Group project

1. Suppose you have a random sample of size $n = 10$ from an $N(\mu, 1)$ distribution with mean $\bar{x}$ and your prior distribution for μ is the mixture distribution

$$\mu \sim 0.2\, N(3.3, 0.37^2) + 0.8\, N(1.1, 0.47^2).$$

(a) Determine the posterior distribution for μ when $\bar{x} = 2.0, 2.3, 2.4, 2.5, 2.8$. Calculate (to three decimal places) the mean, standard deviation and $Pr(\mu > 2.5|\boldsymbol{x})$ for the prior distribution and these posterior distributions. (Hint: Refer to question 11 in the 'Group Exercises')

15 marks

(b) Plot the prior density and these posterior densities on the same graph.

4 marks

(c) Describe the effect of observing these sample means on the posterior distribution by comparing their shape, mean, standard deviation and $Pr(\mu > 2.5|\boldsymbol{x})$ with that of the prior distribution.

6 marks

(d) Plot the posterior weight p_1^* for sample means in the range $\bar{x} \in (0, 30)$ and comment on how p_1^* depends on the sample mean $\bar{x}$. By studying the underlying mathematics, explain the feature you see algebraically.

14 marks

2. 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.

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. An expert says her (prior) beliefs about μ and τ can be summarised as

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

(a) Use a normal probability (q–q) plot to confirm the suitability of the normal distribution as a model for the variation in enzyme measurements. The relevant `R` commands are `qqnorm` and `qqline`.

4 marks

(b) Calculate her prior mean and standard deviation for μ, τ and $\sigma = 1/\sqrt{\tau}$.

6 marks

(c) Determine the (joint) 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.

8 marks

(d) 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.

8 marks

(e) Plot 80%, 90% and 95% prior and posterior confidence regions for $(\mu, \tau)^T$ on the same graph.

2 marks

(f) 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).

5 marks

(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?

2 marks

The hepatologist starts to think about the enzyme levels in the next sample of m patients.

(h)* Determine the predictive distribution for $\bar{Y}$, the mean of this future sample.

2 marks

(i)* Plot the predictive density of $\bar{Y}$ for the case $m = 20$, and determine the 95% prediction interval for $\bar{Y}$.

2 marks

(j)* Verify that the predictive distribution for $V = \sum_{i=1}^{m}(Y_i - \bar{Y})^2/m$, the variance of this future sample, has a scaled F-distribution, that is, $V|\boldsymbol{x} \sim aF_{\nu_1,\nu_2}$ for some choice of a, ν_1 and ν_2. Hints:

1. Recall from MAS2901 that in normal random samples $(m-1)S_u^2/\sigma^2 \sim \chi^2_{m-1}$. The equivalent statement in our Bayesian setting is $mV\tau|\tau \sim \chi^2_{m-1}$.
2. $\chi^2_\nu \equiv Ga(\nu/2, 1/2)$ and $Ga(a,b)/c \equiv Ga(a, bc)$.

3. If $Y \sim aF_{\nu_1,\nu_2}$ then it has density

$$f(y) = \frac{1}{B(\nu_1/2, \nu_2/2)} \left(\frac{\nu_1}{\nu_2 a}\right)^{\nu_1/2} y^{\nu_1/2-1} \left(1 + \frac{\nu_1 y}{\nu_2 a}\right)^{-(\nu_1+\nu_2)/2}, \quad y > 0.$$

9 marks

(k)* For the case $m = 20$, determine the 95% equi-tailed prediction interval for V and hence a 95% confidence interval for $S = \sqrt{V}$, the standard deviation of this future sample.

3 marks

* These questions are quite difficult and are to test good first class students — no help will be given with them.

Presentation

Your report should be clearly written but need not be typed. It should be written as separate answers to each part question, contain the details of any calculations such as those in Qn 2(b) but not any of the `R` commands used to generate numerical or graphical output. Marks will be given for appropriate titling, labelling and annotation of plots.

10 marks