

STK4021 - Applied Bayesian Analysis

Resit Exam 2023 Sample Solutions

Problem 1

(a) The likelihood is

$$L(\theta) = \prod_{i=1}^n \frac{\theta^{1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{\theta}{2} (y_i - \mu)^2 \right\} = (2\pi)^{-n/2} \theta^{n/2} \exp \left\{ -\frac{1}{2} \theta n w_n \right\},$$

so the log-likelihood is

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \theta - \frac{1}{2} \theta n w_n.$$

The derivative is

$$\ell'(\theta) = \frac{n}{2\theta} - \frac{n w_n}{2},$$

and so the maximiser is given by $\hat{\theta} = 1/w_n$.

The Fisher information is given by

$$\mathcal{I}(\theta) = -\mathbb{E}_{\theta} \left[\frac{d^2}{d\theta^2} \log f_{\theta}(y) \right] = \frac{1}{2\theta^2},$$

which yields the normal approximation $\hat{\theta} \approx N(\theta, 2\theta^2/n)$.

(b) The Jeffreys prior is defined by

$$\pi_J(\theta) \propto \sqrt{\mathcal{I}(\theta)} \propto 1/\theta,$$

which is improper as $\int_0^{\infty} d\theta/\theta$ diverges.

(c) We have

$$\begin{aligned} \pi(\theta | y_1, \dots, y_n) &\propto \pi(\theta) \times L(\theta) \\ &\propto \theta^{a-1} \exp \{-b\theta\} \times \theta^{n/2} \exp \left\{ -\frac{1}{2} \theta n w_n \right\} \\ &= \theta^{a+n/2-1} \exp \{-\theta (b + n w_n/2)\}, \end{aligned}$$

so by functional form, $\theta | y_1, \dots, y_n \sim \text{Gamma}(a_n, b_n)$, where

$$a_n = a + n/2, \quad \text{and} \quad b_n = b + n w_n/2.$$

(d) We have

$$\begin{aligned}
\bar{f}(y) &= \int f_\theta(y) \pi(\theta | y_1, \dots, y_n) d\theta \\
&= \int \frac{\theta^{1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{\theta}{2}(y - \mu)^2 \right\} \frac{b_n^{a_n}}{\Gamma(a_n)} \theta^{a_n-1} \exp \{-b_n \theta\} d\theta \\
&= \frac{b_n^{a_n}}{\sqrt{2\pi}\Gamma(a_n)} \int \theta^{a_n+1/2-1} \exp \left\{ -\theta \left(b_n + \frac{1}{2}(y - \mu)^2 \right) \right\} d\theta.
\end{aligned}$$

We recognise the integrand as an unnormalised $\text{Gamma}(a', b')$ density, where $a' = a_n + 1/2$ and $b' = b_n + (y - \mu)^2/2$. Hence, by the normalisation trick,

$$\begin{aligned}
\bar{f}(y) &= \frac{b_n^{a_n}}{\sqrt{2\pi}\Gamma(a_n)} \frac{\Gamma(a_n + 1/2)}{(b_n + \frac{1}{2}(y - \mu)^2)^{a_n+1/2}} \\
&= \frac{\Gamma(a_n + 1/2)}{\Gamma(a_n)} \frac{b_n^{a_n}}{(2\pi)^{1/2}} \left[b_n + \frac{(y - \mu)^2}{2} \right]^{-a_n-1/2} \\
&= \frac{\Gamma(a_n + 1/2)}{\Gamma(a_n)} \left(\frac{1/b_n}{2\pi} \right)^{1/2} \left[1 + \frac{(1/b_n)(y - \mu)^2}{2} \right]^{-a_n-1/2} \\
&= \frac{\Gamma(a_n + 1/2)}{\Gamma(a_n)} \left(\frac{a_n/b_n}{\pi \times 2a_n} \right)^{1/2} \left[1 + \frac{(a_n/b_n)(y - \mu)^2}{2a_n} \right]^{-a_n-1/2} \\
&= \text{St}(y; \mu, a_n/b_n, 2a_n).
\end{aligned}$$

(e) The posterior mean takes the form

$$\mathbb{E}[\theta | y_1, \dots, y_n] = \frac{a_n}{b_n} = \frac{a + n/2}{b + nw_n/2} = \eta \frac{a}{b} + (1 - \eta) \frac{1}{w_n},$$

as required, where $\eta = b/(b + nw_n/2) \in (0, 1)$. As $n \rightarrow \infty$, $\eta \rightarrow 0$. Hence, the second term in $\mathbb{E}[\theta | y_1, \dots, y_n]$ will dominate for large n , so the prior information is washed out and the maximum likelihood estimator takes over.

(f) The result reflects the Bernstein–von Mises theorem, which in particular asserts that the posterior distance between θ and $\hat{\theta}$ tends to zero as $n \rightarrow \infty$.

Problem 2

(a) We have $\mathbb{E}_\theta y_i = (1 - \theta) \times 0 + \theta \times 1 = \theta$. Also, $\mathbb{E}_\theta y_i^2 = (1 - \theta) \times 0^2 + \theta \times 1^2 = \theta$, so that $\text{Var}_\theta y_i = \theta - \theta^2 = \theta(1 - \theta)$.

(b) The posterior density is

$$\begin{aligned}
\pi(\theta | y_1, \dots, y_n) &\propto \pi(\theta) \times \theta^{n\bar{y}}(1 - \theta)^{n(1-\bar{y})} \\
&\propto \theta^{a+n\bar{y}-1}(1 - \theta)^{b+n(1-\bar{y})-1},
\end{aligned}$$

and so by functional form, $\theta | y_1, \dots, y_n \sim \text{Beta}(a_n, b_n)$, where

$$a_n = a + n\bar{y}, \quad \text{and} \quad b_n = b + n(1 - \bar{y}).$$

(c) We have

$$\begin{aligned}
\pi(y_1, \dots, y_n) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a+n\bar{y}-1} (1-\theta)^{b+n(1-\bar{y})-1} d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+n\bar{y})\Gamma(b+n(1-\bar{y}))}{\Gamma(a+b+n)} \int_0^1 \text{Beta}(\theta; a_n, b_n) d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+n\bar{y})\Gamma(b+n(1-\bar{y}))}{\Gamma(a+b+n)},
\end{aligned}$$

as required.

(d) For the mean,

$$\mathbb{E} y_i = \mathbb{E}[\mathbb{E}[y_i | \theta]] = \mathbb{E} \theta = a/(a+b).$$

For the variance,

$$\begin{aligned}
\text{Var } y_i &= \text{Var}(\mathbb{E}[y_i | \theta]) + \mathbb{E}[\text{Var}(y_i | \theta)] = \text{Var } \theta + \mathbb{E} \theta(1-\theta) = \mathbb{E} \theta^2 - (\mathbb{E} \theta)^2 + \mathbb{E} \theta - \mathbb{E} \theta^2 \\
&= \mathbb{E} \theta \{1 - \mathbb{E} \theta\} = \frac{a}{a+b} \left\{ 1 - \frac{a}{a+b} \right\} = \frac{ab}{(a+b)^2}.
\end{aligned}$$

For the covariance,

$$\begin{aligned}
\text{Cov}(y_i, y_j) &= \text{Cov}(\mathbb{E}[y_i | \theta], \mathbb{E}[y_j | \theta]) + \mathbb{E}[\text{Cov}(y_i, y_j | \theta)] = \text{Cov}(\theta, \theta) = \text{Var } \theta \\
&= \frac{ab}{(a+b)^2(a+b+1)}.
\end{aligned}$$

(e) By de Finetti's theorem, there exists a random variable θ such that the y_i are conditionally independent given θ . In particular, $\text{Cov}(y_i, y_j | \theta) = 0$. Hence,

$$\text{Cov}(y_i, y_j) = \text{Cov}(\mathbb{E}[y_i | \theta], \mathbb{E}[y_j | \theta]) + \mathbb{E}[\text{Cov}(y_i, y_j | \theta)] = \text{Var}(\mathbb{E}[y_i | \theta]) \geq 0.$$

Problem 3

(a) The kernel is given by

$$k((x, y), (x', y')) = \pi_{X|Y}(x' | y) \pi_{Y|X}(y' | x').$$

This satisfies

$$\begin{aligned}
\int \pi_{X,Y}(x, y) k((x, y), (x', y')) dx dy &= \int \pi_{X,Y}(x, y) \pi_{X|Y}(x' | y) \pi_{Y|X}(y' | x') dx dy \\
&= \pi_{Y|X}(y' | x') \int \pi_Y(y) \pi_{X|Y}(x' | y) dy \\
&= \pi_{Y|X}(y' | x') \int \pi_{X,Y}(x', y) dy \\
&= \pi_{Y|X}(y' | x') \pi_X(x') \\
&= \pi_{X,Y}(x', y'),
\end{aligned}$$

as required.

- (b) This is different from Gibbs sampling since the acceptance probability is not equal to one. In lectures, the Gibbs sampling kernel was realised as a combination of Metropolis–Hastings kernels, but this requires that the the accept/reject step happens after each component.
- (c) For $g(x, y) = x/y^2$, we can estimate $\mathbb{E} g(X, Y)$ by

$$\hat{g} = \frac{1}{S} \sum_{s=1}^S x_s / y_s^2.$$

We can then estimate the variance using

$$\hat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^S \{x_s / y_s^2 - \hat{g}\}^2.$$

- (d) The chain is irreducible. Assume without loss of generality that $X_1 > 0$. Then for any non-null set $A \subseteq (0, \infty)$, there is a positive probability of hitting A at each odd step of the chain. Similarly, there is a positive probability of hitting all non-null sets $A \subseteq (-\infty, 0)$ for even n .

The chain is 2-periodic since the state changes sign at each iteration.

- (e) Simply letting $X_n = 0$ for all n works. This is certainly not periodic but the chain is bounded and therefore not irreducible on $\mathbb{R}$.