SDS 383C - Statistical Modeling 1: Homework 3

Rahul Nandakumar

Graduate Student, ORIE Program (E-ID: rn9355)

October 20, 2022

1.

Solution: It is given that $\mathbf{y_1}, \ldots \mathbf{y_n} \overset{iid}{\sim} p(\mathbf{y} \mid \boldsymbol{\theta}) = h(\mathbf{y})\exp\{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})\}$

a. We need to prove that

$$\mathbb{E}(\mathbf{y} \mid \boldsymbol{\theta}) = \boldsymbol{\xi}(\boldsymbol{\theta}) = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Using the property of probability distributions, that is the area under the pdf over the support $= 1$

$$\int_{-\infty}^{\infty} p(\mathbf{y} \mid \boldsymbol{\theta})d\mathbf{y} = 1$$

$$\Rightarrow \frac{d}{d\boldsymbol{\theta}} \int_{-\infty}^{\infty} p(\mathbf{y} \mid \boldsymbol{\theta})d\mathbf{y} = 0; \text{ Taking derivative w.r.t } \boldsymbol{\theta} \text{ on both sides}$$

Using Leibniz Integral rule, which is given as follows,

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x,t)dt = f(x,b(x))\frac{d}{dx}b(x) - f(x,a(x))\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}f(x,t)dt$$

When $a(x) = a$ and $b(x) = b$, we get

$$\frac{d}{dx} \int_{a}^{b} f(x,t)dt = \int_{a}^{b} \frac{\partial}{\partial x}f(x,t)dt$$

Using this result in our equation, we get

$$\Rightarrow \frac{d}{d\boldsymbol{\theta}} \int_{-\infty}^{\infty} p(\mathbf{y} \mid \boldsymbol{\theta})d\mathbf{y} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{\theta}}p(\mathbf{y} \mid \boldsymbol{\theta})d\mathbf{y} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{\theta}}h(\mathbf{y})e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})}d\mathbf{y} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} h(\mathbf{y})e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})}\left(\mathbf{y} - \frac{\partial}{\partial \boldsymbol{\theta}}\psi(\boldsymbol{\theta})\right)d\mathbf{y} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \mathbf{y} h(\mathbf{y}) e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})} d\mathbf{y} - \int_{-\infty}^{\infty} h(\mathbf{y}) e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) d\mathbf{y} = 0$$

$$\Rightarrow \mathbb{E}(\mathbf{y}|\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \int_{-\infty}^{\infty} h(\mathbf{y}) e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})} d\mathbf{y} = 0$$

$$\Rightarrow \mathbb{E}(\mathbf{y}|\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}); \text{ Since } \int_{-\infty}^{\infty} h(\mathbf{y}) e^{\boldsymbol{\theta}^T \mathbf{y} - \psi(\boldsymbol{\theta})} d\mathbf{y} = 1. \underline{\text{ Hence proved (a.)}}$$

b. Now we have,
$$p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) = h(\mathbf{y}_0, \lambda) \exp\{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})\}$$

We know that,

$$\int_{-\infty}^{\infty} p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) d\boldsymbol{\theta} = 1$$

$$\Rightarrow \frac{d}{d\boldsymbol{\theta}} \int_{-\infty}^{\infty} p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) d\boldsymbol{\theta} = 0; \text{ Taking derivative on both sides wrt } \boldsymbol{\theta}$$

$$\Rightarrow \frac{d}{d\boldsymbol{\theta}} \int_{-\infty}^{\infty} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} d\boldsymbol{\theta} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{\theta}} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} d\boldsymbol{\theta} = 0; \text{ Applying Leibniz rule as stated above}$$

$$\Rightarrow \int_{-\infty}^{\infty} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} \left(\mathbf{y}_0 - \lambda \frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) d\boldsymbol{\theta} = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \mathbf{y}_0 h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} d\boldsymbol{\theta} - \int_{-\infty}^{\infty} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} \left(\lambda \frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) d\boldsymbol{\theta} = 0$$

Since we know that
$$\int_{-\infty}^{\infty} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} d\boldsymbol{\theta} = 1$$

and
$$\int_{-\infty}^{\infty} h(\mathbf{y}_0, \lambda) e^{\boldsymbol{\theta}^T \mathbf{y}_0 - \lambda \psi(\boldsymbol{\theta})} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) d\boldsymbol{\theta} = \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right)$$

$$\Rightarrow \mathbf{y}_0(1) = \lambda \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) + c$$

$$\Rightarrow \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi(\boldsymbol{\theta}) \right) = \frac{\mathbf{y}_0}{\lambda} + c$$

$$\Rightarrow \mathbb{E}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \frac{\mathbf{y}_0}{\lambda} + c. \underline{\text{ Hence proved (b.)}}$$

Another way to think of this result is to use the hint given that,

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\lambda} \left\{ \mathbf{y}_0 - \left(\mathbf{y}_0 - \lambda \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right\}$$

Let us consider the expectation of this expression on both sides (w.r.t $p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda)$).

$$
\begin{aligned}
\mathbb{E}\left(\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) &= \mathbb{E}\left(\frac{\mathbf{y}_0}{\lambda}\right) - \frac{1}{\lambda}\mathbb{E}\left(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \\
&= \left(\frac{\mathbf{y}_0}{\lambda}\right) - \frac{1}{\lambda}\int_{-\infty}^{\infty}\left(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \times p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda)d\boldsymbol{\theta} \\
&= \left(\frac{\mathbf{y}_0}{\lambda}\right) - \frac{1}{\lambda}\int_{-\infty}^{\infty}\left(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \times h(\mathbf{y}_0, \lambda)\exp\{\boldsymbol{\theta}^T\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})\}d\boldsymbol{\theta} \\
&= \left(\frac{\mathbf{y}_0}{\lambda}\right) - \frac{h(\mathbf{y}_0, \lambda)}{\lambda} \times c
\end{aligned}
$$

Note that this can be done because if we assume $\boldsymbol{\theta}^T\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta}) = k$, then $(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})d\boldsymbol{\theta} = dk$. Now, $\int_{-\infty}^{\infty}\exp\{k\}dk$ equates to a constant as k may occupy any value in the support. Therefore, our final expression becomes,

$$
\mathbb{E}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \frac{\mathbf{y}_0}{\lambda} + c.
$$

c. Now, we need to show that

$$
\mathbb{E}(\boldsymbol{\xi}(\boldsymbol{\theta}) \mid y_{1:n}) = \frac{\mathbf{y_0} + n\bar{\mathbf{y}}}{\lambda + n} + \text{constant}
$$

We have,

$$
\begin{aligned}
p(\boldsymbol{\theta} \mid y_{1:n}) &\propto p(\boldsymbol{\theta}) \times p(y_{1:n} \mid \boldsymbol{\theta}) \\
&\propto p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) \times p(y_{1:n} \mid \boldsymbol{\theta}) \\
&\propto h(\mathbf{y}_0, \lambda)e^{\boldsymbol{\theta}^T\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})} \times \prod_{i=1}^{n} h(\mathbf{y}_i)e^{\boldsymbol{\theta}^T\mathbf{y}_i - \psi(\boldsymbol{\theta})} \\
&\propto h(\mathbf{y}_0, \lambda) \times \prod_{i=1}^{n} h(\mathbf{y}_i) \times e^{\boldsymbol{\theta}^T(\mathbf{y}_0 + \sum_{i=1}^{n}y_i) - \psi(\boldsymbol{\theta})(\lambda + \sum_{i=1}^{n}1)} \\
&\propto h'(\mathbf{y}_0, \lambda)e^{\boldsymbol{\theta}^T(\mathbf{y}_0 + n\bar{\mathbf{y}}) - \psi(\boldsymbol{\theta})(\lambda + n)}
\end{aligned}
$$

We observe that this pdf is similar to the previous question, where $p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) = h(\mathbf{y}_0, \lambda)\exp\{\boldsymbol{\theta}^T\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})\}$. Here, we proved that

$$
\mathbb{E}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \frac{\mathbf{y}_0}{\lambda} + c
$$

By appealing to a similar conjugacy and by comparing the distributions, we can say that for this distribution $p(\boldsymbol{\theta} \mid y_{1:n})$

$$
\mathbb{E}(\boldsymbol{\xi}(\boldsymbol{\theta}) \mid y_{1:n}) = \frac{\mathbf{y_0} + n\bar{\mathbf{y}}}{\lambda + n} + \text{constant}
$$

2. Show that the binomial and negative binomial distributions belong to exponential families.

Solution: The exponential family of distributions is characterized by having the probability distribution in the form,

$$
p(\mathbf{y} \mid \boldsymbol{\theta}) = h(\mathbf{y})\exp\{\phi(\boldsymbol{\theta})^T\mathbf{u}(\mathbf{y}) - \psi(\boldsymbol{\theta})\}
$$

Let us consider a binomial distribution,

$$y_1 \ldots y_n \overset{iid}{\sim} Bin(n, \theta)$$

Let the probability of success is θ. The pmf of this distribution is given by,

$$
\begin{aligned}
p(y|\theta) &= \binom{n}{y} (\theta)^y (1-\theta)^{n-y} \\
&= \binom{n}{y} e^{log\left((\theta)^y (1-\theta)^{n-y}\right)} \\
&= \binom{n}{y} e^{(y)log(\theta)+(n-y)log(1-\theta)} \\
&= \binom{n}{y} e^{(y)(log(\theta)-log(1-\theta))+(n)log(1-\theta)} \\
&= \binom{n}{y} e^{(y)(log\frac{\theta}{1-\theta})+(n)log(1-\theta)}
\end{aligned}
$$

Here, the following similarities are observed with the exponential family of distributions.

$$
\begin{aligned}
\phi(\boldsymbol{\theta})^T &= log\frac{\theta}{1-\theta} \\
\mathbf{u}(\mathbf{y}) &= y \\
h(\mathbf{y}) &= \binom{n}{y} \\
\psi(\boldsymbol{\theta}) &= -nlog(1-\theta)
\end{aligned}
$$

Therefore, we can conclude that the binomial distributions belongs to exponential family of distributions. Now, let us consider the negative binomial distribution. Described as the probability of getting k failures before r successes, this pmf is given by,

$$
\begin{aligned}
p(k|r, \theta) &= \binom{k+r-1}{r-1} (\theta)^r (1-\theta)^k \\
&= \binom{k+r-1}{r-1} e^{log\left((\theta)^r (1-\theta)^k\right)} \\
&= \binom{k+r-1}{r-1} e^{(k)log(1-\theta)+(r)log(\theta)}
\end{aligned}
$$

Here, the following similarities are observed with the exponential family of distributions.

$$
\begin{aligned}
\phi(\boldsymbol{\theta})^T &= log(1-\theta) \\
\mathbf{u}(\mathbf{y}) &= k \\
h(\mathbf{y}) &= \binom{k+r-1}{r-1} \\
\psi(\boldsymbol{\theta}) &= -(r)log(\theta)
\end{aligned}
$$

Therefore, we can conclude that the negative binomial distributions belongs to exponential family of distributions.

3.

Solution:

a. It is given that $y \sim Bin(10, \theta)$. We know that the likelihood function for the binomial distribution is given by

$$p(y_{1:n} \mid \theta) \propto \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{n-y_i}$$

Here n = 10, and it is given that y = 3. Therefore, our likelihood function becomes,

$$p(y \mid \theta) \propto \theta^3(1-\theta)^7$$

It is given that the prior is given by a mixture of two beta's as,

$$p(\theta) = \frac{0.5}{\text{Beta}(10,20)}\theta^{10-1}(1-\theta)^{20-1} + \frac{0.5}{\text{Beta}(20,10)}\theta^{20-1}(1-\theta)^{10-1}$$

We know that,

$$p(\theta \mid y) \propto p(y \mid \theta) \times p(\theta)$$
$$\propto [\theta^3(1-\theta)^7] \times \left[\frac{0.5}{\text{Beta}(10,20)}\theta^9(1-\theta)^{19} + \frac{0.5}{\text{Beta}(20,10)}\theta^{19}(1-\theta)^9\right]$$
$$\propto \frac{0.5}{\text{Beta}(10,20)}\theta^{12}(1-\theta)^{26} + \frac{0.5}{\text{Beta}(20,10)}\theta^{22}(1-\theta)^{16}$$
$$= c \times \left[\frac{0.5}{\text{Beta}(10,20)}\theta^{12}(1-\theta)^{26} + \frac{0.5}{\text{Beta}(20,10)}\theta^{22}(1-\theta)^{16}\right]$$

Let c be a constant. We know that the posterior is proportional to the above expression, however, we are asked to find an exact value for the pdf. For this, we can utilize the idea that the total area under the density = 1, i.e.,

$$\int_0^1 c \times \left[\frac{0.5}{\text{Beta}(10,20)}\theta^{12}(1-\theta)^{26} + \frac{0.5}{\text{Beta}(20,10)}\theta^{22}(1-\theta)^{16}\right] d\theta = 1$$

$$\frac{1}{\int_0^1 \left[\frac{1}{\text{Beta}(10,20)}\theta^{12}(1-\theta)^{26} + \frac{1}{\text{Beta}(20,10)}\theta^{22}(1-\theta)^{16}\right] d\theta} = c'; c' = c \times 0.5$$

Therefore, we get the value of c' as

$$c' = 2.1278644 \times 10^{-3}; \text{ I used Wolfram alpha calculator to obtain the integral value}$$

Now, we can verify that

$$\frac{\text{Beta}(13,27)}{\text{Beta}(20,10)} + \frac{\text{Beta}(23,17)}{\text{Beta}(10,20)} = c' = 2.1278644 \times 10^{-3}$$

Therefore, we can rewrite our posterior as,

$$p(\theta \mid y) = \frac{\frac{\text{Beta}(13,27)}{\text{Beta}(13,27)\text{Beta}(10,20)}\theta^{12}(1-\theta)^{26} + \frac{\text{Beta}(23,17)}{\text{Beta}(23,17)\text{Beta}(20,10)}\theta^{22}(1-\theta)^{16}}{\frac{\text{Beta}(13,27)}{\text{Beta}(20,10)} + \frac{\text{Beta}(23,17)}{\text{Beta}(10,20)}}; \quad \underline{\textit{Hence Proved}}$$

5

b. We plot the posterior superimposed on the prior as follows. The code accompanying this plot is explained and attached with this assignment.
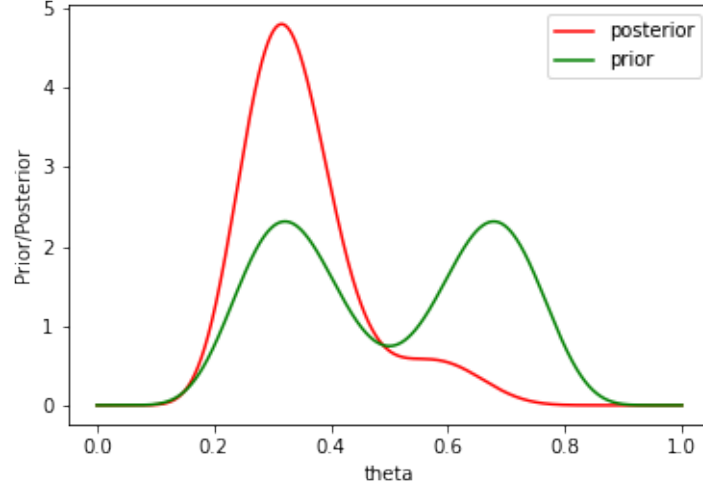

Figure 1: Posterior superimposed on the beta mixture prior

c. We are asked to find a 90% posterior credible intreval for θ. Here, $\alpha = 0.1$. We do this in Python, using the Quantile based credible intrevals, as it was mentioned in the lecture notes that Quantile based credible intrevals are used much more in practice. It is defined as,

$$C_{\theta,\alpha} = \{\theta : Pr(C_{\theta,\alpha \mid y_{1:n}}) = 1 - \alpha\}$$
$$= \{\theta_{n,1-\alpha_1} \leq \theta \leq \theta_{n,\alpha_2}, \alpha_1 + \alpha_2 = \alpha\}$$

We implement a function that calculates the inverse pdf values for our posterior distribution. Now using these values, we consider the 5th quantile and the 95th quantile values for the posterior distribution as this is essentially meaning that the probability of occurence of values in this intreval $\geq 90\%$. In these calculations, we generate 100000 values of θ and find the posterior distribution using the expression as above. By inputting these posterior distribution values, and by calculating the inverse pdf, we can get the values of θ. From this, we can calculate the 5th and 95th quantiles, as here $\alpha_1 = 0.05$. Our calculations leads us to a value of the quantiles as $q_1 = 0.21266799901229458, q_2 = 0.5844785401328614$. The graphical representation of this is also given below.
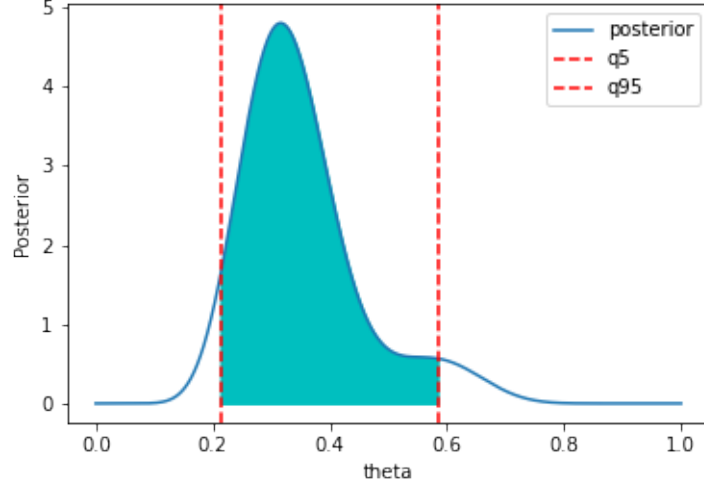
Figure 2: Quantile based credible regions

4.

Solution: We are given that the Jeffry's prior is given by

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{-\frac{1}{2}}$$

Now, we consider a one-one function $\boldsymbol{\psi} = g(\boldsymbol{\theta})$. This implies, $\boldsymbol{\theta} = g^{-1}(\boldsymbol{\psi})$. Now, let us evaluate $\mathbf{I}(\boldsymbol{\psi})$ at this new $\boldsymbol{\theta} = g^{-1}(\boldsymbol{\psi})$. This gives us,

$$
\begin{aligned}
\mathbf{I}(\boldsymbol{\psi}) &= -\mathbb{E}\left(\frac{\partial^2 \log p(y|\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^2}\right) \\
&= -\mathbb{E}\left(\frac{\partial^2 \log p(y|\boldsymbol{\theta} = g^{-1}(\boldsymbol{\psi}))}{\partial \boldsymbol{\theta}^2}\left|\frac{d\boldsymbol{\theta}}{d\boldsymbol{\psi}}\right|^2\right) \\
&= \mathbf{I}(\boldsymbol{\theta})\left|\frac{d\boldsymbol{\theta}}{d\boldsymbol{\psi}}\right|^2
\end{aligned}
$$

Computing the Jeffry's prior on $\boldsymbol{\psi} = g(\boldsymbol{\theta})$ from here, we get,

$$
\begin{aligned}
p(\boldsymbol{\psi}) &\propto |\mathbf{I}(\boldsymbol{\psi})|^{-\frac{1}{2}} \\
&\propto \mathbf{I}(\boldsymbol{\theta})\left|\frac{d\boldsymbol{\theta}}{d\boldsymbol{\psi}}\right|^2 \\
&\propto \mathbf{I}(\boldsymbol{\theta})
\end{aligned}
$$

Thus, we are able to show that Jeffrey's priors satisfies the invariance principle.

5.

Solution: It is given that

$$y_1 \ldots y_n \sim \text{Pois}(\lambda)$$

The likelihood function is given by

$$p(y_{1:n} \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

7

We arrived at the Jeffry's prior for the Poisson distribution as,

$$p(\lambda) \propto \lambda^{\frac{-1}{2}}$$

Now, we can calculate the posterior as follows,

$$p(\lambda \mid y_{1:n}) \propto p(\lambda) \times p(y_{1:n} \mid \lambda)$$

$$\propto \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \times \lambda^{\frac{-1}{2}}$$

$$\propto (e^{-\lambda})^n \times \lambda^{y_1} \ldots \lambda^{y_n} \times \lambda^{\frac{-1}{2}}$$

$$\propto e^{-n\lambda}\lambda^{y_1 + \cdots + y_n - \frac{1}{2}}$$

This posterior looks like the pmf of a Gamma distribution. We need to note that the distribution of the posterior is now in terms of λ. The Gamma distribution is given by,

$$f(x|\alpha, \beta) = \mathrm{Gamma}(\alpha, \beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)}$$

Essentially here, we can see that,

$$x = \lambda$$

$$\beta = n$$

$$\alpha = y_1 + \ldots y_n + \frac{1}{2}$$

Therefore, we can conclude that the posterior is of the form $\mathrm{Gamma}(\alpha, \beta)$, with α and β as described above, and is a proper density.

6.

Solution: It is given that,

$$y_1, \ldots, y_n \sim N(\mu, \sigma^2)$$

with μ being known.

a. The Jeffrey's prior for σ^2 can be calculated as follows. First, we calculate the likelihood function for σ^2. This has been done in previous homeworks, and we get

$$L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

From this, we can arrive at the Fisher information matrix as,

$$\mathbf{I}(\sigma^2) = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\sigma^2)}{\partial(\sigma^2)^2}\right]$$

$$= -\mathbb{E}\left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$= -\left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}n\sigma^2\right]$$

$$= \frac{n}{2\sigma^4}$$

Therefore, the Jeffrey's prior for σ^2 is given as,

$$p(\sigma^2) \propto |\mathbf{I}(\sigma^2)|^{\frac{1}{2}}$$

$$\propto \left| \frac{n}{2\sigma^4} \right|$$

$$\propto \frac{1}{\sigma^2}$$

$b.$ We can compute the corresponding posterior using the following relation.

$$p(\sigma^2|y_{1:n}) \propto p(y_{1:n}|\sigma^2) \times p(\sigma^2)$$

$$\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2} \times \frac{1}{\sigma^2}$$

$$\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma^{1+\frac{2}{n}}} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

$$\propto \left(\frac{1}{\sqrt{2\pi}\sigma^{1+\frac{2}{n}}} \right)^n \prod_{i=1}^{n} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1} e^{\frac{-1}{2\sigma^2} \sum_{i=1}^{n}(y_i - \mu)^2}$$

This looks like the pdf of an inverse gamma distribution. The pdf of the inverse gamma distribution is given by

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x} \right)^{\alpha+1} \exp\left(\frac{-\beta}{x} \right)$$

Our posterior is constructed with the following parameters,

$$\alpha = \frac{n}{2}$$

$$\beta = \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}$$

$c.$ First, we consider a Normal(0,1) distribution and then we draw a sample size of 20 from this distribution. Using this, we can calculate the value of α and β for the posterior. We then plot the distribution of the posterior over the support σ^2 and superimpose it over the prior. The general shape of the prior is given by the following plot, for values of σ^2 between 0 and 1.
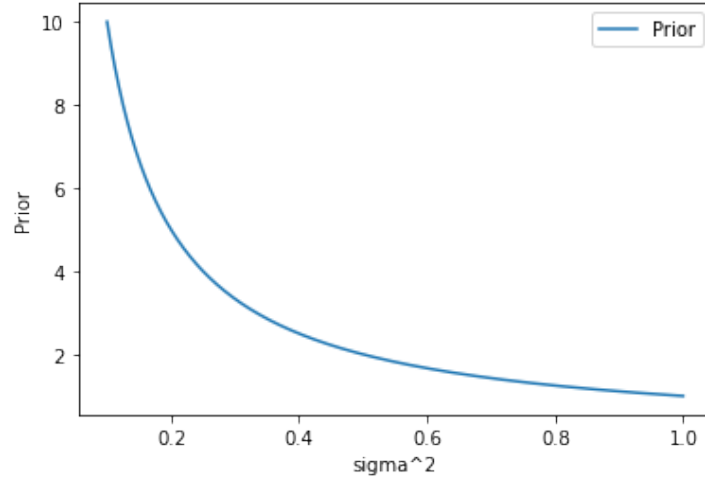
Figure 3: Prior Distribution

We get the values of $\alpha = 10$ and $\beta =$. The distribution of the posterior is illustrated in the graph below. We plot the prior superimposed on the posterior as follows. The code for constructing this plot is explained and attached with the assignment.
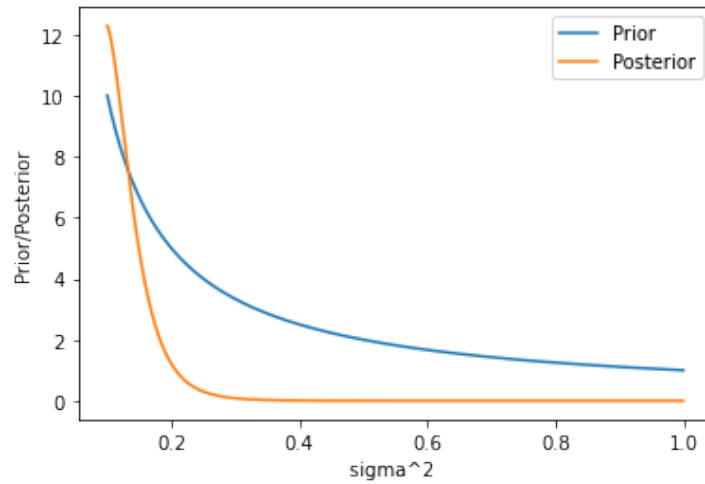

Figure 4: Posterior superimposed on the Normal prior

d. To calculate a 90% centered quantile based credible interval, we make use of the approach followed in Q.3., and we reuse the function defined and implemented in Q.3. Following a similar approach, we get lower and upper bounds of the quantiles as, $(q_5, q_{95}) = (0.10242307810098894, 0.21316013463489708)$. This converts to the following credible region on the posterior.
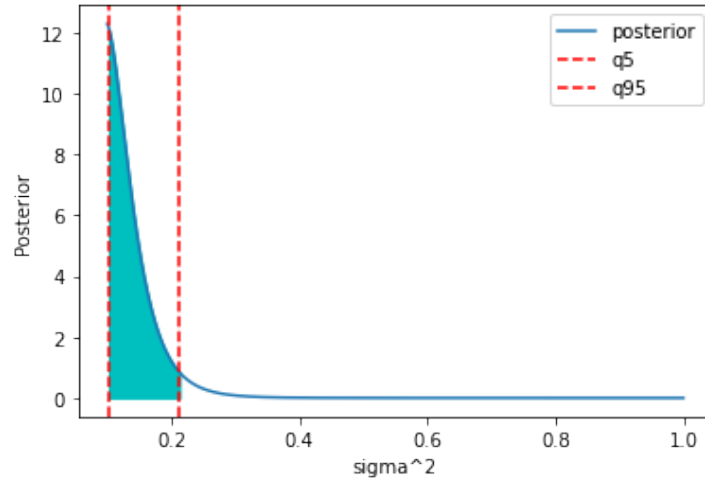
Figure 5: Quantile based credible regions on the posterior

e. The HPD credible regions are defined in a similar fashion to the Quantile based credible regions. Let us consider a 100*(1 - α) posterior credible region to be defined as,

$$C_{\theta,\alpha} = \{\theta : Pr(C_\theta, \alpha \mid y_{1:n}) = 1 - \alpha, p(\theta_1 \mid y_{1:n}) \geq p(\theta_2 \mid y_{1:n}) \forall \theta_1 \in C_\theta, \theta_2 \in C_\theta^c\}$$

For calculating the HPD credible intervals, we make use of the implementation by Packt-Publishing on this link: `https://github.com/PacktPublishing/Bayesian-Analysis-with-Python/blob/master/Chapter%201/hpd%20(1).py`

Using this, we calculate the HPD lower and upper bound values as $(\text{hpd}_l, \text{hpd}_u) = (0.1, 0.19)$. Now, using this interval, we plot the credible region as,
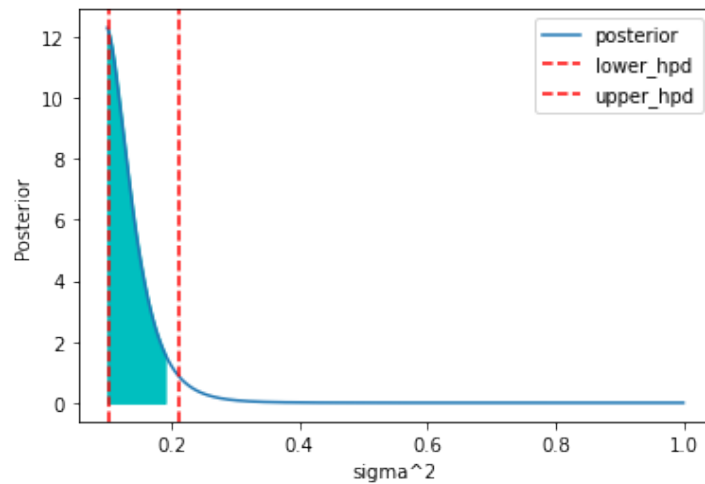

Figure 6: HPD based credible regions on the posterior

The code accomanying these graphs with detailed explanations and comments are attached with this assigment.

11