Group exercises

1. Suppose you have a random sample $x_1, x_2, \ldots, x_n$. In the following models, derive the posterior density and name the posterior distribution (and its parameters).

   (i) $f(x|\theta) = \theta^{x-1}(1-\theta)$, $x = 1, 2, \ldots$ with a $Beta(3,2)$ prior distribution for θ.

   (ii) $f(x|\theta) = \dfrac{e^{-\theta}\theta^x}{x!}$, $x = 0, 1, \ldots$ with a $Exp(2)$ prior distribution for θ.

   (iii) $f(x|\theta) = \dbinom{4}{x}\theta^x(1-\theta)^{4-x}$, $x = 0, 1, 2, 3, 4$ with a $U(0,1)$ prior distribution for θ.

2. Suppose that a random sample $x_1, x_2, \ldots, x_{10}$ is obtained from a uniform $U(0,\theta)$ distribution. Derive the posterior density for θ assuming a gamma $Ga(20,1)$ prior distribution θ. Hint: this distribution will depend on the maximum observed x-value $x_{\max}$.

3. Suppose that a random sample $x_1, x_2, \ldots, x_n$ is obtained from a truncated unit exponential distribution, with density

$$f(x|\theta) = \begin{cases} e^{\theta - x}, & x > \theta \\ 0, & \text{otherwise,} \end{cases}$$

   where $\theta \in \mathbb{R}$. Derive the posterior density for θ assuming a normal $N(b, 1/d)$ prior distribution. Hint: this distribution will depend on the minimum observed x-value $x_{\min}$.

4. The dimensions of a component from a long production run vary according to a $N(\mu, 1)$ distribution, and the mean dimension μ varies from production run to production run according to a $N(10, 1/4)$ distribution. From one production run 12 components are drawn at random and their average dimension is found to be $10\frac{1}{3}$. On this information what is the probability that the mean component dimension is at least 10?

5. A trucking company owns a large fleet of well-maintained trucks. Suppose that breakdowns occur at random times. The owner of the company is interested in learning about the daily rate θ at which breakdowns occur. It is known that the number of breakdowns X on a typical day has a Poisson distribution with mean θ. The owner has some knowledge about the rate parameter θ based on the observed number of breakdowns in previous years and expresses these prior beliefs using a

$Ga(4, 2)$ distribution. The daily number of truck breakdowns are obtained on five consecutive days as 3, 2, 3, 1, and 2. Assuming these data are a random sample, determine the posterior distribution of θ.

6. Suppose that the time in minutes required to serve a customer at a certain facility has an exponential distribution for which the parameter θ is unknown, and that the prior distribution of θ is a gamma distribution with mean 0.2 and standard deviation 1. If the average time required to serve a random sample of 20 customers is observed to be 3.8 minutes, determine the posterior distribution of θ.

7. The following data is the time intervals (in minutes) between eruptions of the "Old Faithful" geyser. Note that the average time between eruptions is 67.38 minutes, that is, 1.123 hours. If the eruptions occur randomly in time according to a Poisson process with a rate of θ eruptions per hour then the times between eruptions will form a random sample from an exponential distribution with rate θ. Making this assumption, establish a prior distribution for θ with $E(\theta) = 1$ and $Var(\theta) = 1$. Determine the posterior distribution for θ. Comment on your analysis.

70	64	72	76	80	48	88	53	71	56	69	72	76	54	76
65	54	86	40	87	49	76	51	77	49	71	78	80	51	82
49	80	43	83	49	75	47	78	71	69	63	64	82	68	71
71	63	79	66	75	56	83	67	65	77	72	79	73	53	69
53	78	55	67	68	73	53	70	69	66	79	48	90	49	78
52	79	49	75	75	50	87	40	76	57	71	70	69	72	51
84	43	73	73	70	84	71	79	58	73					

8. The negative binomial distribution is used to model scenarios in which we observe the number of independent (success-fail) trials needed before we see a successful trial; the trials must have the same success probability. Suppose $x_1, x_2, \dots, x_n$ are a random sample from a negative binomial $NegBin(k, \theta)$ distribution, where k is known.

   (i) Verify that the congugate prior distribution is a Beta distribution.

   (ii) Determine the choice of parameters g and h for the $Beta(g, h)$ distribution that give it maximal variance.
   Hint: reparameterise the distribution in terms of its mean $m = g/(g + h)$ and $s = g + h$; determine the choice of m and s that maximises the variance of the distribution, and hence the choice of g and h.

   Determine the posterior distribution for θ assuming

   (iii) vague prior knowledge;

   (iv) a very large sample.

9. The Rayleigh distribution is often used to measure variability in magnetic resonance imaging (MRI). Suppose $x_1, x_2, \dots, x_n$ are a random sample from a Rayleigh $R(\theta)$ distribution. Determine the posterior distribution for θ assuming

   (i) vague prior knowledge;

(ii) a very large sample.

10. The Pareto distribution is often used to model data in many areas, ranging from the wealth of individuals to the sizes of meteorites. Suppose $x_1, x_2, \ldots, x_n$ are a random sample from a Pareto $Pa(1, \theta)$ distribution. Determine the posterior distribution for θ assuming

    (i) vague prior knowledge;

    (ii) a very large sample.

11. Suppose that a random sample of size $n = 10$ from an $N(\mu, 1)$ distribution has mean $\bar{x} = 2.5$. The prior distribution for μ is the mixture distribution

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

    (i) Determine the posterior distribution for μ. Note that, for this model

$$f_i(\boldsymbol{x}) = \frac{\pi_i(\mu)\, f(\boldsymbol{x}|\mu)}{\pi_i(\mu|\boldsymbol{x})} \propto \frac{\sqrt{d_i}}{\sqrt{D_i}} \exp\left\{\frac{1}{2}\left[D_i B_i^2 - d_i b_i^2\right]\right\},$$

    where the constant of proportionality doesn't depend on component prior i. In order to get accurate values for the posterior weights, you should calculate posterior component means and standard deviations to at least 4 dp.

    (ii) Calculate the prior and posterior mean and standard deviation.

    (iii) Plot the prior and posterior densities for μ.

    (iv) Calculate the prior and posterior probability that μ exceeds 2.5. Comment on the effect of incorporating the data.

12. Consider the exponential model and gamma mixture prior distribution described in Example 1.13. Define your own R functions to calculate the first component weight (p_1^*) and the posterior mean and standard deviation as functions of the sample mean $\bar{x}$. Investigate the behaviour of these functions and give an intuitive explanation of their general properties.

13. Suppose that you have a random sample $x_1, x_2, \ldots, x_n$ from an $N(\mu, \sigma^2)$ distribution and take a $NGa(b, c, g, h)$ prior distribution for $(\mu, \tau)^T$, where $\tau = 1/\sigma^2$. Verify that the posterior mean for μ is greater than the prior mean if and only if the sample mean is greater than the prior mean.

14. Suppose that you have a random sample from an $N(\mu, 1/\tau)$ distribution and believe that the conjugate $NGa(b, c, g, h)$ prior distribution is appropriate for $(\mu, \tau)^T$. Let $t_{\nu,\alpha}$ and $\chi^2_{\nu,\alpha}$ denote the upper α-points of the t_ν and χ^2_ν distributions respectively. Determine

    (i) the equi-tailed 95% Posterior confidence interval for μ;

    (ii) the equi-tailed 95% Posterior confidence interval for τ (hint: if $W \sim Ga(a, b)$ then $2bW \sim \chi^2_{2a}$);

    (iii) a 95% Posterior confidence interval for $\sigma = 1/\sqrt{\tau}$;

(iv) the 95% Posterior HDI for μ.

(v) Why is it not straightforward to determine the 95% HDI for τ? Determine the 95% HDI for τ when the size of the random sample is large.

Compare your answers with the equivalent 95% frequentist confidence intervals:

μ: $(\bar{x} - t_{n-1,0.025} s_u/\sqrt{n}, \bar{x} + t_{n-1,0.025} s_u/\sqrt{n})$
τ: $(\chi^2_{n-1,0.975}/\{(n-1)s_u^2\}, \chi^2_{n-1,0.025}/\{(n-1)s_u^2\})$
σ: $\left(\sqrt{\{(n-1)s_u^2\}/\chi^2_{n-1,0.025}}, \sqrt{\{(n-1)s_u^2\}/\chi^2_{n-1,0.975}}\right)$.

15. Suppose that you have a random sample $x_1, x_2, \ldots, x_n$ from an $N(\mu, 1/\tau)$ distribution and the sample size n is sufficiently large that the posterior distribution for $(\mu, \tau)^T$ is close to its asymptotic form. Determine

(i) the 95% HDI for μ;

(ii) the 95% HDI for τ;

(iii) the 95% HDI for $\sigma = 1/\sqrt{\tau}$.

Compare your answers with the equivalent 95% frequentist confidence intervals:

μ: $(\bar{x} - 1.96s/\sqrt{n}, \bar{x} + 1.96s/\sqrt{n})$
τ: $(1/s^2 - 1.96\sqrt{2}/(\sqrt{n}s^2), 1/s^2 + 1.96\sqrt{2}/(\sqrt{n}s^2))$
σ: $(s - 1.96s/\sqrt{2n}, s + 1.96s/\sqrt{2n})$.

16. Suppose that you have a random sample $x_1, x_2, \ldots, x_n$ from a $Ga(\alpha, \lambda)$ distribution.

(i) Determine the asymptotic posterior distribution for $\boldsymbol{\theta} = (\alpha, \lambda)^T$. Hint: you should define the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$ in terms of the equations they must satisfy, with these involving the sample mean $\bar{x}$, the geometric mean $\bar{x}_g$ and the digamma function $\psi(x) = \Gamma'(x)/\Gamma(x)$.

(ii) Suppose a sample of size $n = 100$ gives sample mean $\bar{x} = 3.0$ and geometric mean $\bar{x}_g = 2.5$ leading to maximum likelihood estimates $\hat{\alpha} = 2.8983$ and $\hat{\lambda} = 0.9661$. Determine the asymptotic posterior distribution for $\boldsymbol{\theta}$. Also calculate the asymptotic posterior correlation between α and λ. Hint: you will need to use the R functions `digamma` for $\psi(\cdot)$ and `trigamma` for $\psi'(\cdot)$.

17. In a calibration experiment, the times between successive emissions from a radioactive source were measured using two techniques: one (X) is very accurate and the other (Y) less precise. Suppose that the observed times $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are a (bivariate) random sample from a distribution in which $X \sim Exp(\lambda)$ and $Y|X = x \sim N(x, \sigma^2)$. Determine the posterior distribution for $(\lambda, \sigma)^T$ when the sample size is large.

18. Consider the Metropolis–Hastings scheme described in Example 4.5 which simulates realisations from the standard normal distribution.

(i) Investigate the relationship between the acceptance rate of the proposals (acc) and the lag 1 autocorrelation (r_1) of the output as the size of the innovation a takes values $1, 2, \ldots, 8$.

(ii) How accurate are your acceptance rates (for each a)? Hint: consider the Bernoulli observations which describe whether a proposal is accepted at each iteration. This model scenario was considered in one of your MAS2903 exercises. Use the asymptotic distribution of the true acceptance probability based on these underlying Bernoulli observations.

(iii) For each value of a, obtain a sample of 1000 realisations which are almost uncorrelated (with $|r_1| < 0.02$) and assess whether each sample is plausibly from the standard normal distribution using a Q–Q plot.

19. Suppose the posterior distribution is a standard normal distribution, with density $\phi(\cdot)$. Construct a Metropolis–Hastings algorithm which samples this posterior distribution by using a normal random walk proposal with standard deviation k.

20. Suppose the posterior distribution is $\theta|\boldsymbol{x} \sim Ga(G, H)$ distribution, with G and H known. Suppose you want to construct a Metropolis–Hastings algorithm which samples this posterior distribution by using a (skewed) $Exp(a/\theta)$ proposal distribution, where $a > 0$, i.e an exponential distribution with its parameter $\lambda = a/\theta$.

    (i) Determine the mean of the proposal distribution.

    (ii) What good feature does this proposal distribution have?

    (iii) Write down the steps in this Metropolis–Hastings algorithm to simulate realisations from the posterior distribution.

21. Suppose the posterior distribution is $\theta|\boldsymbol{x} \sim \textit{Inv-Chi}(G, H)$ distribution, with G and H known. Suppose you want to construct a Metropolis–Hastings algorithm which samples this posterior distribution by using a (skewed) log normal $LN(\log\theta, a^2)$ proposal distribution, where $a > 0$.

    (i) Determine the mean of the proposal distribution.

    (ii) What good feature does this proposal distribution have?

    (iii) Write down the steps in this Metropolis–Hastings algorithm to simulate realisations from the posterior distribution.

22. Rework Qn 21 for the case where interest is about $\beta = \log\theta$: suppose the posterior distribution is $\theta|\boldsymbol{x} \sim \textit{Inv-Chi}(G, H)$ distribution, with G and H known.

    (i) Use equation (2.1) to determine the posterior density of $\beta = \log\theta$.

    Suppose you want to construct a Metropolis–Hastings algorithm which samples the posterior distribution for β by using a normal random walk with variance a^2.

    (ii) Write down the steps in this Metropolis–Hastings algorithm to simulate realisations from the posterior distribution.

    (iii) Show that the acceptance probabilities in these two Metropolis–Hastings algorithms are the same, that is $\alpha(\theta, \theta^*) = \alpha(\beta, \beta^*)$, where $\beta = \log\theta$ and $\beta^* = \log\theta^*$.

23. The Weibull distribution is commonly used to model lifetime data. Suppose we have a random sample $x_1, x_2, \dots, x_n$ from a Weibull $Wei(\beta, \lambda)$ distribution, with parameters $\beta > 0$ and $\lambda > 0$. Suppose that the prior distribution has $\beta \sim Ga(a, b)$ and $\lambda \sim Ga(c, d)$, independently, for known values a, b, c and d.
    (i) Determine the posterior density for $(\beta, \lambda)^T$ up to a multiplicative constant.
    (ii) Determine the posterior conditional densities for $\beta|\lambda$ and $\lambda|\beta$.
    (iii) Write down the steps in a Metropolis–Hastings algorithm to simulate realisations from the posterior distribution. Your algorithm should have separate steps for each parameter and use normal random walks with variances Σ_β and Σ_λ respectively.

24. The von Mises distribution is commonly used to model circular data, that is, data on the circle such as wind directions. Suppose we have a random sample $x_1, x_2, \dots, x_n$ from a von Mises $vM(\mu, \lambda)$ distribution with mean direction $\mu \in (0, 2\pi)$ and concentration parameter $\lambda > 0$. Suppose that the prior distribution has $\mu \sim U(0, 2\pi)$ and $\lambda \sim Ga(g, h)$, independently, for known values of g and h.
    (i) Determine the posterior density for $(\mu, \lambda)^T$ up to a multiplicative constant.
    (ii) Determine the posterior densities for $\mu|\lambda$ and $\lambda|\mu$.
    (iii) Write down the steps in Metropolis-Hastings algorithm to simulate realisations from the posterior distribution. Your algorithm should have separate steps for each parameter and use normal random walks with variances Σ_μ and Σ_λ respectively.
    (iv) Using the trigonometric identities $\cos(A - B) = \cos A \cos B + \sin A \sin B$ and $A \cos x + B \sin x = C \cos(x - D)$ where $C = \sqrt{A^2 + B^2}$ and $D = \arctan(B/A)$, show that the update for μ can be achieved using a (more efficient) Gibbs step.

25. A drug company wants to assess the level of side effects from a new drug. A random sample of n people are given the drug and note is taken on the dose level (X) and whether they suffer side effects ($Y = 1$ if yes and $Y = 0$ if no). It is decided that the relationship between dose level and side effects can be described using a linear probit regression model in which

$$Pr(Y = 1|X = x) = \Phi(\beta + \theta x),$$

    where $\Phi(\cdot)$ is the standard normal distribution function. Suppose that the prior distribution has $\beta \sim N(a, 1/b^2)$ and $\theta \sim N(c, 1/d^2)$, independently for known values of a, b, c and d.
    (i) Using the data $(y_1, x_1), \dots, (y_n, x_n)$, determine the likelihood function $f(\mathbf{y}|\beta, \theta)$ and hence the posterior density for $(\beta, \theta)^T$ up to a multiplicative constant.
    (ii) Determine the posterior densities for $\beta|\theta$ and $\theta|\beta$.
    (iii) Write down the steps in Metropolis-Hastings algorithm to simulate realisations from the posterior distribution. Your algorithm should have separate steps for each parameter and use normal random walks with variances Σ_β and Σ_θ respectively.

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