MAS3902

Bayesian Inference

Solutions: Problems Class 1

Semester 1, 2019–20

10. (i) The likelihood function is

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

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

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

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

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

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

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

where

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

Now

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

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

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

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

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

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

where

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

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

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

where

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

The likelihood function is

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

Therefore, for $i = 1, 2$

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

Hence

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

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

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

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

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

(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

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

Also

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

whence the variance of the posterior distribution is

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

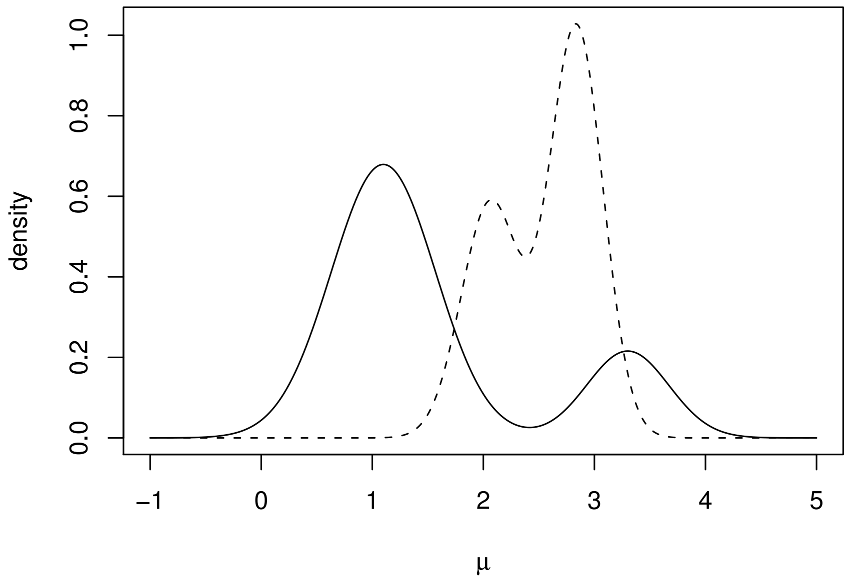

Figure 1: Prior (dashed) and posterior (solid) densities for μ

(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.

(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

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

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

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

Therefore

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

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

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