Solutions for Week 4 & 5

1. (a) Let $\boldsymbol{y} = \{y_1, \ldots, y_N\}$. By Bayes' theorem, the posterior distribution is $\pi(p \mid \boldsymbol{y}) \propto \pi(\boldsymbol{y} \mid p)\pi(p)$. The likelihood function is

$$\begin{aligned}\pi(\boldsymbol{y} \mid p) &= \prod_{i=1}^{N} \binom{n}{y_i} p^{y_i}(1-p)^{n-y_i} \\ &\propto p^{\sum y_i}(1-p)^{nN-\sum y_i}\end{aligned}$$

The prior density is proportional to $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$. The posterior density is therefore

$$\begin{aligned}\pi(p \mid \boldsymbol{y}) &\propto p^{\sum y_i}(1-p)^{nN-\sum y_i} p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{\sum y_i+\alpha-1}(1-p)^{nN-\sum y_i+\beta-1}\end{aligned}$$

This has the functional form of the Beta distribution, so

$$p \mid \boldsymbol{y} \sim \text{Beta}\left(\sum_{i=1}^{N} y_i + \alpha,\, nN - \sum_{i=1}^{N} y_i + \beta\right).$$

(b) Suppose we observe a new data point, y_{N+1}. Note that

$$\pi(p \mid \boldsymbol{y}, y_{N+1}) \propto \pi(y_{N+1} \mid p, \boldsymbol{y})\pi(p \mid \boldsymbol{y})\pi(\boldsymbol{y}) \propto \pi(y_{N+1} \mid p)\pi(p \mid \boldsymbol{y}).$$

We have

$$\pi(y_{N+1} \mid \theta) = \binom{n}{y_{N+1}} p^{y_{N+1}}(1-p)^{n-y_{N+1}},$$

and therefore

$$\begin{aligned}\pi(p \mid \boldsymbol{y}, y_{N+1}) &\propto p^{y_{N+1}}(1-p)^{n-y_{N+1}} p^{\sum_{i=1}^{N} y_i+\alpha-1}(1-p)^{nN-\sum_{i=1}^{N} y_i+\beta-1} \\ &= p^{\sum_{i=1}^{N+1} y_i+\alpha-1}(1-p)^{n(N+1)-\sum_{i=1}^{N+1} y_i+\beta-1}\end{aligned}$$

This has the functional form of the Beta distribution, so

$$p \mid \boldsymbol{y}, y_{N+1} \sim \text{Beta}\left(\sum_{i=1}^{N+1} y_i + \alpha,\, n(N+1) - \sum_{i=1}^{N+1} y_i + \beta\right).$$

This is the same distribution as in (a), but with N replaced by $N+1$.

(c) This is a very similar exercise to (a), but with the likelihood function now being

$$\begin{aligned}\pi(\boldsymbol{y}, y_{N+1} \mid p) &= \prod_{i=1}^{N+1} \binom{n}{y_i} p^{y_i} (1-p)^{n-y_i} \\ &\propto p^{\sum_{i=1}^{N+1} y_i} (1-p)^{n(N+1) - \sum_{i=1}^{N+1} y_i}.\end{aligned}$$

Similar calculation as in part (a) leads to

$$p \mid \boldsymbol{y}, y_{N+1} \sim \text{Beta}\left(\sum_{i=1}^{N+1} y_i + \alpha,\, n(N+1) - \sum_{i=1}^{N+1} y_i + \beta \right).$$

This means that the posterior distribution is the same regardless of if we observe all the data in one go, or if we collect more data as we go along. This is similar to human reasoning, where we can update our beliefs based on evidence as we observe it.

2. (a) By Bayes' theorem, the posterior distribution is

$$\pi(\lambda \mid \boldsymbol{y}) \propto \pi(\boldsymbol{y} \mid \lambda) \pi(\lambda).$$

The likelihood function is the product of the N density functions

$$\pi(\boldsymbol{y} \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \propto \lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda}.$$

The prior distribution is

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Therefore, the posterior density

$$\begin{aligned}\pi(\lambda \mid \boldsymbol{y}) &\propto \lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &= \lambda^{\sum_{i=1}^{N} y_i + \alpha - 1} e^{-(N+\beta)\lambda}.\end{aligned}$$

This has the functional form of a Gamma distribution, and so the posterior distribution is

$$\lambda \mid \boldsymbol{y} \sim \text{Gamma}\left(\sum_{i=1}^{N} y_i + \alpha,\, N + \beta \right).$$

(b) With $\alpha = 1$, the posterior mean is

$$\mathbb{E}(\lambda \mid \boldsymbol{y}) = \left(\frac{\sum_{i=1}^{N} y_i + 1}{N + \beta} \right).$$

and the posterior variance is

$$Var(\lambda \mid \boldsymbol{y}) = \left(\frac{\sum_{i=1}^{N} y_i + 1}{(N + \beta)^2} \right).$$

Large value of β (i.e., having a stronger prior belief that λ is small) decreases both the posterior mean and the posterior variance. Small values of β leads to higher posterior mean and variance. Additionally, small values of β makes the prior less informative, allowing the data to have a greater influence on the posterior.

(c) The posterior predictive distribution for a new observation $\tilde{y}$ is

$$\pi(\tilde{y} \mid \boldsymbol{y}) = \int \pi(\tilde{y} \mid \lambda)\pi(\lambda \mid \boldsymbol{y})\, d\lambda.$$

The first term in the integral is the likelihood function evaluated at $\tilde{y}$ and the second term is the posterior distribution, i.e. the information we've learnt about λ so far. We can now work through the algebra to derive the terms for this:

$$\begin{aligned}\pi(\tilde{y} \mid \boldsymbol{y}) &= \int \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!} \frac{(\beta+N)^{\sum y+\alpha}}{\Gamma(\alpha+\sum_{i=1}^N y_i)} \lambda^{\sum_{i=1}^N y_i+\alpha-1} e^{-(N+\beta)\lambda}\, d\lambda \\ &= \frac{(\beta+N)^{\sum y+\alpha}}{\tilde{y}!\Gamma(\alpha+\sum_{i=1}^N y_i)} \int \lambda^{\tilde{y}+\sum_{i=1}^N y_i+\alpha-1} e^{-(N+\beta+1)\lambda}\, d\lambda.\end{aligned}$$

We can turn the integrand into the density function of Gamma $\left(\tilde{y}+\sum_{i=1}^N y_i+\alpha,\ N+\beta+1\right)$, by multiplying inside the integrating constant, and dividing by it outside.

$$\begin{aligned}\pi(\tilde{y} \mid \boldsymbol{y}) &= \frac{(\beta+N)^{\sum y+\alpha}}{\tilde{y}!\Gamma(\alpha+\sum_{i=1}^N y_i)} \frac{\Gamma(\tilde{y}+\sum_{i=1}^N y_i+\alpha)}{(N+\beta+1)^{\tilde{y}+\sum_{i=1}^N y_i+\alpha}} \int \frac{(N+\beta+1)^{\tilde{y}+\sum_{i=1}^N y_i+\alpha}}{\Gamma(\tilde{y}+\sum_{i=1}^N y_i+\alpha)} \\ &\qquad \times \lambda^{\tilde{y}+\sum_{i=1}^N y_i+\alpha-1} e^{-(N+\beta+1)\lambda}\, d\lambda \\ &= \frac{(\beta+N)^{\sum y+\alpha}}{\tilde{y}!\Gamma(\alpha+\sum_{i=1}^N y_i)} \frac{\Gamma(\tilde{y}+\sum_{i=1}^N y_i+\alpha)}{(N+\beta+1)^{\tilde{y}+\sum_{i=1}^N y_i+\alpha}} \\ &= \frac{\Gamma(\tilde{y}+\sum_{i=1}^N y_i+\alpha)}{\tilde{y}!\Gamma(\alpha+\sum_{i=1}^N y_i)} \left(\frac{N+\beta}{N+\beta+1}\right)^{\sum_{i=1}^N y_i+\alpha} \left(\frac{1}{N+\beta+1}\right)^{\tilde{y}}.\end{aligned}$$

This is the probability mass function of Negative Binomial distribution with $p = (N+\beta)/(N+\beta+1)$ and $r = \sum_{i=1}^N y_i + \alpha$, evaluated at $\tilde{y}$.

3. To find density of Y, we start with finding its cumulative distribution function. We first consider the case where h is a strictly increasing function

$$\pi(Y \le y) = \pi(h(X) \le y) = \pi(X \le h^{-1}(y)) = F_X(h^{-1}(y)),$$

where $F_X(\cdot)$ is the CDF for X. Now we take derivative with respect to y and obtain

$$\pi(y) = \frac{\partial}{\partial y} F_X(h^{-1}(y)) = \pi(h^{-1}(y)) \frac{\partial h^{-1}(y)}{\partial y} = \pi(x)\frac{\partial x}{\partial y}.$$

If h is strictly decreasing, almost the same argument leads to

$$\pi(y) = -\pi(x)\frac{\partial x}{\partial y}.$$

Therefore, combining these two cases, we have

$$\pi(y) = \pi(x)\Big|\frac{\partial x}{\partial y}\Big|.$$

Now, let $X \sim \text{Exp}(1)$. Since $Y = \sqrt{X}$, we have $X = Y^2$. Note that $\pi(x) = e^{-x}$ and Differentiating with respect to Y gives

$$\frac{dx}{dy} = 2y, \quad \text{so} \quad \Big|\frac{dx}{dy}\Big| = 2y.$$

Therefore,

$$\pi(y) = e^{-x}2y = e^{-y^2}2y, \qquad y > 0.$$

4. The density function for the exponential distribution is

$$\pi(x \mid \lambda) = \lambda e^{-\lambda x}, \quad x > 0.$$

(a) To find an invariant (Jefferys) prior, we need to compute the Fisher information. Note that

$$\log \pi(X \mid \lambda) = \log \lambda - \lambda X.$$

Differentiating twice gives,

$$\begin{aligned} \frac{\partial \log \pi(X \mid \lambda)}{\partial \lambda} &= \frac{1}{\lambda} - X \\ \frac{\partial^2 \log \pi(X \mid \lambda)}{\partial \lambda^2} &= -\frac{1}{\lambda^2}. \end{aligned}$$

The Fisher information is therefore

$$\mathbb{E}\left(-\frac{\partial^2 \log \pi(X \mid \lambda)}{\partial \lambda^2}\right) = \frac{1}{\lambda^2}.$$

By Jeffreys' theorem, an invariant prior distribution is

$$\pi(\lambda) \propto \frac{1}{\lambda}, \lambda > 0.$$

(b) By Bayes' theorem, the posterior distribution is $\pi(\lambda \mid \boldsymbol{x}) \propto \pi(\boldsymbol{x} \mid \lambda)\pi(\lambda)$, where $\boldsymbol{x} = (x_1, \ldots, x_n)$. Using the invariant prior distribution gives

$$\pi(\lambda \mid \boldsymbol{x}) \propto \lambda^{n-1} \exp\left\{-\lambda \sum_{i=1}^{n} x_i\right\}.$$

This function form shows the posterior distribution is a Gamma distribution

$$\lambda \mid \boldsymbol{x} \sim \text{Gamma}(n, \sum_{i=1}^{n} x_i).$$

(c) Consider the integral of the prior distribution over the domain of the parameter $\lambda \in (0, \infty)$:

$$\int_0^\infty \frac{1}{\lambda}\, d\lambda = \infty,$$

which suggests that it is an improper prior distribution. Note that in this case the posterior distribution is still a valid distribution and therefore it is fine to use an improper prior.