

Bayesian Inference: Assignment 3

Answers to the starred (*) questions should be submitted to Canvas by no later than **4pm, Friday 22nd March 2024**. Solutions to non-assessed questions will be made available in class and/or in Canvas.

Question 1* (30 marks total)

A hepatologist is interested in the levels of the liver enzyme *ornithine carbonyltransferase* in patients suffering from acute viral hepatitis. She collects measurements from a random sample of patients and the logarithm of their enzyme measurements are given in the following table. They are also available in the R datafile `hepatitis` in the `nclbayes` package; alternatively, you can download these data from a webpage by typing (but **not** copy-and-pasting!):

```
1 hepatitis = read.table("https://www.mas.ncl.ac.uk/~nlf8/teaching/mas3902/hepatitis.txt")
```

2.64	2.51	2.20	2.53	2.02	2.47	2.75	2.77	2.91	2.45
2.25	1.96	2.22	2.23	1.98	2.70	2.61	2.76	2.03	2.38
2.62	2.28	2.47	3.04	1.91	2.71	2.89	2.70	2.29	2.50

Assume that the enzyme measurement varies according to a $N(\mu, 1/\tau)$ distribution. After extensiver discussions with an expert, prior beliefs about μ and τ are summarised as

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim NGa(2.6, 1, 5, 0.4).$$

- Calculate the expert's prior mean and standard deviation for μ , τ and $\sigma = 1/\sqrt{\tau}$.
- Determine the posterior distribution for $(\mu, \tau)^T$ after combining the hepatologist's prior beliefs with the data. Calculate the posterior mean and standard deviation of μ , τ and σ .
- From the lecture notes, we know that $\mu|\mathbf{x} \sim t_{2G}(B, H/GC)$ and $\tau|\mathbf{x} \sim Ga(G, H)$ (where expressions for B , C , G and H are given in the notes). Show that $\sigma|\mathbf{x} \sim Inv\text{-}Chi(G, H)$.
- Plot the (marginal) prior and posterior densities for μ on the same graph. Construct similar plots for τ and σ . Also produce contour plots of the (joint) prior and posterior densities for $(\mu, \tau)^T$ on the same graph.
- Plot 80%, 90% and 95% prior and posterior confidence regions for $(\mu, \tau)^T$ on the same graph.
- Use these plots and your calculations to comment on the main changes in the hepatologist's beliefs about μ , τ and σ after incorporating the data. Include a comment on the prior-to-posterior change in the dependence structure (contour shape) of (μ, τ) and on their confidence regions for (μ, τ) .

- (g) The hepatologist is particularly interested in whether the population mean level μ is larger than 2.7. Determine the prior and posterior probabilities for $\mu > 2.7$. Have the data been informative?
- (h) The hepatologist starts to think about the enzyme level Y in the next patient admitted to the study. Write down the predictive distribution for this future observation, and find the predictive probability that $Y > 2.7$.

Question 2

Suppose you are interested in the aerobic capacity (ml/kg/min) of athletes, measurements on which are assumed Normally distributed with mean μ and precision τ .

Prior specification:

- (a) In a previous study, the aerobic capacity of athletes was assumed Normal with known standard deviation $\sigma = \sqrt{3}$ ml/kg/min; the mean μ was given a Normal prior such that $\mu \sim N(71, 1)$. Assuming the same values for prior mean and variance in the current study, find b and c such that

$$\mu|\tau \sim N\left(b, \frac{1}{c\tau}\right).$$

- (b) Unlike the previous study, in the current study we are assuming *both* μ and τ (and hence $\sigma = 1/\sqrt{\tau}$) are unknown. We assume a Gamma prior for the precision τ , that is,

$$\tau \sim Ga(g, h).$$

Discussions with a sports scientist suggest that $E(\tau) = Var(\tau) = 2$. Find g and h .

- (c) Using results from Chapter 2, write down the joint prior for $(\mu, \tau)^T$.

Posterior distribution:

The aerobic capacity of 50 randomly chosen athletes was measured, giving a mean of $\bar{x} = 71.5$ ml/kg/min and standard deviation $s = 2.3$ ml/kg/min.

- (d) Using results from Chapter 2, find the posterior distribution for $(\mu, \tau)^T$.
- (e) Complete the following table:

μ	Prior distribution: $\mu \sim$ Mean: Variance:	Posterior distribution: $\mu \mathbf{x} \sim$ Mean: Variance:
τ	Prior distribution: $\tau \sim$ Mean: Variance:	Posterior distribution: $\tau \mathbf{x} \sim$ Mean: Variance:
σ	Prior distribution: $\sigma \sim$ Mean: Variance:	Posterior distribution: $\sigma \mathbf{x} \sim$ Mean: Variance:

- (f) Sketch the prior and posterior for μ on the same graph.
- (g) Which of the plots in Figure 1 (a) – left or right – correspond to the prior (dashed) and posterior (bold) distributions for $(\mu, \tau)^T$? Explain your answer.

Confidence intervals and regions:

- (h) Find the 95% highest density intervals for $\mu|\mathbf{x}$. You might find the following output from R useful:

```
> qt(0.975, 54)
[1] 2.004879
```

- (i) Why are highest density intervals for τ and σ not as straightforward to obtain? Use a Normal approximation to obtain approximate 95% confidence intervals for $\tau|\mathbf{x}$ and $\sigma|\mathbf{x}$.
- (j) Describe what the plot in Figure 1(b) shows.

Prediction:

Suppose we are interested in the aerobic capacity of the next athlete to join the study.

- (k) Using results from Chapter 2, find the predictive distribution for the aerobic capacity of this new athlete.
- (l) Find the 95% highest density interval associated with the predictive distribution in (i).

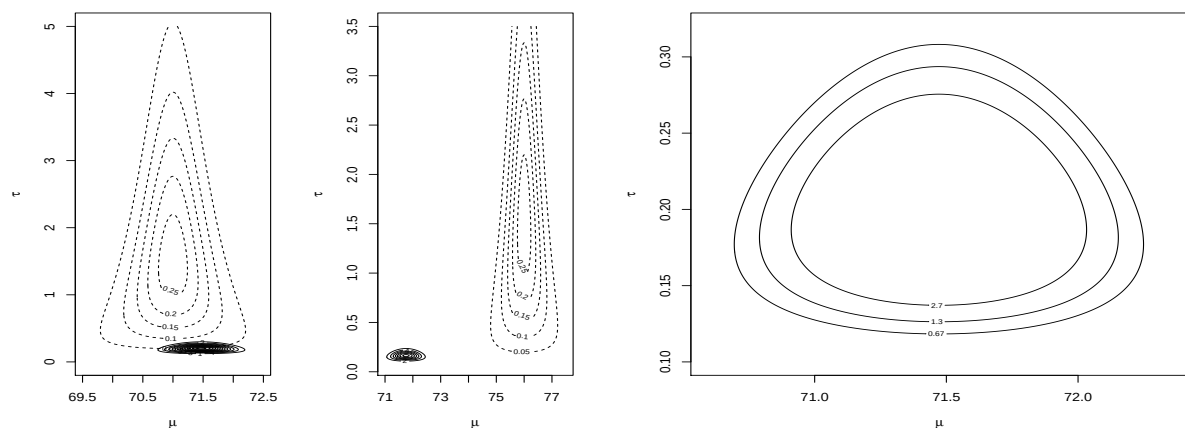


Figure 1: (a) Prior (dashed) and posterior (solid) densities (left); (b) contours from a bivariate density (right).

Solutions

Question 2

(a) From the previous study, we have $X_i|\mu \sim N(\mu, 3)$ (so $\tau = 1/3$); we also have that $\mu \sim N(71, 1)$. In *our* study, we assume that $\mu|\tau \sim N(b, 1/c\tau)$.

- Keeping the prior means the same gives $b = 71$
- Keeping the prior variances the same gives $1/c\tau = 1$, that is,

$$\frac{1}{c \left(\frac{1}{3}\right)} = 1,$$

and so $c = 3$.

(b) We assume $\tau \sim Ga(g, h)$.

- Using $E(\tau) = g/h = 2$ gives $g = 2h$.
- Using $Var(\tau) = g/h^2 = 2$ gives

$$\frac{2h}{h^2} = 2 \longrightarrow h = 1;$$

therefore $g = 2$.

(c) So, from (a) and (b) we have that

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim NGa(71, 3, 2, 1).$$

(d) From chapter 2, we know that

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} | \mathbf{x} \sim NGa(B, C, G, H),$$

with B , C , G , and H as defined on page 32 of the lecture notes. With $b = 71$, $c = 3$, $g = 2$, $h = 1$ and $n = 50$, $\bar{x} = 71.5$, $s = 2.3$, we have

$$B = 71.4717, C = 53, G = 27, H = 133.6038.$$

Therefore,

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} | \mathbf{x} \sim NGa(71.4717, 53, 27, 133.6038),$$

(e) For example, we have $\mu | \mathbf{x} \sim t_{2G}(B, H/GC) = t_{54}(71.4717, 0.0934)$, with

$$E(\mu | \mathbf{x}) = 71.4717; \quad Var(\mu | \mathbf{x}) = \frac{54 \times 0.0934}{52} = 0.09699,$$

using the formula sheet. Similarly, for the prior and the other parameters, gives:

μ	Prior distribution: $\mu \sim t_4(71, 0.1667)$ Mean: 71 Variance: 0.3334	Posterior distribution: $\mu \mathbf{x} \sim t_{54}(71.4717, 0.0934)$ Mean: 71.4717 Variance: 0.09699
τ	Prior distribution: $\tau \sim Ga(2, 1)$ Mean: 2 Variance: 2	Posterior distribution: $\tau \mathbf{x} \sim Ga(27, 133.6038)$ Mean: 0.2021 Variance: 0.00153
σ	Prior distribution: $\sigma \sim InvChi(2, 1)$ Mean: 0.8862 Variance: 0.2146	Posterior distribution: $\sigma \mathbf{x} \sim InvChi(27, 133.6038)$ Mean: 2.2560 Variance: 0.0491

(g) Must be the left-hand plot – the modes are at the incorrect place for μ in both prior and posterior in the right-hand plot.

(h) Using the fact that $\mu|\mathbf{x} \sim t_{54}(71.4717, 0.0934)$, we can find

$$71.4717 \pm 2.00489\sqrt{0.0934} \longrightarrow (70.859, 72.8442).$$

(i) Not usually symmetric! For $\tau|\mathbf{x} \sim Ga(27, 133.6038)$, we have $E(\tau|\mathbf{x}) = 0.2021$ and $Var(\tau|\mathbf{x}) = 0.00153$. Therefore, approximately, we have

$$\tau|\mathbf{x} \sim N(0.2021, 0.00153),$$

giving $0.2021 \pm 1.96 \times \sqrt{0.00153} \longrightarrow (0.1254, 0.2788)$. Similarly for $\sigma|\mathbf{x}$ we have $(1.8217, 2.6903)$.

(j) 80%, 90% and 95% credible regions for $(\mu, \tau)^T$.

(k) $Y|\mathbf{x} \sim t_{54}(71.4717, 5.0417)$.

(l) $71.4717 \pm 2.004879 \times \sqrt{5.0417} \longrightarrow (66.9699, 75.9734)$ ml/kg/min..