



Newcastle
University

MAS3902

Bayesian Inference

Solutions

Semester 1, 2018–19

© Prof R. J. Boys

School of Mathematics, Statistics & Physics

1. This question requires the ability to recognise distributional forms. A typical problem with this question is a confusion between the values of the observed data and the sample space (eg. $x = 1, 2, \dots$ in (i)). This question looks at a general case in which the data is $x_1, x_2, \dots, x_n$.

(i) The likelihood function is

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

The prior distribution is $\theta \sim \text{Beta}(3, 2)$, with density

$$\pi(\theta) = \frac{\theta^2(1-\theta)}{B(3, 2)}, \quad 0 < \theta < 1.$$

Therefore, using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^2(1-\theta) \times \theta^{n(\bar{x}-1)}(1-\theta)^n, \quad 0 < \theta < 1 \\ &\propto \theta^{n(\bar{x}-1)+2}(1-\theta)^{n+1}, \quad 0 < \theta < 1, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim \text{Beta}\{n(\bar{x} - 1) + 3, n + 2\}$.

(ii) The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = k e^{-n\theta} \theta^{n\bar{x}}$$

where k is constant w.r.t. θ . The prior distribution is $\theta \sim \text{Exp}(2)$, with density

$$\pi(\theta) = 2e^{-2\theta}, \quad \theta > 0.$$

Therefore, using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto e^{-2\theta} \times k e^{-n\theta} \theta^{n\bar{x}}, \quad \theta > 0 \\ &\propto \theta^{n\bar{x}} e^{-(n+2)\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim \text{Ga}(n\bar{x} + 1, n + 2)$.

(iii) The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \binom{4}{x_i} \theta^{x_i} (1-\theta)^{4-x_i} = k \theta^{n\bar{x}} (1-\theta)^{n(4-\bar{x})}$$

where k is constant w.r.t. θ . The prior distribution is $\theta \sim U(0, 1)$, with density

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

Therefore, using Bayes Theorem, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto 1 \times k\theta^{n\bar{x}}(1-\theta)^{n(4-\bar{x})}, \quad 0 < \theta < 1 \\ &\propto \theta^{n\bar{x}}(1-\theta)^{n(4-\bar{x})}, \quad 0 < \theta < 1,\end{aligned}$$

that is, $\theta|\mathbf{x} \sim \text{Beta}\{n\bar{x} + 1, n(4 - \bar{x}) + 1\}$.

2. The likelihood function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^{10} \frac{1}{\theta}, \quad x_1 < \theta, x_2 < \theta, \dots, x_{10} < \theta \\ &= \theta^{-10}, \quad x_{\max} < \theta.\end{aligned}$$

The prior distribution is $\theta \sim \text{Ga}(20, 1)$, with density

$$\begin{aligned}\pi(\theta) &= \frac{1^{20}\theta^{20-1}e^{-\theta}}{\Gamma(20)}, \quad \theta > 0 \\ &\propto \theta^{19}e^{-\theta}, \quad \theta > 0.\end{aligned}$$

Therefore, using Bayes Theorem, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{19}e^{-\theta} \times \theta^{-10}, \quad \theta > x_{\max} \\ &\propto \theta^9e^{-\theta}, \quad \theta > x_{\max}.\end{aligned}$$

Note that this posterior distribution is a gamma $\text{Ga}(10, 1)$ distribution, truncated below so that only values larger than $x_{\max}$ are possible.

3. The likelihood function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n e^{\theta-x_i}, \quad x_1 > \theta, x_2 > \theta, \dots, x_n > \theta \\ &= e^{n(\theta-\bar{x})}, \quad x_{\min} > \theta.\end{aligned}$$

The prior distribution is $\theta \sim N(b, 1/d)$ and so, using Bayes Theorem, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \exp\{-d(\theta-b)^2/2\} \times e^{n(\theta-\bar{x})}, \quad \theta < x_{\min} \\ &\propto \exp\{-[d(\theta-b)^2 - 2n\theta]/2\}, \quad \theta < x_{\min} \\ &\propto \exp\{-[d(\theta^2 - 2b\theta + b^2) - 2n\theta]/2\}, \quad \theta < x_{\min} \\ &\propto \exp\{-[d\theta^2 - 2(db + n)\theta]/2\}, \quad \theta < x_{\min} \\ &\propto \exp\left\{-\frac{d}{2}\left(\theta - \frac{db+n}{d}\right)^2\right\}, \quad \theta < x_{\min}.\end{aligned}$$

Note that this posterior distribution is a normal $N(b + n/d, 1/d)$ distribution, truncated above so that only values smaller than $x_{\min}$ are possible.

4. From Example 1.3 (with $\tau = 1$, $b = 10$, $d = 4$, $n = 12$, $\bar{x} = 31/3$), the posterior distribution is $\mu|\mathbf{x} \sim N(B, 1/D)$, where

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} = \frac{40 + 124}{4 + 12} = 10.25$$

and

$$D = d + n\tau = 4 + 12 = 16 = \frac{1}{0.25^2},$$

that is, $\mu|\mathbf{x} \sim N(10.25, 0.25^2)$. Therefore $Pr(\mu > 10|\mathbf{x}) = 0.8413$ using $1 - \text{pnorm}(10, 10.25, 0.25)$

5. The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^5 \frac{\theta^{x_i} e^{-\theta}}{x_i!} \propto \theta^{x_1 + \dots + x_5} e^{-5\theta} = \theta^{11} e^{-5\theta}.$$

Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^3 e^{-2\theta} \times \theta^{11} e^{-5\theta}, \quad \theta > 0 \\ &\propto \theta^{14} e^{-7\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim Ga(G = 15, H = 7)$.

6. The gamma prior distribution for θ has mean 0.2 and standard deviation 1, whence $\theta \sim Ga(0.04, 0.2)$. Also, we have $n = 20$ observations with mean $\bar{x} = 3.8$ from an $Exp(\theta)$ distribution.

The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{20} \theta e^{-\theta x_i} \propto \theta^{20} e^{-20\bar{x}\theta} = \theta^{20} e^{-76\theta}.$$

Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{-0.96} e^{-0.2\theta} \times \theta^{20} e^{-76\theta}, \quad \theta > 0 \\ &\propto \theta^{19.04} e^{-76.2\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim Ga(G = 20.04, H = 76.2)$. This posterior distribution has mean 0.263 and standard deviation 0.059 (both to 3 d.p.). Clearly, the data has changed our (posterior) beliefs about the mean service rate θ : we now believe that θ takes a larger value (the mean has increased) and we are much more certain about likely values of θ (the variance is greatly reduced).

7. From the data we have $n = 100$ and $\bar{x} = 1.123$ hours. If the data are a random sample from an exponential distribution then the mathematics used to derive the posterior distribution are more straightforward if we take a gamma prior distribution for θ . (Note we

could use any continuous distribution on the positive real line. Other possibilities include a log-normal distribution, a Weibull distribution and a Pareto distribution.) A gamma $Ga(g, h)$ prior distribution has mean g/h and variance g/h^2 and so to make $E(\theta) = 1$ and $Var(\theta) = 1$ we need to take $g = h = 1$.

The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{100} \theta e^{-\theta x_i} \propto \theta^{100} e^{-100\bar{x}\theta} = \theta^{100} e^{-112.3\theta}.$$

Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^0 e^{-\theta} \times \theta^{100} e^{-112.3\theta}, \quad \theta > 0 \\ &\propto \theta^{100} e^{-113.3\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim Ga(G = 101, H = 113.3)$.

This distribution has mean 0.8914 and standard deviation 0.0887. The posterior mean is a little lower than the prior mean. Also the posterior standard deviation is considerably smaller than the prior standard deviation and so the data have been very informative about θ . Thus the posterior distribution will not be sensitive to minor changes to the prior mean and variance, that is, the posterior parameters G and H will be fairly similar for different (but similar) values of g and h . Also, given the amount of information in the data, it is likely that any other choice of prior distribution (with profile similar to the gamma distribution) will result in a very similar posterior distribution. Incidentally, by the Central Limit Theorem, this gamma posterior distribution is well approximated with a normal distribution with the same mean and variance.

8. (i) If the prior distribution is $\theta \sim Beta(g, h)$ then, using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{g-1} (1-\theta)^{h-1} \times \prod_{i=1}^n \theta^k (1-\theta)^{x_i-k}, \quad 0 < \theta < 1 \\ &\propto \theta^{g+nk-1} (1-\theta)^{h+n\bar{x}-nk-1}, \quad 0 < \theta < 1 \\ &\text{i.e. } \theta|\mathbf{x} \sim Beta(G = g + nk, H = h + n\bar{x} - nk). \end{aligned}$$

Therefore, both the prior and posterior distributions are Beta distributions and so the Beta is the conjugate prior distribution.

(ii) As the beta distribution restricts values to the range $(0,1)$, there is a finite upper limit to the variance. If $\theta \sim Beta(g, h)$ then it has variance $Var(\theta) = gh/\{(g+h)^2(g+h+1)\}$. Note that this distribution requires its parameters $g > 0$ and $h > 0$. If we now transform the parameters to $m = g/(g+h)$ and $s = g+h$ then $Var(\theta) = m(1-m)/(s+1)$. Note that, as this is a 1-1 transformation, we have a free choice of values for m and s subject to the restrictions $0 < m < 1$ and $s > 0$. Clearly, this ratio is maximised when the numerator is maximised and the denominator is minimised. Now $m(1-m)$ is maximised

at $m = 1/2$, and $(s + 1)$ is minimised by taking s as small as possible. In terms of g and h , these conditions are $g = h$ and $g + h \rightarrow 0$, that is, $g = h \rightarrow 0$.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g = h \rightarrow 0$.

(iii) We make the Beta prior distribution vague by making it have as large a variance as possible. This is achieved, from (ii) above, by letting $g = h \rightarrow 0$. Hence, using the result in (i), the posterior distribution assuming vague prior information is

$$\theta|\mathbf{x} \sim \text{Beta}(nk, n\bar{x} - nk).$$

(iv) The asymptotic posterior distribution (as $n \rightarrow \infty$) is

$$\theta|\mathbf{x} \sim N(\hat{\theta}, J(\hat{\theta})^{-1}),$$

where

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) = \frac{nk}{\theta^2} + \frac{n(\bar{x} - k)}{(1 - \theta)^2}.$$

Now

$$\begin{aligned} \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 & \implies \frac{nk}{\hat{\theta}} - \frac{n(\bar{x} - k)}{1 - \hat{\theta}} = 0 \\ & \implies \hat{\theta} = \frac{k}{\bar{x}} \\ & \implies J(\hat{\theta}) = \frac{n\bar{x}^3}{k(\bar{x} - k)} \\ & \implies J(\hat{\theta})^{-1} = \frac{k(\bar{x} - k)}{n\bar{x}^3}. \end{aligned}$$

Therefore, for large n , the posterior distribution for θ is

$$\theta|\mathbf{x} \sim N\left(\frac{k}{\bar{x}}, \frac{k(\bar{x} - k)}{n\bar{x}^3}\right) \quad \text{approximately.}$$

9. (i) The likelihood function is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n 2x_i \theta e^{-\theta x_i^2} \propto \theta^n \exp \left\{ -\theta \sum_{i=1}^n x_i^2 \right\} = \theta^n \exp \{ -n\bar{x}^2 \theta \},$$

where $\bar{x}^2 = \sum_{i=1}^n x_i^2 / n$ is the mean of the squared observations. Therefore the conjugate prior distribution is a Gamma $Ga(g, h)$ distribution. Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) & \propto \pi(\theta) f(\mathbf{x}|\theta) \\ & \propto \frac{h^g \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}^2 \theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim Ga(G = g + n, H = h + n\bar{x}^2)$.

We saw in Example 1.6 that to make a gamma distribution vague, we take $g \rightarrow 0$ and $h \rightarrow 0$. Therefore using a vague prior distribution gives the posterior distribution as $\theta|\mathbf{x} \sim Ga(n, n\bar{x}^2)$.

(ii) The asymptotic posterior distribution (as $n \rightarrow \infty$) is

$$\theta|\mathbf{x} \sim N(\hat{\theta}, J(\hat{\theta})^{-1}),$$

where

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) = \frac{n}{\theta^2}.$$

Now

$$\begin{aligned} \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 &\implies \frac{n}{\hat{\theta}} - n\bar{x}^2 = 0 \\ &\implies \hat{\theta} = 1/\bar{x}^2 \\ &\implies J(\hat{\theta})^{-1} = \frac{1}{n} (\bar{x}^2)^2. \end{aligned}$$

Therefore, for large n , the posterior distribution for θ is

$$\theta|\mathbf{x} \sim N\left(1/\bar{x}^2, \frac{1}{n} (\bar{x}^2)^2\right) \quad \text{approximately.}$$

Note that the posterior distribution is centred on the likelihood mode and its variance tends to zero as $n \rightarrow \infty$.

10. (i) The likelihood function is

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

where $\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}$ is the geometric mean of the observations. Therefore the conjugate prior distribution is a Gamma $Ga(g, h)$ distribution. Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \frac{h^g \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n \log \bar{x}_g \theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\mathbf{x} \sim Ga(G = g + n, H = h + n \log \bar{x}_g)$.

We saw in Example 1.6 that to make a gamma distribution vague, we take $g \rightarrow 0$ and $h \rightarrow 0$. Therefore using a vague prior distribution gives the posterior distribution as $\theta|\mathbf{x} \sim Ga(n, n \log \bar{x}_g)$.

(ii) The asymptotic posterior distribution (as $n \rightarrow \infty$) is

$$\theta|\mathbf{x} \sim N(\hat{\theta}, J(\hat{\theta})^{-1}),$$

where

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) = \frac{n}{\theta^2}.$$

Now

$$\begin{aligned}
\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 & \implies \frac{n}{\hat{\theta}} - n \log \bar{x}_g = 0 \\
& \implies \hat{\theta} = 1 / \log \bar{x}_g \\
& \implies J(\hat{\theta})^{-1} = \frac{1}{n(\log \bar{x}_g)^2}.
\end{aligned}$$

Therefore, for large n , the posterior distribution for θ is

$$\theta|\mathbf{x} \sim N\left(\frac{1}{\log \bar{x}_g}, \frac{1}{n(\log \bar{x}_g)^2}\right) \quad \text{approximately.}$$

Note that the posterior distribution is centred on the likelihood mode and its variance tends to zero as $n \rightarrow \infty$.

11. (i) We will consider the general case in which the prior distribution is a mixture distribution with component distributions

$$\pi_1(\mu) = N(b_1, 1/d_1) \quad \text{and} \quad \pi_2(\mu) = N(b_2, 1/d_2)$$

and (prior) weights p_1 and p_2 .

We have already seen that combining a random sample of size n from a normal $N(\mu, 1/\tau)$ distribution (with h known) with a normal $N(b, 1/d)$ prior distribution results in a normal $N(B, 1/D)$ posterior distribution where

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} \quad \text{and} \quad D = d + n\tau.$$

Therefore, the (overall) posterior distribution will be a mixture distribution with component distributions

$$\pi_1(\mu|\mathbf{x}) = N(B_1, 1/D_1) \quad \text{and} \quad \pi_2(\mu|\mathbf{x}) = N(B_2, 1/D_2)$$

where

$$B_i = \frac{d_i b_i + n\tau\bar{x}}{d_i + n\tau}, \quad D_i = d_i + n\tau.$$

We now calculate the (posterior) weights p_1^* and $p_2^* = 1 - p_1^*$, which will depend on both prior information and the data. We have

$$p_1^* = \frac{p_1 f_1(\mathbf{x})}{p_1 f_1(\mathbf{x}) + p_2 f_2(\mathbf{x})} \quad \text{and} \quad p_2^* = 1 - p_1^*$$

where

$$f_i(\mathbf{x}) = \frac{\pi_i(\mu) f(\mathbf{x}|\mu)}{\pi_i(\mu|\mathbf{x})}.$$

The likelihood function is

$$f(\mathbf{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\}.$$

Therefore, for $i = 1, 2$

$$\begin{aligned}
f_i(\mathbf{x}) &= \frac{\pi_i(\mu) f(\mathbf{x}|\mu)}{\pi_i(\mu|\mathbf{x})} \\
&= \frac{\left(\frac{d_i}{2\pi}\right)^{1/2} \exp\left\{-\frac{d_i}{2}(\mu - b_i)^2\right\} \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}[s^2 + (\bar{x} - \mu)^2]\right\}}{\left(\frac{D_i}{2\pi}\right)^{1/2} \exp\left\{-\frac{D_i}{2}(\mu - B_i)^2\right\}} \\
&= \frac{c \sqrt{d_i}}{\sqrt{D_i}} \exp\left\{-\frac{1}{2}[d_i(\mu - b_i)^2 + n\tau(\bar{x} - \mu)^2 - D_i(\mu - B_i)^2]\right\} \\
&= \dots \\
&= \frac{c \sqrt{d_i}}{\sqrt{D_i}} \exp\left\{\frac{1}{2}[D_i B_i^2 - d_i b_i^2]\right\}.
\end{aligned}$$

Hence

$$\begin{aligned}
(p_1^*)^{-1} - 1 &= \frac{p_2 f_2(\mathbf{x})}{p_1 f_1(\mathbf{x})} \\
&= \frac{p_2 \frac{c \sqrt{d_2}}{\sqrt{D_2}} \exp\left\{\frac{1}{2}[D_2 B_2^2 - d_2 b_2^2]\right\}}{p_1 \frac{c \sqrt{d_1}}{\sqrt{D_1}} \exp\left\{\frac{1}{2}[D_1 B_1^2 - d_1 b_1^2]\right\}} \\
&= \frac{p_2 \sqrt{D_1 d_2}}{p_1 \sqrt{D_2 d_1}} \exp\left\{\frac{1}{2}[D_2 B_2^2 - d_2 b_2^2 - D_1 B_1^2 + d_1 b_1^2]\right\}.
\end{aligned}$$

In the numerical case mentioned, with $n = 10$, $\bar{x} = 2.5$, $\tau = 1$, (prior) component parameters $b_1 = 3.3$, $d_1 = 1/0.37^2$, $b_2 = 1.1$, $d_2 = 1/0.47^2$ and (prior) weights $p_1 = 0.2$, $p_2 = 0.8$, we have

$$\begin{aligned}
B_1 &= \frac{d_1 b_1 + n\tau \bar{x}}{d_1 + n\tau} = 2.8377, & D_1 &= d_1 + n\tau = \frac{1}{0.2404^2}, \\
B_2 &= \frac{d_2 b_2 + n\tau \bar{x}}{d_2 + n\tau} = 2.0637, & D_2 &= d_2 + n\tau = \frac{1}{0.2624^2},
\end{aligned}$$

and $p_1^* = 0.6150$, $p_2^* = 0.3850$. Therefore, the posterior distribution is

$$\pi(\mu|\mathbf{x}) = 0.6150 N(2.8377, 0.2404^2) + 0.3850 N(2.0637, 0.2624^2).$$

(ii) The prior mean is

$$E(\mu) = p_1 E_1(\mu) + p_2 E_2(\mu) = 0.2 \times 3.3 + 0.8 \times 1.1 = 1.54.$$

Also

$$E(\mu^2) = p_1 E_1(\mu^2) + p_2 E_2(\mu^2) = 0.2 \times (0.37^2 + 3.3^2) + 0.8 \times (0.47^2 + 1.1^2) = 3.3501$$

whence the variance of the prior distribution is

$$Var(\mu) = E(\mu^2) - E(\mu)^2 = 3.3501 - 1.54^2 = 0.9785 = 0.9892^2.$$

The posterior mean is

$$E(\mu|\mathbf{x}) = p_1^* E_1(\mu|\mathbf{x}) + p_2^* E_2(\mu|\mathbf{x}) = 0.6150 \times 2.8377 + 0.3850 \times 2.0637 = 2.5397.$$

Also

$$\begin{aligned} E(\mu^2|\mathbf{x}) &= p_1^* E_1(\mu^2|\mathbf{x}) + p_2^* E_2(\mu^2|\mathbf{x}) \\ &= 0.6150 \times (0.2404^2 + 2.8377^2) + 0.3850 \times (0.2624^2 + 2.0637^2) \\ &= 6.6540 \end{aligned}$$

whence the variance of the posterior distribution is

$$Var(\mu|\mathbf{x}) = E(\mu^2|\mathbf{x}) - E(\mu|\mathbf{x})^2 = 6.6540 - 2.5397^2 = 0.2039 = 0.4516^2.$$

(iii) The prior and posterior distributions are bi-modal (see Figure 1). However, observing $\bar{x} = 2.5$ has transferred more weight to the component with the larger mean.

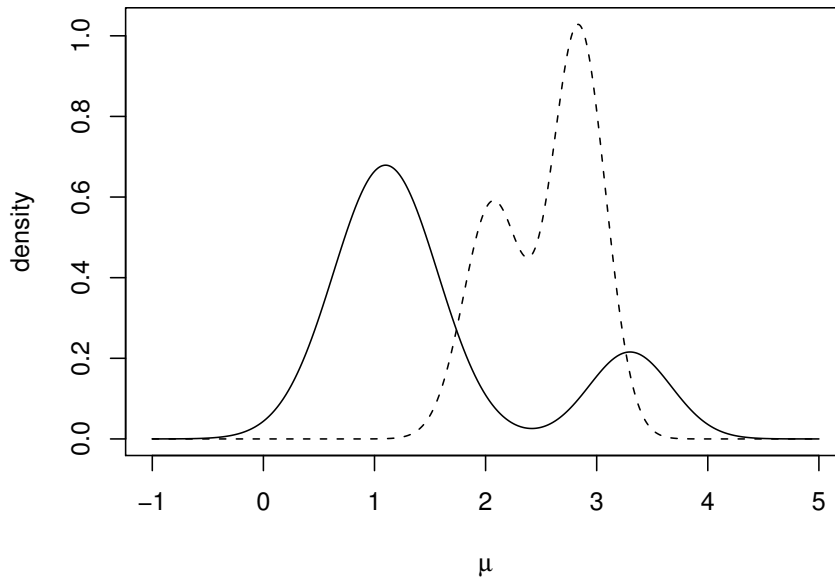


Figure 1: Prior (dashed) and posterior (solid) densities for μ

(iv) The prior probability that μ exceeds 2.5 is

$$\begin{aligned} Pr(\mu > 2.5) &= \int_{2.5}^{\infty} \pi(\mu) d\mu \\ &= \int_{2.5}^{\infty} \{p_1 \pi_1(\mu) + p_2 \pi_2(\mu)\} d\mu \\ &= p_1 \int_{2.5}^{\infty} \pi_1(\mu) d\mu + p_2 \int_{2.5}^{\infty} \pi_2(\mu) d\mu \\ &= p_1 Pr_1(\mu > 2.5) + p_2 Pr_2(\mu > 2.5) \\ &= 0.1981 \end{aligned}$$

using the command `0.2*(1-pnorm(2.5,3.3,0.37))+0.8*(1-pnorm(2.5,1.1,0.47))`

Similarly, the posterior probability that μ exceeds 2.5 is

$$\begin{aligned} Pr(\mu > 2.5|\mathbf{x}) &= \int_{2.5}^{\infty} \pi(\mu|\mathbf{x}) d\mu \\ &= p_1^* Pr_1(\mu > 2.5|\mathbf{x}) + p_2^* Pr_2(\mu > 2.5|\mathbf{x}) \\ &= 0.5843 \end{aligned}$$

using the command

`0.6150*(1-pnorm(2.5,2.8377,0.2404))+0.3850*(1-pnorm(2.5,2.0637,0.2624))`

Note that incorporating the data has substantially increased the probability of μ exceeding 2.5.

13. From the results in section 2.2, we have

$$E(\mu|\mathbf{x}) - E(\mu) > 0 \iff \frac{bc + n\bar{x}}{c + n} - b > 0 \iff bc + n\bar{x} - b(c + n) > 0 \iff \bar{x} - b > 0.$$

Therefore

$$E(\mu|\mathbf{x}) > E(\mu) \iff \bar{x} > E(\mu).$$

Notice that we can write the posterior mean as a convex combination of the prior and sample means:

$$B = \alpha b + (1 - \alpha)\bar{x} \quad \text{where} \quad \alpha = \frac{c}{c + n} \in (0, 1).$$

14. The relevant results from section 2.2 are

$$\tau|\mathbf{x} \sim Ga(G, H) \quad \text{and} \quad \mu|\mathbf{x} \sim t_{2G} \left(B, \frac{H}{GC} \right)$$

where

$$B = \frac{bc + n\bar{x}}{c + n}, \quad C = c + n, \quad G = g + \frac{n}{2}, \quad H = h + \frac{cn(\bar{x} - b)^2}{2(c + n)} + \frac{ns^2}{2}.$$

Rewriting these non-standard posterior distributions in terms of more well-known distributions (and using the hint) we have, given $\mathbf{x}$

$$\tau^* = 2H\tau \sim \chi_{2G}^2 \quad \text{and} \quad \mu^* = \frac{\mu - B}{\sqrt{H/(GC)}} \sim t_{2G}.$$

(i) A 95% Bayesian confidence interval for μ is any interval (a, b) for which

$$Pr(a < \mu < b|\mathbf{x}) = 0.95,$$

or equivalently

$$Pr \left(\frac{a - B}{\sqrt{H/(GC)}} < \frac{\mu - B}{\sqrt{H/(GC)}} < \frac{b - B}{\sqrt{H/(GC)}} \middle| \mathbf{x} \right) = 0.95,$$

that is,

$$Pr\left(\frac{a-B}{\sqrt{H/(GC)}} < \mu^* < \frac{b-B}{\sqrt{H/(GC)}} \middle| \mathbf{x}\right) = 0.95$$

As $\mu^* \sim t_{2G}$ we obtain the equi-tailed 95% Bayesian confidence interval by taking the end-points in the probability above as the lower and upper 2.5% values of the t_{2G} distribution. Let $t_{\nu,\alpha}$ be the upper α -percentile of the t_{ν} -distribution. Then the equi-tailed 95% Bayesian confidence interval for μ is (a, b) where

$$\begin{aligned} \frac{a-B}{\sqrt{H/(GC)}} &= t_{2G,0.975} = -t_{2G,0.025}, & \frac{b-B}{\sqrt{H/(GC)}} &= t_{2G,0.025} \\ \implies a &= B - t_{2G,0.025} \sqrt{\frac{H}{GC}}, & b &= B + t_{2G,0.025} \sqrt{\frac{H}{GC}}. \end{aligned}$$

(ii) A 95% Bayesian confidence interval for τ is any interval (a, b) for which

$$Pr(a < \tau < b | \mathbf{x}) = 0.95,$$

or equivalently

$$Pr(2Ha < 2H\tau < 2Hb | \mathbf{x}) = 0.95,$$

that is,

$$Pr(2Ha < \tau^* < 2Hb | \mathbf{x}) = 0.95.$$

As $\tau^* \sim \chi_{2G}^2$ we obtain the equi-tailed 95% Bayesian confidence interval by taking the end-points in the probability above as the lower and upper 2.5% values of the χ_{2G}^2 distribution. Let $\chi_{\nu,\alpha}^2$ is the upper α -percentile of the χ_{ν}^2 -distribution. Then the equi-tailed 95% Bayesian confidence interval for τ is (a, b) where

$$\begin{aligned} 2Ha &= \chi_{2G,0.975}^2, & 2Hb &= \chi_{2G,0.025}^2 \\ \implies a &= \frac{1}{2H} \chi_{2G,0.975}^2, & b &= \frac{1}{2H} \chi_{2G,0.025}^2. \end{aligned}$$

(iii) We can get a 95% Bayesian confidence interval for $\sigma = 1/\sqrt{\tau}$ from a 95% Bayesian confidence interval for τ as follows. Suppose (a, b) is a 95% Bayesian confidence interval for τ . Then

$$\begin{aligned} Pr(a < \tau < b | \mathbf{x}) &= 0.95 & \implies & Pr(\sqrt{a} < \sqrt{\tau} < \sqrt{b} | \mathbf{x}) = 0.95 \\ & & \implies & Pr\left(\frac{1}{\sqrt{b}} < \frac{1}{\sqrt{\tau}} < \frac{1}{\sqrt{a}} \middle| \mathbf{x}\right) = 0.95 \\ & & \implies & Pr\left(\frac{1}{\sqrt{b}} < \sigma < \frac{1}{\sqrt{a}} \middle| \mathbf{x}\right) = 0.95. \end{aligned}$$

Therefore, a 95% Bayesian confidence interval for σ is $(1/\sqrt{b}, 1/\sqrt{a})$. Using the equi-tailed 95% Bayesian confidence interval for τ in (ii) gives a 95% Bayesian confidence interval for σ as

$$\left(\frac{\sqrt{2H}}{\sqrt{\chi_{2G,0.025}^2}}, \frac{\sqrt{2H}}{\sqrt{\chi_{2G,0.975}^2}} \right).$$

(iv) The interval in (i) above is an 95% HDI for μ because all the values in the interval are more likely than those outside the interval – we know this because the t_{2G} -distribution is symmetric about zero.

(v) It is not straightforward to determine the 95% HDI for τ as its posterior distribution is not a symmetric distribution. Thus the interval in (ii) is not an HDI. However, for large n (in which case G is large), we can use the Central Limit Theorem to approximate the posterior distribution for τ by a normal distribution with the equivalent mean and variance. Hence

$$\tau \sim Ga(G, H) \implies \tau \sim N\left(\frac{G}{H}, \frac{G}{H^2}\right) \text{ approximately.}$$

Therefore, for large n , an approximate 95% HDI for τ is

$$\left(\frac{G}{H} - \frac{1.96\sqrt{G}}{H}, \frac{G}{H} + \frac{1.96\sqrt{G}}{H} \right).$$

We can also say that, as $G \simeq n/2$ and $H \simeq ns^2/2$ for large n , an approximate 95% HDI for τ is

$$\left(\frac{1}{s^2} - 1.96\sqrt{\frac{2}{ns^4}}, \frac{1}{s^2} + 1.96\sqrt{\frac{2}{ns^4}} \right).$$

This interval can be found via an alternate route, namely via the asymptotic posterior distribution: the solution to Example 3.1 shows that for large n , $\tau|\mathbf{x} \sim N(1/s^2, 2/(ns^4))$, approximately, and the HDI above follows directly from this distribution.

Frequentist confidence intervals are based on the pivotal quantities

$$\frac{\bar{X} - \mu}{S_u/\sqrt{n}} \sim t_{n-1} \quad \text{and} \quad \frac{(n-1)S_u^2}{\sigma^2} \sim \chi_{n-1}^2,$$

where $\bar{X}$ and S_u^2 are random variables and μ and $\sigma^2 = 1/\tau$ are fixed quantities. The corresponding frequentist versions of the 95% confidence intervals above are

$$\mu: (\bar{x} - t_{n-1,0.025}S_u/\sqrt{n}, \bar{x} + t_{n-1,0.025}S_u/\sqrt{n})$$

$$\tau: (\chi_{n-1,0.975}^2/\{(n-1)s_u^2\}, \chi_{n-1,0.025}^2/\{(n-1)s_u^2\})$$

$$\sigma: (\sqrt{\{(n-1)s_u^2\}/\chi_{n-1,0.025}^2}, \sqrt{\{(n-1)s_u^2\}/\chi_{n-1,0.975}^2}).$$

These intervals would take the same numerical values as their (exact) Bayesian counterparts if $c = h = 0$ and $g = -1/2$. This is almost the case of vague prior knowledge

except that this requires $g = 0$: note that $g = -1/2$ is not an admissible value as $g > 0$. Even when the Bayesian and frequentist intervals take similar values, the interpretation of the intervals is very different.

15. From Example 3.1, the asymptotic posterior distribution for (μ, τ) in this scenario is

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} \Big| \mathbf{x} \sim N \left\{ \begin{pmatrix} \bar{x} \\ \frac{1}{s^2} \end{pmatrix}, \begin{pmatrix} \frac{s^2}{n} & 0 \\ 0 & \frac{2}{ns^4} \end{pmatrix} \right\},$$

or equivalently

$$\mu | \mathbf{x} \sim N(\bar{x}, s^2/n), \quad \tau | \mathbf{x} \sim N\{1/s^2, 2/(ns^4)\}, \quad \text{independently.}$$

(i) The 95% HDI for μ is

$$\left(\bar{x} - \frac{1.96 s}{\sqrt{n}}, \bar{x} + \frac{1.96 s}{\sqrt{n}} \right).$$

(ii) The 95% HDI for τ is

$$\left(\frac{1}{s^2} - \frac{1.96\sqrt{2}}{\sqrt{n}s^2}, \frac{1}{s^2} + \frac{1.96\sqrt{2}}{\sqrt{n}s^2} \right).$$

(iii) Asymptotic theory tells us that the (asymptotic) posterior distribution for σ is a normal distribution. To determine the mean and variance of this distribution we have to rework the results in Example 3.1 but in terms of σ rather than τ . This involves taking derivatives of the log-likelihood with respect to σ , showing that the m.l.e.s are $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = s$ and then determining the second derivatives of the log-likelihood to show that the observed information matrix is

$$J(\hat{\mu}, \hat{\sigma}) = \begin{pmatrix} n/s^2 & 0 \\ 0 & 2n/s^2 \end{pmatrix}.$$

Therefore the asymptotic posterior distribution for (μ, σ) is

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \Big| \mathbf{x} \sim N \left\{ \begin{pmatrix} \bar{x} \\ s \end{pmatrix}, \begin{pmatrix} \frac{s^2}{n} & 0 \\ 0 & \frac{s^2}{2n} \end{pmatrix} \right\},$$

or equivalently

$$\mu | \mathbf{x} \sim N(\bar{x}, s^2/n), \quad \sigma | \mathbf{x} \sim N\{s, s^2/(2n)\}, \quad \text{independently.}$$

Notice that we get the same marginal distribution for $\mu | \mathbf{x}$ for both parameterisations. Thus, the 95% HDI for σ is

$$\left(s - \frac{1.96 s}{\sqrt{2n}}, s + \frac{1.96 s}{\sqrt{2n}} \right).$$

In large samples, frequentist confidence intervals are determined by using the sampling distribution of the maximum likelihood estimators: $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1})$. Using some results in Example 3.1, we have

$$\begin{pmatrix} \hat{\mu} \\ \hat{\tau} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu \\ \tau \end{pmatrix}, \begin{pmatrix} \frac{1}{n\tau} & 0 \\ 0 & \frac{2\tau^2}{n} \end{pmatrix} \right\},$$

or equivalently

$$\hat{\mu} \sim N\{\mu, 1/(n\tau)\}, \quad \hat{\tau} \sim N(\tau, 2\tau^2/n), \quad \text{independently.}$$

Typically, the variances of these distributions are approximated by replacing the parameters with their ML estimates, giving (approximate) 95% equi-tailed confidence intervals

$$\begin{aligned} \mu : \quad \hat{\mu} \pm \frac{1.96}{\sqrt{n\hat{\tau}}} &= \bar{x} \pm \frac{1.96 s}{\sqrt{n}} \\ \tau : \quad \hat{\tau} \pm \frac{1.96\sqrt{2}\hat{\tau}}{\sqrt{n}} &= \frac{1}{s^2} \pm \frac{1.96\sqrt{2}}{\sqrt{n}s^2}. \end{aligned}$$

We can show that Fisher's information matrix is

$$I(\mu, \sigma) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2 \end{pmatrix}$$

and so the asymptotic distribution of the ML estimators for (μ, σ) is

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{pmatrix} \right\},$$

or equivalently

$$\hat{\mu} \sim N(\mu, \sigma^2/n), \quad \hat{\sigma} \sim N\{\sigma, \sigma^2/(2n)\}, \quad \text{independently.}$$

Again, approximating the variance of the distribution for $\hat{\sigma}$ by replacing the parameters with their ML estimates, gives the (approximate) 95% equitailed confidence interval

$$\sigma : \quad \hat{\sigma} \pm \frac{1.96 \hat{\sigma}}{\sqrt{2n}} = s \pm \frac{1.96 s}{\sqrt{2n}}.$$

Notice that these frequentist intervals take the same numerical values as their Bayesian counterparts. However, the interpretation of the Bayesian and the frequentist intervals are very different.

16. (i) The likelihood function is

$$f(\mathbf{x}|\alpha, \lambda) = \prod_{i=1}^n \frac{\lambda^\alpha x_i^{\alpha-1} e^{-\lambda x_i}}{\Gamma(\alpha)} = \frac{\lambda^{n\alpha} \bar{x}_g^{n(\alpha-1)} e^{-n\bar{x}\lambda}}{\Gamma(\alpha)^n},$$

where $\bar{x}_g = \sqrt[n]{\prod x_i}$, and so the log-likelihood function is

$$\log f(\mathbf{x}|\alpha, \lambda) = n\alpha \log \lambda + n(\alpha - 1) \log \bar{x}_g - n\bar{x}\lambda - n \log \Gamma(\alpha).$$

Therefore

$$\begin{aligned} \frac{\partial}{\partial \alpha} \log f(\mathbf{x}|\alpha, \lambda) &= n \log \lambda + n \log \bar{x}_g - n\psi(\alpha) \\ \frac{\partial}{\partial \lambda} \log f(\mathbf{x}|\alpha, \lambda) &= \frac{n\alpha}{\lambda} - n\bar{x} \end{aligned}$$

and so the maximum likelihood estimates $(\hat{\alpha}, \hat{\lambda})$ satisfy

$$\log \hat{\lambda} + \log \bar{x}_g - \psi(\hat{\alpha}) = 0 \quad \text{and} \quad \frac{\hat{\alpha}}{\hat{\lambda}} - \bar{x} = 0,$$

that is,

$$\log \hat{\lambda} + \log \bar{x}_g - \psi(\hat{\lambda}\bar{x}) = 0 \quad \text{and} \quad \hat{\alpha} = \hat{\lambda}\bar{x}.$$

The second derivatives are

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} \log f(\mathbf{x}|\alpha, \lambda) &= -n\psi'(\alpha) \\ \frac{\partial^2}{\partial \lambda^2} \log f(\mathbf{x}|\alpha, \lambda) &= -\frac{n\alpha}{\lambda^2} \\ \frac{\partial^2}{\partial \alpha \partial \lambda} \log f(\mathbf{x}|\alpha, \lambda) &= \frac{n}{\lambda}. \end{aligned}$$

Therefore

$$J_{11}(\hat{\alpha}, \hat{\lambda}) = n\psi'(\hat{\alpha}), \quad J_{12}(\hat{\alpha}, \hat{\lambda}) = -n/\hat{\lambda}, \quad J_{22}(\hat{\alpha}, \hat{\lambda}) = n\hat{\alpha}/\hat{\lambda}^2$$

and so

$$J(\hat{\alpha}, \hat{\lambda}) = \begin{pmatrix} n\psi'(\hat{\alpha}) & -n/\hat{\lambda} \\ -n/\hat{\lambda} & n\hat{\alpha}/\hat{\lambda}^2 \end{pmatrix} = n \begin{pmatrix} \psi'(\hat{\alpha}) & -1/\hat{\lambda} \\ -1/\hat{\lambda} & \hat{\alpha}/\hat{\lambda}^2 \end{pmatrix}$$

whence

$$J(\hat{\alpha}, \hat{\lambda})^{-1} = \frac{1}{n \left\{ \frac{\hat{\alpha}\psi'(\hat{\alpha})}{\hat{\lambda}^2} - \frac{1}{\hat{\lambda}^2} \right\}} \begin{pmatrix} \hat{\alpha}/\hat{\lambda}^2 & 1/\hat{\lambda} \\ 1/\hat{\lambda} & \psi'(\hat{\alpha}) \end{pmatrix} = \frac{1}{n\{\hat{\alpha}\psi'(\hat{\alpha}) - 1\}} \begin{pmatrix} \hat{\alpha} & \hat{\lambda} \\ \hat{\lambda} & \hat{\lambda}^2\psi'(\hat{\alpha}) \end{pmatrix}.$$

Therefore, the asymptotic posterior distribution (as $n \rightarrow \infty$) is

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \Big| \mathbf{x} \sim N \left\{ \begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix}, \frac{1}{n\{\hat{\alpha}\psi'(\hat{\alpha}) - 1\}} \begin{pmatrix} \hat{\alpha} & \hat{\lambda} \\ \hat{\lambda} & \hat{\lambda}^2\psi'(\hat{\alpha}) \end{pmatrix} \right\}.$$

(ii) We have $n = 100$, $\bar{x} = 3.0$ and $\bar{x}_g = 2.5$ leading to maximum likelihood estimates $\hat{\alpha} = 2.8983$ and $\hat{\lambda} = 0.9661$. Note that the value of $\hat{\lambda}$ (and hence that of $\hat{\alpha}$) can be found using the R code

```
zerofn=function(x){(log(x)+log(2.5)-digamma(3*x))^2}
optimise(zerofn,interval=c(0,5))
```

Now $\psi'(\hat{\alpha}) = 0.4112474$ and so

$$\frac{1}{n\{\hat{\alpha}\psi'(\hat{\alpha}) - 1\}} \begin{pmatrix} \hat{\alpha} & \hat{\lambda} \\ \hat{\lambda} & \hat{\lambda}^2\psi'(\hat{\alpha}) \end{pmatrix} = \frac{1}{19.1918} \begin{pmatrix} 2.8983 & 0.9661 \\ 0.9661 & 0.3838 \end{pmatrix} = \begin{pmatrix} 0.1510 & 0.0503 \\ 0.0503 & 0.0200 \end{pmatrix}.$$

Therefore, the asymptotic posterior distribution (as $n \rightarrow \infty$) is

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \Big| \mathbf{x} \sim N \left\{ \begin{pmatrix} 2.8983 \\ 0.9661 \end{pmatrix}, \begin{pmatrix} 0.1510 & 0.0503 \\ 0.0503 & 0.0200 \end{pmatrix} \right\}.$$

The asymptotic posterior correlation between α and λ is

$$\text{Corr}(\alpha, \lambda | \mathbf{x}) = \frac{\text{Cov}(\alpha, \lambda | \mathbf{x})}{\text{SD}(\alpha | \mathbf{x}) \text{SD}(\lambda | \mathbf{x})} = \frac{0.0505}{\sqrt{0.1510 \times 0.0200}} = 0.9189,$$

that is, there is very strong positive posterior correlation between the parameters.

17. The likelihood function is

$$\begin{aligned} f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= \prod_{i=1}^n \frac{\lambda}{\sqrt{2\pi}\sigma} \exp \left\{ -\lambda x_i - \frac{(y_i - x_i)^2}{2\sigma^2} \right\} \\ &= \lambda^n (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -n\lambda \bar{x} - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i)^2 \right\} \end{aligned}$$

and so the log-likelihood function is

$$\log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) = n \log \lambda - n \log \sigma - n\bar{x}\lambda - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i)^2 - \frac{n}{2} \log(2\pi).$$

Taking derivatives, we have

$$\begin{aligned} \frac{\partial}{\partial \lambda} \log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= \frac{n}{\lambda} - n\bar{x} \\ \frac{\partial}{\partial \sigma} \log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - x_i)^2. \end{aligned}$$

Thus the maximum likelihood estimates are

$$\hat{\lambda} = \frac{1}{\bar{x}} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}.$$

Note that technically we've only shown that these are turning points and not the maximum points.

The second order derivatives are

$$\begin{aligned} \frac{\partial^2}{\partial \lambda^2} \log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= -\frac{n}{\lambda^2} \\ \frac{\partial^2}{\partial \sigma \partial \lambda} \log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= 0 \\ \frac{\partial^2}{\partial \sigma^2} \log f(\mathbf{x}, \mathbf{y} | \lambda, \sigma) &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (y_i - x_i)^2. \end{aligned}$$

Therefore the observed information matrix has elements

$$\begin{aligned} J_{11}(\hat{\lambda}, \hat{\sigma}) &= \frac{n}{\hat{\lambda}^2} = n\bar{x}^2 \\ J_{12}(\hat{\lambda}, \hat{\sigma}) &= 0 \\ J_{22}(\hat{\lambda}, \hat{\sigma}) &= -\frac{n}{\hat{\sigma}^2} + \frac{3}{\hat{\sigma}^4} \sum_{i=1}^n (y_i - x_i)^2 = -\frac{n}{\hat{\sigma}^2} + \frac{3n\hat{\sigma}^2}{\hat{\sigma}^4} = \frac{2n}{\hat{\sigma}^2} \end{aligned}$$

with inverse

$$J(\lambda, \sigma)^{-1} = \begin{pmatrix} 1/(n\bar{x}^2) & 0 \\ 0 & \sum_{i=1}^n (y_i - x_i)^2 / (2n^2) \end{pmatrix}.$$

Therefore, for large n , the (approximate) posterior distribution is

$$\begin{pmatrix} \lambda \\ \sigma \end{pmatrix} \Big| \mathbf{x} \sim N_2 \left\{ \begin{pmatrix} 1/\bar{x} \\ \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \end{pmatrix}, \begin{pmatrix} 1/(n\bar{x}^2) & 0 \\ 0 & \sum_{i=1}^n (y_i - x_i)^2 / (2n^2) \end{pmatrix} \right\}.$$

Putting this another way, for large n

$$\lambda | \mathbf{x} \sim N\{1/\bar{x}, 1/(n\bar{x}^2)\}, \quad \sigma | \mathbf{x} \sim N \left\{ \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}, \frac{1}{2n^2} \sum_{i=1}^n (y_i - x_i)^2 \right\},$$

independently.

18. (i) A sample can be obtained using `metropolis(N=10000, initial=0, a=1)` and changing the value of a in each run to get 8 samples. The acceptance rates (acc) and lag 1 autocorrelations (r_1) are given in Table 1. These values are plotted in Figure 2. The plot clearly shows that as the innovation size a increases, the acceptance rate decreases. Also the lag 1 autocorrelations decrease and then increase as a increases. The asymptotic “optimal” acceptance rate of 0.234 is achieved by taking $a \simeq 7$. However this does not minimise the lag 1 autocorrelations: this is done at $a \simeq 4$.

a	1	2	3	4	5	6	7	8
acc	0.802	0.631	0.496	0.389	0.315	0.271	0.235	0.194
r_1	0.876	0.697	0.593	0.564	0.598	0.652	0.675	0.725

Table 1: Table of acceptance rates (acc) and lag 1 autocorrelations (r_1) for various values of a

(ii) Each acceptance rate is calculated as the proportion of times the proposal is accepted in a run of N iterations. If we denote the actual probability of acceptance as θ and $X_i = 1, 0$ depending on whether or not the proposal at iteration i is accepted then $X_i | \theta \sim \text{Bin}(1, \theta)$, independently. This scenario is covered in one of the examples in your MAS2903 notes: the asymptotic posterior distribution for θ is $\theta | \mathbf{x} \sim N(\bar{x}, \bar{x}(1 - \bar{x})/N)$, where $\bar{x}$ is the proportion of accepted proposals. Thus, taking a 95% HDI as representing accuracy, we have that the accuracy is $\pm 1.96 \sqrt{\bar{x}(1 - \bar{x})/N}$. The acceptance rates and their accuracy are given in Table 2. It shows that all the acceptance rates have an accuracy of within 0.01.

a	1	2	3	4	5	6	7	8
acc	0.802	0.631	0.496	0.389	0.315	0.271	0.235	0.194
accuracy	0.0078	0.0095	0.0098	0.0096	0.0091	0.0087	0.0083	0.0078

Table 2: Table of acceptance rates (acc) and their accuracy for various values of a

(iii) We consider first the case with the largest lag 1 autocorrelation: $a = 1$. Running this code for $N = 40k$ iterations and then thinning by taking every `thin = 40` iterates

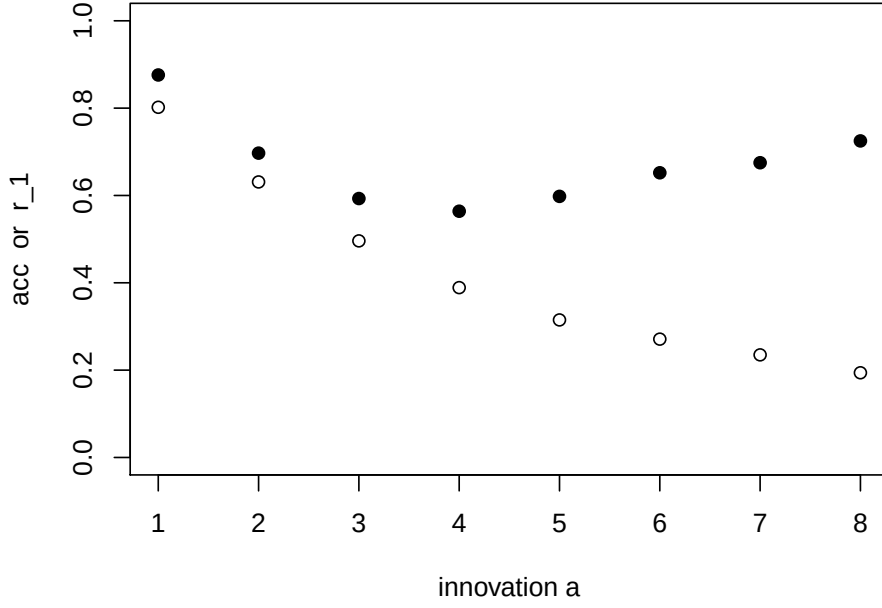


Figure 2: A plot of acceptance rates acc (open circles) and lag 1 autocorrelations r_1 (solid circles) for various values of a

produces a sample of 1000 with $|r_1| \simeq 0.01$. This is sufficiently low to consider this sample to be almost uncorrelated. Using these values of N and `thin` on runs of the code for different values of a gives samples with similarly low lag 1 autocorrelations. The Q–Q plots to assess normality for these samples (using `qqnorm`) is given in Figure 3. These plots clearly show points lying very close to a straight line and so it is reasonable to conclude that each sample is representative of the standard normal distribution.

19. We have $\pi(\theta|\mathbf{x}) = \phi(\theta)$ and the proposal distribution is $\theta^*|\theta \sim N(\theta, k^2)$ for some choice of $k > 0$. As this proposal is symmetric, the acceptance probability is

$$\begin{aligned} \alpha(\theta, \theta^*) &= \min \left\{ 1, \frac{\pi(\theta^*|\mathbf{x})}{\pi(\theta|\mathbf{x})} \right\} = \min \left\{ 1, \frac{\phi(\theta^*)}{\phi(\theta)} \right\} \\ &= \min \left\{ 1, e^{(\theta^2 - \theta^{*2})/2} \right\}. \end{aligned}$$

Therefore the algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\theta^{(0)} = 0$ say.
2. Generate a proposed value $\theta^* \sim N(\theta^{(j-1)}, k^2)$.
3. Evaluate the acceptance probability $\alpha(\theta^{(j-1)}, \theta^*) = \min \left\{ 1, \frac{\phi(\theta^*)}{\phi(\theta^{(j-1)})} \right\}$
4. Set $\theta^{(j)} = \theta^*$ with probability $\alpha(\theta^{(j-1)}, \theta^*)$, and set $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
5. Change the counter from j to $j + 1$ and return to step 2.

20. (i) The proposal distribution has mean $E(\theta^*|\theta) = \theta/a$.

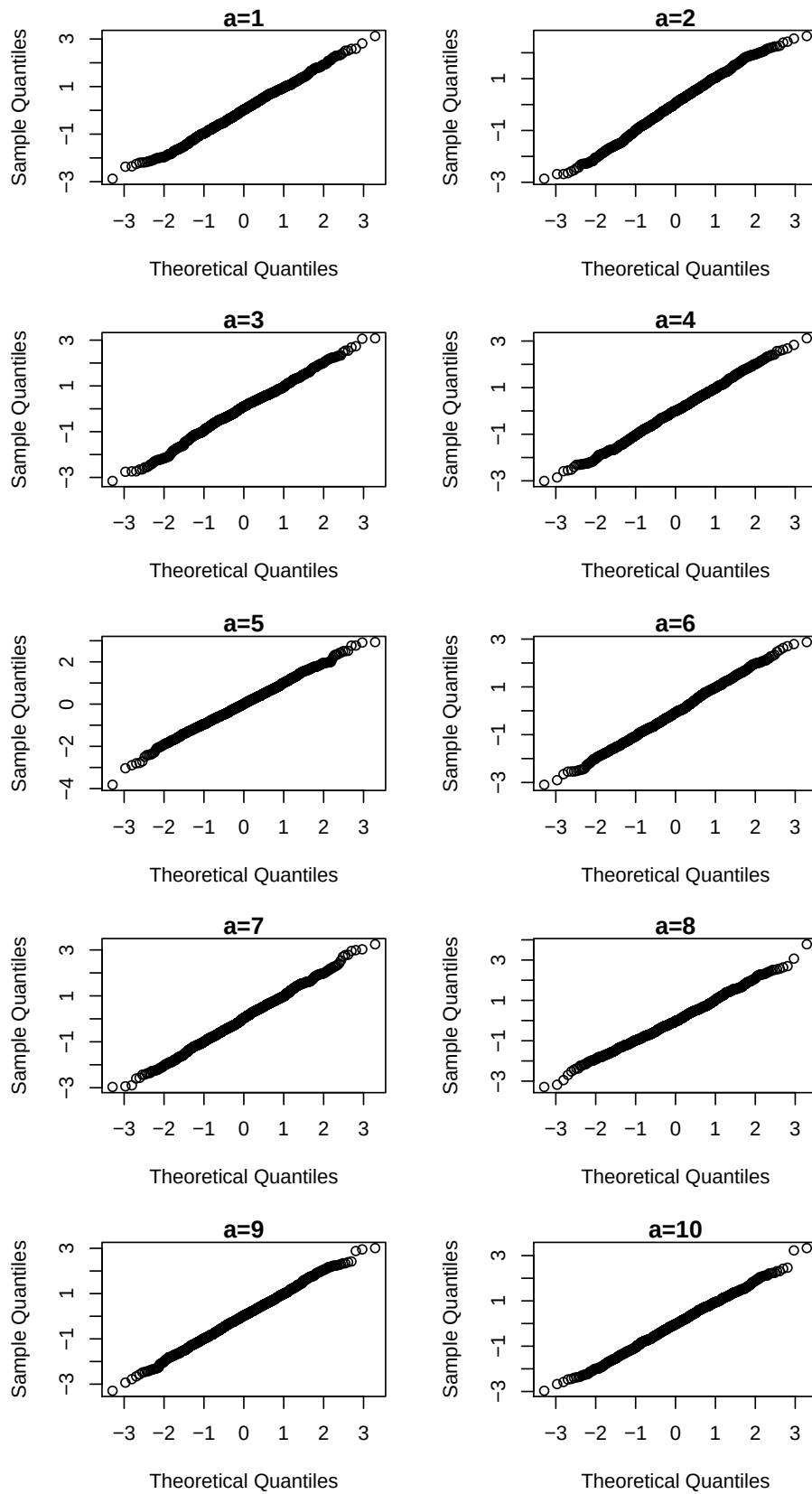


Figure 3: Q-Q plots to assess the normality of the posterior samples for various values of a

(ii) This is a sensible proposal distribution as it only generates positive values for θ^* .

(iii) The posterior density is

$$\pi(\theta|\mathbf{x}) = \frac{H^G \theta^{G-1} e^{-H\theta}}{\Gamma(G)}, \quad \theta > 0$$

and the proposal distribution has density

$$q(\theta^*|\theta) = \frac{a}{\theta} e^{-a\theta^*/\theta}, \quad \theta^* > 0.$$

The acceptance probability for the proposed value θ^* is

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*|\mathbf{x}) q(\theta|\theta^*)}{\pi(\theta|\mathbf{x}) q(\theta^*|\theta)} \right\}.$$

Now

$$\begin{aligned} \frac{\pi(\theta^*|\mathbf{x}) q(\theta|\theta^*)}{\pi(\theta|\mathbf{x}) q(\theta^*|\theta)} &= \frac{\theta^{*G-1} e^{-H\theta^*} \times \theta^{*-1} e^{-a\theta/\theta^*}}{\theta^{G-1} e^{-H\theta} \times \theta^{-1} e^{-a\theta^*/\theta}} \\ &= \left(\frac{\theta^*}{\theta} \right)^{G-2} \exp \left\{ -H(\theta^* - \theta) - a \left(\frac{\theta}{\theta^*} - \frac{\theta^*}{\theta} \right) \right\}. \end{aligned}$$

Therefore the Metropolis–Hastings algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\theta^{(0)} = 0$ from say the prior distribution.
2. Generate a proposed value $\theta^* \sim \text{Exp}(a/\theta)$.
3. Evaluate the acceptance probability

$$\alpha(\theta^{(j-1)}, \theta^*) = \min \left\{ 1, \left(\frac{\theta^*}{\theta^{(j-1)}} \right)^{G-2} \exp \left\{ -H(\theta^* - \theta^{(j-1)}) - a \left(\frac{\theta^{(j-1)}}{\theta^*} - \frac{\theta^*}{\theta^{(j-1)}} \right) \right\} \right\}.$$

4. Set $\theta^{(j)} = \theta^*$ with probability $\alpha(\theta^{(j-1)}, \theta^*)$, and set $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
5. Change the counter from j to $j + 1$ and return to step 2.

21. (i) The proposal distribution has mean

$$E(\theta^*|\theta) = \exp \left\{ \log \theta + \frac{a^2}{2} \right\} = \theta e^{a^2/2}.$$

(ii) This is a sensible proposal distribution as it only generates positive values for θ^* .

(iii) The posterior density is

$$\pi(\theta|\mathbf{x}) = \frac{2H^G \theta^{-2G-1} e^{-H/\theta^2}}{\Gamma(G)}, \quad \theta > 0$$

and the proposal distribution has density

$$q(\theta^*|\theta) = \frac{1}{\theta^* \sqrt{2\pi a^2}} \exp \left\{ -\frac{(\log \theta^* - \log \theta)^2}{2a^2} \right\}, \quad \theta^* > 0.$$

The acceptance probability for the proposed value θ^* is

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*|\mathbf{x}) q(\theta|\theta^*)}{\pi(\theta|\mathbf{x}) q(\theta^*|\theta)} \right\}.$$

Now

$$\begin{aligned} \frac{\pi(\theta^*|\mathbf{x}) q(\theta|\theta^*)}{\pi(\theta|\mathbf{x}) q(\theta^*|\theta)} &= \frac{\theta^{*-2G-1} e^{-H/\theta^{*2}} \times \frac{1}{\theta} \exp \left\{ -\frac{(\log \theta - \log \theta^*)^2}{2a^2} \right\}}{\theta^{-2G-1} e^{-H/\theta^2} \times \frac{1}{\theta^*} \exp \left\{ -\frac{(\log \theta^* - \log \theta)^2}{2a^2} \right\}} \\ &= \left(\frac{\theta}{\theta^*} \right)^{2G} \exp \left\{ -H \left(\frac{1}{\theta^{*2}} - \frac{1}{\theta^2} \right) \right\}. \end{aligned}$$

Therefore the Metropolis–Hastings algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\theta^{(0)}$ from say the prior distribution.
2. Generate a proposed value $\theta^* \sim LN(\log \theta^{(j-1)}, a^2)$.
3. Evaluate the acceptance probability

$$\alpha(\theta^{(j-1)}, \theta^*) = \min \left\{ 1, \left(\frac{\theta^{(j-1)}}{\theta^*} \right)^{2G} \exp \left\{ -H \left(\frac{1}{\theta^{*2}} - \frac{1}{\theta^{(j-1)2}} \right) \right\} \right\}.$$

4. Set $\theta^{(j)} = \theta^*$ with probability $\alpha(\theta^{(j-1)}, \theta^*)$, and set $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
5. Change the counter from j to $j + 1$ and return to step 2.

22. (i) From the notes, equation (2.1) tells us that if X is a random variable with probability density function $f_X(x)$ and g is a bijective (1–1) function then the random variable $Y = g(X)$ has probability density function

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Here we have $\beta = \log \theta$. This is a bijective function with inverse $\theta = e^\beta$. Therefore the posterior density for β is, for $\beta \in \mathbb{R}$

$$\begin{aligned} \pi(\beta|\mathbf{x}) &= \pi_\theta(\theta = e^\beta|\mathbf{x}) \left| \frac{d}{d\beta} e^\beta \right| \\ &= \frac{2H^G e^{-(2G+1)\beta} e^{-He^{-2\beta}}}{\Gamma(G)} \times e^\beta \\ &= \frac{2H^G e^{-2G\beta} e^{-He^{-2\beta}}}{\Gamma(G)}. \end{aligned}$$

(ii) The proposal distribution is a symmetric normal random walk and so the acceptance probability for a proposed value β^* is

$$\alpha(\beta, \beta^*) = \min \left\{ 1, \frac{\pi(\beta^*|\mathbf{x})}{\pi(\beta|\mathbf{x})} \right\}.$$

Now

$$\frac{\pi(\beta^*|\mathbf{x})}{\pi(\beta|\mathbf{x})} = \frac{e^{-2G\beta^*} e^{-He^{-2\beta^*}}}{e^{-2G\beta} e^{-He^{-2\beta}}} = \exp \{ -2G(\beta^* - \beta) - H(e^{-2\beta^*} - e^{-2\beta}) \}.$$

Therefore the Metropolis–Hastings algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\beta^{(0)}$ from say the prior distribution.
2. Generate a proposed value $\beta^* \sim N(\beta^{(j-1)}, a^2)$.
3. Evaluate the acceptance probability

$$\alpha(\beta^{(j-1)}, \beta^*) = \min \left(1, \exp \left\{ -2G(\beta^* - \beta^{(j-1)}) - H(e^{-2\beta^*} - e^{-2\beta^{(j-1)}}) \right\} \right).$$

4. Set $\theta^{(j)} = \theta^*$ with probability $\alpha(\theta^{(j-1)}, \theta^*)$, and set $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
5. Change the counter from j to $j + 1$ and return to step 2.

(iii) We have that

$$\begin{aligned} \alpha(\theta, \theta^*) &= \left(\frac{\theta}{\theta^*} \right)^{2G} \exp \left\{ -H \left(\frac{1}{\theta^{*2}} - \frac{1}{\theta^2} \right) \right\} \\ &= \left(\frac{e^\beta}{e^{\beta^*}} \right)^{2G} \exp \left\{ -H \left(\frac{1}{e^{2\beta^*}} - \frac{1}{e^{2\beta}} \right) \right\} \\ &= \exp \{ -2G(\beta^* - \beta) - H(e^{-2\beta^*} - e^{-2\beta}) \} \end{aligned}$$

and so $\alpha(\theta, \theta^*) = \alpha(\beta, \beta^*)$.

Therefore, using this log normal proposal on is the same as using a normal random walk on the log-scale. Often using this log normal proposal is referred to as a log-normal random walk.

23. This is in the specimen exams.

24. (i) The likelihood function is

$$\begin{aligned} f(\mathbf{x}|\mu, \lambda) &= \prod_{i=1}^n \frac{1}{2\pi I_0(\lambda)} \exp \{ \lambda \cos(x_i - \mu) \} \\ &= (2\pi)^{-n} I_0(\lambda)^{-n} \exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu) \right\}. \end{aligned}$$

Using Bayes' Theorem, the posterior density is, for $0 \leq \mu \leq 2\pi$, $\lambda > 0$

$$\begin{aligned}\pi(\mu, \lambda | \mathbf{x}) &\propto \pi(\mu, \lambda) f(\mathbf{x} | \mu, \lambda) \\ &\propto \frac{1}{2\pi} \times \lambda^{g-1} e^{-h\lambda} \times (2\pi)^{-n} I_0(\lambda)^{-n} \exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu) \right\} \\ &\propto \lambda^{g-1} I_0(\lambda)^{-n} \exp \left\{ \lambda \left[\sum_{i=1}^n \cos(x_i - \mu) - h \right] \right\}.\end{aligned}$$

(ii) The conditional posterior density for μ is

$$\begin{aligned}\pi(\mu | \lambda, \mathbf{x}) &= \frac{\pi(\mu, \lambda | \mathbf{x})}{\pi(\lambda | \mathbf{x})} \\ &\propto \pi(\mu, \lambda | \mathbf{x}) \\ &\propto \exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu) \right\}, \quad 0 \leq \mu \leq 2\pi.\end{aligned}$$

The conditional posterior density for λ is

$$\begin{aligned}\pi(\lambda | \mu, \mathbf{x}) &= \frac{\pi(\mu, \lambda | \mathbf{x})}{\pi(\mu | \mathbf{x})} \\ &\propto \pi(\mu, \lambda | \mathbf{x}) \\ &\propto \lambda^{g-1} I_0(\lambda)^{-n} \exp \left\{ \lambda \left[\sum_{i=1}^n \cos(x_i - \mu) - h \right] \right\}, \quad \lambda > 0.\end{aligned}$$

(iii) The proposal distributions are symmetric normal random walks and so the proposal ratios are one. Therefore the Metropolis–Hastings algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise chain to $(\mu^{(0)}, \lambda^{(0)})$ from say the prior distribution.
2. Generate a proposed value $\mu^* \sim N(\mu^{(j-1)}, \Sigma_\mu)$.
3. Evaluate the acceptance probability

$$\alpha_1(\mu^{(j-1)}, \mu^*) = \min \left(1, \frac{\pi(\mu^* | \lambda^{(j-1)}, \mathbf{x})}{\pi(\mu^{(j-1)} | \lambda^{(j-1)}, \mathbf{x})} \right).$$

4. Set $\mu^{(j)} = \mu^*$ with probability $\alpha_1(\mu^{(j-1)}, \mu^*)$, and set $\mu^{(j)} = \mu^{(j-1)}$ otherwise.
5. Generate a proposed value $\lambda^* \sim N(\lambda^{(j-1)}, \Sigma_\lambda)$.
6. Evaluate the acceptance probability

$$\alpha_2(\lambda^{(j-1)}, \lambda^*) = \min \left(1, \frac{\pi(\lambda^* | \mu^{(j)}, \mathbf{x})}{\pi(\lambda^{(j-1)} | \mu^{(j)}, \mathbf{x})} \right).$$

7. Set $\lambda^{(j)} = \lambda^*$ with probability $\alpha_2(\lambda^{(j-1)}, \lambda^*)$ and set $\lambda^{(j)} = \lambda^{(j-1)}$ otherwise.

8. Change the counter from j to $j + 1$ and return to Step 2.

Now if $0 \leq \mu^* \leq 2\pi$

$$\begin{aligned}\frac{\pi(\mu^*|\lambda, \mathbf{x})}{\pi(\mu|\lambda, \mathbf{x})} &= \frac{\exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu^*) \right\}}{\exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu) \right\}} \\ &= \exp \left\{ \lambda \sum_{i=1}^n [\cos(x_i - \mu^*) - \cos(x_i - \mu)] \right\},\end{aligned}$$

otherwise the ratio is zero.

If $\lambda^* > 0$

$$\begin{aligned}\frac{\pi(\lambda^*|\mu, \mathbf{x})}{\pi(\lambda|\mu, \mathbf{x})} &= \frac{\lambda^{*g-1} I_0(\lambda^*)^{-n} \exp \left\{ \lambda^* \left[\sum_{i=1}^n \cos(x_i - \mu) - h \right] \right\}}{\lambda^{g-1} I_0(\lambda)^{-n} \exp \left\{ \lambda \left[\sum_{i=1}^n \cos(x_i - \mu) - h \right] \right\}} \\ &= \left(\frac{\lambda^*}{\lambda} \right)^{g-1} \left\{ \frac{I_0(\lambda)}{I_0(\lambda^*)} \right\}^n \exp \left\{ (\lambda^* - \lambda) \left[\sum_{i=1}^n \cos(x_i - \mu) - h \right] \right\},\end{aligned}$$

otherwise the ratio is zero.

(iv) The conditional posterior density for μ is

$$\pi(\mu|\lambda, \mathbf{x}) \propto \exp \left\{ \lambda \sum_{i=1}^n \cos(x_i - \mu) \right\}, \quad 0 \leq \mu \leq 2\pi.$$

Now, using the trigonometric identities

$$\begin{aligned}\sum_{i=1}^n \cos(x_i - \mu) &= \sum_{i=1}^n (\cos x_i \cos \mu - \sin x_i \sin \mu) \\ &= C \cos \mu - S \sin \mu \quad \text{where} \quad C = \sum_{i=1}^n \cos x_i, \quad S = \sum_{i=1}^n \sin x_i \\ &= k \cos(\mu - A) \quad \text{where} \quad A = \arctan(S/C), \quad k = \sqrt{C^2 + S^2}\end{aligned}$$

and so

$$\pi(\mu|\lambda, \mathbf{x}) \propto \exp \{ k\lambda \cos(\mu - A) \}, \quad 0 \leq \mu \leq 2\pi,$$

that is, $\mu|\lambda, \mathbf{x} \sim \nu M(A, k\lambda)$. Therefore we can use a Gibbs step which simulates from the full conditional distribution for μ rather than use a less efficient M-H step – this is less efficient as some proposals are not accepted.

25. (i) The likelihood function is

$$\begin{aligned}f(\mathbf{y}|\mu, \lambda) &= \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \{1 - \Phi(\beta + \theta x_i)\}^{1-y_i} \\ &= \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i}.\end{aligned}$$

Using Bayes' Theorem, the posterior density is, for $(\beta, \theta)^T \in \mathbb{R}^2$

$$\begin{aligned}\pi(\beta, \theta | \mathbf{y}) &\propto \pi(\beta, \theta) f(\mathbf{y} | \beta, \theta) \\ &\propto b\phi\{b(\beta - a)\} \times d\phi\{d(\theta - c)\} \times \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i} \\ &\propto \phi\{b(\beta - a)\} \phi\{d(\theta - c)\} \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i},\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density.

(ii) The conditional posterior density for β is, for $\beta \in \mathbb{R}$

$$\begin{aligned}\pi(\beta | \theta, \mathbf{y}) &= \frac{\pi(\beta, \theta | \mathbf{y})}{\pi(\theta | \mathbf{y})} \\ &\propto \pi(\beta, \theta | \mathbf{y}) \\ &\propto \phi\{b(\beta - a)\} \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i}.\end{aligned}$$

The conditional posterior density for θ is, for $\theta \in \mathbb{R}$

$$\begin{aligned}\pi(\theta | \beta, \mathbf{y}) &= \frac{\pi(\beta, \theta | \mathbf{y})}{\pi(\beta | \mathbf{y})} \\ &\propto \pi(\beta, \theta | \mathbf{y}) \\ &\propto \phi\{d(\theta - c)\} \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i}.\end{aligned}$$

(iii) The proposal distributions are symmetric normal random walks and so the proposal ratios are one. Therefore the Metropolis–Hastings algorithm is

1. Initialise the iteration counter to $j = 1$, and initialise chain to $(\beta^{(0)}, \theta^{(0)})$ from say the prior distribution.
2. Generate a proposed value $\beta^* \sim N(\beta^{(j-1)}, \Sigma_\beta)$.
3. Evaluate the acceptance probability

$$\alpha_1(\beta^{(j-1)}, \beta^*) = \min \left(1, \frac{\pi(\beta^* | \theta^{(j-1)}, \mathbf{y})}{\pi(\beta^{(j-1)} | \theta^{(j-1)}, \mathbf{y})} \right).$$

4. Set $\beta^{(j)} = \beta^*$ with probability $\alpha_1(\beta^{(j-1)}, \beta^*)$, and set $\beta^{(j)} = \beta^{(j-1)}$ otherwise.
5. Generate a proposed value $\theta^* \sim N(\theta^{(j-1)}, \Sigma_\theta)$.
6. Evaluate the acceptance probability

$$\alpha_2(\theta^{(j-1)}, \theta^*) = \min \left(1, \frac{\pi(\theta^* | \beta^{(j)}, \mathbf{y})}{\pi(\theta^{(j-1)} | \beta^{(j)}, \mathbf{y})} \right).$$

7. Set $\theta^{(j)} = \theta^*$ with probability $\alpha_2(\theta^{(j-1)}, \theta^*)$ and set $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
8. Change the counter from j to $j + 1$ and return to Step 2.

Also

$$\frac{\pi(\beta^*|\theta, \mathbf{y})}{\pi(\beta|\theta, \mathbf{y})} = \frac{\phi\{b(\beta^* - a)\} \prod_{i=1}^n \Phi(\beta^* + \theta x_i)^{y_i} \Phi(-\beta^* - \theta x_i)^{1-y_i}}{\phi\{b(\beta - a)\} \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i}}$$

and

$$\frac{\pi(\theta^*|\beta, \mathbf{y})}{\pi(\theta|\beta, \mathbf{y})} = \frac{\phi\{d(\theta^* - c)\} \prod_{i=1}^n \Phi(\beta + \theta^* x_i)^{y_i} \Phi(-\beta - \theta^* x_i)^{1-y_i}}{\phi\{d(\theta - c)\} \prod_{i=1}^n \Phi(\beta + \theta x_i)^{y_i} \Phi(-\beta - \theta x_i)^{1-y_i}}.$$