_1, p_2) \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3)$ is conjugate for this problem. Since we assume the data comes from a multinomial distribution:

$f(\boldsymbol{x}|p_1, p_2) \propto p_1^{x1}p_2^{x_2}(1-p_1-p_2)^{x_3}$

$f(p_1, p_2) \propto p_1^{\alpha_1-1}p_2^{\alpha_2-1}(1-p_1-p_2)^{\alpha_3-1}$

$\therefore f(p_1, p_2|\boldsymbol{x})$

$\propto f(\boldsymbol{x}|p_1, p_2)f(p_1, p_2)$

$= p_1^{x1}p_2^{x_2}(1-p_1-p_2)^{x_3} \times p_1^{\alpha_1-1}p_2^{\alpha_2-1}(1-p_1-p_2)^{\alpha_3-1}$

$= p_1^{(x_1+\alpha_1)-1}p_2^{(x_2+\alpha_2)-1}(1-p_1-p_2)^{(x_3+\alpha_3)-1}$

$\therefore (p_1, p_2|\boldsymbol{x}) \sim Dirichlet(x_1+\alpha_1, x_2+\alpha_2, x_3+\alpha_3)$

Hence, the prior is indeed conjugate for this problem.

b)

To solve for the posterior modes, we need to make use of the method of Lagrange multipliers (since we wish to solve a constrained optimisation problem) on the log-likelihood function. Denoting $1 - p_1 - p_2 = p_3$:

$$\log\big(f(p_1, p_2|\boldsymbol{x})\big)$$

$$= \log\left(\frac{\Gamma(x_1 + \alpha_1 + x_2 + \alpha_2 + x_3 + \alpha_3)}{\Gamma(x_1 + \alpha_1) + \Gamma(x_2 + \alpha_2) + \Gamma(x_3 + \alpha_3)} p_1^{(x_1+\alpha_1)-1} p_2^{(x_2+\alpha_2)-1} p_3^{(x_3+\alpha_3)-1}\right)$$

$$= \log(c) + (x_1 + \alpha_1 - 1)\log(p_1) + (x_2 + \alpha_2 - 1)\log(p_2) + (x_3 + \alpha_3 - 1)\log(p_3), \qquad (c = \text{constant})$$

$$= c + (x_1 + \alpha_1 - 1)\log(p_1) + (x_2 + \alpha_2 - 1)\log(p_2) + (x_3 + \alpha_3 - 1)\log(p_3)$$

We have the constraint $p_1 + p_2 + p_3 = 1 \rightarrow 1 - p_1 - p_2 - p_3 = 0$. Let $\lambda = 1 - p_1 - p_2 - p_3$, then we have:

$$\mathcal{L}(p_1, p_2, p_3, \lambda)$$

$$= c + (x_1 + \alpha_1 - 1)\log(p_1) + (x_2 + \alpha_2 - 1)\log(p_2) + (x_3 + \alpha_3 - 1)\log(p_3) + \lambda(1 - p_1 - p_2 - p_3)$$

$$\frac{\partial \mathcal{L}}{\partial p_1} = \frac{x_1 + \alpha_1 - 1}{p_1} - \lambda = 0 \rightarrow p_1 = \frac{x_1 + \alpha_1 - 1}{\lambda}$$

$$\frac{\partial \mathcal{L}}{\partial p_2} = \frac{x_2 + \alpha_2 - 1}{p_2} - \lambda = 0 \rightarrow p_2 = \frac{x_2 + \alpha_2 - 1}{\lambda}$$

$$\frac{\partial \mathcal{L}}{\partial p_3} = \frac{x_3 + \alpha_3 - 1}{p_3} - \lambda = 0 \rightarrow p_3 = \frac{x_3 + \alpha_3 - 1}{\lambda}$$

Using $p_1 + p_2 + p_3 = 1$:

$$\frac{x_1 + \alpha_1 - 1}{\lambda} + \frac{x_2 + \alpha_2 - 1}{\lambda} + \frac{x_3 + \alpha_3 - 1}{\lambda} = 1$$

$$\therefore \lambda = \sum_{i=1}^{3}(x_i + a_i - 1)$$

Using $x_1 = 94,\ x_2 = 258,\ x_3 = 822$:

$$\rightarrow mode(p_1|\boldsymbol{x}) = \frac{x_1 + \alpha_1 - 1}{\sum_{i=1}^{3}(x_i + a_i - 1)} = \frac{\alpha_1 + 93}{\alpha_1 + \alpha_2 + \alpha_3 + 1171}$$

$$\rightarrow mode(p_2|\boldsymbol{x}) = \frac{x_2 + \alpha_2 - 1}{\sum_{i=1}^{3}(x_i + a_i - 1)} = \frac{\alpha_2 + 257}{\alpha_1 + \alpha_2 + \alpha_3 + 1171}$$

c)

$$mode(p_1|\boldsymbol{x}) = \frac{\alpha_1 + 93}{\alpha_1 + \alpha_2 + \alpha_3 + 1171}$$

The effect of the prior on the posterior mode of p_1 is like observing $(\alpha_1 - 1), (\alpha_2 - 1), (\alpha_3 - 1)$ counts in each category (Often/Sometimes/Never) prior to the current experiment.

In particular, for our non-informative prior $Dirichlet(1,1,1)$, then in terms of the posterior mode, it is like observing 0 counts in each category prior to the current experiment.

d)

$$f(p_1, p_2|\boldsymbol{x}') \propto p_1^{(x_1' + \alpha_1) - 1} p_2^{(x_2' + \alpha_2) - 1} (1 - p_1 - p_2)^{(x_3' + \alpha_3) - 1}$$

$$\therefore (p_1, p_2|\boldsymbol{x}) \sim Dirichlet(x_1' + \alpha_1, x_2' + \alpha_2, x_3' + \alpha_3)$$

Using $x_1' = 267,\ x_2' = 155,\ x_3' = 390$:

$$mode(p_1|\boldsymbol{x}') = \frac{x_1' + \alpha_1 - 1}{\sum_{i=1}^{3}(x_i' + a_i - 1)} = \frac{\alpha_1 + 266}{\alpha_1 + \alpha_2 + \alpha_3 + 811}$$

$$mode(p_2|\boldsymbol{x}') = \frac{x_2' + \alpha_2 - 1}{\sum_{i=1}^{3}(x_i' + a_i - 1)} = \frac{\alpha_2 + 154}{\alpha_1 + \alpha_2 + \alpha_3 + 811}$$

$$mode(p_1|\boldsymbol{x}') = \frac{x_1' + \alpha_1 - 1}{\sum_{i=1}^{3}(x_i' + a_i - 1)} = \frac{\alpha_1 + 389}{\alpha_1 + \alpha_2 + \alpha_3 + 811}$$

Q8.1

a)

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, \qquad x \in (0,1), \qquad \alpha,\beta > 0$$

$$\log\big(f(x)\big) = \log\left(\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\right)$$

$$\log\big(f(x)\big) = (\alpha-1)\log(x) + (\beta-1)\log(1-x) - \log\big(\mathrm{B}(\alpha,\beta)\big)$$

$$\frac{\partial \log\big(f(x)\big)}{\partial x} = \frac{\alpha-1}{x} - \frac{\beta-1}{1-x} = 0$$

$$(\alpha-1)(1-x) = (\beta-1)x$$

$$\alpha-1 = (\alpha+\beta-2)x$$

$$\rightarrow x^* = \frac{\alpha-1}{\alpha+\beta-2}$$

If $\alpha < 1, \beta > 1 \rightarrow \alpha - 1 < 0$ but we don't know whether $\alpha + \beta - 2 < 0$ or > 0 so ignore this case.

If $\alpha > 1, \beta < 1 \rightarrow \alpha - 1 > 0$ but we don't know whether $\alpha + \beta - 2 < 0$ or > 0 so ignore this case.

If $\alpha, \beta > 1 \rightarrow \alpha - 1 > 0, \alpha + \beta - 2 > 0 \rightarrow x^* > 0$

If $\alpha, \beta < 1 \rightarrow \alpha - 1 < 0, \alpha + \beta - 2 < 0 \rightarrow x^* < 0$

$$\frac{\partial^2 \log\big(f(x)\big)}{\partial x^2} = -\frac{\alpha-1}{x^2} - \frac{\beta-1}{(1-x)^2}$$

If $\alpha, \beta < 1 \rightarrow \frac{\partial^2 \log(f(x))}{\partial x^2} > 0 \; \forall \, x \in (0,1) \rightarrow \log\big(f(x)\big)$ is concave up or ∪ shaped so x^* is not a maximum.

If $\alpha, \beta > 1 \rightarrow \frac{\partial^2 \log(f(x))}{\partial x^2} < 0 \; \forall \, x \in (0,1) \rightarrow \log\big(f(x)\big)$ is concave down or ∩ shaped so x^* is a maximum.

See attached code $Q8\_1.m$

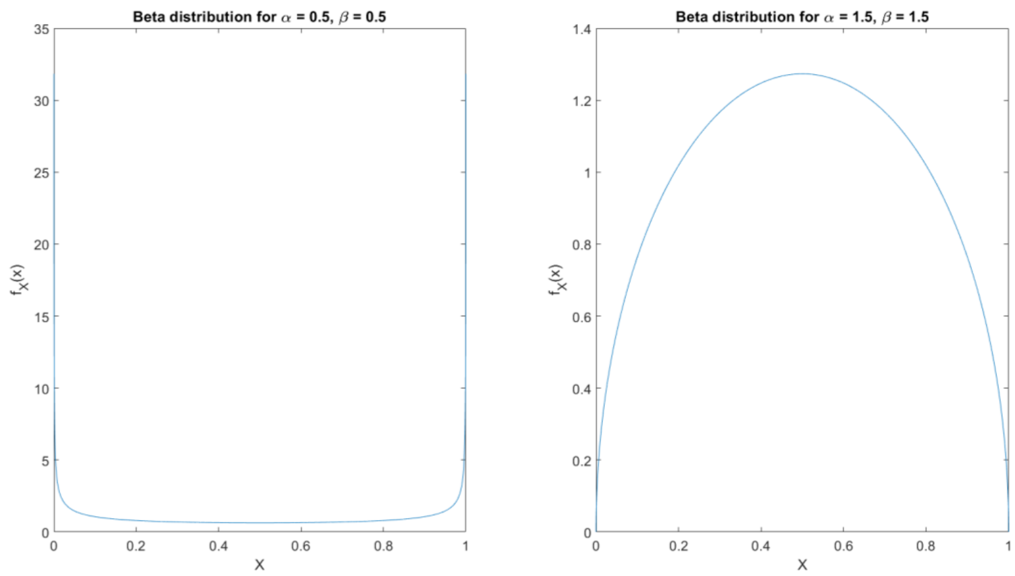

Figure 3: Plot comparing pdfs of Beta distribution with $\boldsymbol{\alpha}, \boldsymbol{\beta} < 1$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} > 1$.

b)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

$$\mathbb{E}[X] = \int_0^1 x \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\, dx\,, \qquad \big(X \sim Beta(\alpha,\beta), x \in (0,1)\big)$$

$$\mathbb{E}[X] = \int_0^1 \frac{x^{(\alpha+1)-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\, dx$$

$$\mathbb{E}[X] = \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} \int_0^1 \frac{x^{(\alpha+1)-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha+1,\beta)}\, dx$$

The integral expression is just the integral of the pdf a $Beta(\alpha+1,\beta)$ random variable, so we have:

$$\int_0^1 \frac{x^{(\alpha+1)-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha+1,\beta)}\, dx = 1$$

$$\therefore\ \mathbb{E}[X] = \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} \cdot 1 = \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)}$$

Recall that $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$:

$$\mathbb{E}[X] = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

Using the property that $\Gamma(\lambda+1)=\lambda\Gamma(\lambda)$ if $\lambda \in \mathbb{R}^{+}$:

$$\mathbb{E}[X]=\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \qquad (\alpha,\beta\in\mathbb{R}^{+}\rightarrow\alpha+\beta\in\mathbb{R}^{+})$$

$$\therefore\ \mathbb{E}[X]=\frac{\alpha}{\alpha+\beta}$$

Q8.17

a)

$$f(\boldsymbol{y}|\boldsymbol{x})=\int f(\boldsymbol{y},\boldsymbol{\theta}|\boldsymbol{x})\,d\boldsymbol{\theta}\,, \qquad \text{(marginal pdf from joint pdf)}$$

$$f(\boldsymbol{y}|\boldsymbol{x})=\int f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})f(\boldsymbol{\theta}|\boldsymbol{x})\,d\boldsymbol{\theta}\,, \qquad \text{(joint pdf} = \text{conditional pdf} \times \text{marginal pdf)}$$

$$f(\boldsymbol{y}|\boldsymbol{x})=\int f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{x})\,d\boldsymbol{\theta}\,, \qquad (\boldsymbol{y}\text{ independent of }\boldsymbol{x})$$

b)

See attached code $Q8\_17.m$:

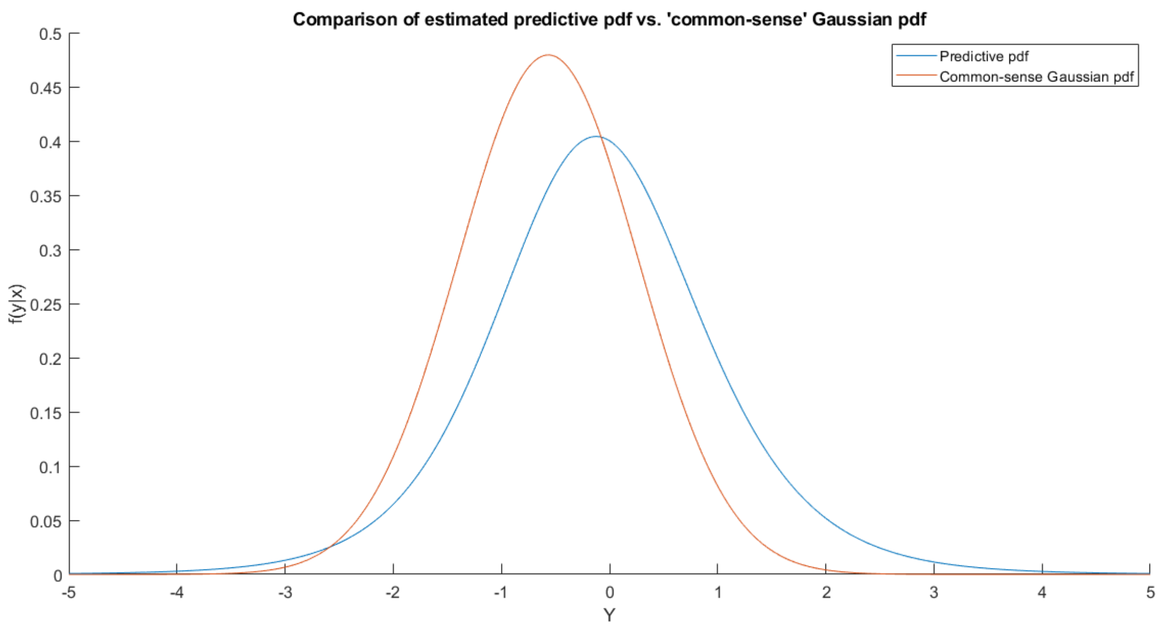

Figure 4: Plot comparing pdfs after generating $N=1000$ samples from Gibbs Sampler.

Bonus Question

b)

See attached code $Q\_bonus.m$ for parts **b)-d)**.

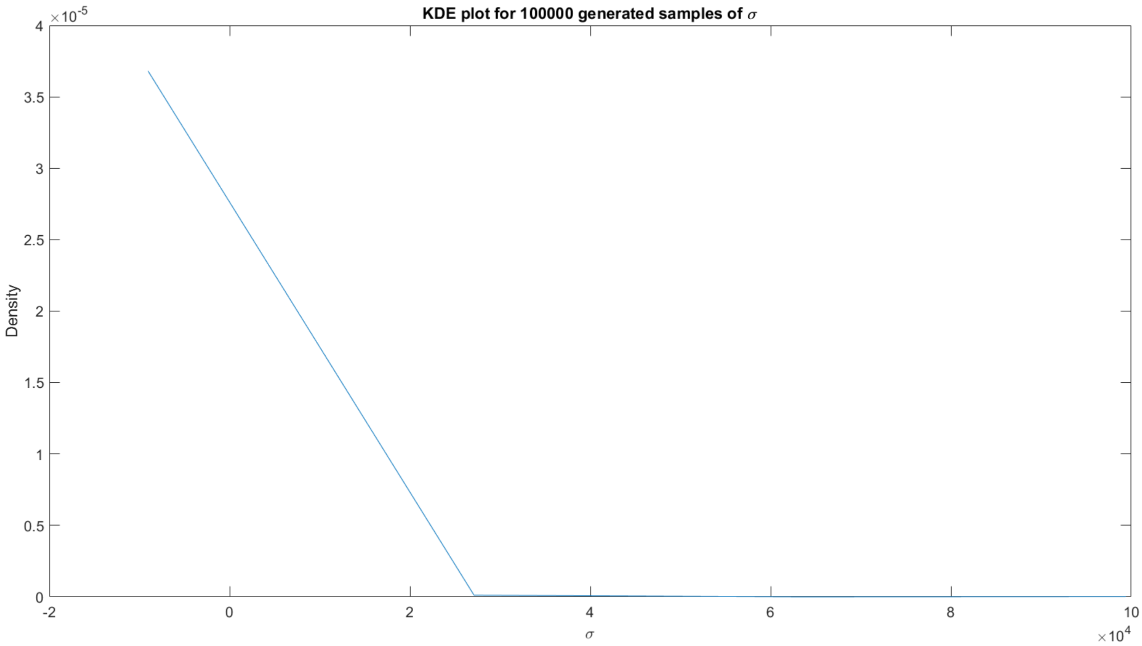

Figure 5: Plot of a KDE of the sample with bandwidth $= 2^2$.

c)

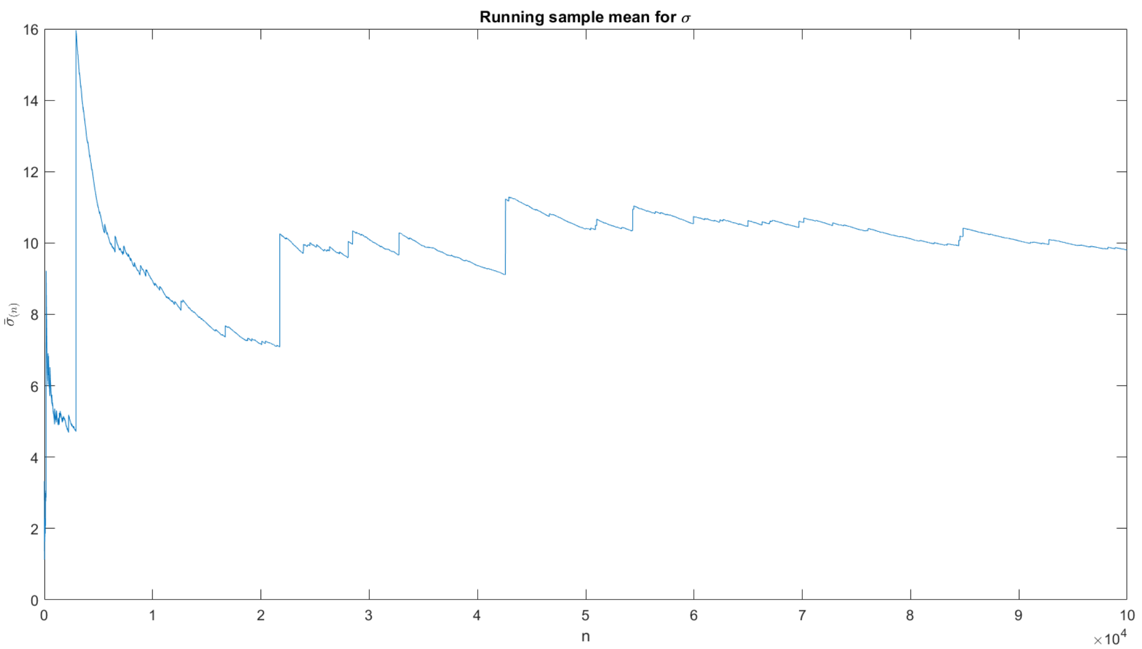

Figure 6: Plot of the running sample mean $\overline{\sigma}_{(n)},\ n = 1, 2, \dots, 100000.$

d)

From figure 6, it appears that the sample mean does settle/converge to a value as n increases. In the code, this value was $\bar{\sigma}_{(100000)} \approx 8.9424$ which changes slightly each time a new run is performed.

Q8.12

c)

See attached code $Q8\_12.m$.

d)

After implementing the Gibbs sampler, the constructed 95% credible intervals for p and λ based on $N = 10000$ generated samples are given by $(0.2043, 0.4115)$ and $(1.4205, 2.6245)$ respectively. It is evident that both these credible intervals contain the true respective values $p = 0.3$ and $\lambda = 2$.