

MA40189 - Solution Sheet Two

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 conditionally independent given λ so that $f(x | \lambda) = \prod_{i=1}^n f(x_i | \lambda)$ where $x = (x_1, \dots, x_n)$. Suppose that $\lambda \sim \text{Gamma}(\alpha, \beta)$ and $X_i | \lambda \sim \text{Exp}(\lambda)$ where λ represents the rate so that $E(X_i | \lambda) = \lambda^{-1}$.

- (a) Show that $\lambda | x \sim \text{Gamma}(\alpha + n, \beta + n\bar{x})$.

As $X_i | \lambda \sim \text{Exp}(\lambda)$ then, with $x = (x_1, \dots, x_n)$,

$$f(x | \lambda) = \lambda^n e^{-n\bar{x}\lambda}$$

which, when viewed as a function of λ , is a kernel of a Gamma distribution. We expect that if the prior is a general member of the Gamma family we will have conjugacy. We now confirm this using $\lambda \sim \text{Gamma}(\alpha, \beta)$. The posterior is

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

which is a kernel of a $\text{Gamma}(\alpha + n, \beta + n\bar{x})$ density, that is $\lambda | x \sim \text{Gamma}(\alpha + n, \beta + n\bar{x})$.

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

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

where $c = \frac{\beta}{\beta + n\bar{x}}$. Hence $E(\lambda | X)$ is a weighted average of the prior mean, $E(\lambda) = \frac{\alpha}{\beta}$, and the maximum likelihood estimate, \bar{x}^{-1} , of λ .

A water company is interested in the failure rate of water pipes. They ask two groups of engineers about their prior beliefs about the failure rate. The first group believe the mean failure rate is $\frac{1}{8}$ with coefficient of variation $\frac{1}{\sqrt{11}}$, whilst the second group believe the mean is $\frac{1}{11}$ with coefficient of variation $\frac{1}{2}$.

[Note: The coefficient of variation is the standard deviation divided by the mean.]

Let X_i be the time until water pipe i fails and assume that the X_i follow the exponential likelihood model described above. A sample of five of pipes is taken and the following times to failure were observed: 8.2, 9.2, 11.2, 9.8, 10.1.

- (c) Find the appropriate members of the conjugate families the prior statements of the two groups of engineers represent. In each case find the posterior mean and variance. Approximating the posterior by $N(E(\lambda|x), Var(\lambda|x))$, where $x = (x_1, \dots, x_5)$, estimate, in each case, the probability that the failure rate is less than 0.1.

For a $Gamma(\alpha, \beta)$ distribution, the mean is $\frac{\alpha}{\beta}$ whilst the coefficient of variation is $\sqrt{\frac{\alpha}{\beta^2}} \times \frac{\beta}{\alpha} = \frac{1}{\sqrt{\alpha}}$.

Suppose that the first group of engineers choose the prior $\lambda \sim Gamma(\alpha_1, \beta_1)$ then $\frac{\alpha_1}{\beta_1} = \frac{1}{8}$ and $\frac{1}{\sqrt{\alpha_1}} = \frac{1}{\sqrt{11}}$. Hence, $\alpha_1 = 11$ and $\beta_1 = 88$.

If the second group of engineers choose the prior $\lambda \sim Gamma(\alpha_2, \beta_2)$ then $\frac{\alpha_2}{\beta_2} = \frac{1}{11}$ and $\frac{1}{\sqrt{\alpha_2}} = \frac{1}{2}$. Hence, $\alpha_2 = 4$ and $\beta_2 = 44$.

We observe $n = 5$ and $\bar{x} = 9.7$ so that the first group of engineers have a posterior given by $Gamma(11 + 5, 88 + 5(9.7)) = Gamma(16, 136.5)$ whilst the second group of engineers have a posterior given by $Gamma(4 + 5, 44 + 5(9.7)) = Gamma(9, 92.5)$. We summarise the results in the following table.

Prior	Posterior	$E(\lambda X)$	$Var(\lambda X)$	$P(\lambda < 0.1 X)$
$Gamma(11, 88)$	$Gamma(16, 136.5)$	0.11722	0.00086	0.2776
$Gamma(4, 44)$	$Gamma(9, 92.5)$	0.09730	0.00105	0.5319

- (d) How do you expect any differences between the engineers to be reconciled as more data becomes available?

The maximum likelihood estimate of λ is $\frac{1}{9.7} = 0.10309$ which is closest to the second group of engineers who had a higher initial variance, corresponding to more uncertainty, for λ than the first group. Thus, the posterior for the second group is closer to the likelihood than for the first. We can also observe this by calculating the corresponding weights of prior mean and maximum likelihood estimate in the posterior means as per part (b). For the first group, we have $c = \frac{\beta_1}{\beta_1 + n\bar{x}} = \frac{88}{136.5} = 0.64469$ so that the posterior mean attaches a heavier weight to the prior mean whereas for the second group of engineers the weight is $c = \frac{\beta_2}{\beta_2 + n\bar{x}} = \frac{44}{92.5} = 0.47568$ so that a heavier weight is attached to the maximum likelihood estimate.

As more and more data are collected, the influence of the prior in the posterior will decrease and the data will take over. This can be seen clearly in the posterior mean calculation in part (b). Assuming \bar{x} stabilises then, as $n \rightarrow \infty$, $c \rightarrow 0$ and the posterior mean will tend towards the maximum likelihood estimate.

2. Let x be the number of successes in n independent Bernoulli trials, each one having unknown probability θ of success. It is judged that θ may be

modelled by a $Unif(0, 1)$ distribution so

$$f(\theta) = 1, \quad 0 < \theta < 1.$$

An extra trial, z is performed, independent of the first n given θ , but with probability $\frac{\theta}{2}$ of success. The full data is thus (x, z) where $z = 1$ if the extra trial is a success and 0 otherwise.

(a) Show that

$$f(\theta | x, z = 0) = c\{\theta^{\alpha-1}(1-\theta)^{\beta-1} + \theta^{\alpha-1}(1-\theta)^{\beta}\}$$

where $\alpha = x + 1$, $\beta = n - x + 1$ and $c = \frac{1}{B(\alpha, \beta) + B(\alpha, \beta + 1)}$.

As $X | \theta \sim Bin(n, \theta)$ then

$$f(x | \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad (1)$$

whilst as $Z | \theta \sim Bernoulli(\frac{\theta}{2})$ then

$$\begin{aligned} f(z = 0 | \theta) &= P(Z = 0 | \theta) = 1 - \frac{\theta}{2} \\ &= \frac{1}{2}(2 - \theta) = \frac{1}{2}\{1 + (1 - \theta)\}. \end{aligned} \quad (2)$$

As X and Z are conditionally independent given θ , from (1) and (2), we have

$$\begin{aligned} f(x, z = 0 | \theta) &= f(x | \theta)f(z = 0 | \theta) \\ &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \times \frac{1}{2}\{1 + (1 - \theta)\}. \end{aligned} \quad (3)$$

As $f(\theta) = 1$ then, using (3), the posterior is

$$\begin{aligned} f(\theta | x, z = 0) &\propto f(x, z = 0 | \theta)f(\theta) \\ &\propto \theta^x (1-\theta)^{n-x} + \theta^x (1-\theta)^{n-x+1}. \end{aligned}$$

Letting $\alpha = x + 1$, $\beta = n - x + 1$ we have

$$f(\theta | x, z = 0) = c\{\theta^{\alpha-1}(1-\theta)^{\beta-1} + \theta^{\alpha-1}(1-\theta)^{\beta}\} \quad (4)$$

where

$$\begin{aligned} c^{-1} &= \int_0^1 \{\theta^{\alpha-1}(1-\theta)^{\beta-1} + \theta^{\alpha-1}(1-\theta)^{\beta}\} d\theta \\ &= \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta + \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta} d\theta \\ &= B(\alpha, \beta) + B(\alpha, \beta + 1). \end{aligned} \quad (5)$$

(b) Hence show that

$$E(\theta | X, Z = 0) = \frac{(x+1)(2n-x+4)}{(n+3)(2n-x+3)}.$$

From (4) we have

$$\begin{aligned}
E(\theta | X, Z = 0) &= \int_0^1 \theta \times c \{ \theta^{\alpha-1} (1-\theta)^{\beta-1} + \theta^{\alpha-1} (1-\theta)^\beta \} d\theta \\
&= c \left\{ \int_0^1 \theta^\alpha (1-\theta)^{\beta-1} d\theta + \int_0^1 \theta^\alpha (1-\theta)^\beta d\theta \right\} \\
&= c \{ B(\alpha+1, \beta) + B(\alpha+1, \beta+1) \}. \tag{6}
\end{aligned}$$

Note that $B(\alpha, \beta+1) = \frac{\Gamma(\alpha)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)} = \frac{\beta}{\alpha+\beta} B(\alpha, \beta)$ so that, from (5),

$$\begin{aligned}
c &= \frac{1}{B(\alpha, \beta) + B(\alpha, \beta+1)} = \frac{1}{B(\alpha, \beta) + \frac{\beta}{\alpha+\beta} B(\alpha, \beta)} \\
&= \frac{\alpha + \beta}{B(\alpha, \beta)(\alpha + 2\beta)}. \tag{7}
\end{aligned}$$

Now, $B(\alpha+1, \beta) = \frac{\alpha}{\alpha+\beta} B(\alpha, \beta)$ and $B(\alpha+1, \beta+1) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)} B(\alpha, \beta)$ so that

$$\begin{aligned}
B(\alpha+1, \beta) + B(\alpha+1, \beta+1) &= \left\{ \frac{\alpha}{\alpha+\beta} + \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)} \right\} B(\alpha, \beta) \\
&= \frac{\alpha(\alpha+2\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)} B(\alpha, \beta). \tag{8}
\end{aligned}$$

Substituting (8) and (7) into (6) gives

$$E(\theta | X, Z = 0) = \frac{\alpha(\alpha+2\beta+1)}{(\alpha+\beta+1)(\alpha+2\beta)} = \frac{(x+1)(2n-x+4)}{(n+3)(2n-x+3)}$$

as $\alpha = x+1$ and $\beta = n-x+1$.

(c) **Show that, for all x , $E(\theta | X, Z = 0)$ is less than $E(\theta | X, Z = 1)$.**

In this case $f(z = 1 | \theta) = \frac{\theta}{2}$ so that

$$f(x, z = 1 | \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \times \frac{1}{2} \theta. \tag{9}$$

Now, as $f(\theta) \propto 1$, $f(\theta | x, z = 1) \propto f(x, z = 1 | \theta)$ which, from viewing (9) as a function of θ , we observe as a kernel of a $Beta(x+2, n-x+1)$ distribution, so that $\theta | x, z = 1 \sim Beta(x+2, n-x+1)$. Hence, $E(\theta | X, Z = 1) = \frac{x+2}{n+3}$. Now, from part (b),

$$E(\theta | X, Z = 0) = \frac{(x+1)(2n-x+4)}{(n+3)(2n-x+3)} = \frac{x+1}{n+3} \left(1 + \frac{1}{2n-x+3} \right)$$

Hence

$$\begin{aligned}
E(\theta | X, Z = 0) < E(\theta | X, Z = 1) &\Leftrightarrow \frac{x+1}{2n-x+3} < 1 \\
&\Leftrightarrow x < n+1
\end{aligned}$$

which is true as $x \in \{0, 1, \dots, n\}$.

3. Let X_1, \dots, X_n be conditionally independent given θ , so $f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$ where $x = (x_1, \dots, x_n)$, with each $X_i|\theta \sim N(\mu, \theta)$ where μ is known.

(a) Let $s(x) = \sum_{i=1}^n (x_i - \mu)^2$. Show that we can write

$$f(x|\theta) = g(s, \theta)h(x)$$

where $g(s, \theta)$ depends upon $s(x)$ and θ and $h(x)$ does not depend upon θ but may depend upon x . The equation shows that $s(X) = \sum_{i=1}^n (X_i - \mu)^2$ is sufficient for X_1, \dots, X_n for learning about θ .

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

where $g(s, \theta) = \theta^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta}s(x)\right\}$ and $h(x) = (2\pi)^{-\frac{n}{2}}$.

(b) An inverse-gamma distribution with known parameters $\alpha, \beta > 0$ is judged to be the prior distribution for θ . So,

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\beta/\theta}, \quad \theta > 0.$$

i. Show that the distribution of the precision $\tau = \frac{1}{\theta}$ is $Gamma(\alpha, \beta)$.

Let $Y \sim Inv-Gamma(\alpha, \beta)$ so that $f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} e^{-\beta/y}$, $y > 0$ and consider the distribution of $Z = \frac{1}{Y}$. The density function of Z , $f_Z(z)$, is

$$\begin{aligned} f_Z(z) &= \frac{\partial}{\partial z} P(Z \leq z) = \frac{\partial}{\partial z} P\left(\frac{1}{Y} \leq z\right) \\ &= \frac{\partial}{\partial z} P\left(Y \geq \frac{1}{z}\right) \\ &= -\frac{\partial}{\partial z} P\left(Y \leq \frac{1}{z}\right) \\ &= -\frac{\partial y}{\partial z} f_Y\left(\frac{1}{z}\right) \end{aligned}$$

where the last equation follows using the chain rule. Now, as $y = \frac{1}{z}$, we have

$$\begin{aligned} f_Z(z) &= -\left(-\frac{1}{z^2}\right) f_Y\left(\frac{1}{z}\right) = \frac{1}{z^2} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{z}\right)^{-(\alpha+1)} e^{-\beta/(1/z)} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} \end{aligned}$$

which is the density function of the $Gamma(\alpha, \beta)$ distribution. Hence, $Z \sim Gamma(\alpha, \beta)$. It is immediate that if $\theta \sim Inv-Gamma(\alpha, \beta)$ then $\tau \sim Gamma(\alpha, \beta)$.

ii. Find the posterior distribution of θ given $x = (x_1, \dots, x_n)$.

$$\begin{aligned} f(\theta|x) &\propto f(x|\theta)f(\theta) \\ &= (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta}s(x)\right\} \times \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta} \\ &\propto \theta^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta}s(x)\right\} \times \theta^{-(\alpha+1)}e^{-\beta/\theta} \\ &= \theta^{-(\alpha+\frac{n}{2}+1)} \exp\left\{-\frac{\beta+\frac{s(x)}{2}}{\theta}\right\} \end{aligned}$$

which is a kernel of an *Inv-Gamma* $(\alpha + \frac{n}{2}, \beta + \frac{s(x)}{2})$. Hence, we have the posterior $\theta|x \sim \text{Inv-Gamma}(\alpha + \frac{n}{2}, \beta + \frac{s(x)}{2})$.

iii. 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, $s(x)/n$, of θ .

As $\theta|x \sim \text{Inv-Gamma}(\alpha + \frac{n}{2}, \beta + \frac{s(x)}{2})$ we have that¹

$$\begin{aligned} E(\theta|X) &= \frac{\beta + \frac{s(x)}{2}}{\alpha + \frac{n}{2} - 1} \\ &= \frac{(\alpha - 1)\frac{\beta}{\alpha - 1} + \frac{n}{2}\frac{s(x)}{n}}{\alpha + \frac{n}{2} - 1} = c\left(\frac{\beta}{\alpha - 1}\right) + (1 - c)\left(\frac{s(x)}{n}\right) \end{aligned}$$

where $c = \frac{\alpha - 1}{\alpha + \frac{n}{2} - 1}$. Hence $E(\theta|X)$ is a weighted average of the prior mean, $E(\theta) = \frac{\beta}{\alpha - 1}$, and the maximum likelihood estimate, $\frac{s(x)}{n}$, of θ .

4. Suppose that X_1, \dots, X_n are identically distributed discrete random quantities taking k possible values with probabilities $\theta_1, \dots, \theta_k$. Inference is required about $\theta = (\theta_1, \dots, \theta_k)$ where $\sum_{j=1}^k \theta_j = 1$.

(a) Assuming that the X_i s are independent given θ , explain why

$$f(x|\theta) \propto \prod_{j=1}^k \theta_j^{n_j} \tag{10}$$

where $x = (x_1, \dots, x_n)$ and n_j is the number of x_i s observed to take the j th possible value.

This is just a generalisation of Bernoulli trials. Let $I_{\{x_i, j\}}$ denote the indicator function which is equal to one if x_i is in the j th class and zero otherwise. Then

$$P(X_i = x_i | \theta) = \prod_{j=1}^k \theta_j^{I_{\{x_i, j\}}}$$

¹Note that we need $\alpha + \frac{n}{2} > 1$ for $E(\theta|X)$ to be finite. For any $\alpha > 0$ then $n > 2$ will ensure this. For the decomposition as a weighted average we implicitly assume $\alpha > 1$ so that $E(\theta)$ is finite.

and, by the conditional independence,

$$f(x|\theta) = \prod_{i=1}^n P(X_i = x_i | \theta) = \prod_{i=1}^n \prod_{j=1}^k \theta_j^{I_{\{x_i, j\}}}$$

which gives (10) as we thus get a contribution of θ_j for each x_i which lands in the j th class and there are a total of n_j of these, $j = 1, \dots, k$. Notice that if we are only told the n_j s rather than which specific x_i contributed to each n_j then the likelihood is

$$f(x|\theta) = \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k \theta_j^{n_j}$$

which is the Multinomial distribution.

- (b) **Suppose that the prior for θ is Dirichlet distributed with known parameters $a = (a_1, \dots, a_k)$ so**

$$f(\theta) = \frac{1}{B(a)} \prod_{j=1}^k \theta_j^{a_j-1}$$

where $B(a) = B(a_1, \dots, a_k) = \frac{\prod_{j=1}^k \Gamma(a_j)}{\Gamma(\sum_{j=1}^k a_j)}$. **Show that the posterior for θ given x is Dirichlet with parameters $a + n = (a_1 + n_1, \dots, a_k + n_k)$.**

$$\begin{aligned} f(\theta|x) &\propto f(x|\theta)f(\theta) \\ &\propto \left\{ \prod_{j=1}^k \theta_j^{n_j} \right\} \times \left\{ \prod_{j=1}^k \theta_j^{a_j-1} \right\} \\ &= \prod_{j=1}^k \theta_j^{a_j+n_j-1} \end{aligned}$$

which is a kernel of the Dirichlet with parameter $a + n = (a_1 + n_1, \dots, a_k + n_k)$ so that the posterior $\theta|x$ is Dirichlet with parameter $a + n$.

The Multinomial distribution is the multivariate generalisation of the Binomial ($k = 2$ for the Multinomial gives the Binomial) and the Dirichlet the multivariate generalisation of the Beta ($k = 2$ for the Dirichlet gives the Beta). It is straightforward to obtain the moments of the Dirichlet distribution. For example,

$$E \left(\prod_{j=1}^k \theta_j^{m_j} \right) = \int_{\theta} \prod_{j=1}^k \theta_j^{m_j} \times \frac{1}{B(a)} \prod_{j=1}^k \theta_j^{a_j-1} d\theta \quad (11)$$

$$= \frac{1}{B(a)} \int_{\theta} \prod_{j=1}^k \theta_j^{a_j+m_j-1} d\theta \quad (12)$$

$$= \frac{B(a+m)}{B(a)} \quad (13)$$

where $a + m = (a_1 + m_1, \dots, a_k + m_k)$. Notice that the integral in (11) is k -dimensional (as θ is k -dimensional) and (13) follows as the integral in (12) is a

kernel of the Dirichlet distribution with parameter $a + m$. In particular, taking $m_j = 1$ and $m_{j'} = 0$ for $j' \neq j$ we have

$$\begin{aligned}
 E(\theta_j) &= \frac{B(a+m)}{B(a)} = \frac{\left\{ \prod_{j' \neq j}^k \Gamma(a_{j'}) \right\} \Gamma(a_j + 1)}{\Gamma((\sum_{j=1}^k a_j) + 1)} \times \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \\
 &= \frac{\Gamma(a_j + 1)}{\Gamma((\sum_{j=1}^k a_j) + 1)} \times \frac{\Gamma(\sum_{j=1}^k a_j)}{\Gamma(a_j)} \\
 &= \frac{a_j}{\sum_{j=1}^k a_j}.
 \end{aligned}$$

As $\theta | x$ is Dirichlet with parameter $a + n$ we have

$$\begin{aligned}
 E(\theta_j | x) &= \frac{a_j + n_j}{\sum_{j=1}^k a_j + \sum_{j=1}^k n_j} \\
 &= \frac{\tilde{a}}{\tilde{a} + \tilde{n}} \left(\frac{a_j}{\tilde{a}} \right) + \frac{\tilde{n}}{\tilde{a} + \tilde{n}} \left(\frac{n_j}{\tilde{n}} \right)
 \end{aligned}$$

where $\tilde{a} = \sum_{j=1}^k a_j$ and $\tilde{n} = \sum_{j=1}^k n_j$. The posterior mean for θ_j is a weighted average of its prior mean, $\frac{a_j}{\tilde{a}}$, and the classical maximum likelihood estimate of θ_j , $\frac{n_j}{\tilde{n}}$. Notice that \tilde{a} controls the weight of the prior mean in the posterior mean and is often said to represent the prior strength.