QUEEN MARY, UNIVERSITY OF LONDON
MTH6102: Bayesian Statistical Methods

Solutions of exercise sheet 8

2023-2024

This is assessed and counts for 4% of the module total. The deadline for submission is Monday the **27th November at 11am**.

Submit the R code used as an R script file (with extension .R) or screenshot. But you need to write the answers and report the R output in a separate file. This can be a pdf document or a clearly legible image of hand-written work.

1. **10 marks.** For the geometric model of exercise sheet 5, question 2, suppose that for a prior distribution, we assign a Beta(α, β) distribution. We take $\alpha = 1$ and choose β so that the prior probability $P(q \leq 0.4)$ is 0.9.

   (a) Find the value of β that satisfies this requirement.

   (b) Use R to simulate a large sample from this beta distribution, and so check your calculation.

   Solution:

   (a) A beta prior distribution for q has pdf

   $$p(q) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)}, \; 0 \leq q \leq 1$$

   where B is the beta function with

   $$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

   When $\alpha = 1$, we get

   $$\begin{aligned} p(q) = \frac{(1-q)^{\beta-1}}{B(1,\beta)} &= \frac{\Gamma(1+\beta)}{\Gamma(1)\Gamma(\beta)}(1-q)^{\beta-1} \\ &= \frac{\beta\Gamma(\beta)}{\Gamma(\beta)}(1-q)^{\beta-1}. \end{aligned}$$

   And so,

   $$p(q) = \beta(1-q)^{\beta-1}.$$

   For any $c \in \mathbb{R}$, the cdf is

   $$F(c) = P(q \leq c) = \int_0^c \beta(1-q)^{\beta-1}\, dq = \Big[-(1-q)^{\beta}\Big]_0^c = 1-(1-c)^{\beta}.$$

We want to find $\beta > 0$ such that

$$F(0.4) = P(q \leq 0.4) = 0.9.$$

$$0.9 = P(q \leq 0.4) = F(0.4) = 1 - (0.6)^{\beta}$$

$$(0.6)^{\beta} = 0.1$$

$$\beta = \frac{\log(0.1)}{\log(0.6)} = 4.51.$$

(b) The R code to check calculation is

```
> alpha=1
> beta=4.51
> prior_sample=rbeta(1000,shape1=alpha,shape2=beta)
> mean(prior_sample <=0.4)
[1] 0.917
```

Alternatively, one could use

```
> pbeta(0.4, shape1=alpha,shape2=beta)
[1] 0.9001238
```

2. **45 marks.** Let $x_1, \ldots, x_n$ iid from $N(\mu, \sigma^2)$ where $\sigma^2 = 1$.

(a) Show that the Jeffreys prior for the normal likelihood is

$$p(\mu) = c_1 \sqrt{n/\sigma^2}, \quad \mu \in \mathbb{R}$$

for some constant $c_1 > 0$.

(b) Is this a proper prior or improrer prior? Explain.

(c) Derive the posterior density for μ under the normal likelihood $N(\mu, \sigma^2)$ and Jeffreys prior for μ. Plot the density. Use the same dataset `x` that you used in problem 2, exercise sheet 5.

(d) Simulate 1,000 draws from the posterior derived in (c) and plot a histogram of the simulated values.

(e) Let $\theta = \exp(\mu)$. Find the posterior density of θ analytically and plot the density.

(f) Estimate θ by Monte Carlo integration.

(g) Compute a 95% equal tail interval for θ analytically and by simulation.

Solution:

(a) Let $x = (x_1, \ldots, x_n)$, where each x_i is assumed to be generated from $N(\mu, \sigma^2)$ with pdf

$$p(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x_i - \mu)^2}{2\sigma^2}\}.$$

The likelihood function is

$$
\begin{aligned}
p(x \mid \mu) &= \prod_{i=1}^{n} p(x_i \mid \mu, \sigma^2) \qquad (1) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x_i-\mu)^2}{2\sigma^2}\} \\
&= (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}.
\end{aligned}
$$

The Jeffreys prior for the normal likelihood with known variance is defined to be

$$p(\mu) \propto \sqrt{I(\mu)},$$

where $I(\mu) = -E[\frac{d^2}{d\mu^2}\log p(X \mid \mu)]$ is the Fisher information function and $X = (X_1, \ldots, X_n)$. We have

$$
\begin{aligned}
\log p(X \mid \mu &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2, \\
\frac{d}{d\mu}\log p(X \mid \mu) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu), \\
\frac{d^2}{d\mu^2}\log p(X \mid \mu) &= -\frac{n}{\sigma^2}.
\end{aligned}
$$

Thus,

$$I(\mu) = -E[\frac{d^2}{d\mu^2}\log p(X \mid \mu)] = \sqrt{n/\sigma^2}.$$

Therefore,

$$p(\mu) \propto \sqrt{n/\sigma^2}, \quad \mu \in \mathbb{R}. \qquad (2)$$

(b) This is an improper prior because the integral $\int_{\mathbb{R}} p(\mu)\, d\mu$ is not finite.

(c) The posterior density, $p(\mu \mid x)$ under the normal likelihood in (1) and Jeffreys prior in (2) is

$$p(\mu \mid x) \propto p(\mu) \propto p(x \mid \mu).$$

Using the likelihood in (1), Jeffreys prior in (2), and ignoring the terms that do not depend on μ, we get

$$
\begin{aligned}
p(\mu \mid x) &\propto \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\} \\
&= \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{n\bar{x}\mu}{\sigma} - \frac{n\mu^2}{2\sigma^2}\} \qquad \text{(ignoring constant terms)} \\
&\propto \exp\{\frac{n\bar{x}\mu}{\sigma} - \frac{n\mu^2}{2\sigma^2}\} \\
&= \exp\{-\frac{n}{2\sigma^2}(\mu^2 - 2\bar{x}\mu)\} \qquad \text{(completing the square)} \\
&\propto \exp\{-\frac{1}{2(\sigma^2/n)}(\mu-\bar{x})^2\}.
\end{aligned}
$$

Hence, the posterior density, $p(\mu \mid x)$, is $N(\bar{x}, \frac{\sigma^2}{n})$. We also have $\bar{x} = 2.46$, $\sigma = 1$ and $n = 35$, so $p(\mu \mid x)$, is $N(2.46, \frac{1}{35})$. Figure 1 shows the plot of the posterior density, $N(\bar{x}, \frac{\sigma^2}{n})$.

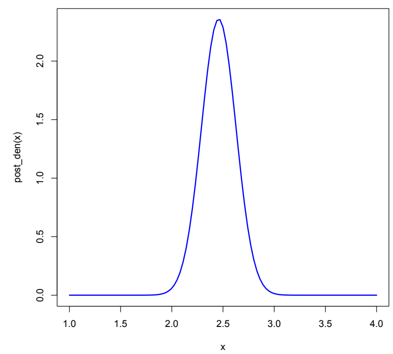

Figure 1: Plot of the posterior density, $N(\bar{x}, \frac{\sigma^2}{n})$, of μ.

(d) Figure 2 shows the histogram of 1000 realisations from $N(\bar{x}, \frac{\sigma^2}{n})$, of μ.

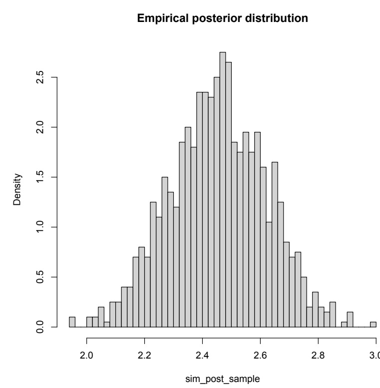

Figure 2: Histogram of 1000 iid realisations from $N(\bar{x}, \frac{\sigma^2}{n})$, of μ.

(e) If we let $\theta = g(\mu)$, then $g(\mu)$ is an increasing function with inverse $\mu = g^{-1}(\theta) = \ln(\theta)$ and $\frac{d}{d\theta} g^{-1}(\theta) = \frac{1}{\theta}$. Note that the support of θ is the interval $(0, \infty)$. Applying the transformation of variables, for $\theta \in (0, \infty)$, we get

$$\begin{aligned} p(\theta \mid x) &= \frac{d}{d\theta} g^{-1}(\theta) p(g^{-1}(\theta) \mid x) \\ &= \frac{1}{\theta} \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\{-\frac{(\ln\theta - \bar{x})^2}{2(\sigma^2/n)}\}. \end{aligned}$$

Hence, the posterior density of θ is

$$p(\theta \mid x) = \frac{1}{\theta} \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\{-\frac{(\ln\theta - \bar{x})^2}{2(\sigma^2/n)}\}, \quad \theta > 0.$$

This is known as the lognormal distribution with parameters $\bar{x}$ and σ^2/n. Figure 3 gives the plot of the density $p(\theta \mid x)$ with parameters $\bar{x} = 2.46$ and $\sigma^2/n = 1/35$.

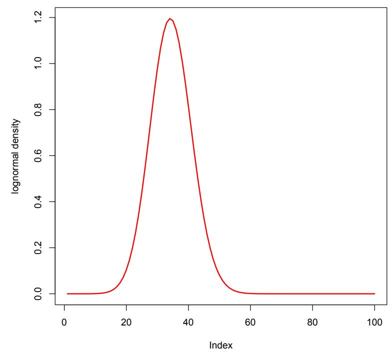

Figure 3: Plot of the posterior pdf $p(\theta \mid x)$ with with parameters $\bar{x} = 2.46$, $\sigma^2/n = 1/35$

(f) We can use the posterior mean to estimate $\theta = g(\mu) = \exp(\mu)$ given by

$$I = E(\theta) = E(g(\mu)) = \int_{\mathbb{R}} g(\mu)p(\mu \mid x)\, d\mu.$$

Hence, to estimate I using Monte Carlo integration we simulate N iid observations $M_i, i = 1, \ldots, N$ from the posterior density $N(\bar{x}, \frac{\sigma^2}{n})$, and estimate I by

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} g(M_i) = \frac{1}{N}\sum_{i=1}^{N} \exp(M_i).$$

In R,

```
> sim_mu_post_sample=rnorm(1000,mean=post_mean,sd=sqrt(post_variance
  ))
> mean(exp(sim_mu_post_sample))
[1] 11.83156
```

So, $\hat{I} = 11.83$ is the Monte Carlo estimate for the posterior mean of θ.

(g) First, a 95% equal tail interval for μ is

$$[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}],$$

which is the same with a 95% confidence interval for μ. Since $g(\cdot)$ is monotone increasing, a 95% equal tail interval for $\theta = g(\mu) = \exp(\mu)$ is

$$[\exp\{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}\}, \exp\{\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\}].$$

Using the data, the 95% equal tail interval for $\theta = g(\mu) = \exp(\mu)$ is $[8.39, 16.27]$. Using simulations, we first note that $\theta_i = \exp(M_i)$, $i = 1, \ldots, N$ can be viewed as an iid sample from θ. Hence, a 95% equal tail interval for $\theta = \exp(\mu)$ can be obtained by sorting the simulated values $\theta_i = \exp(M_i)$, $i = 1, \ldots, N$, and finding the 0.025 and 0.975 quantiles. In R,

```
> quantile(exp(sim_mu_post_sample),c(0.025,0.975))
     2.5%     97.5%
 8.501876 16.224258
```

3. **45 marks.** Consider the 10 Bernoulli(q) observations: 0,1,0,1,0,0,0,0,0,0. Plot the posterior for q using these priors: Beta(0.5,0.5), Beta(1,1), Beta(10,10), Beta(100,100). Comment on the effect of the prior on the posterior.

   Solution: Let $x = (0,1,0,1,0,0,0,0,0,0)$. Then each observation x_i is a Bernoulli with pdf

   $$p(x_i \mid q) = q^{x_i}(1-q)^{1-x_i}, \quad x_i = 0,1, i = 1,\ldots,10.$$

   A beta prior distribution beta(α,α) for q has pdf

   $$p(q) = \frac{q^{\alpha-1}(1-q)^{\alpha-1}}{B(\alpha,\alpha)},\ 0 \le q \le 1,$$

   where B is the beta function. The posterior density of q, $p(q \mid x)$, is

   $$\begin{aligned} p(q \mid x) &\propto q^{\sum_{i=1}^{10} x_i}(1-q)^{n-\sum_{i=1}^{10} x_i} q^{\alpha-1}(1-q)^{\alpha-1} \\ &= q^{k+\alpha-1}(1-q)^{n-k+\alpha-1}, \end{aligned}$$

   where $k = \sum_{i=1}^{10} x_i = 2$ is the number of successes. Hence, the posterior density is beta$(2+\alpha, 8+\alpha)$. We also have that the prior mean and the prior variance of q are, respectively

   $$E(q) = \frac{1}{2}, \quad \mathrm{var}(q) = \frac{1}{4(2\alpha+1)}.$$

   Hence, as α increases, the prior standard deviation decreases and the prior becomes more informative. On the other hand, as α decreases, the prior standard deviation increases and the prior becomes less informative (wider, more flat).

   Figure 4 shows the posterior for q using these priors: Beta(0.5,0.5) (left upper panel), Beta(1,1) (right upper panel), Beta(10,10) (left lower panel) and Beta(100,100) (right lower panel), respectively. We observe that as the prior density becomes more informative (smaller standard deviation, larger α) the posterior mode moves toward the prior mode. On the other hand, as the prior becomes less informative (more flat), the posterior mode moves towards the mle and away from the prior mode.

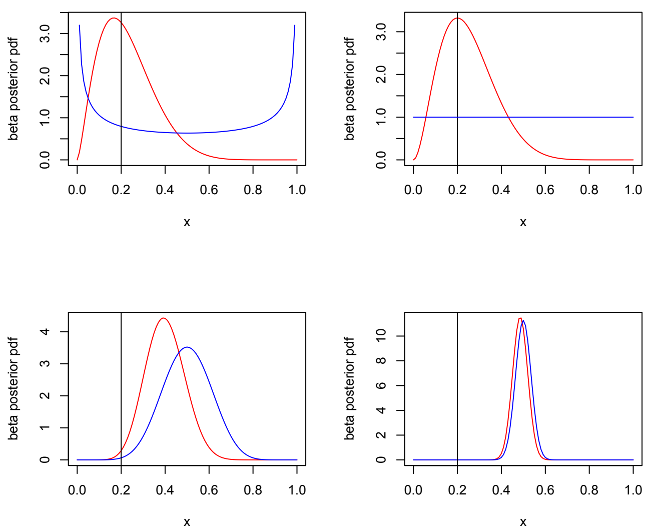

Figure 4: Plot of the beta posterior density beta($2+\alpha, 8+\beta$) using these priors: Beta(0.5,0.5), Beta(1,1), Beta(10,10), Beta(100,100) (from left to right). Red denotes the posterior density, blue the prior density and the black vertical line represents the MLE which is 0.2.