

MA40189 - Solution Sheet Five

Simon Shaw, s.shaw@bath.ac.uk
<https://people.bath.ac.uk/masss/ma40189.html>

2021/22 Semester II

1. Let X_1, \dots, X_n be exchangeable so that the X_i are conditionally independent given a parameter θ . For each of the following distributions for $X_i | \theta$ find the Jeffreys prior and the corresponding posterior distribution for θ .

- (a) $X_i | \theta \sim \text{Bern}(\theta)$.

The likelihood is

$$\begin{aligned} f(x | \theta) &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} \\ &= \theta^{n\bar{x}} (1 - \theta)^{n - n\bar{x}}. \end{aligned}$$

As θ is univariate

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x | \theta) &= \frac{\partial^2}{\partial \theta^2} \{n\bar{x} \log \theta + (n - n\bar{x}) \log(1 - \theta)\} \\ &= -\frac{n\bar{x}}{\theta^2} - \frac{(n - n\bar{x})}{(1 - \theta)^2}. \end{aligned}$$

The Fisher information is thus

$$\begin{aligned} I(\theta) &= -E \left\{ -\frac{n\bar{X}}{\theta^2} - \frac{(n - n\bar{X})}{(1 - \theta)^2} \middle| \theta \right\} \\ &= \frac{nE(\bar{X} | \theta)}{\theta^2} + \frac{\{n - nE(\bar{X} | \theta)\}}{(1 - \theta)^2} \\ &= \frac{n}{\theta} + \frac{n}{1 - \theta} = \frac{n}{\theta(1 - \theta)}. \end{aligned}$$

The Jeffreys prior is then

$$f(\theta) \propto \sqrt{\frac{n}{\theta(1 - \theta)}} \propto \theta^{-\frac{1}{2}} (1 - \theta)^{-\frac{1}{2}}.$$

This is the kernel of a $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ distribution. This makes it straightforward to obtain the posterior using the conjugacy of the Beta distribution relative to the Bernoulli likelihood. The posterior is $\text{Beta}(\frac{1}{2} + n\bar{x}, \frac{1}{2} + n - n\bar{x})$. (If you are not sure why, see Solution Sheet Four Question 1 (a) with $\alpha = \beta = \frac{1}{2}$.)

(b) $X_i | \theta \sim Po(\theta)$.

The likelihood is

$$\begin{aligned} f(x | \theta) &= \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &= \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^n x_i!}. \end{aligned}$$

As θ is univariate

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x | \theta) &= \frac{\partial^2}{\partial \theta^2} \left(n\bar{x} \log \theta - n\theta - \sum_{i=1}^n \log x_i! \right) \\ &= -\frac{n\bar{x}}{\theta^2}. \end{aligned}$$

The Fisher information is thus

$$\begin{aligned} I(\theta) &= -E \left(-\frac{n\bar{X}}{\theta^2} \mid \theta \right) \\ &= \frac{nE(\bar{X} | \theta)}{\theta^2} = \frac{n}{\theta}. \end{aligned}$$

The Jeffreys prior is then

$$f(\theta) \propto \sqrt{\frac{n}{\theta}} \propto \theta^{-\frac{1}{2}}.$$

This is often expressed as the improper $Gamma(\frac{1}{2}, 0)$ distribution. This makes it straightforward to obtain the posterior using the conjugacy of the Gamma distribution relative to the Poisson likelihood. The posterior is $Gamma(\frac{1}{2} + n\bar{x}, n)$. (If you are not sure why, see Solution Sheet Four Question 4 (a) with $\alpha = \frac{1}{2}$, $\beta = 0$.)

(c) $X_i | \theta \sim Maxwell(\theta)$, the Maxwell distribution with parameter θ so that

$$f(x_i | \theta) = \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \theta^{\frac{3}{2}} x_i^2 \exp \left\{ -\frac{\theta x_i^2}{2} \right\}, \quad x_i > 0$$

and $E(X_i | \theta) = 2\sqrt{\frac{2}{\pi\theta}}$, $Var(X_i | \theta) = \frac{3\pi-8}{\pi\theta}$.

The likelihood is

$$\begin{aligned} f(x | \theta) &= \prod_{i=1}^n \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \theta^{\frac{3}{2}} x_i^2 \exp \left\{ -\frac{\theta x_i^2}{2} \right\} \\ &= \left(\frac{2}{\pi} \right)^{\frac{n}{2}} \theta^{\frac{3n}{2}} \left(\prod_{i=1}^n x_i^2 \right) \exp \left\{ -\frac{\theta}{2} \sum_{i=1}^n x_i^2 \right\}. \end{aligned}$$

As θ is univariate

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x | \theta) &= \frac{\partial^2}{\partial \theta^2} \left(\frac{n}{2} \log \frac{2}{\pi} + \frac{3n}{2} \log \theta + \sum_{i=1}^n \log x_i^2 - \frac{\theta}{2} \sum_{i=1}^n x_i^2 \right) \\ &= -\frac{3n}{2\theta^2}. \end{aligned}$$

The Fisher information is

$$I(\theta) = -E \left(-\frac{3n}{2\theta^2} \middle| \theta \right) = \frac{3n}{2\theta^2}.$$

The Jeffreys prior is then

$$f(\theta) \propto \sqrt{\frac{3n}{2\theta^2}} \propto \theta^{-1}.$$

Thus is often expressed as the improper *Gamma*(0,0) distribution. This makes it straightforward to obtain the posterior using the conjugacy of the Gamma distribution relative to the Maxwell likelihood. The posterior is *Gamma*($\frac{3n}{2}, \frac{1}{2} \sum_{i=1}^n x_i^2$). (If you are not sure why, see Solution Sheet Four Question 1 (c) with $\alpha = \beta = 0$.)

2. Let X_1, \dots, X_n be exchangeable so that the X_i are conditionally independent given a parameter λ . Suppose that $X_i | \lambda \sim \text{Exp}(\lambda)$ where λ represents the rate so that $E(X_i | \lambda) = \lambda^{-1}$.

- (a) Show that $X_i | \lambda \sim \text{Exp}(\lambda)$ is a member of the 1-parameter exponential family. Hence, write down a sufficient statistic $t(X)$ for $X = (X_1, \dots, X_n)$ for learning about λ .

We write

$$f(x_i | \lambda) = \lambda e^{-\lambda x_i} = \exp \{-\lambda x_i + \log \lambda\}$$

which is of the form $\exp\{\phi_1(\lambda)u_1(x_i) + g(\lambda) + h(x_i)\}$ where $\phi_1(\lambda) = -\lambda$, $u_1(x_i) = x_i$, $g(\lambda) = \log \lambda$ and $h(x_i) = 0$. From the Proposition given in Lecture 9 (see p26 of the on-line notes) a sufficient statistic is

$$t(X) = \sum_{i=1}^n u_1(X_i) = \sum_{i=1}^n X_i.$$

- (b) Find the Jeffreys prior and comment upon whether or not it is improper. Find the posterior distribution for this prior.

The likelihood is

$$f(x | \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda n \bar{x}}. \quad (1)$$

As θ is univariate

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x | \theta) &= \frac{\partial^2}{\partial \theta^2} (n \log \lambda - \lambda n \bar{x}) \\ &= -\frac{n}{\lambda^2}. \end{aligned}$$

The Fisher information is

$$I(\lambda) = -E \left(-\frac{n}{\lambda^2} \middle| \lambda \right) = \frac{n}{\lambda^2}.$$

The Jeffreys prior is thus

$$f(\lambda) \propto \sqrt{\frac{n}{\lambda^2}} \propto \lambda^{-1}. \quad (2)$$

This is the improper $Gamma(0, 0)$ distribution. The posterior can be obtained using the conjugacy of the Gamma distribution relative to the Exponential likelihood. It is $Gamma(n, n\bar{x})$. (If you are not sure why, see Solution Sheet Three Question 1 (a) with $\alpha = \beta = 0$.)

(c) Consider the transformation $\phi = \log \lambda$.

i. By expressing $L(\lambda) = f(x | \lambda)$ as $L(\phi)$ find the Jeffreys prior for ϕ .

Inverting $\phi = \log \lambda$ we have $\lambda = e^\phi$. Substituting this into (1) we have

$$L(\phi) = e^{n\phi} \exp\{-n\bar{x}e^\phi\}.$$

We have

$$\begin{aligned} \frac{\partial^2}{\partial \phi^2} \log L(\phi) &= \frac{\partial^2}{\partial \phi^2} (n\phi - n\bar{x}e^\phi) \\ &= -n\bar{x}e^\phi. \end{aligned}$$

The Fisher information is

$$\begin{aligned} I(\phi) &= -E(-n\bar{X}e^\phi | \phi) \\ &= nE(X_i | \phi)e^\phi. \end{aligned}$$

Now $E(X_i | \phi) = E(X_i | \lambda) = \lambda^{-1} = e^{-\phi}$ so that $I(\phi) = n$. The Jeffreys prior is then

$$f_\phi(\phi) \propto \sqrt{n} \propto 1 \quad (3)$$

which is the improper uniform distribution on $(-\infty, \infty)$.

ii. By transforming the distribution of the Jeffreys prior for λ , $f(\lambda)$, find the distribution of ϕ .

From (2) we have that $f_\lambda(\lambda) \propto \lambda^{-1}$. Using the familiar change of variables formula,

$$f_\phi(\phi) = \left| \frac{\partial \lambda}{\partial \phi} \right| f_\lambda(e^\phi)$$

with $\frac{\partial \lambda}{\partial \phi} = e^\phi$ as $\lambda = e^\phi$, we have that

$$f_\phi(\phi) \propto |e^\phi| e^{-\phi} = 1$$

which agrees with (3). This is an illustration of the invariance to reparameterisation of the Jeffreys prior.

3. The Jeffreys prior for Normal distributions. In Lecture 13 we showed that for an exchangeable collection $X = (X_1, \dots, X_n)$ with $X_i | \theta \sim N(\theta, \sigma^2)$ where σ^2 is known the Jeffreys prior for θ is $f(\theta) \propto 1$.

- (a) Consider, instead, that $X_i | \theta \sim N(\mu, \theta)$ where μ is known. Find the Jeffreys prior for θ .

The likelihood is

$$\begin{aligned} f(x | \theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{1}{2\theta} (x_i - \mu)^2 \right\} \\ &= (2\pi)^{-\frac{n}{2}} \theta^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2 \right\}. \end{aligned}$$

As θ is univariate

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x | \theta) &= \frac{\partial^2}{\partial \theta^2} \left\{ -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \theta - \frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ &= \frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

The Fisher information is

$$\begin{aligned} I(\theta) &= -E \left\{ \frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^n (X_i - \mu)^2 \middle| \theta \right\} \\ &= -\frac{n}{2\theta^2} + \frac{1}{\theta^3} \sum_{i=1}^n E \{ (X_i - \mu)^2 | \theta \} \\ &= -\frac{n}{2\theta^2} + \frac{n}{\theta^2} = \frac{n}{2\theta^2}. \end{aligned}$$

The Jeffreys prior is then

$$f(\theta) \propto \sqrt{\frac{n}{2\theta^2}} = \theta^{-1}.$$

- (b) Now suppose that $X_i | \theta \sim N(\mu, \sigma^2)$ where $\theta = (\mu, \sigma^2)$. Find the Jeffreys prior for θ .

The likelihood is

$$f(x | \theta) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

so that

$$\log f(x | \theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

As $\theta = (\mu, \sigma^2)$ then the Fisher Information matrix is

$$I(\theta) = - \begin{pmatrix} E \left\{ \frac{\partial^2}{\partial \mu^2} \log f(x | \theta) \middle| \theta \right\} & E \left\{ \frac{\partial^2}{\partial \mu \partial \sigma^2} \log f(x | \theta) \middle| \theta \right\} \\ E \left\{ \frac{\partial^2}{\partial \mu \partial \sigma^2} \log f(x | \theta) \middle| \theta \right\} & E \left\{ \frac{\partial^2}{\partial (\sigma^2)^2} \log f(x | \theta) \middle| \theta \right\} \end{pmatrix}$$

Now,

$$\begin{aligned}\frac{\partial}{\partial \mu} \log f(x|\theta) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu); \\ \frac{\partial}{\partial (\sigma^2)} \log f(x|\theta) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2;\end{aligned}$$

so that

$$\begin{aligned}\frac{\partial^2}{\partial \mu^2} \log f(x|\theta) &= -\frac{n}{\sigma^2}; \\ \frac{\partial^2}{\partial \mu \partial (\sigma^2)} \log f(x|\theta) &= -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu); \\ \frac{\partial^2}{\partial (\sigma^2)^2} \log f(x|\theta) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2.\end{aligned}$$

Noting that $E(X_i - \mu|\theta) = 0$ and $E\{(X_i - \mu)^2|\theta\} = \sigma^2$, the Fisher information matrix is

$$I(\theta) = -\begin{pmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} - \frac{n}{\sigma^4} \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

The Jeffreys prior is

$$\begin{aligned}f(\theta) &\propto \sqrt{|I(\theta)|} \\ &= \sqrt{\frac{n^2}{2\sigma^6}} \propto \sigma^{-3}.\end{aligned}$$

(c) **Comment upon your answers for these three Normal cases.**

Suppose $\theta = (\mu, \sigma^2)$ and $X_i|\theta \sim N(\mu, \sigma^2)$. If μ is unknown and σ^2 known then the Jeffreys prior is $f(\mu) \propto 1$. If μ is known and σ^2 is unknown then the Jeffreys prior is $f(\sigma^2) \propto \sigma^{-2}$. When both μ and σ^2 are unknown then the Jeffreys prior is $f(\mu, \sigma^2) \propto \sigma^{-3}$. Jeffreys himself found this inconsistent, arguing that $f(\mu, \sigma^2) \propto \sigma^{-2}$, the product of the priors $f(\mu)$ and $f(\sigma^2)$. Jeffreys' argument was that ignorance about μ and σ^2 should be represented by independent ignorance priors for μ and σ^2 separately. However, it is not clear under what circumstances this prior judgement of independence should be imposed.

4. **Consider, given θ , a sequence of independent Bernoulli trials with parameter θ . We wish to make inferences about θ and consider two possible methods. In the first, we carry out n trials and let X denote the total number of successes in these trials. Thus, $X|\theta \sim Bin(n, \theta)$ with**

$$f_X(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \dots, n.$$

In the second method, we count the total number Y of trials up to and including the r th success so that $Y|\theta \sim Nbin(r, \theta)$, the negative binomial distribution, with

$$f_Y(y|\theta) = \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}, \quad y = r, r+1, \dots$$

- (a) **Obtain the Jeffreys prior distribution for each of the two methods. You may find it useful to note that $E(Y | \theta) = \frac{r}{\theta}$.**

For $X | \theta \sim \text{Bin}(n, \theta)$ we have

$$\log f_X(x | \theta) = \log \binom{n}{x} + x \log \theta + (n - x) \log(1 - \theta)$$

so that

$$\frac{\partial^2}{\partial \theta^2} \log f_X(x | \theta) = -\frac{x}{\theta^2} - \frac{(n - x)}{(1 - \theta)^2}.$$

The Fisher information is

$$\begin{aligned} I(\theta) &= -E \left\{ -\frac{X}{\theta^2} - \frac{(n - X)}{(1 - \theta)^2} \middle| \theta \right\} \\ &= \frac{n}{\theta} + \frac{n}{1 - \theta} = \frac{n}{\theta(1 - \theta)}. \end{aligned}$$

The Jeffreys prior is thus

$$f(\theta) \propto \sqrt{\frac{n}{\theta(1 - \theta)}} \propto \theta^{-\frac{1}{2}}(1 - \theta)^{-\frac{1}{2}}$$

which is a kernel of the $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ distribution.

For $Y | \theta \sim \text{Nbin}(r, \theta)$ we have

$$\log f_Y(y | \theta) = \log \binom{y - 1}{r - 1} + r \log \theta + (y - r) \log(1 - \theta)$$

so that

$$\frac{\partial^2}{\partial \theta^2} \log f_Y(y | \theta) = -\frac{r}{\theta^2} - \frac{(y - r)}{(1 - \theta)^2}.$$

The Fisher information is

$$\begin{aligned} I(\theta) &= -E \left\{ -\frac{r}{\theta^2} - \frac{(Y - r)}{(1 - \theta)^2} \middle| \theta \right\} \\ &= \frac{r}{\theta^2} + \frac{r(\frac{1}{\theta} - 1)}{(1 - \theta)^2} \\ &= \frac{r}{\theta^2} + \frac{r}{\theta(1 - \theta)} = \frac{r}{\theta^2(1 - \theta)}. \end{aligned}$$

The Jeffreys prior is thus

$$f(\theta) \propto \sqrt{\frac{r}{\theta^2(1 - \theta)}} \propto \theta^{-1}(1 - \theta)^{-\frac{1}{2}}$$

which can be viewed as the improper $\text{Beta}(0, \frac{1}{2})$ distribution.

- (b) **Suppose we observe $x = r$ and $y = n$. For each method, calculate the posterior distribution for θ with the Jeffreys prior. Comment upon your answers.**

We summarise the results in a table.

Prior	Likelihood	Prop Likelihood	Posterior
$Beta(\frac{1}{2}, \frac{1}{2})$	$X \theta \sim Bin(n, \theta)$	$\theta^x (1 - \theta)^{n-x}$	$Beta(\frac{1}{2} + x, \frac{1}{2} + n - x)$
$Beta(0, \frac{1}{2})$	$Y \theta \sim Nbin(r, \theta)$	$\theta^r (1 - \theta)^{y-r}$	$Beta(r, \frac{1}{2} + y - r)$

Notice that if $x = r$ and $y = n$ then the two approaches have proportional likelihoods: in both cases we observed x successes in n trials but $\theta | x \sim Beta(\frac{1}{2} + x, \frac{1}{2} + n - x)$ and $\theta | y \sim Beta(x, \frac{1}{2} + n - x)$. Although the observed data is the same, Jeffreys' approach yields different posterior distributions which seems to contradict the notion of an noninformative prior. This occurs because Jeffreys' prior violates the **likelihood principle**. In short this principle states that the likelihood contains all the information about the data x so that two likelihoods contain the same information if they are proportional. In this case, the two likelihoods are proportional but yield different posterior distributions. Many classical methods, such as confidence intervals, violate the likelihood principle but Bayesian statistics (using proper prior distributions) does not.