MA40189 - Solution Sheet Four

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

2021/22 Semester II

1. **Let $X_1, \ldots, X_n$ be exchangeable so that the X_i are conditionally independent given a parameter θ.**

   (a) **Let $X_i \,|\, \theta \sim Bern(\theta)$.**

   i. **Show that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family and for $X = (X_1, \ldots, X_n)$ state the sufficient statistic for learning about θ.**

   Notice that we can write

$$\begin{aligned} f(x_i \,|\, \theta) &= \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \exp\left\{\left(\log\frac{\theta}{1-\theta}\right)x_i + \log(1-\theta)\right\} \end{aligned}$$

   so that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family with $\phi_1(\theta) = \log\frac{\theta}{1-\theta}$, $u_1(x_i) = x_i$, $g(\theta) = \log(1-\theta)$ and $h(x_i) = 0$. Notice that, from Proposition 1 (see Lecture 11), $t_n = [n, \sum_{i=1}^n X_i]$ is a sufficient statistic.

   ii. **By viewing the likelihood as a function of θ, which generic family of distributions (over θ) is the likelihood a kernel of?**

   The likelihood, without expressing in the explicit exponential family form, is

$$f(x \,|\, \theta) = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}$$

   which, viewing as a function of θ, we immediately recognise as a Beta kernel (in particular, a Beta$(n\bar{x}+1, n-n\bar{x}+1)$).

   iii. **By first finding the corresponding posterior distribution for θ given $x = (x_1, \ldots, x_n)$, show that this family of distributions is conjugate with respect to the likelihood $f(x \,|\, \theta)$.**

   Taking $\theta \sim Beta(\alpha, \beta)$ we have that

$$\begin{aligned} f(\theta \,|\, x) &\propto \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}} \times \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^{\alpha+n\bar{x}-1}(1-\theta)^{\beta+n-n\bar{x}-1} \end{aligned}$$

   so that $\theta \,|\, x \sim Beta(\alpha + n\bar{x}, \beta + n - n\bar{x})$. Thus, the prior and the posterior are in the same family giving conjugacy.

   Deriving the results directly from exponential family representation

Expressed in the 1-parameter exponential family form the likelihood is

$$f(x \,|\, \theta) \;=\; \exp\left\{\left(\log \frac{\theta}{1-\theta}\right)\sum_{i=1}^{n} x_i + n\log(1-\theta)\right\}$$

from which we immediately observe the sufficient statistic $t_n = [n, \sum_{i=1}^{n} x_i]$. Viewing $f(x \,|\, \theta)$ as a function of θ the natural conjugate prior is a member of the 2-parameter exponential family of the form

$$f(\theta) \;=\; \exp\left\{a\left(\log \frac{\theta}{1-\theta}\right) + d\log(1-\theta) + c(a,d)\right\}$$

where $c(a, d)$ is the normalising constant. Hence,

$$\begin{aligned} f(\theta) &\propto \exp\left\{a\left(\log \frac{\theta}{1-\theta}\right) + d\log(1-\theta)\right\} \\ &= \theta^{a}(1-\theta)^{d-a} \end{aligned}$$

which we recognise as a kernel of a Beta distribution. The convention is to label the hyperparameters as α and β so that we put $\alpha = \alpha(a, d) = a + 1$ and $\beta = \beta(a, d) = d - a + 1$ (equivalently, $a = a(\alpha, \beta) = \alpha - 1$, $d = d(\alpha, \beta) = \beta + \alpha - 2$). The conjugate prior distribution is $\theta \sim Beta(\alpha, \beta)$.

(b) **Let $X_i \,|\, \theta \sim N(\mu, \theta)$ with μ known.**

i. **Show that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family and for $X = (X_1, \ldots, X_n)$ state the sufficient statistic for learning about θ.**

Writing the normal density as an exponential family (parameter θ as μ is a known constant) we have

$$f(x_i \,|\, \theta) \;=\; \exp\left\{-\frac{1}{2\theta}(x_i-\mu)^2 - \frac{1}{2}\log\theta - \log\sqrt{2\pi}\right\}$$

so that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family. The sufficient statistic is $t_n = [n, \sum_{i=1}^{n}(x_i - \mu)^2]$. Note that, expressed explicitly as a 1-parameter exponential family, the likelihood for $x = (x_1, \ldots, x_n)$ is

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

so that the natural conjugate prior has the form

$$\begin{aligned} f(\theta) &= \exp\left\{-a\frac{1}{\theta} - d\log\theta + c(a,d)\right\} \\ &\propto \theta^{-d}\exp\left\{-a\frac{1}{\theta}\right\} \end{aligned}$$

which we recognise as a kernel of an Inverse-Gamma distribution.

ii. **By viewing the likelihood as a function of θ, which generic family of distributions (over θ) is the likelihood a kernel of?**

In conventional form,

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

which, viewing $f(x \,|\, \theta)$ as a function of θ, we recognise as a kernel of an Inverse-Gamma distribution (in particular, an Inv-gamma$(\frac{n-2}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2)$).

iii. **By first finding the corresponding posterior distribution for θ given $x = (x_1, \ldots, x_n)$, show that this family of distributions is conjugate with respect to the likelihood $f(x \,|\, \theta)$.**

Taking $\theta \sim$ Inv-gamma(α, β) we have

$$\begin{aligned} f(\theta \,|\, x) \quad &\propto \quad \theta^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \times \theta^{-(\alpha+1)} \exp\left\{-\frac{\beta}{\theta}\right\} \\ &= \quad \theta^{-(\alpha+\frac{n}{2}+1)} \exp\left\{-\left(\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{1}{\theta}\right\} \end{aligned}$$

which we recognise as a kernel of an Inverse-Gamma distribution so that $\theta \,|\, x \sim$ Inv-gamma$(\alpha+\frac{n}{2}, \beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2)$. Hence, the prior and posterior are in the same family giving conjugacy.

(c) **Let $X_i \,|\, \theta \sim Maxwell(\theta)$, 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 $E(X_i \,|\, \theta) = 2\sqrt{\frac{2}{\pi\theta}}$, $Var(X_i \,|\, \theta) = \frac{3\pi-8}{\pi\theta}$.

i. **Show that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family and for $X = (X_1, \ldots, X_n)$ state the sufficient statistic for learning about θ.**

Writing the Maxwell density in exponential family form we have

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

so that $f(x_i \,|\, \theta)$ belongs to the 1-parameter exponential family. The sufficient statistic is $t_n = [n, \sum_{i=1}^{n} x_i^2]$. Note that, expressed explicitly as a 1-parameter exponential family, the likelihood for $x = (x_1, \ldots, x_n)$ is

$$f(x \,|\, \theta) \quad = \quad \exp\left\{-\theta\sum_{i=1}^{n}\frac{x_i^2}{2} + \frac{3n}{2}\log\theta + \sum_{i=1}^{n}\log x_i^2 + \frac{n}{2}\log\frac{2}{\pi}\right\}$$

so that the natural conjugate prior has the form

$$\begin{aligned} f(\theta) &= \exp\{-a\theta + d\log\theta + c(a,d)\} \\ &\propto \theta^d e^{-a\theta} \end{aligned}$$

which we recognise as a kernel of a Gamma distribution.

ii. **By viewing the likelihood as a function of θ, which generic family of distributions (over θ) is the likelihood a kernel of?**

In conventional form,

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

which, viewing $f(x\,|\,\theta)$ as a function of θ, we recognise as a kernel of a Gamma distribution (in particular, Gamma$(\frac{3n+2}{2}, \frac{1}{2}\sum_{i=1}^{n} x_i^2)$).

iii. **By first finding the corresponding posterior distribution for θ given $x = (x_1, \ldots, x_n)$, show that this family of distributions is conjugate with respect to the likelihood $f(x\,|\,\theta)$.**

Taking $\theta \sim$ Gamma(α, β) we have

$$\begin{aligned} f(\theta\,|\,x) &\propto \theta^{\frac{3n}{2}} \exp\left\{-\left(\frac{\sum_{i=1}^{n} x_i^2}{2}\right)\theta\right\} \times \theta^{\alpha-1}e^{-\beta\theta} \\ &= \theta^{\alpha+\frac{3n}{2}-1} \exp\left\{-\left(\beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right)\theta\right\} \end{aligned}$$

which, of course, is a kernel of a Gamma distribution so that $\theta\,|\,x \sim$ Gamma$(\alpha+\frac{3n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2)$. The prior and the posterior are in the same family giving conjugacy.

2. **Let $X_1, \ldots, X_n$ be exchangeable so that the X_i are conditionally independent given a parameter θ. Suppose that $X_i\,|\,\theta$ is geometrically distributed with probability density function**

$$f(x_i\,|\,\theta) = (1-\theta)^{x_i-1}\theta, \quad x_i = 1, 2, \ldots.$$

(a) **Show that $f(x\,|\,\theta)$, where $x = (x_1, \ldots, x_n)$, belongs to the 1-parameter exponential family. Hence, or otherwise, find the conjugate prior distribution and corresponding posterior distribution for θ.**

As the X_i are exchangeable then

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

and so belongs to the 1-parameter exponential family. The conjugate prior is of the form

$$\begin{aligned} f(\theta) &\propto \exp\{a\log(1-\theta) + b\log\theta\} \\ &= \theta^b(1-\theta)^a \end{aligned}$$

which is a kernel of a Beta distribution. Letting $\alpha = b+1$, $\beta = a+1$ then we have $\theta \sim Beta(\alpha, \beta)$.

$$\begin{aligned} f(\theta \,|\, x) &\propto f(x \,|\, \theta) f(\theta) \\ &\propto \theta^n (1-\theta)^{(n\bar{x}-n)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \end{aligned}$$

which is a kernel of a $Beta(\alpha+n, \beta+n\bar{x}-n)$ so that $\theta \,|\, x \sim Beta(\alpha+n, \beta+n\bar{x}-n)$.

(b) **Show that the posterior mean for θ can be written as a weighted average of the prior mean of θ and the maximum likelihood estimate, $\bar{x}^{-1}$.**

$$\begin{aligned} E(\theta \,|\, X) &= \frac{\alpha+n}{(\alpha+n)+(\beta+n\bar{x}-n)} \\ &= \frac{\alpha+n}{\alpha+\beta+n\bar{x}} \\ &= \left(\frac{\alpha+\beta}{\alpha+\beta+n\bar{x}}\right)\left(\frac{\alpha}{\alpha+\beta}\right) + \left(\frac{n\bar{x}}{\alpha+\beta+n\bar{x}}\right)\left(\frac{1}{\bar{x}}\right) \\ &= \lambda E(\theta) + (1-\lambda)\bar{x}^{-1}. \end{aligned}$$

(c) **Suppose now that the prior for θ is instead given by the probability density function**

$$f(\theta) = \frac{1}{2B(\alpha+1,\beta)}\theta^{\alpha}(1-\theta)^{\beta-1} + \frac{1}{2B(\alpha,\beta+1)}\theta^{\alpha-1}(1-\theta)^{\beta},$$

where $B(\alpha, \beta)$ denotes the Beta function evaluated at α and β. Show that the posterior probability density function can be written as

$$f(\theta \,|\, x) = \lambda f_1(\theta) + (1-\lambda) f_2(\theta)$$

where

$$\lambda = \frac{(\alpha+n)\beta}{(\alpha+n)\beta + (\beta - n + \sum_{i=1}^{n} x_i)\alpha}$$

and $f_1(\theta)$ and $f_2(\theta)$ are probability density functions.

$$\begin{aligned} f(\theta \,|\, x) &\propto f(x \,|\, \theta) f(\theta) \\ &= \theta^n(1-\theta)^{(n\bar{x}-n)}\left\{\frac{\theta^{\alpha}(1-\theta)^{\beta-1}}{B(\alpha+1,\beta)} + \frac{\theta^{\alpha-1}(1-\theta)^{\beta}}{B(\alpha,\beta+1)}\right\} \\ &= \frac{\theta^{\alpha_1}(1-\theta)^{\beta_1-1}}{B(\alpha+1,\beta)} + \frac{\theta^{\alpha_1-1}(1-\theta)^{\beta_1}}{B(\alpha,\beta+1)} \end{aligned}$$

where $\alpha_1 = \alpha + n$ and $\beta_1 = \beta + n\bar{x} - n$. Finding the constant of proportionality we observe that $\theta^{\alpha_1}(1-\theta)^{\beta_1-1}$ is a kernel of a $Beta(\alpha_1+1, \beta_1)$ and $\theta^{\alpha_1-1}(1-\theta)^{\beta_1}$ is a kernel of a $Beta(\alpha_1, \beta_1+1)$. So,

$$f(\theta \,|\, x) = c\left\{\frac{B(\alpha_1+1,\beta_1)}{B(\alpha+1,\beta)} f_1(\theta) + \frac{B(\alpha_1,\beta_1+1)}{B(\alpha,\beta+1)} f_2(\theta)\right\}$$

where $f_1(\theta)$ is the density function of $Beta(\alpha_1+1, \beta_1)$ and $f_2(\theta)$ the density function of $Beta(\alpha_1, \beta_1+1)$. Hence,

$$c^{-1} = \frac{B(\alpha_1+1,\beta_1)}{B(\alpha+1,\beta)} + \frac{B(\alpha_1,\beta_1+1)}{B(\alpha,\beta+1)}$$

so that $f(\theta \,|\, x) = \lambda f_1(\theta) + (1-\lambda) f_2(\theta)$ with

$$\begin{aligned}
\lambda &= \frac{\frac{B(\alpha_1+1,\beta_1)}{B(\alpha+1,\beta)}}{\frac{B(\alpha_1+1,\beta_1)}{B(\alpha+1,\beta)} + \frac{B(\alpha_1,\beta_1+1)}{B(\alpha,\beta+1)}} \\
&= \frac{\frac{\alpha_1(\alpha+\beta)B(\alpha_1,\beta_1)}{\alpha(\alpha_1+\beta_1)B(\alpha,\beta)}}{\frac{\alpha_1(\alpha+\beta)B(\alpha_1,\beta_1)}{\alpha(\alpha_1+\beta_1)B(\alpha,\beta)} + \frac{\beta_1(\alpha+\beta)B(\alpha_1,\beta_1)}{\beta(\alpha_1+\beta_1)B(\alpha,\beta)}} \\
&= \frac{\alpha_1\beta}{\alpha_1\beta + \beta_1\alpha} \\
&= \frac{(\alpha+n)\beta}{(\alpha+n)\beta + (\beta + \sum_{i=1}^n x_i - n)\alpha}.
\end{aligned}$$

3. **Let $X_1, \ldots, X_n$ be exchangeable so that the X_i are conditionally independent given a parameter θ. Suppose that $X_i \,|\, \theta$ is distributed as a double-exponential distribution with probability density function**

$$f(x_i \,|\, \theta) = \frac{1}{2\theta} \exp\left\{-\frac{|x_i|}{\theta}\right\}, \quad -\infty < x_i < \infty$$

for $\theta > 0$.

(a) **Find the conjugate prior distribution and corresponding posterior distribution for θ following observation of $x = (x_1, \ldots, x_n)$.**

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

which, when viewed as a function of θ, is a kernel of $Inv\text{-}gamma(n-1, \sum_{i=1}^n |x_i|)$. We thus take $\theta \sim Inv\text{-}gamma(\alpha, \beta)$ as the prior so that

$$\begin{aligned}
f(\theta \,|\, x) &\propto \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^n |x_i|\right\} \frac{1}{\theta^{\alpha+1}} \exp\left\{-\frac{\beta}{\theta}\right\} \\
&= \frac{1}{\theta^{\alpha+n+1}} \exp\left\{-\frac{1}{\theta}\left(\beta + \sum_{i=1}^n |x_i|\right)\right\}
\end{aligned}$$

which is a kernel of $Inv\text{-}gamma(\alpha+n, \beta + \sum_{i=1}^n |x_i|)$. Thus, with respect to $X \,|\, \theta$, the prior and posterior are in the same family, showing conjugacy, with $\theta \,|\, x \sim Inv\text{-}gamma(\alpha+n, \beta + \sum_{i=1}^n |x_i|)$.

(b) **Consider the transformation $\phi = \theta^{-1}$. Find the posterior distribution of $\phi \,|\, x$.**

We have $\phi = g(\theta)$ where $g(\theta) = \theta^{-1}$ so that $\theta = g^{-1}(\phi) = \phi^{-1}$. Transforming $f_\theta(\theta \,|\, x)$ to $f_\phi(\phi \,|\, x)$ we have

$$\begin{aligned} f_\phi(\phi \,|\, x) &= \left|\frac{\partial \theta}{\partial \phi}\right| f_\theta(g^{-1}(\phi) \,|\, x) \\ &\propto \left|\frac{-1}{\phi^2}\right| \frac{1}{\frac{1}{\phi}^{\alpha+n+1}} \exp\left\{-\frac{1}{\frac{1}{\phi}}\left(\beta + \sum_{i=1}^n |x_i|\right)\right\} \\ &= \phi^{\alpha+n-1} \exp\left\{-\phi\left(\beta + \sum_{i=1}^n |x_i|\right)\right\} \end{aligned}$$

which is a kernel of a $Gamma(\alpha + n, \beta + \sum_{i=1}^n |x_i|)$ distribution. That is $\phi \,|\, x \sim Gamma(\alpha + n, \beta + \sum_{i=1}^n |x_i|)$. The result highlights the relationship between the Gamma and Inv-Gamma distributions shown on question 3(b)(i) of Question Sheet Two.

4. **Let $X_1, \ldots, X_n$ be a finite subset of a sequence of infinitely exchangeable random quantities with joint density function**

$$f(x_1, \ldots, x_n) = n! \left(1 + \sum_{i=1}^n x_i\right)^{-(n+1)}.$$

Show that they can be represented as conditionally independent and exponentially distributed.

Using de Finetti's Representation Theorem (Theorem 2 of the on-line notes), the joint distribution has an integral representation of the form

$$f(x_1, \ldots, x_n) = \int_\theta \left\{\prod_{i=1}^n f(x_i \,|\, \theta)\right\} f(\theta)\, d\theta.$$

If $X_i \,|\, \theta \sim \text{Exp}(\theta)$ then

$$\prod_{i=1}^n f(x_i \,|\, \theta) = \prod_{i=1}^n \theta \exp(-\theta x_i) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right).$$

Notice that, viewed as a function of θ, this looks like a kernel of Gamma$(n+1, \sum_{i=1}^n x_i)$. The result holds if we can find an $f(\theta)$ such that

$$n! \left(1 + \sum_{i=1}^n x_i\right)^{-(n+1)} = \int_\theta \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) f(\theta)\, d\theta.$$

The left hand side looks like the normalising constant of a Gamma$(n+1, 1+\sum_{i=1}^n x_i)$ (as $n! = \Gamma(n+1)$) and if $f(\theta) = \exp(-\theta)$ then the integrand on the right hand side is a kernel of a Gamma$(n+1, 1+\sum_{i=1}^n x_i)$. So, if $\theta \sim$ Gamma$(1, 1)$ then $f(\theta) = \exp(-\theta)$ and we have the desired representation.

5. **Let $X_1, \ldots, X_n$ be exchangeable so that the X_i are conditionally independent given a parameter θ. Suppose that $X_i \,|\, \theta$ is distributed as a Poisson distribution with mean θ.**

(a) **Show that, with respect to this Poisson likelihood, the gamma family of distributions is conjugate.**

$$\begin{aligned} f(x \mid \theta) &= \prod_{i=1}^{n} P(X_i = x_i \mid \theta) \\ &\propto \prod_{i=1}^{n} \theta^{x_i} \exp\{-\theta\} \\ &= \theta^{n\bar{x}} \exp\{-n\theta\}. \end{aligned}$$

As $\theta \sim Gamma(\alpha, \beta)$ then

$$\begin{aligned} f(\theta \mid x) &\propto f(x \mid \theta) f(\theta) \\ &\propto \theta^{n\bar{x}} \exp\{-n\theta\} \theta^{\alpha-1} \exp\{-\beta\theta\} \\ &= \theta^{\alpha+n\bar{x}-1} \exp\{-(\beta+n)\theta\} \end{aligned}$$

which is a kernel of a $Gamma(\alpha + n\bar{x}, \beta + n)$ distribution. Hence, the prior and posterior are in the same family giving conjugacy.

(b) **Interpret the posterior mean of θ paying particular attention to the cases when we may have weak prior information and strong prior information.**

$$\begin{aligned} E(\theta \mid X) &= \frac{\alpha + n\bar{x}}{\beta + n} \\ &= \frac{\beta\left(\frac{\alpha}{\beta}\right) + n\bar{x}}{\beta + n} \\ &= \lambda\left(\frac{\alpha}{\beta}\right) + (1-\lambda)\bar{x} \end{aligned}$$

where $\lambda = \frac{\beta}{\beta+n}$. Hence, the posterior mean is a weighted average of the prior mean, $\frac{\alpha}{\beta}$, and the data mean, $\bar{x}$, which is also the maximum likelihood estimate.

Weak prior information corresponds to a large variance of θ which can be viewed as small β (β is the inverse scale parameter). In this case, more weight is attached to $\bar{x}$ than $\frac{\alpha}{\beta}$ in the posterior mean.

Strong prior information corresponds to a small variance of θ which can be viewed as large β (once again, β is the inverse scale parameter). In this case, more weight is attached to $\frac{\alpha}{\beta}$ than $\bar{x}$ in the posterior mean which thus favours the prior mean.

(c) **Suppose now that the prior for θ is given hierarchically. Given λ, θ is judged to follow an exponential distribution with mean $\frac{1}{\lambda}$ and λ is given the improper distribution $f(\lambda) \propto 1$ for $\lambda > 0$. Show that**

$$f(\lambda \mid x) \propto \frac{\lambda}{(n+\lambda)^{n\bar{x}+1}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

$\theta \,|\, \lambda \sim Exp(\lambda)$ so $f(\theta \,|\, \lambda) = \lambda \exp\{-\lambda\theta\}$.

$$\begin{aligned} f(\lambda, \theta \,|\, x) &\propto f(x \,|\, \theta, \lambda) f(\theta, \lambda) \\ &= f(x \,|\, \theta) f(\theta \,|\, \lambda) f(\lambda) \\ &\propto \left(\theta^{n\bar{x}} \exp\{-n\theta\}\right) (\lambda \exp\{-\lambda\theta\}) \\ &= \lambda\theta^{n\bar{x}} \exp\{-(n+\lambda)\theta\}. \end{aligned}$$

Thus, integrating out θ,

$$\begin{aligned} f(\lambda \,|\, x) &\propto \int_0^\infty \lambda\theta^{n\bar{x}} \exp\{-(n+\lambda)\theta\}\, d\theta \\ &= \lambda \int_0^\infty \theta^{n\bar{x}} \exp\{-(n+\lambda)\theta\}\, d\theta \end{aligned}$$

As the integrand is a kernel of a $Gamma(n\bar{x}+1, n+\lambda)$ distribution we thus have

$$\begin{aligned} f(\lambda \,|\, x) &\propto \frac{\lambda\Gamma(n\bar{x}+1)}{(n+\lambda)^{n\bar{x}+1}} \\ &\propto \frac{\lambda}{(n+\lambda)^{n\bar{x}+1}}. \end{aligned}$$